Construction of TP Quantitative Models Based on Diatoms and Cladocera for the Irish Ecoregion Using Palaeolimnological Techniques

Thesis submitted for the degree of Doctor of Philosophy

by

Guangjie Chen

Msc., Bsc. (Lanzhou University, China)

Supervisor: Dr. Catherine Dalton

Department of Geography

Mary Immaculate College

University of Limerick

Submitted to the University of Limerick, August 2006

A thousand-mile journey is started by taking the first step.

Confucius (551-479 BC)

千里之行，始于足下

孔子（公元前 551-479 年）

Abstract

Title of thesis: Construction of TP Quantitative Models Based on Diatoms and Cladocera for the Irish Ecoregion Using Palaeolimnological Techniques

Author of thesis: Guangjie Chen

This study uses palaeolimnological techniques and numerical methods to construct TP transfer functions for the Irish Ecoregion based on both phytoplankton (diatoms) and zooplankton (Cladocera) remains and applies them to the sediments of seven impacted lakes to identify the pre-impacted reference conditions and help inform lake restoration targets.

A 72-lake diatom training-set and a 33-lake Cladocera training-set were developed along a TP gradient (max. 142.3 μg l^{-1} TP). Seventeen related environmental variables were available for data exploration. A variety of exploratory and multivariate data analyses were used to investigate environmental and biological data structure and their relationships using the R program. Detrended Correspondence Analysis (DCA) revealed a high degree of species heterogeneity in the diatom data compared to the Cladocera data. Both datasets were used to assess the viability of the Irish Lake Typology physico-chemical classification scheme using hierarchical clustering method and both organisms provided good biological verification.

Ordination analyses showed that nutrient gradient was among the most significant variables in determining both the diatom assemblages of 72 lakes and the Cladocera assemblages of 33 lakes. TP transfer functions were constructed using Weighted Averaging (WA)-related and linear modelling methods. A diatom TP transfer function produced a jack-knifed coefficient of determination (r^2_{jack}) of 0.743 with a root mean squared error (RMSEP) of 0.213 based on untransformed diatom data from 70 lakes. A TP transfer function built on square root transformed Cladocera data from 31 lakes yielded an r^2_{jack} of 0.729 with a RMSEP of 0.206. A sub-set of lakes where both diatom and Cladocera data were available were examined to compare the predicted TP from both models and they displayed a strong correlation (r = 0.685) for log-transformed TP.

TP models based on both indicators were applied in top-bottom analyses of seven impacted lakes. Diatom and Cladocera results indicated a similar trend of nutrient enrichment between the current (top samples) and reference status (bottom samples) for most of the lakes. Therefore the use of combined TP transfer functions based on diatom and Cladocera provided a comprehensive insight into the reference conditions and ecological status for lake restoration due to their differential positions in the community dynamics and distinctive sensitivity to water quality of lakes.

Declaration

I hereby declare that this thesis represents my work and has not been submitted in whole or in part, by me or another person for the purpose of obtaining any other qualification.

Signed: _____

Date: _____

Acknowledgements

I would like to thank my supervisor Catherine Dalton for her kind support and patience in supervising my Ph.D. work during the past three years. Her encouragement and advice not only improved the quality of this thesis but also will benefit my future research work. This work was funded by the Irish Environmental Protection Agency (EPA) (2002_PHD2_34). This work was also linked with the IN-SIGHT project funded by EPA/ERTDI of Ireland (2002-W-LS/7) and led by David Taylor of Trinity College Dublin.

I would also like to thank Manel Leira of for his kind help with the joint fieldwork with the IN-SIGHT project and diatom taxonomy and harmonization. Thanks to Ken Irvine of Trinity College Dublin for additional supervision of my Ph.D. work and also Elvira de Eyto of Marine Institute for their support on Cladocera work. I would also like to express my appreciation to Alice Wemaëre (EPA), Gavin Simpson, Thomas Davidson (both from University College London) and Eileen Cox (Natural History Museum of London) for their help with my research work. Thanks also go to Jonathan Holmes (University College London) and his family and also Xiaozhong Huang for their kindness during my stay in London. I have to say thanks to Hucai Zhang and Fahu Chen for their guidance and enthusiasm during my postgraduate studies in Lanzhou University, which have greatly benefited my Ph.D. work.

I can't forget the friendly support of Mary Immaculate College: thanks particularly go to the Reception, Inter-library Loan, Security, IT, Maintenance offices and also Des McCafferty and Shane Cullinane of Geography Department. In addition, I would like to thank all those who have contributed to the open-source program R and related packages. Their works have made my data analyses much easier. I have to thank my brother Guangyong for all kinds of help from reference collection to encouragement. I will also miss many friends in Limerick and they have provided a friendly environment for me and my family.

Finally this thesis is dedicated to my wife SUN Ying and daughter CHEN Qi. I couldn't finish this thesis without their sacrifice, endurance and love for me!

Thank you!

Lough McNean (August 2003)

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 EU-Water Framework Directive (WFD)

The deterioration of surface water quality, degradation of aquatic ecosystems and unsustainable use of water resources have become serious environmental problems at both regional and global scales. Effective water policy and legislation are urgently needed to protect the water quality and ensure the sustainable use of water resources. The establishment and implementation of the European Union Water Framework Directive (WFD) aims to improve the water quality throughout Europe and achieve good water status for all surface waters of member states by 2015 (European Council, 2000). Ecological status, an expression of the quality of the structure and functioning of aquatic ecosystems associated with surface waters, is adopted by this Directive to assess the quality of surface waters. This criterion of ecological quality is composed of three integrated groups of quality elements: physico-chemical, hydromorphological and biological. All three quality elements are required for the monitoring and assessment of water quality, however, more emphasis is placed on the biological elements which are supported by physico-chemical and hydromorphological elements (Pollard & Huxham, 1998; Irvine *et al.*, 2002). The ecological status of surface waters are categorised on the basis of the degree of deviation from the type-specific pre-impact status under undisturbed conditions (European Council, 2000). Therefore characterisation of surface water body types and the determination of pre-impact reference conditions are the two important steps in the ecological assessment of surface waters.

Lakes are important components of inland surface waters and they play significant roles in freshwater ecosystem and biodiversity. Principal factors used for classifying lakes are physical and chemical factors that determine the characteristics of the lake and hence the biological population structure and composition (European Council, 2000). Several physical and chemical factors have been outlined as significant for establishing lake types in Ecoregion 17 (Ireland and North Ireland) on the basis of Annex II and XI of the WFD (Irvine *et al.*, 2002). An Irish Lake Typology scheme comprising 13 typology classes was proposed by the Irish Environmental Protection Agency (EPA) mainly based on alkalinity, mean depth and lake area (Free *et al.*, 2005). Establishment of reference

conditions for type-specific lakes can be achieved through several methods, including the use of historical records, hindcasting models and palaeolimnological techniques (European Council, 2000; Irvine *et al.*, 2002). In particular palaeolimnological methods have been recognised and applied as vigorous tools in tracking the history of water quality (e.g. Anderson, 1995b; Smol, 2002). They also have been used for identifying reference lakes and determining reference conditions (e.g. Pollard & Huxham, 1998; Bennion *et al.*, 2004a). The IN-SIGHT project, Identification of Reference Status for Irish Lake Typologies using Palaeolimnological Methods and Techniques, was initiated to aid the implementation of the WFD in Ireland. It aims to verify and establish type-specific biological reference conditions of 35 Irish lakes covering the majority of Irish lake types using palaeolimnological methods (Taylor *et al.*, 2006).

1.2 Palaeolimnological Methods

Palaeolimnology is the study of lake sediments to reconstruct the history of lake environment, catchment development and climate change. Lake deposits are reliable archives of the physical, chemical and biological environments of lakes (Cohen, 2003). Included in the lake sediments are remains of organisms that lived in the water (Frey, 1988b). The biological remains in lake sediments can represent the integrated communities of the whole lake habitats as seasonal and spatial variation of biological elements are reduced (Frey, 1988a; Smol, 2002). Diatoms and Cladocera are among the most important organisms in lake sediments in terms of abundance, diversity and indicator sensitivity.

Biological elements for assessing the ecological status of lakes include the composition and abundance of phytoplankton and benthic invertebrate fauna (European Council, 2000). Diatoms are significant components of phytoplankton in lake systems (Round, 1981; Reynolds, 1984; Wetzel, 2001) and they have become one of the most important environmental proxies during recent decades because of their sensitivity to habitat environment, relatively high productivity and diversity, and also well-defined taxonomy (Dixit *et al.*, 1992; Cox, 1996; Battarbee *et al.*, 2001; Mackay *et al.*, 2003). Therefore different diatom communities in lake sediments can indicate the hydrochemical conditions where they lived, including pH (Birks *et al.*, 1990; Stevenson *et al.*, 1991), nutrients (Hall & Smol, 1999), alkalinity (Brugam, 1983) and salinity (Fritz, 1990).

Cladocera are small crustaceans living in both pelagic and benthic habitats and they are found in most lakes and ponds (Wetzel, 2001; Brönmark & Hansson, 2005). Chydorids, the benthic and littoral Cladocera, are attached to vegetation or bottom-dwelling and they have been used in ecological quality monitoring of lakes (de Eyto & Irvine, 2002; de Eyto *et al.*, 2003). Planktonic cladocerans, such as *Danphia*, are significant for the ecology of lakes as they can both reduce the amount of phytoplankton and serve as prey for invertebrate and vertebrate predators (Lampert & Sommer, 1997; Dodson & Frey, 2001). Cladocera remains in lake sediments have been used to reconstruct eutrophication (Whiteside, 1970; Hofmann, 1996), acidification (Nilssen & Sandoy, 1990; Simpson, 2005a), fish density (Jeppesen *et al.*, 1996) and submerged macrophytes (Thoms *et al.*, 1999).

With the development of advanced multivariate statistics and their application in ecological and environmental data, numerical methods, particularly the transfer function method, has been successfully applied in palaeolimnology. They can provide quantitative information on environmental reconstruction (including coefficient of determination and error estimates) in addition to the indicator species and assemblage approach (Birks, 1995; ter Braak, 1995; Birks, 2005b). A transfer function is a multivariate calibration model to predict the environmental conditions from the biological remains in surface sediments of lakes. Transfer functions based on diatoms have been successfully developed to infer pH (ter Braak & van Dam, 1989; Birks *et al.*, 1990; Dixit *et al.*, 1993), TP (Hall & Smol, 1992; Bennion, 1994; Dixit & Smol, 1994), air temperature (Lotter *et al.*, 1997) and salinity (Fritz, 1990; Sylvestre *et al.*, 2001). Other biological indicators are also used to quantify species-environment relationships, including algae (King *et al.*, 2000; DeNicola *et al.*, 2004), chrysophycean cysts (Facher & Schmidt, 1996), zooplankton (Jeppesen *et al.*, 1996; Davidson, 2005), Cladocera (Brodersen *et al.*, 1998; Bos & Cumming, 2003), Chironomids (Lotter *et al.*, 1999; Brooks *et al.*, 2001) and non-marine Ostracoda (Mezquita *et al.*, 2005).

Each biotic indicator can provide an independent line of ecological information but one single indicator may give misleading environmental information. This can be caused by complex physical processes during sedimentation (e.g. taphonomy) and differential sensitivity and responses to the environmental gradients of interest (Lowe & Walker, 1997; Lotter, 2005). The use of two or more indicators (multi-proxy), instead of only one indicator (single-proxy), can provide consensus results through cross-validation and

also help to identify the weakness of each proxy as every proxy has its own strength and weakness in palaeolimnology and environmental reconstruction (Lotter, 2005). Diatom assemblages in surface sediments can represent all the habitats in the lake (DeNicola, 1986; Cameron, 1995) and can also provide integrated information on water quality of the lake (Anderson, 1990). However, surface sediment diatoms can display spatial and temporal variability (Owen & Crossley, 1992) and dissimilarity to live diatoms in their asynchronous responses and inconsistent magnitudes of response to changes in water quality (Cameron, 1995; Dokulil & Teubner, 2005). Also disproportional abundance of certain diatom groups in surface sediments can reduce the credibility of the sediment diatom community (Haberyan, 1990), e.g. the overabundance of small *Fragilaria* sp. In sediments of shallow lakes (Bennion *et al.*, 1996; Sayer, 2001). Faithful representation of live littoral and planktonic Cladocera has also been revealed by surface sediments in several studies (Frey, 1960; Davidson, 2005; Kattel *et al.*, 2006). However, due to complex *in situ* physical and chemical processes affecting the distribution and abundance of Cladocera remains (Korhola & Rautio, 2001), differential habitat preferences (particularly for chydorids) (Hofmann, 1987b; Hann, 1989) and differential preservation between the different taxonomic groups of Cladocera (Frey, 1986), caution should be taken in interpretation of sedimentary Cladocera data. Therefore the combination of two or more biological indicators can help reinforce the credibility of other indicator(s) and improve the accuracy of environmental reconstruction (Jeppesen *et al.*, 2001; Bennion *et al.*, 2004b; Lotter, 2005). Thus in the same lake environment different biological organisms (e.g. diatoms and Cladocera) can show inconsistency in environmental reconstruction and the use of two or more biological indicators can reduce the danger of misinterpretation of their assemblage changes.

The use of palaeolimnological methods can be used to directly reconstruct past biological assemblages and indirectly to reconstruct past nutrient levels quantitatively through transfer functions (Battarbee *et al.*, 2005). When a large number of lakes are included for ecological assessment, down-core analysis of lake sediments would be time- and labour-intensive. The use of top and bottom approach is a viable alternative method with the assumption that the top and bottom samples of the core can represent the present-day and reference conditions respectively (Cumming *et al.*, 1992). Bottom samples (e.g. from >30 cm sediment depth of lakes) may represent pre-industrial (i.e., pre-1850) conditions (Dixit *et al.*, 1999). As biological assemblages from different habitats of the lakes are integrated and accumulated continuously in the sediment

samples, spatial and temporal variability can be minimised in the top-bottom analysis. This method has been successfully applied in 257 lakes in U.S. (Dixit *et al.*, 1999), 50 lakes in Canada (Reavie *et al.*, 2002), 219 lakes in U.K. (Bennion *et al.*, 2004a) and 35 lakes in Ireland (Leira *et al.*, 2006).

1.3 Eutrophication of Irish Lakes

Surface waters account for around 2% of Irish lands in comparison with ca. 1% for Europe (Reynolds, 1998). The geographical distribution of Irish lakes is very uneven with ca. 70% of lakes located in the western counties and ca. 22% of lakes in the north midlands (Allott *et al.*, 1998). Only 100 Irish lakes have a surface area greater than 1 km^2 (100 ha) with the majority of lakes less than 0.05 km^2 (5 ha) (Toner *et al.*, 2005). Comprehensive water quality monitoring of Irish lakes was first conducted by Flanagan & Toner (1975) and this survey of 53 Irish lakes revealed that eight lakes were excessively productive. Systematic and long-term lake monitoring has been implemented across Ireland by the Irish Environmental Protection Agency (EPA) since 1982 (Toner *et al.*, 1986; Clabby *et al.*, 1992; Bowman *et al.*, 1996; Lucey *et al.*, 1999; McGarrigle *et al.*, 2002; Toner *et al.*, 2005).

Water quality of Irish lakes was observed to be relatively stable during the period of 1982-2003 and over 50% of lakes surveyed in the 2001-3 period were classified as oligotrophic (Toner *et al.*, 2005). With regard to combined surface areas of lakes, an increase in the percentage surface area assigned to the oligotrophic group was observed in comparison to those surveyed in 1995-97 and 1998-00. However, among the 27 large lakes with surface area of above 7.5 km^2, five of them still display a high degree of nutrient enrichment and nine of them show increased phytoplankton in the 2001-3 period (Toner *et al.*, 2005). As the majority of the Irish lakes are relatively shallow with a small surface area (Toner, 1977), they are more liable to water pollution due to their relatively small lake volume and strong water-sediment interaction (Scheffer, 1998; Wetzel, 2001).

Eutrophication has been identified as the principal pressure on lake water quality in Ireland since 1970s (Flanagan & Toner, 1975; Toner, 1977; Bowman *et al.*, 1996). The principal sources of nutrient enrichment in Ireland are diffuse agricultural activities and

point source discharges of domestic and industrial wastes (Bowman & Clabby, 1998). Over three-quarters of Irish lands are used for agriculture and forestry and around 90% of pasture and crop land is cultivated as grasslands (Jennings *et al.*, 2003). The diffuse nutrient transport from grasslands has been identified as the biggest threat to the ecological status of Irish lakes (Morgan, 1977; Allott *et al.*, 1998). Intensive cattle farming and high density of pig and poultry production are also responsible for the decline of water quality of Irish lakes (Allott *et al.*, 1998). In summary eutrophication has become the main concern in protecting and improving water quality of Irish lakes (Irvine *et al.*, 2002; Jennings *et al.*, 2003; Toner *et al.*, 2005) and is the focus of this study.

1.4 Research Rationale and Thesis Structure

The hypotheses used in this study are that biological assemblages are sensitive to and can be quantitatively correlated with environmental gradients in lake waters, particularly the nutrient gradient. The remains of biological assemblages in lake sediments can faithfully represent the live biotic assemblages in lake waters. This study aims to employ palaeolimnological and statistical methods to construct relationships between biological assemblages and total phosphorus (TP) for a suite of Irish lakes. Diatoms and Cladocera are the two biological indicators used in this study. As insufficient biological data were examined for verifying the Irish Lake Typology, biological classification of surface sediment fossil assemblages in this study will help assess the viability of the Irish Lake Typology classification scheme.

Firstly the research background and rationale for this study are introduced in Chapter 1. In Chapter 2 a comprehensive literature review starts with lake nutrients and phosphorus is highlighted due to its significant role in the growth of aquatic organisms and lake eutrophication. The use of two freshwater organisms (diatoms and Cladocera) as indicators of nutrient levels is summarized particularly in the context of palaeolimnology. The transfer function technique for developing TP inference models is introduced and TP transfer functions based on diatoms and Cladocera are summarized and evaluated on the basis of the published studies. Details of the study sites and research methods used in this study are outlined in Chapter 3. This includes the study area description, sampling of 75 lakes and laboratory analyses for diatoms and

Cladocera. In addition the numerical methods used in this study are detailed, including ordination analysis, cluster analysis and transfer function modelling methods.

A 72-lake diatom training set and a 33-lake Cladocera training set are developed in Chapters 4 and 5 respectively. The pattern and distribution of related environment and biological data are explored and summarized. The reliability of surface sediment Cladocera is assessed through comparison with contemporary Cladocera communities from six lakes. Relationships between environmental variables and biological assemblages are examined using constrained ordination analysis. Biological classification of lakes is produced and compared with the physico-chemical lake typology classification for both training sets. Lakes with similar physical and chemical characteristics are supposed to support aquatic organisms with similar assemblage structure and abundance. All the data analyses in Chapters 4 and 5 are performed in the R program.

After examining the soundness of TP in determining the biological assemblages, three transfer function methods are used to develop diatom- and Cladocera-based TP inference models and their performances are compared and evaluated in Chapter 6. TP predicted by diatom and Cladocera models are compared for 29 lakes to validate the performances of both models. TP models based on diatoms and Cladocera are applied in TP reconstructions for seven lakes using top-bottom approach in Chapter 7 and this would help identify the pre-impact conditions and set lake restoration targets. This multi-proxy analysis enables the cross-validation of TP reconstructions by indicating consensus in reconstructed nutrient level and identifying dissimilar responses between both organisms to nutrient level. Lastly in Chapter 8 the research results are summarized and concluded and suggestions on future directions are also given.

Chapter 2: Literature Review

This chapter introduces the nutrient status of lakes in the context of freshwater ecology and lake management. Nutrient dynamics, classification of nutrient levels and causes of nutrient enrichment (eutrophication) of lakes are outlined. This is followed by a review of responses of two biotic organisms (diatoms and Cladocera) to nutrient dynamics. Finally the use of the transfer function method in quantifying species-TP relationships is summarized. Problems and potential solutions for applying this method are also discussed. This literature review aims to explore the current research context before developing a community-based training-set for the Irish Ecoregion.

2.1 Nutrient Status and Eutrophication of Lakes

2.1.1 Lake Nutrients

Nutrients are chemical elements that organisms require for cell growth and reproduction (Lampert & Sommer, 1997). The quantity of nutrients in lake water is mainly determined by bedrock type, vegetation cover, soils, lake size, and human activities in the catchment area (Brönmark & Hansson, 2005). Soils and rocks supply most of the ions required for organism growth (Wetzel, 2001). Primary producers including algae can absorb and concentrate the nutrients in their cells and they are then the providers of nutrients for herbivores (e.g. crustaceans) through the food web (Lampert & Sommer, 1997). Predatory organisms, like fish, consume the herbivores and therefore, nutrients are transported upwards through the food chain within the lake. Nutrients are constantly recycled through the ecosystem by processes like decomposition, absorption and excretion of organisms (Wetzel, 2001).

As all organisms are composed of the same major elements (C, N, P) and their balance affects the reproduction, nutrient cycling and food web dynamics (Elser & Urabe, 1999). Terrestrial food webs are carbon dependent and freshwater food webs are based on more nutrients (Elser *et al.*, 2000). The relative amount of elements C, N, and P in planktonic organisms was 106: 16:1 (by atoms) (often called the Redfield ratio) (Brönmark &

Hansson, 2005). Although silicon is required by all phytoplankton in protein and carbohydrate synthesis, it is of major significance to diatomaceous algae, chrysophytes and some higher aquatic plants (Reynolds, 1984; Wetzel, 2001). A comparison of the relative amounts of different elements required for plant growth with supplies available in fresh waters illustrates the importance of phosphorus (P) and to a lesser extent nitrogen (N) (see Table 2.1).

Table 2.1 Relative availability (A) and demand (D) of essential elements required for aquatic plant and algae (All the values are relative to phosphorus which has a value of 1; the higher the ratio of A/D, the greater the relative availability of that element; modified after Wetzel (2001) and Brönmark & Hansson (2005)).

Element	Availability(A)	Demand(D)	A/D	Function
P	1	1	1	DNA, RNA, ATP, enzymes
Zn	0.07	0.04	2	Enzyme activator
N	23	8.75	3	Amino-acids and proteins
Ca	40	8	5	Cell membrane
Cu	0.05	0.006	8	Enzymes
K	20	6	11	Enzyme activator
Mn	0.9	0.3	11	Photosynthesis, enzymes
Mo	0.001	0.0004	11	Enzymes
C	1200	81.25	15	Photosynthesis
Mg	22	1.4	16	Chlorophyll, energy transfer
Na	32	0.5	64	Cell membrane
Co	0.02	0.0002	100	Vitamin B12
Si	268	0.7	383	Cell wall (diatoms)
Fe	54	0.06	900	Enzymes
O	89000000	1006.25	88447	Basic for metabolism

The 15 elements listed in Table 2.1 limit growth of algae and other plants in lakes. P is the scarcest element with the lowest availability relative to the demand (1:1). P and N are among the most studied elements in ecological stoichiometry and chemical compositions of organisms (Wetzel, 2001). Nitrogen is absorbed by algae in the synthesis of amino-acids and proteins and can enter the lakes by precipitation, nitrogen fixation and catchment drainage (Reynolds, 1984; Brönmark & Hansson, 2005). Nitrogen can particularly limit phytoplankton production in eutrophic lakes with relative high phosphate concentrations and correspondingly low N:P ratios (Reynolds, 1984). However, nitrogen is generally not the main limiting nutrient for freshwater organisms and its concentration is less strongly connected to nutrient status of lakes than phosphorus (Brönmark & Hansson, 2005). Phosphorus is naturally less abundant than nitrogen and its concentration is often reduced to very low levels by plant uptake during the growing season in lakes (Moss, 1998). Phosphorus is the focus in this study.

2.1.1.1 Phosphorus

No other element has been studied as intensively as phosphorus in fresh waters, because of its major role in biological metabolism and also its low natural supply relative to demand (Wetzel, 2001). Phosphorus is essential for all organisms since it is used in fundamental processes such as storage and transfer of genetic information (DNA and RNA), cell metabolism (various enzymes), and in the energy system of the cells (adenosine triphosphate, ATP) (Moss, 1998; Brönmark & Hansson, 2005).

Phosphorus exists in lakes not only as inorganic ions (like PO_4^{3-}, $H_2PO_4^-$ and HPO_4^{2-}), but also in inorganic polymers, organic phosphorus compounds, living organisms and dead detritus. Phosphorus is taken up as phosphate (PO_4^{3-}), the only inorganic fraction of phosphorus of importance for organisms (Moss, 1998). The majority of phosphorus, greater than 80%-90%, is bound up in organic phosphorus (i.e. incorporated in organisms). Of the total organic phosphorus, at least 70% is within the particulate organic fraction, and the remainder is present as dissolved or colloidal organic phosphorus (Wetzel, 2001). The sum of organic and inorganic phosphorus is called total phosphorus (TP) and it is widely used for indicating and classifying the fertility of lakes (Brönmark & Hansson, 2005).

Lake sediments serve as phosphorus sinks, as particle detritus and dead organisms containing P continuously settle (Guy *et al.*, 1994b). Generally there is more phosphorus in sediments than in lake waters (Wetzel, 2001). However, the exchange of phosphorus between sediments and the overlying water is a major component for the phosphorus cycling in lakes and many factors and processes can affect such exchange (Scheffer, 1998). In shallow lakes where thermal stratification is weak or non-existent, intensive sediment-water interaction can enable a rapid return of nutrient materials into the water column and therefore increase the phosphorus loads in lake waters (Gibson *et al.*, 1996; Søndergaard *et al.*, 1999). During the summer the relatively high temperatures of sediments in shallow lakes can increase the mineralization rates and lead to a greater release of nutrients from sediments (Jeppesen *et al.*, 1997). The flux of phosphorus between the surface sediment and the water is also determined by oxygen availability, which controls the precipitation of phosphate. In lakes where stratification occurs, the

oxygen concentration is low and this will lead to phosphorus release to the water (Wetzel, 2001). Few organisms can survive with low oxygen levels and therefore there is limited consumption of phosphorus in the water column. The increase of pH caused by the consumption of CO_2 by high algal biomass in eutrophic lakes can also lead to the release of phosphorus from the sediment which can promote higher algae production (Brönmark & Hansson, 2005). However, algae are producers of oxygen through photosynthesis and therefore can conversely limit the release of phosphorus to the water (Wetzel, 2001).

2.1.1.2 Classification of Nutrient Status

Phosphorus generally limits the growth of algae and plants in fresh waters and thus is a main determinant for primary production. Its concentration is also relatively easy to measure. These features have led to classification of lakes mainly based on phosphorus concentration. The most commonly used trophic classification scheme was proposed by the OECD (1982). Lakes with low total phosphorus (TP) concentration (annual mean values below 10 $\mu g\ l^{-1}$), therefore with low productivity, are classified as 'oligotrophic', and 'ultraoligotrophic' lakes have TP values less than 4 $\mu g\ l^{-1}$. Lakes with medium nutrient content (annual mean TP values in the range of 10-35 $\mu g\ l^{-1}$) are categorized as 'mesotrophic' and those with higher TP values (between 35 and 100 $\mu g\ l^{-1}$) as 'eutrophic'. Lakes with extremely high TP concentration (above 100 $\mu g\ l^{-1}$) are classified as 'hypereutrophic'.

Many lakes, particularly deep stratified lakes, display a winter maximum and summer minimum in TP concentrations because of the continuous loss of nutrients from the epilimnion (the stratum of warm, well-mixed water above the thermocline) to the hypolimnion (the water below the thermocline) during the summer (Guy *et al.*, 1994a). However, total phosphorus concentrations of some lakes, particularly shallow lakes, have a different seasonal regime because of their mixed water column and intensive sediment-water contact (Scheffer, 1998). In addition Gibson *et al.* (1996) found that the annual range of TP concentrations increases as the annual maximum TP increases. However, lakes often have small variation in phosphorus concentrations in comparison with the strong seasonality of some other indicator parameters like chlorophyll-*a*. TP is

generally a reliable indicator of the nutrient level of lakes (Bennion & Smith, 2000; Jennings *et al.*, 2003).

In practice this fixed OECD classification boundaries have been modified from country to country or for different types of lakes (Smol, 2002; Søndergaard *et al.*, 2005). TP classification has been complemented by other trophic parameters like algal biomass (often estimated from chlorophyll *a* concentrations) and water transparency (OECD, 1982). A combined classification scheme using TP and Chlorophyll-*a* is practiced in Ireland (Toner *et al*., 2005). An open boundary system for trophic classification was also proposed by OECD (1982) with probabilistic parameters. Despite its arbitrariness the fixed boundary system with defined boundary is easy to apply and the resultant trophic category can be more accurate when all the trophic parameters are considered in comparison to the open boundary system (OECD, 1982).

2.1.2 Eutrophication

Eutrophication refers to the nutrient enrichment (mainly P and N) of water bodies (OECD, 1982). Small concentration of nutrients can affect aquatic systems dramatically. Algal and cyanobacterial (blue-green algal) blooms, excessive aquatic macrophyte growth and deepwater oxygen depletion are the major symptoms of eutrophication (Smol, 2002). In Ireland eutrophication has been identified as the most common water quality problem (McGarrigle, 2001; Toner *et al*., 2005).

Nutrients like P are transported into the lake water mainly from catchment areas or via atmospheric deposition (Jennings *et al*., 2003). In pristine systems, the input of phosphorus from the catchment is determined by the flow of water through the drainage system and the underlying geology. Most lakes have natural TP concentrations of between 10 and 100 μg l^{-1}. Phosphorus levels in lake waters are generally lowest in mountain areas of crystalline bedrock and increase in lowland waters derived from sedimentary rock deposits (Wetzel, 2001). Atmospheric deposition can be a significant source of phosphorus for some freshwater systems, particularly in oligotrophic lakes (Gibson *et al.*, 1995).

Natural eutrophication is a slow process in the geological history of a lake when the lake basin is gradually filled with sediments (Wetzel, 2001). The reduced volume of lake water leads to an increase in the trophic level, even if there is no change in nutrient load. Marked natural eutrophication periods are rare and can be caused by dramatic events, such as forest fire and tree die-off, or rapid climatic events, such as droughts. However, eutrophication as a result of natural disturbance is generally much less frequent and less pervasive compared with the anthropogenic impacts since the Industrial Revolution, particularly during the recent decades (Smol, 2002).

Human activities have been changing the landscape of the earth over thousands of years. The contribution of different nutrient sources will vary with the type and intensity of land use and management practice in the catchment. For inland waters, the increased intensity of deforestation and agricultural activity has greatly changed catchment areas. Nutrients are mainly lost by the erosion of surface soils from cultivated lands and run-off of animal manure and, particularly during the past decades, artificial fertilizers (Mason, 2002). Deforestation can increase the erosion rate of surface soils and some types of forest management (e.g., fertilization) increase the input of nutrients from the catchment (Jennings *et al.*, 2003).

During the past decades an accelerated rate of industrialization and urbanization has significantly increased the input of phosphorus and other nutrients into fresh waters around the world. Industrial wastes and storm drainage have been producing large amount of nutrients from the urban sources. Domestic sewage, as well as the widespread use of phosphorus-containing detergents, results primarily in increased phosphorus loading in the waters and soils (McGarrigle, 2001). However, much of the effluent in urban areas is from point sources, which can be controlled and managed through improving the wastewater treatment facilities and the introduction of nutrient-free detergents. More challenging is the control and management of the diffuse sources (e.g., fertilizer run-off from agriculture) and this is the main cause of lake eutrophication in Ireland (Jennings *et al.*, 2003; Toner *et al.*, 2005).

2.2 Biological Indicators of Lake Nutrient Status

The EU-Water Framework Directive (WFD) has emphasized the ecological monitoring, assessment and restoration of surface waters as its priority subject (Irvine *et al.*, 2002). Biological organisms (mainly plankton, macrophytes, benthos and fish) have proven to be reliable and significant indicators of nutrient status of lakes (Mason, 2002), including both eutrophication (Smol, 2002) and oligotrophication ((Battarbee *et al.*, 2005). At a regional scale, the response model of lake communities to nutrient enrichment is remarkably similar, and the degree of replacement of one species group by another is a good indication of the degree of enrichment (Stoermer, 1984). Diatoms (Hall & Smol, 1999) and cladocerans (Korhola & Rautio, 2001) are among the most sensitive and reliable biological proxies for lake eutrophication and were therefore selected for this study.

2.2.1 Diatoms (Bacillariophyceae)

Diatoms are microscopic unicellular algae, with sizes ranging from approximately 5 to 500 μm and belong to the class Bacillariophyceae (Barber & Haworth, 1981). The most striking feature of the diatom cell is the highly resistant siliceous wall (termed frustule and also described as cell or valve) enclosing the living contents (Round *et al.*, 1990). Diatoms are usually composed of two valves with a series of linking bands (or girdle) sitting in between the valves (Cox, 1996). These siliceous cells are commonly abundant and diverse in fresh waters and the cell walls are preserved long after death of the cell and decay of its organic contents (Barber & Haworth, 1981). Diatoms are photosynthetic plants and their reproduction is predominantly asexual (Barber & Haworth, 1981; Krammer & Lange-Bertalot, 2000). Diatoms can live in a variety of habitats: free-living in the open water (planktonic), mobile or immobile on the bottom (benthic) or attached to a substrate (epiphytic when attached to other plants and epilithic when attached to rocks) (Barber & Haworth, 1981; Round *et al*., 1990). The growth and behaviour of planktonic diatom populations are strongly influenced by the availability of silica and the stability of the water column, as well as several other factors including light intensity, nutrient level and grazing pressure (Reynolds, 1984; Round *et al*., 1990). The benthos is generally more diverse than the plankton in regard to species richness and the life forms present, and the epiphytic and epilithic diatoms are commonly best developed in submerged habitats (Round *et al*., 1990).

Diatoms occur either in centric or innate forms based on the cell shapes and arrangement of valve markings (Cox, 1996; Krammer & Lange-Bertalot, 2000). Cells in girdle view are usually approximately rectangular but cells in valve view can have much more variable outlines. In comparison with diatom valves the cell girdle is generally less used for taxonomic identification due to its inadequate features available under the light microscope (Krammer & Lange-Bertalot, 2000). The taxonomy of diatoms is mainly based on the morphology of silica walls, with families and genera being determined on symmetry, shape and the arrangement of valve markings (Barber & Haworth, 1981; Cox, 1996). The key features for identifying diatom frustule morphology include valve shapes, striae patterns and features, the raphe and axial areas etc (Barber & Haworth, 1981). The nomenclature adopted in this study mainly followed those of Krammer & Lange-Bertalot (2000). The Bacillariophyceae are divided into two Orders based on symmetry and the arrangement of the rib systems and areolae. One is the centrale, whose valves are either circular or showing a symmetrical centre in the middle of valve. The other is pennale, whose valves are always elongated and structures are oriented around a median axis. Most of the centrales are in planktonic forms while the pennales are mainly found in the littoral regions of lakes (Round *et al.*, 1990; Krammer & Lange-Bertalot, 2000). The first comprehensive investigation of freshwater diatoms in the Irish Ecoregion identified 765 taxa belonging to 48 families following the nomenclature of Hustedt (1927-1966, 1930) (Foged, 1977).

2.2.1.1 Diatoms and Nutrient Dynamics of Lakes

Diatoms have been widely applied in a broad range of subjects from environmental to earth sciences due to their unique features (Stoermer & Smol, 1999). Diatoms have one of the shortest generation times of all biological indicators and therefore they can respond rapidly to environmental change (Stevenson & Pan, 1999). They occur in all types of aquatic environments and contribute significantly to the primary productivity of aquatic ecosystems as photoautotrophs (Round *et al.*, 1990). The taxonomy of diatoms are well documented and they can also be easily sampled and processed for identification (Cox, 1996).

Diatoms are well suited to studies of eutrophication of surface waters (Hall & Smol, 1999). They are early and sensitive indicators of environmental change as their growth

is directly controlled by nutrient and light availability (Tilman *et al.*, 1982). Diatoms can respond quickly to lake eutrophication because they migrate and replicate rapidly (Dixit *et al.*, 1992). Individual species of diatoms have narrow optima and tolerances for many environmental variables including nutrient concentrations (Van Dam *et al.*, 1994). Oligotrophic lakes often have a diatom flora dominated by species of *Cyclotella* and *Tabellaria*, while *Asterionella formosa*, *Fragilaria crotonensis*, *Stephanodiscus astraea* and *Melosira granulata* are usually the dominant taxa in eutrophic lakes (Mason, 2002). Different diatom assemblages may prefer distinct habitats of the lake, ranging from plants (epiphyton), rocks (epilithon), sand (epipsammon) and mud (epipelon) of the littoral area to the open water (plankton). Therefore an assemblage shift from littoral to planktonic can be an indication of the lake eutrophication (e.g., Osborne & Moss, 1977). Also changes in diatom assemblages can be caused by biotic interactions with other algae, zooplankton and fish, and the signals of nutrient enrichment can be extracted from diatom assemblages through the food web and trophic structure of lakes (Dixit *et al.*, 1992).

2.2.1.2 The Use of Diatom Remains to Infer Past Nutrient Levels of Lakes

Planktonic and benthic diatoms are sedimentated on the beds of most water bodies in the form of whole or broken frustules (Krammer & Lange-Bertalot, 2000). Diatoms are well preserved in the sediments in high abundance and diversity because their cells are resistant to decay (Round, 1981). One of the most common palaeolimnological applications of diatoms has been to investigate eutrophication history of individual lakes (e.g., Hall & Smol, 1999; Smol, 2002). Qualitative interpretation of eutrophication history can be inferred from the sedimentary diatoms based on ecological information of individual indicator taxa provided by contemporary phycological surveys (e.g., Van Dam *et al.*, 1994). Diatoms in the sediments can be analysed to reconstruct the change of habitat environment during the sedimentation period based on their qualitative ecological information (Dixit *et al.*, 1992).

One of the most obvious features in the response of diatom assemblages to eutrophication is the shift of dominant species from benthic and epiphytic to planktonic. In the English Lake District a rise of nutrient levels from about 1850 was recorded by a shift away from benthic diatoms and a rise of planktonic *Asterionella formosa* in the

lake sediments, while the replacement of *A. formosa* with *Cyclotella* spp. Around 1945–65 and return of epiphytic *Achnanthes minutissima* since 1990 corresponded to an improvement of water quality (Barker *et al.*, 2005). Another indication of changes in nutrient level is the replacement of dominant diatom taxa by other taxa from lower or higher trophic levels in the food chain. The dominance of diatom assemblages shifted among the oligotrophic *Cyclotella comensis* and *C. ocellata*, to oligo-mesotrophic diatom taxa (e.g. *Cyclotella pseudostelligera*, *Fragilaria crotonensis*) and mesotrophic-eutrophic taxa *Stephanodiscus minutulus*/*parvus* in lake sediments from Alaska (Finney *et al.*, 2002). The succession in the diatom assemblage structure can be used to track the history of nutrient dynamics. A sediment core of Lough Augher in Northern Ireland, which has experienced eutrophication from 1900 until 1972-73, displayed a sedimentary diatom assemblage succession with a shift from mesotrophic planktonic forms (*Aulacoseira ambigua*, *Asterionella formosa*, *Fragilaria crotonensis*) to a variety of small *Stephanodiscus* spp., typical of very eutrophic conditions (e.g. *S. parvus*, *S. hantzschii*) (Anderson *et al.*, 1990).

2.2.2 Cladocera (Water Fleas)

Cladocerans are a group of small, transparent and discus-shaped crustaceans with adult size ranging from 0.2 to 18 mm (Dodson & Frey, 2001) (see Figure 2.1). Their general shape and jerky swimming account for their common name, ‘water fleas’. They can live in almost any kind of freshwater habitat, from large lakes to small ponds. Most cladocerans have a transparent clear-to-yellow carapace or shell that is attached to the back of the neck, wrapping around the body. The shape of the shell in lateral view can be various: oval, circular, elongated or angular (Pennak, 1989). There are often different types of surface markings on the shell, like reticulation and striation. The head is usually more or less dome-shaped but can have long, even pointed extensions. The most conspicuous internal structure of the head is a large black compound eye and a small black ocellus (simple eye) for most species as shown in Figure 2.1. Like other crustaceans, cladocerans have five pairs of appendages on the head part of the body. The cladoceran body posterior to the head is comprised of a thorax and an abdomen hanging within the carapace (Scourfield & Harding, 1966). The part of the body at the end of the abdomen and posterior to the anus is called postabdomen and it ends in a pair of claws. Cladocera reproduce either sexually or asexually, depending on environmental

conditions (Dodson & Frey, 2001). Most species reproduce parthenogenetically most of the time and female offspring are most often developed. Ephippia containing resting eggs produced by sexual reproduction are often found in harsh environments and they can withstand severe conditions like drying and freezing (Lynch, 1980).

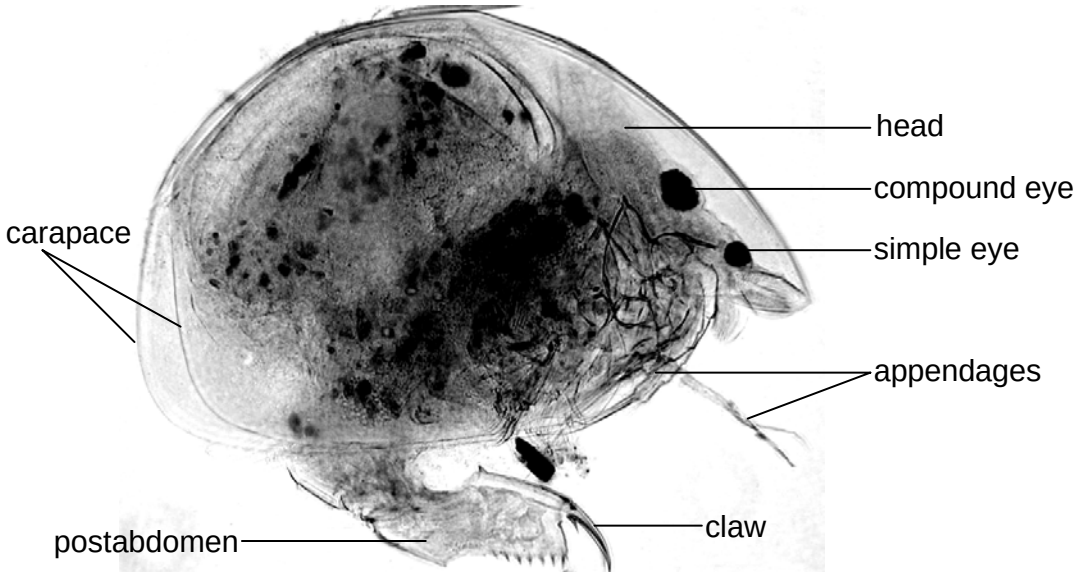

Figure 2.1 Morphological features of whole body of Cladocera species (*Alona guttata*)

Cladocera are now classified as an order and grouped into four suborders, in contrast to the previous classification scheme, which divided it into four orders, as the orders are probably not closely related and the recent molecular evidence suggests that cladocerans are descended from a common evolutionary ancestral group (Korhola & Rautio, 2001). The four suborders (Anomopoda, Ctenopoda, Onychophoda and Haplopoda) are composed of 11 well-defined families, of which three families are commonly found in freshwaters. Members of the Daphniidae and Bosminidae families are primarily planktonic (inhabiting in the open water) and they are pelagic filter feeders mainly on algae and to some extent on bacteria (Dodson & Frey, 2001). While members of Chydoridae family prefer the lake bottom surfaces (e.g. on plant, sand, mud and rock), and they typically feed by crawling along surfaces, scraping and filtering food particles (Freyer, 1968). The relatively uniform habitats in open water offer less scope of diversification for Daphniidae and Bosminidae in comparison to bottom-dwelling and crawling Chydoridae and therefore the diversity within the Chydoridae is higher than the other two families (Freyer, 1968). The carapace and headshield of chydorids are physically the strongest among all Cladocera families and this may indicate an evolutionary adaptation of the Chydoridae to the benthic habitats (Freyer, 1968). More

than 90 Cladocera taxa are known to inhabit European inland waters (Korhola & Rautio, 2001) while 41 Chydorid and 16 planktonic Cladocera species were found in the Irish freshwaters (Duigan, 1992; Irvine *et al.*, 2001).

2.2.2.1 Cladocera and Nutrient Dynamics of Lakes

Cladocerans are a dominant component of the zooplankton and littoral micro-crustaceans of most freshwater lakes. They occupy an intermediate trophic status in food webs and nutrient dynamics. Zooplankton N:P ratio is important in understanding effects of food quality on secondary production in aquatic ecosystems (Sterner & Hessen, 1994). Freshwater zooplankton with low body N: P ratios and high phosphorus demands for growth, particularly *Daphnia*, are sensitive to the P content of their food and suffer decreased growth and reproduction when consuming food with low P content (Urabe *et al.*, 1997). Other herbivorous cladocerans like *Bosmina* are less limited by phosphorus than *Daphnia*, but competition between *Bosmina* and *Daphnia* might shift in favour *Bosmina* under P-limited conditions (Sterner & Hessen, 1994; Schulz & Sterner, 1999). Loss of macrophyte habitat, increased predation from fish, and hypoxia at the sediment–water interface have been identified as potential negative effects of eutrophication on littoral zoobenthos like chydorids (Vadeboncouer *et al.*, 2003). At an early stage of nutrient enrichment the increase in macrophyte diversity and food availability may produce higher abundance and species richness of chydorids (de Eyto, 2000). With eutrophication macrophytes may die off and a shift from a plant-dominated lake to a phytoplankton-dominated lake can affect the littoral communities due to the loss of macrophytes (Scheffer, 1998).

Cladocerans have proved to be sensitive to nutrient dynamics and they have been studied to track the patterns of lake eutrophication (Korhola & Rautio, 2001; Shumate *et al.*, 2002). A shift of cladoceran taxa from *Bosmina longispina* to *B. longirostris* has been documented in a number of early studies and has been attributed to nutrient enrichment (e.g., Hasler, 1947; Edmondson *et al.*, 1956; Beeton, 1965). *Daphnia* exhibited improved growth in response to the phosphate addition in three hypertrophic Dutch lakes (DeMott *et al.*, 2001). Chydorid distribution from 66 shallow lakes across Europe showed that the proportional abundance of *Chydorus sphaericus* increased with decreasing water quality, while a concurrent increase in species like *Alona rectangula*

and *Pleuroxus uncinatus* and decrease in *Alonella excisa* and *Alonopsis elongata* could also be linked with eutrophication (de Eyto *et al.*, 2003).

2.2.2.2 The Use of Cladocera Remains to Infer Past Nutrient Levels

Cladoceran groups are well preserved in lake sediments compared to other zooplankton (Harmsworth, 1968). However, Cladocera individuals rarely remain intact in the sediment after death and the various exoskeleton components are usually separated from each other (Frey, 1960). Chitin, the main component of the Cladocera skeleton, is very inert chemically and it is preserved due to its resistance to biological degradation (Korhola & Rautio, 2001). Hard-shelled forms of Cladocera, such as Chydoridae, are well preserved, whereas soft-shelled chitinous taxa, such as Daphniidae, are often represented by smaller fragments (mainly postabdominal claws) and resting eggs (Jeppesen *et al.*, 2001; Korhola & Rautio, 2001). Some skeletons of Chydoridae and Bosminidae are completely preserved and their remains provide a good index of the original live assemblages (Frey, 1976). It was first demonstrated by Frey (1960) that cladoceran remains in lake sediments closely reflected the living population in terms of species presence and their relative abundance. The validity of the use of cladoceran remains was confirmed by other studies on fossil and live cladoceran assemblages (e.g., Davidson, 2005; Kattel *et al.*, 2006).

Cladocera remains have proved to be good indicators in reconstructing anthropogenic impacts on lake ecosystems, including nutrient enrichment (Jeppesen *et al*., 2001). Alteration in the Cladocera community structure can be a good indicator of nutrient level change (Hofmann, 1987b). An early study by Whiteside (1970) investigated the relationship between lake type and Chydorid assemblages as reflected by the chydorids in surface sediments from 77 Danish lakes. Good correlations between chydorids and three lake types (clear-water lakes, ponds and bogs, polluted lakes) were revealed. Broderson *et al*. (1998) examined the Chydorid assemblages of surface sediments from 32 Danish lakes and confirmed that chydorid remains in surface sediments are valuable indicators for the trophic status of lakes. A clear community response in Cladocera to experimental eutrophication was demonstrated in laminated sediments (Hann *et al.*, 1994). In addition the indicator or dominant species of Cladocera can be used for disclosing the change in nutrient status. The abrupt increase of *Chydorus sphaericus* and

decrease of macrophyte-associated *Alona affinis* in absolute abundances corresponded to accelerated eutrophication in Lake Apopka, Florida (Shumate *et al.*, 2002). In Ireland cladoceran assemblages from a sediment core from Lough Ennell showed significant changes indicating a radical alteration in trophic status of the lake (Murray & Douglas, 1977). The relative abundance of *C. sphaericus* increased from 8% at 60 cm depth to 30% at the core top with a concurrent decrease in *Alonella excisa*, an indicator species for nutrient-poor waters.

However, the relationship between Cladocera communities and total phosphorus is likely complicated by other factors, such as fish predation and macrophyte cover (e.g. Jeppesen *et al.*, 2001; Davidson, 2005). The greater the predation pressure by fish, the more biased is the structure and abundance of zooplankton community in comparison with the assemblage under no predation pressure. Fish usually select larger prey, whereas with invertebrate predation, small-sized individuals are effectively removed (Brooks & Dodson, 1965). Both live and surface sediment zooplankton assemblages (mainly composed of Cladocera) from 39 shallow lakes were shown to be strongly influenced by fish density and submerged macrophyte abundance (Davidson, 2005). Submerged macrophytes can have significant impact on the trophic structure, nutrient dynamics and water clarity of shallow lakes (Scheffer *et al.*, 1993). It was found that the provision of a refuge from fish predation for large-bodied cladocerans among stands of submerged macropytes is an important stabilising mechanism against nutrient-induced phytoplankton increases in clear water shallow lakes (Stansfield *et al.*, 1997). Also significant positive association of Cladocera abundance with increasing macrophyte coverage were apparent throughout the summer. A positive relationship was observed between the relative abundance of surface sediment chydorids and the average cover of submerged macrophytes (Thoms *et al.*, 1999). The complicated relationship between cladoceran assemblages and nutrient level is one of the main reasons that cladocerans have not been as widely used as some other biological organisms (e.g. diatoms) for tracking lake eutrophication.

The response of Cladocera assemblages to eutrophication may not be reflected in lake sediments. Cores of two North German lakes were subject to postglacial eutrophication as indicated by the chironomid assemblages but no such trend was reflected in the Cladocera assemblages (Hofmann, 1987b). Fossil Cladocera records showed that they were more sensitive indicators of predation than of lake trophic status although the

increase in nutrient level was reflected by increased abundance of *C. sphaericus* (Brugam & Speziale, 1983). Therefore further research work is needed to improve the confidence with which Cladocera remains can be applied in ecological investigation and monitoring. The combination of Cladocera and other biological records like diatoms can help assess the validity of the uniform responses of these indicators to the environmental variables (Jeppesen *et al*., 2001).

2.3 Quantification of Species-Nutrient Relationship

As the distribution and abundance of organisms are influenced by chemical, physical and biological characteristics of the habitat environment, the relationship between present-day distribution and abundance of biota and the environment can be used to reconstruct past environmental conditions on the basis of fossil communities. There are a great number of environmental variables that influence species distribution and abundance, as well as the number of biological taxa considered for building such relationships. Before environmental conditions can be reconstructed, the ecological optima and tolerances of these indicators have to be estimated. Several methods have been used to extract the ecological information between biota and environmental variables. Field investigation and laboratory experiment are typical approaches to collect data (Hall & Smol, 1999). However, the use of surface-sediment training sets has proved to be the most powerful one for calibrating the distribution and abundance of taxa (Smol, 2002).

The surface-sediment transfer function method (see Figure 2.2) involves sampling a range of lakes (training set) for indicator species and a suite of environmental variables, which are then related using statistical techniques (Smol, 2002). Ecological information from the water column and littoral habitats are accumulated and integrated in lake sediments (Frey, 1986; Smol, 2002). Such accumulation of fossil assemblages in surface sediments over time also minimizes the effect of annual or seasonal patterns during biological monitoring (Brodersen *et al.*, 1998). The profundal zone of a lake normally consists of exposed fine sediment free of vegetation with high variety and abundance of biological community remains. The sediment sample from the deepest part of lake is assumed to represent the general sedimentary characteristics of the lake and catchment (Anderson, 1995b; Smol, 2002). Transfer function can be applied in the fossil biological

assemblage of lake sediments to quantitatively reconstruct the environmental variables of interest (see Figure 2.2).

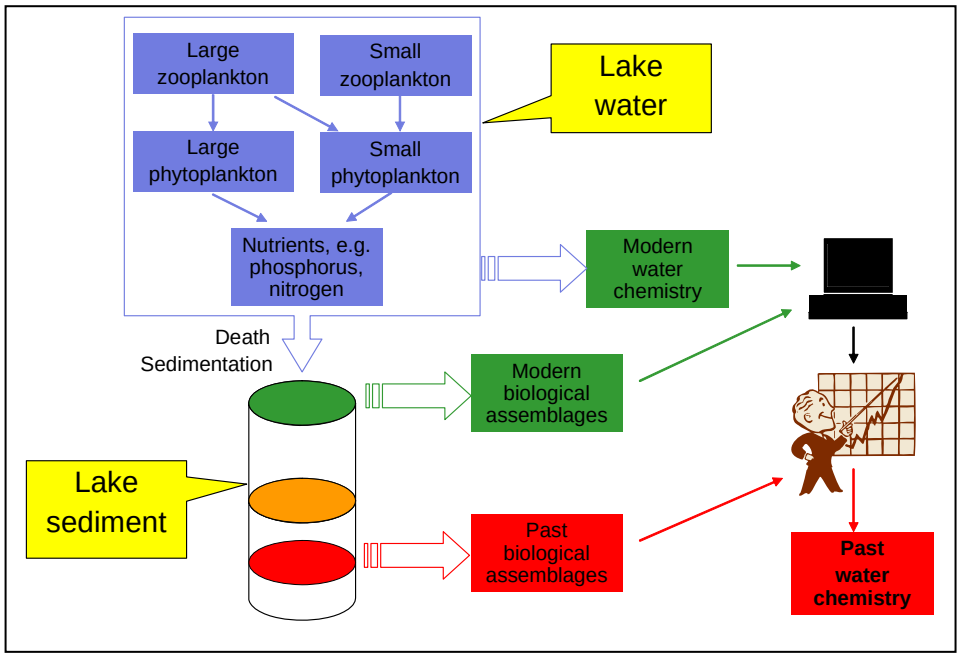

Figure 2.2 Outline of transfer function development and their use for reconstruction.

Several major assumptions for quantitative reconstruction based on training sets are discussed by Birks (1995, 1998). The taxa used in the training sets should have a systematic response to the habitat environment where they live. The environmental variable of interest must be an important determinant of the ecosystem while other variables have negligible influences on the system. In addition the statistical methods should be adequate for numerical modelling. It is assumed thta ecological optima and tolerances of modern and fossil taxa be the same and evolutionary processes will have neglectable influence on palaeolimnological assessment. The test data for model evaluation has to be independent of the training data (Telford & Birks, 2005). These assumptions have proved to be safe as shown by lots of studies on dominant biological indicators and important environmental variables and for the time frames of most environmental studies (Smol, 2002).

Numerical methods for transfer function construction are depending on the species response along the environment gradient (Birks, 1995, 1998). The Gaussian unimodal model is a good approximation for biological data that span long gradients and therefore unimodal-based modelling techniques (e.g., Weighted Averaging (WA) and Weighted Averaging Partial Least Square (WA-PLS)) are appropriate for transfer function

development. Linear-based techniques (e.g., Partial Least Square (PLS)) are appropriate when species are generally behaving monotonically along a short gradient. Details of the numerical methods are discussed and detailed in Chapter 3.

2.3.1 Diatom-based TP Transfer Functions

Diatom-based TP transfer functions can provide quantitative estimates of historical phosphorus concentration from the sedimentary sequence of diatoms. They have been one of the most important palaeolimnological techniques during the last two decades (Anderson, 1997a). At least twenty-five diatom-based TP transfer functions have been constructed during the past twenty years around the world. These diatom training sets are generally composed of lakes from distinct eco-climatic regions, containing various lake types (e.g. lake depth and area) and spanning different TP gradient lengths.

Different diatom training sets can incorporate various lake types in terms of lake depth, alkalinity, lake area, etc. Reavie & Smol (2001) selected 64 South Ontario lakes across the range of nutrient level from oligotrophic to eutrophic, with an average maximum depth of 26.5 m. Thirty-one shallow ponds with maximum depth generally smaller than 3 m were selected in Southeast England (Bennion, 1994). Two training sets from Ontario, Canada (Reavie & Smol, 2001; Werner & Smol, 2005) were mainly based on lakes with moderate to high alkalinities. The lakes selected by Kauppila *et al*. (2002) excluded the very small or large lakes, therefore this training set was composed of 68 medium-sized lakes with a size range of 12-973 ha. While 33 reservoirs and lakes with the mean size of 6240 ha and a much wider size range of 14-13 882 ha were selected in the diatom TP training set from Australia (Tibby, 2004), because 29 of the 33 sites have large dams as a result of river regulation.

Lakes located within the same ecoregion generally have similar conditions like climate, vegetation, soil and underlying geology. Therefore a regional TP transfer function can be applied to lake sediments from the same ecoregion with potentially high predictability. However, a regional training set often has uneven distribution of sites along the TP gradient and has difficulty covering the full range of the environment gradient. The datasets from Southeast England (Bennion, 1994) and Denmark (Bradshaw *et al.*, 2002) included many sites with TP over 50 µg l^{-1} and thus were

strongly biased towards eutrophic lakes. The training sets developed for such areas are commonly region-specific, and are less applicable to other regions with different physico-chemical characteristics. Some data sets selected lakes spreading over several ecoregions with different geology, climate, vegetation and soil (e.g., the 55 lakes spanning three different eco-regions investigated by Ramstack *et al*. (2003)). Such a selection can enable investigators to construct and compare the different eco-region models and help identify the main impact factors. The combination of several regional datasets, like the Northwest Europe dataset consisting of 164 lakes from 6 regional datasets and covering a TP range of 5-1190 μg l^{-1} (Bennion *et al.*, 1996), can form a large-scale training set that can have a broader range of diatom species distribution and a longer TP gradient.

It is not always true that the bigger the size of the training set or the larger the area it covers, the greater its robustness and predictability will be. With an increase in training set size and geographical distribution, the heterogeneity within it will also increase and the prediction errors rise. The ecological response in lakes to the same environment variable can vary on a continental, regional or even smaller scale. Furthermore, a short environmental gradient can still occur in a large-scale training set and this short gradient can reduce the accuracy of the transfer function produced. For example, the training set of 257 lakes in Northeastern United states only has a TP range of 3-48μg l^{-1} (Dixit *et al.*, 1999).

2.3.2 Cladocera-based TP Transfer Functions

Four transfer functions based on cladoceran assemblages have been developed to infer the total phosphorus (TP) in central Europe and North America. Brodersen *et al*. (1998) investigated the water chemistry and surface-sediment cladocerans in 32 Danish lakes. The strong relationship between species data and TP values enabled the establishment of a chydorid-based WA model to infer TP and estimate changes in nutrient concentrations in Danish lakes that have occurred since the mid-1960s. A statistically significant relationship between benthic cladoceran assemblages and epilimnetic TP concentrations was built for 68 small alpine lakes in Switzerland (Lotter *et al.*, 1998). In central British Columbia in Canada, 53 lakes were sampled for water chemistry and surface sediment Cladocera in addition to a suit of physical and spatial explanatory variables (Bos &

Cumming, 2003). Total phosphorus was identified to have the most significant relationship with Cladocera assemblages. Predictive models were developed to estimate TP from species abundance data using weighted averaging techniques along a TP gradient of 5-146 μg l^{-1}. All these studies confirm that Cladocera remains are reliable and sensitive indicators for trophic status and quantitative TP inference models based on Cladocera are useful tools in indicating the nutrient status of lakes quantitatively.

2.3.3 Problems with and Improvements in TP Transfer Functions

The transfer function method has been found powerful in quantifying the diatom-TP relationship but the sources of errors have been observed in many studies (e.g. Anderson, 1995a; Bennion *et al.*, 1996; Sayer, 2001). Intra-annual variability in nutrient concentration is high and tends to be greatest in most nutrient-enriched waters (Gibson *et al.*, 1996; Bennion & Smith, 2000). Strong seasonality of planktonic diatoms, with main growth periods in spring and autumn in meso- to eutrophic lakes were observed by Bradshaw *et al.* (2002). This implied that the measurement of TP, which is most relevant for diatom ecology, may not be adequately assessed by single or a few measurements during the year. In practice the water sampling strategy used in investigations is not always the same. Annual mean TP values were the most commonly used, while spring, summer and autumn TP measurements were also used in many training sets. Within a given training set, surface sediment samples are often taken randomly through the year and this has proved influential in the calibration models (Sayer, 2001). Improvement for training sets may be achieved by standardizing water chemistry and sediment sampling methods (Bennion & Smith, 2000; Sayer, 2001). The predictability of the models can be improved by conducting a whole-year water-quality monitoring program with a high sampling frequency (Bennion & Smith, 2000; Kauppila *et al.*, 2002). Furthermore, annual means rather than winter-spring means can provide more appropriate estimates of TP due to the importance of internal cycling of nutrients in summer particularly for shallow lakes (Bennion & Smith, 2000).

Lake selection is important for the robustness of the training sets. Exclusion of acidic and 'brown water' lakes could possibly reduce secondary gradients in the data set without compromising the applicability of the TP inference model (Kauppila *et al.*, 2002). Also typically at least 40 or so lakes should be included, but generally the more

the better (Smol, 2002). There is evidence to suggest that diatom-TP training sets may be much less effective when applied to shallow lakes (Bennion *et al.*, 2001; Sayer, 2001). In shallow lakes macrophytes can grow over the whole lake and key factors controlling the diatom assemblages such as light and substrate can complicate the relationship between diatoms and trophic status. In addition the prevalence of non-planktonic diatoms mainly driven by habitat availability in shallow lakes is poorly related to phosphorus and this can reduce the accuracy of predictability of diatom-TP models (Bennion *et al.*, 1995; Sayer, 2001; Bennion *et al.*, 2005). It has been suggested that this conflicting influence of habitat availability in training sets might be overcome by developing habitat-specific datasets based on sampling contemporary planktonic, epiphytic and benthic communities (Bennion *et al.*, 1995; Sayer, 2001). A planktonic diatom-based TP inference model was developed to minimise the influence of diatoms from other habitats (Bradshaw *et al.*, 2002).

To improve the accuracy and predictability of TP inference models, a full range of surface water TP should be included. In practice this is not always possible due to geological and geographical features, lake types and human activity in different training sets lakes. Training sets from Northwest Europe generally included a large number of lakes with annual mean TP above 100 μg l^{-1} (e.g., Bennion, 1995; Broderson *et al*., 1998). Other training sets were mainly composed of lakes with relative short TP gradient and strong bias towards the lower end of the gradient (e.g., Hall & Smol, 1996; Enache & Prairie, 2002). A relatively short nutrient gradient may contribute to the low levels of model predictability for some training sets, such as 54 lakes from Ontario with a TP gradient of 2.7-24.3 μg l^{-1} (Hall & Smol, 1996). It is suggested that TP reconstructions would have better accuracy for the TP range which is near or within the main TP gradient of inference models (Reavie & Smol, 2001).

As diatoms are not the only algal group and represent only one of several trophic levels in lakes, their use alone cannot determine the complex interactions among different communities and trophic levels (Hall & Smol, 1999). It remains unclear whether cladocerans are directly sensitive to TP or whether they are influenced more profoundly by concomitant changes in primary producer communities or substrate availability (Lotter *et al.*, 1998; de Eyto & Irvine, 2002). The presence and intensity of predation and submerged macrophyte coverage have also proved to be significant factors in controlling the Cladocera assemblage structure and abundance (Brooks & Dodson,

1965; Thoms *et al.*, 1999; Davidson, 2005). Cladocera-based transfer functions have also been successfully developed for other environmental variables, including submerged macrophytes and fish community (Jeppesen *et al.*, 1996), summer temperature (Lotter *et al.*, 1997) and water level (Korhola, 1999).

Different organisms and their fossil counterparts can respond quite differently to eutrophication and it would be misleading to rely on a single proxy indicator when using the palaeolimnological record to infer past environmental change. Further investigations that compare fossil records along defined gradients of environmental change or within specific experimental designs were suggested to more clearly assign mechanistic explanations to the observed fossil trends (Jeppesen *et al.*, 2001). Multiple-proxy analysis could encompass multiple trophic levels, incorporate broad change in the overall diversity of habitats, and benefit the study of eutrophication (Sayer, 2001). This current project aims to combine the diatoms and Cladoceran remains to construct the TP inference models for the Irish Ecoregion. The combination of two or more biological indicators, including diatoms, cladocerans and chironomids, can improve the accuracy and predictability of inference models for nutrient status and dynamics in lakes (Hall & Smol, 1999).

Chapter 3: Study Sites and Methods

This chapter first introduces the study area, site selection, lake typologies and sampling of 75 lakes from the Irish Ecoregion included in this study. Then details of laboratory analyses for both diatoms and Cladocera are described, including sample processing, slide preparation, species identification and counting. Then the main numerical methods selected for this study are summarized. Multivariate analysis and the transfer function method are detailed along with related software and packages.

3.1 Study Sites and Samples

3.1.1 Study Area and Site Selection

Ireland is located between 51˚20′-55˚20′ N latitude and between 5˚20′-10˚40′ W longitude in the Northwest of Europe. Its climate is dominated by the moderating Atlantic Ocean and generally has a relative warm climate with an annual average temperature of around 9 °C. Minimum air temperature falls below zero on about 40 days per year in inland areas and on less than 10 days per year in most coastal areas (see *www.met.ie*). Annual average rainfall between 1972-94 was approximately at 1200-1400 mm $year^{-1}$ in the lowland areas of Western Ireland and 800-1000 mm $year^{-1}$ in Eastern Ireland except for the mountain areas (Jordan, 1997).

Lakes were selected for sampling according to the research aim and methods, environmental data availability and the feasibility of fieldwork. For this study the key criterion for lake selection was the TP status of lake waters. Nutrient levels have been found to contribute significantly to the performance of the community-based transfer functions as reviewed in Chapter 2. Seventy-five lakes across the Irish Ecoregion were selected for coring in this study, including 72 lakes for the diatom training set, 33 lakes for the Cladocera training set and seven lakes for top-bottom analysis. Seventy-four lakes are included in both training sets with a TP range of 0-142.3 μg l^{-1} and site location of these training set lakes are shown in Figure 3.1. Summary information of all the 75 study lakes is shown in Table 3.1.

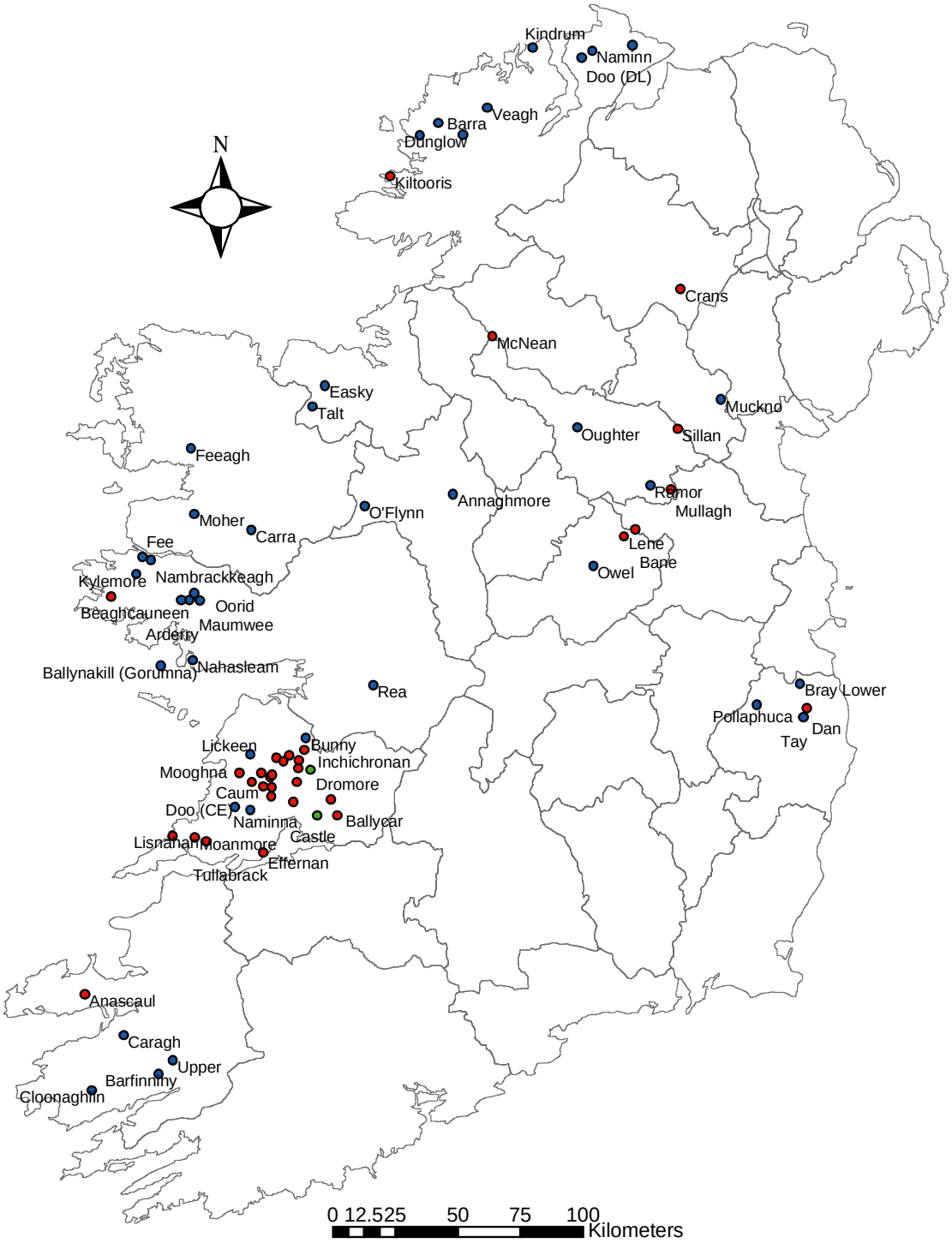

Figure 3.1 Location map of 74 study lakes included in the diatom and Cladocera training sets for the Irish Ecoregion.

Table 3.1 Summary information of 75 lakes included in this study (H-Area= Hydrometric Area; ● = analyzed by the author; ■ = analyzed by Manel Leira; - = no analysis)

Lake Name	County	H-Area	Grid Reference	Sampler, sampling year	Sampling Equipment	Diatom Analysis	Cladocera Analysis
Anascaul	Kerry	22	Q 585 052	Chen, 2004	Gravity Corer	●	●
Annaghmore	Roscommon	26	M 900 837	IN-SIGHT, 2003	Gravity Corer	■	-
Arderry	Galway	31	L 995 457	IN-SIGHT, 2003	Gravity Corer	■	-
Atedaun	Clare	27	R 295 885	IN-SIGHT, 2004	Gravity Corer	●	●
Ballyallia	Clare	28	R 342 809	Chen, 2004	Gravity Corer	●	●
Ballybeg	Clare	27	R 330 739	IN-SIGHT, 2004	Gravity Corer	●	●
Ballycar	Clare	27	R 414 690	Leira, 2002	Echman Grab	-	●
Ballynakill	Galway	31	L 856 225	IN-SIGHT, 2003	Gravity Corer	■	-
Ballyteige	Clare	27	R 348 888	Chen, 2004	Gravity Corer	●	●
Bane	Westmeath	7	N 550 712	IN-SIGHT, 2003	Gravity Corer	■	●
Barfinnihy	Kerry	21	V 850 768	IN-SIGHT, 2003	Gravity Corer	■	-
Barra	Donegal	38	B 935 120	IN-SIGHT, 2003	Gravity Corer	■	-
Beaghcauneen	Galway	32	L 680 472	Chen, 2004	Gravity Corer	●	●
Bray Lower	Wicklow	10	O 137 161	EAIL, 1996-97	Echman Grab	●	-
Bunny	Clare	27	R 375 967	IN-SIGHT, 2003	Gravity Corer	■	-
Caragh	Kerry	22	V 725 905	EAIL, 1996-97	Echman Grab	●	-
Carra	Mayo	30	M 180 710	Leira, 2002	Echman Grab	■	-
Castle	Clare	27	R 486 690	Leira, 2002	Echman Grab	●	●
Caum	Clare	23	R 182 810	Chen, 2004	Gravity Corer	●	●
Cloonaghlin	Kerry	21	V 610 709	IN-SIGHT, 2003	Gravity Corer	■	-
Crans	Tyrone	3	H 711 568	IN-SIGHT, 2004	Gravity Corer	■	●
Cullaun	Clare	27	R 315 905	IN-SIGHT, 2003	Gravity Corer	■	●
Cullaunyheeda	Clare	27	R 464 747	Leira, 2002	Echman Grab	■	●
Dan	Wicklow	10	O 150 040	IN-SIGHT, 2003	Gravity Corer	■	-
Doo	Clare	28	R 120 721	EAIL, 1996-97	Echman Grab	●	-
Doo	Donegal	39	C 359 394	IN-SIGHT, 2003	Gravity Corer	■	-
Dromore	Clare	27	R 346 859	Leira, 2002	Echman Grab	●	●
Drumanure	Clare	28	R 215842	Chen, 2004	Gravity Corer	●	●
Dunglow	Donegal	38	B 782 117	IN-SIGHT, 2003	Gravity Corer	■	-
Easky	Sligo	35	G 442 225	IN-SIGHT, 2003	Gravity Corer	■	-
Effernan	Clare	27	R 222 558	Chen, 2004	Gravity Corer	●	●
Egish	Monaghan	36	H 795 132	IN-SIGHT, 2004	Gravity Corer	■	●
Fad Inishowen	Donegal	40	C 539 439	IN-SIGHT, 2003	Gravity Corer	■	-
Fee	Galway	32	L 790 613	IN-SIGHT, 2003	Gravity Corer	■	-
Feeagh	Mayo	32	F 965 000	IN-SIGHT, 2003	Gravity Corer	■	-
Garvillaun	Clare	28	R 248 829	Chen, 2004	Gravity Corer	●	●
Gortaganniv	Clare	27	R 251 759	Chen, 2004	Gravity Corer	●	●
Inchichronan	Clare	27	R 391853	Leira, 2002	Echman Grab	-	●
Inchiquin	Clare	21	R 268 897	IN-SIGHT, 2004	Gravity Corer	■	●
Keel	Donegal	38	B 847 162	IN-SIGHT, 2003	Gravity Corer	■	-
Kiltooris	Donegal	38	G 676 972	IN-SIGHT, 2003	Gravity Corer	■	●
Kindrum	Donegal	38	C 185 430	IN-SIGHT, 2003	Gravity Corer	■	-
Kylemore	Galway	32	L 770 552	IN-SIGHT, 2003	Gravity Corer	■	-
Lene	Westmeath	7	N 510 685	IN-SIGHT, 2003	Gravity Corer	■	●
Lickeen	Clare	28	R 176 909	EAIL, 1996-97	Echman Grab	●	-
Lisnahan	Clare	28	Q 900 617	Chen, 2004	Gravity Corer	●	●
Maumwee	Galway	30	L 977 484	EAIL, 1996-97	Echman Grab	●	-
McNean	Leitrim	36	H 040 400	IN-SIGHT, 2003	Gravity Corer	■	●
Moanmore	Clare	28	Q 979 611	Chen, 2004	Gravity Corer	●	●
Moher	Mayo	32	L 977 766	EAIL, 1996-97	Echman Grab	●	-
Mooghna	Clare	28	R 137 842	Chen, 2004	Gravity Corer	●	●
Morgans	Clare	27	R 255 835	Chen, 2004	Gravity Corer	●	●
Muckanagh	Clare	27	R 370 925	IN-SIGHT, 2003	Gravity Corer	■	●

Lake Name	County	H-Area	Grid Reference	Sampler, sampling year	Sampling Equipment	Diatom Analysis	Cladocera Analysis
Muckno	Monaghan	6	H 856 175	EAIL, 1996-97	Echman Grab	●	-
Mullagh	Cavan	7	N 677 855	IN-SIGHT, 2004	Gravity Corer	■	●
Nahasleam	Galway	31	L 971 244	IN-SIGHT, 2003	Gravity Corer	■	-
Nambrackkeagh	Galway	32	L 821 603	IN-SIGHT, 2003	Gravity Corer	■	-
Naminn	Donegal	40	C 396 419	IN-SIGHT, 2003	Gravity Corer	■	-
Naminna	Clare	28	R 176 710	IN-SIGHT, 2003	Gravity Corer	■	-
O'Flynn	Roscommon	26	M 585 795	IN-SIGHT, 2003	Gravity Corer	■	-
Oorid	Galway	31	L 930 460	IN-SIGHT, 2003	Gravity Corer	■	-
Oughter	Cavan	36	H 342 075	EAIL, 1996-97	Echman Grab	●	-
Owel	Westmeath	26	N 400 581	EAIL, 1996-97	Echman Grab	●	-
Pollaphuca	Wicklow	9	N 985 086	EAIL, 1996-97	Echman Grab	●	-
Ramor	Cavan	7	N 603 868	EAIL, 1996-97	Echman Grab	●	-
Rea	Galway	29	M 615 155	IN-SIGHT, 2003	Gravity Corer	■	-
Rosconnell	Clare	28	R 222 793	Chen, 2004	Gravity Corer	●	●
Rushaun	Clare	28	R 253 791	Chen, 2004	Gravity Corer	●	●
Shindilla	Galway	31	L 960 460	IN-SIGHT, 2003	Gravity Corer	■	-
Sillan	Monaghan	36	H 700 070	IN-SIGHT, 2004	Gravity Corer	■	●
Talt	Sligo	34	G 398 150	IN-SIGHT, 2003	Gravity Corer	■	-
Tay	Wicklow	10	O 160 750	IN-SIGHT, 2003	Gravity Corer	■	●
Tullabrack	Clare	28	R 018 597	Chen, 2004	Gravity Corer	●	●
Upper	Kerry	22	V 900 817	IN-SIGHT, 2003	Gravity Corer	■	-
Veagh	Donegal	38	C 022 215	IN-SIGHT, 2003	Gravity Corer	■	-

Geographically these 75 lakes are widely located in 12 counties of Ireland but most of the lakes are situated along the west coast, including 28 in Co. Clare, 11 in Co. Galway and 9 in Co. Donegal (see Figure 3.1 and Table 3.1). Geologically the most common bedrock types are Carboniferous Limestone, Granite, Shale, Silurian Quartzite and Sandstone. In Co. Clare Carboniferous Limestone and Shale are the two dominant bedrock types in the catchment area of study lakes. For lakes from Co. Galway the catchment areas are dominated by several bedrock types including Granite, Ordovician and Schist and Gneiss. The Geology of Co. Donegal lakes is dominated by Granite, Schist and Quartzite. Sandstone is predominant in the catchment areas of lakes in Co. Kerry. Silurian Quartzite and Granite are prevalent in the catchment areas of Co. Caven and Co. Wicklow respectively. The 75 lakes are located in 21 of the 40 Hydrometric Areas across the Irish Ecoregion (see Table 3.1). Each Hydrometric area comprises a single river catchment or a group of smaller catchments (Toner *et al.*, 2005). Several Hydrometric Areas are well represented like Area 27 (14 lakes), Area 28 (12 lakes), Area 32 (6 lakes) and Area 38 (6 lakes). While the Hydrometric Areas 3, 6, 9, 18, 34 and 39 only contain one lake each.

Twelve commonly used and potentially significant physico-chemical variables including TP were used in this study, including total phosphorus (TP), Chlorophyll-*a*, pH, alkalinity, conductivity, catchment area, lake area and catchment area: lake area ratio, mean and maximum depth and altitude. All the physical and hydrochemical data of the 75 lakes are listed in Appendex B and D. The hydrochemical and physical data were provided mainly through links with other projects (Irvine *et al.*, 2001; Wemaëre, 2005; Taylor *et al.*, 2006) and the Irish Environmental Protection Agency (EPA). Other publications were also referred to for information used in this study (e.g. Flanagan & Toner, 1975; King & Champ, 2000; Free, 2002; McGarrigle *et al.*, 2002; Toner *et al.*, 2005). Chemical data for 40 lakes were directly provided by the Irish EPA (see Appendix D). Chemical data for the other lakes were mainly collected during the summer season for most of the lakes with a sampling frequency varying from one to nine times per year (see Appendix D).

Land cover data from the CORINE 2000 database were provided by the Irish EPA. CORINE (CO-oRdination of INformation on the Environment programme) Land Cover project uses a unique combination of satellite images and other data to reveal all kinds of information on land resources across Europe. The Irish National CORINE database includes 41 land cover classes based on a 25 ha minimum mapping unit. Land cover data of this study contained most of the 41 land cover classes, and were summarized into five groups (urban, forestry, pasture, peatland and agriculture) for data analysis (see Appendix C). Data for Ballycar, Castle, Rosconnell and Caragh were collected from Wemaëre (2005) and Irvine *et al.* (2001) but were based on the CORINE 1990 dataset.

An Irish lake typology scheme including thirteen typology classes was proposed by the Irish EPA (Free *et al.*, 2005) mainly based on alkalinity, mean depth and lake area (see Table 3.2). All 13 lake types were represented by the 75 study sites with 3-11 lakes in each category. Typology class 13 was identified for lakes only based at altitudes of > 300 m (Free *et al.*, 2005) and Lough Bray Lower with an altitude of 378 m fell into that category. This lake was re-categorised as type 3 in this study to make it more comparable with other lakes due to its low alkalinity, deep waters and relatively small area. The data set contains 31 lakes with low alkalinity (< 20 mg l^{-1} $CaCO_3$), 24 lakes with moderate alkalinity (20-100 mg l^{-1} $CaCO_3$) and 20 lakes with high alkalinity (> 100 mg l^{-1} $CaCO_3$) (see Table 3.2). Lake Type 4 (low alkalinity, deep and large)

contained 11 lakes, more than any other lake type. The lowest number of lakes occur in Type 2 (low alkalinity, shallow and large), Type 6 (moderate alkalinity, shallow and large), Type 10 (high alkalinity, shallow, large) and Type 11 (high alkalinity, deep and small) respectively.

Table 3.2 Classification of the 75 study lakes based on the Irish Lake Typology scheme (Free *et al.*, 2005) (see Appendix A for site names)

Alkalinity	<20 mg l^{-1} $CaCO_3$				20 – 100 mg l^{-1} $CaCO_3$				>100 mg l^{-1} $CaCO_3$			
Mean Depth	≤4 m		>4 m		≤4 m		>4 m		≤4 m		>4 m	
Lake Area	≤50 ha	>50 ha	≤50 ha	>50 ha	≤50 ha	>50 ha	≤50 ha	>50 ha	≤50 ha	>50 ha	≤50 ha	>50 ha
Typology Class	1	2	3	4	5	6	7	8	9	10	11	12
No. of Lakes	9	3	8	11	8	3	5	8	5	4	3	8
Lakes (Site Code)	CAU HOH MAU MOA NAB NAH NAM NAN TUL	BAR DOC EAS	ANS BAF BEA BRL DOO FAD KEE TAY	ARD CAR CLO DAN FEE FEG KYL OOR SHI UPE VEA	DRU GAR GOR KIL LIS MOR MUL RUS	LIC OUG RAM	BAL CRA EFF MOO ROS	DUN EGI KIN MCN MUN OWE POL TAL	ATE BAB BAC BAT CAS	BUN CAA MUC OFL	BAA CUL DRO	ANN BAN CUY INC INQ LEN REA SIL

Water column stratification is generally restricted to the summer months in small and relatively deep lakes in Ireland whereas shallow lakes do not show an annual cycle of thermal stratification (Irvine *et al.*, 2001). Winter stratification is uncommon as Irish lakes are seldom ice covered due to their proximity to the Atlantic Ocean. Of the 31 lakes surveyed in 1996-7 (13 of which are included in this study) almost all were found to stratify during the summer (Irvine *et al*., 2001). Either stable or temporary summer stratification was found in six small lakes in Co. Clare despite the windy and temperate weather conditions (Allott, 1986). The occurrence of summer stratification can influence the oxygen availability and internal loading of nutrients (like phosphorus) in deep waters and therefore affect the growth of plants and animals within lakes.

3.1.2 Sediment Sampling and Sub-sampling

Fifty-eight of the 75 study lakes were sampled in the summer of 2003 and 2004 either by the author or through the joint fieldwork with the IN-SIGHT project (Taylor *et al.*, 2006) (see Table 3.1). Eleven lakes were sampled in 1996-7 by Norman Allott but not analysed by the Ecological Assessment of Irish Lakes (EAIL) project (Irvine *et al*., 2001) and also six lakes were sampled by Manel Leira in 2002.

A Renberg gravity corer (HTH Teknik, Vårvågen 37, SE-95149 Luleå) was used to extract sediments for the 58 lakes in 2003 and 2004 (see Table 3.1). This device can close completely under water so that sediments are not lost during retrieval (Renberg, 1991). Bathymetric surveys were carried out using an echo sounder and a portable GPS when the bathymetry data were not available or not accurate enough to determine the deepest point for sediment coring. Sediment cores were sub-sampled using a vertical extruder immediately after coring in the field. Sediments were sectioned at 0.5 cm interval for the top 5 cm and at 1.0 cm interval for the rest of the depth. The sediment cores from Ballyallia and Anascaul were sub-sampled every 1 cm due to technical problems. Sub-samples were placed in sealed and labelled sampling bags and stored in cool box, and kept out of direct sunlight before being stored in a refrigerator at the Sediment Laboratory of Geography Department, Mary Immaculate College for subsequent analyses. Seventeen samples provided by Norman Allott and Manel Leira (see Table 3.1) were taken using an Echman Grab in the profoundal area of the lakes. Surface sediments of the top 2-3 cm were removed and sealed in screw-top vials and were stored as above.

3.2 Laboratory Analyses

3.2.1 Diatom Analysis

Surface sediments from the top 0.5 cm for 55 lakes, 1 cm for two lakes (Ballyallia and Anascaul) and ca. 2-3 cm for 15 lakes sampled with Echman Grab (see Table 3.1) were processed for diatom analysis. Core bottom samples from seven lakes were also processed.

3.2.1.1 Slide Preparation

Procedures for slide preparation adopted here mainly followed those of Battarbee *et al*. (2001). The key focus during diatom slide preparation is that no diatom valves are lost or damaged. Care was taken to avoid contamination of the samples. The detailed procedures are summarised as follows: oxidation of organic matter of around 0.1 g wet sediment with 5ml 30% hydrogen peroxide (H_2O_2) in the water bath at 80°C for around 4 hours, adding 5-10 drops of 10% hydrochloric acid (HCl) to eliminate the remaining H_2O_2 and carbonates, washing with distilled water and centrifuging at 1200 rpm for 4 minutes for 4 times, drying 5 ml of the slurry on a coverslip, and mounting with Naphrax® (a resin of high refractive index) on a hotplate at 100-150°C for around 10 minutes.

3.2.1.2 Valve Identification and Counting

A Meiji ML2000 microscope with phase contrast and 100× oil immersion objective and 1000× magnification was used for diatom counts. The phase contrast condenser can increase the contrast between the mountant and the diatom cells. A Nikon Coolpix 4500 digital camera attached to the microscope was used to facilitate diatom identification and counting. Images captured were used for taxonomic verification and harmonisation as well as records for unknown taxa. At least 300-500 valves were counted for surface and bottom samples. A continuous transect from the edge to center of the diatom slides were counted. This potentially includes equal portions of slides to avoid any sorting caused by evaporation and is generally thought to be representative of the whole slide (Battarbee, 1986).

Identification of diatoms was mainly based on cell shape and the detailed structure of the silicious wall (termed frustule). Diatom nomenclature and taxonomy mainly followed the series of books by Krammer & Lange-Bertalot (1986, 1988, 1991a, 1991b, 2000), together with a wide range of references (Foged, 1977; Stevenson *et al.*, 1991; Lange-Bertalot & Metzeltin, 1996; Prygiel & Coste, 2000; Håkansson, 2002). An unpublished French diatom list is used for coding the diatom taxa in this study (see appendix E). The single valve is used as the basic counting unit, so the complete frustule was counted as two and chains of frustules as the total number of the individual

valves. When counting fragments of valves, fragments are included in the counts only if no double or multiple counting takes place as suggested by Battarbee (1986), Other strategies included counting the broken valves only if they represent approximately three-quarters of the valves.

Quality control on diatom identification has been fulfilled through cross-verification and discussion within the IN-SIGHT project. Diatom harmonization was conducted by the author and Dr. Leira on the combined diatom training set. Sub-species were merged for several species like *Cocconeis placentula* and *Stephanodiscus hantzschii* (see Figure 3.3) mainly due to their similar ecological requirement for growth (Van Dam *et al.*, 1994). Some species that have different forms were also merged under the same species, including the two common forms of *Aulacoseira subarctica* (Gibson *et al.*, 2003) in this training set (Figure 3.3). Taxa that were difficult to distinguish and also occurred with a relatively low abundance were merged with species with close morphology, including *Cyclotella karmmeri* (merged with *C. kutzingiana*), *Stephanodiscus rotula* (merged with *S. neoastreae*), *Cyclotella praetermissa* (merged with *C. radiosa*) and *Navicula phyllepta* (merged with *N. gregaria*). Synonymous species were also merged, e.g. *Navicula heimansii* (synonymous with *N. leptostriata*), *Achnanthes impexa* (synonymous with *Navicula impexa*). Girdle views of taxa of some genera occurred in relatively high abundance, particularly those of *Fragilaria* and *Eunotia* (see Figure 3.3). Some taxa in girdle view were difficult to identify and were therefore counted as unidentified within the genus.

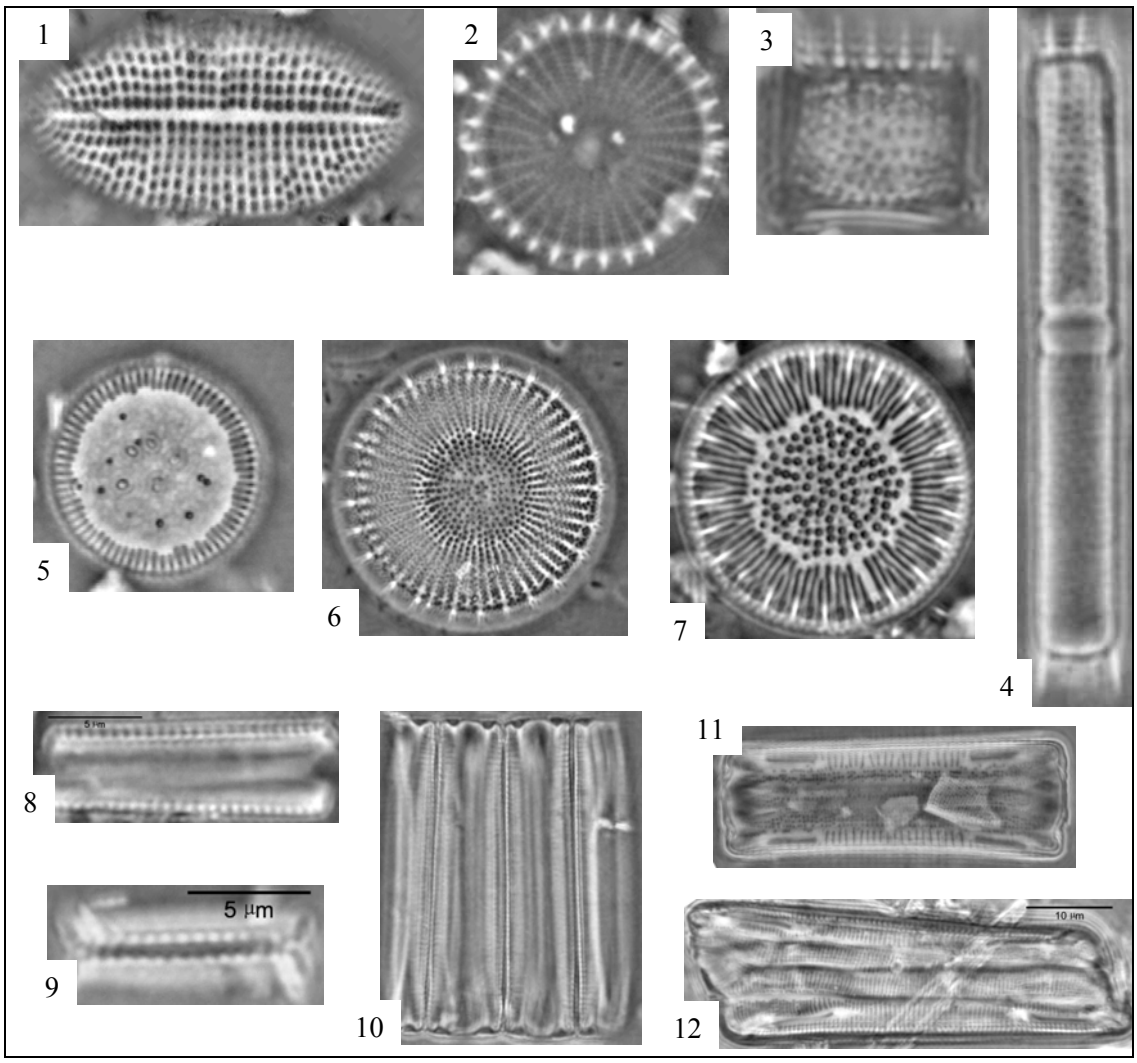

Figure 3.2 Images of some diatom taxa in the 72-lake diatom training set (1 *Cocconeis placentula var. euglypta*, 2 *Stephanodiscus hantzschii*, 3&4 *Aulacoseira subarctica*, 5 *Cyclotella kutzingiana*, 6 *Stephanodiscus neoastreae*, 7 *Cyclotella radiosa*, 8&9 *Fragilaria exigua* (girdle view), 10 *Fragilaria virens* (girdle view), 11&12 *Eunotia incisa* (girdle view)).

3.2.2 Cladocera Analysis

3.2.2.1 Water Samples

For Cladocera taxonomic identification fresh samples were taken from the water column, rocks and plants from around 10 lakes. Samples were washed gently under a tap through a 53µm mesh sieve into a beaker until the rinsing water came through clear. Care was taken to collect all the material rinsed from the mesh and the volume of the final solution did not exceed that of standard solution of around 50 ml. The samples were then preserved in 70-90% industrial alcohol, which prevents the decay of

Cladocera bodies or parts. A pipette with an aperture large enough to accommodate the intact Cladocera exoskeletons was used to absorb 5 ml of solution into a counting chamber for taxa identification and enumeration.

3.2.2.2 Sedimentary Samples

Surface sediments from the top 1 cm from 28 lakes sampled with a gravity corer and the top 2-3 cm from five lakes sampled with an Echman Grab (see Table 3.1) were used for Cladocera analysis. Core bottom samples from seven lakes were also processed.

The sediment digestion method for Cladocera analysis mainly followed that of Frey (1986). Approximately 5 g of wet sediments were deflocculated with 50 ml 10% potassium hydroxide (KOH) in a 250 ml beaker, heated on a hot plate to 65-70 ºC for approximately one hour and stirred gently. Five ml of 10% HCl was added to remove carbonate from calcareous sediments after the contents of the beaker were allowed to cool. After CO_2 bubbles were released, samples were filtered through a sieve with a 53μm mesh size, which would retain small exoskeleton parts. Remains retained on the screen were carefully transferred into vials, diluted to a volume of around 5-10 ml and then kept in the refrigerator for further analysis.

Permanent slides for Cladocera analysis were prepared by transferring 0.05 ml aliquot of the well-shaken concentrate to a glass slide through a precision pipette. The slide was placed on a hotplot at 35-50 ºC until almost all of the liquid was evaporated. Care was taken to prevent complete drying. Two drops of Glycerine jelly were then dripped onto the residue and a pin was used to mix the jelly and the sample residue gently for a homogeneous distribution. Gram's safranine solution was used to stain and mark the fragments on the slides. A 22 mm ×22 mm coverslip was then mounted and nail varnish was employed around the edges to seal the cover slip. The slides were left overnight to dry and then were ready for identification and counting.

3.2.2.3 Identification and Counting

Examination of intact exoskeleton components from water samples greatly helps with the identification of the sedimentary remains. Taxonomy guides of freshwater Cladocera, mostly from Europe, have been referred to (Scourfield & Harding, 1966; Smirnov, 1974; Amoros, 1984; Margaritora, 1985; Pennak, 1989; Duigan, 1992; Alonso, 1996; Smirnov, 1996; Dodson & Frey, 2001). Lists of Cladocera taxa belonging to Chydoridae and other families were used to falicitate sedimentary Cladocera identification and counting (Duigan, 1992; de Eyto, 2000; Irvine *et al.*, 2001).

Identification of the fossil Cladocera remains is complicated by the fact that only fragments of Cladocera bodies are preserved and not all Cladocera are preserved equally well in the sediments (Frey, 1960). Hard-shelled forms of cladocerans, such as chydorids, are well preserved, whereas soft-shelled chitinous taxa, such as *Daphina*, is represented by smaller fragments (Jeppesen *et al.*, 2001). These may also be caused by various factors including fish predation and complex physical and chemical operations on the bodies during the transportation and burial processes (Frey, 1986). Headshields (see Figure 3.4 and Figure 3.5), postabdomens (Figure 3.6) and shells or carapaces (Figure 3.7) are the main identifiable features for sedimentary Chydoridae and Bosminidae, as well as the claws for Daphniidae (Figure 3.6).

There is not a comprehensive taxonomy key to sedimentary Cladocera, but the pioneering works by D.G. Frey (1958, 1959, 1960, 1962a, 1962b, 1964, 1965) significantly improved the identification of sedimentary fragments and enabled Cladocera remains to be a significant proxy in paleoecology and paleolimnology. For example, the arrangement of head pores on the head shield has been an important key for sedimentary Chydoridae and Bosminidae (see Figure 3.4 and Figure 3.5). Works by many other researchers have also greatly helped with sediment Cladocera taxonomy (Goulden & Frey, 1963; Deevey & Deevey, 1971; Lieder, 1983; Melo & Hebert, 1994; Bos, 2001). A Cladocera species code developed by the author was used in this study for data analysis (see Appendix F).

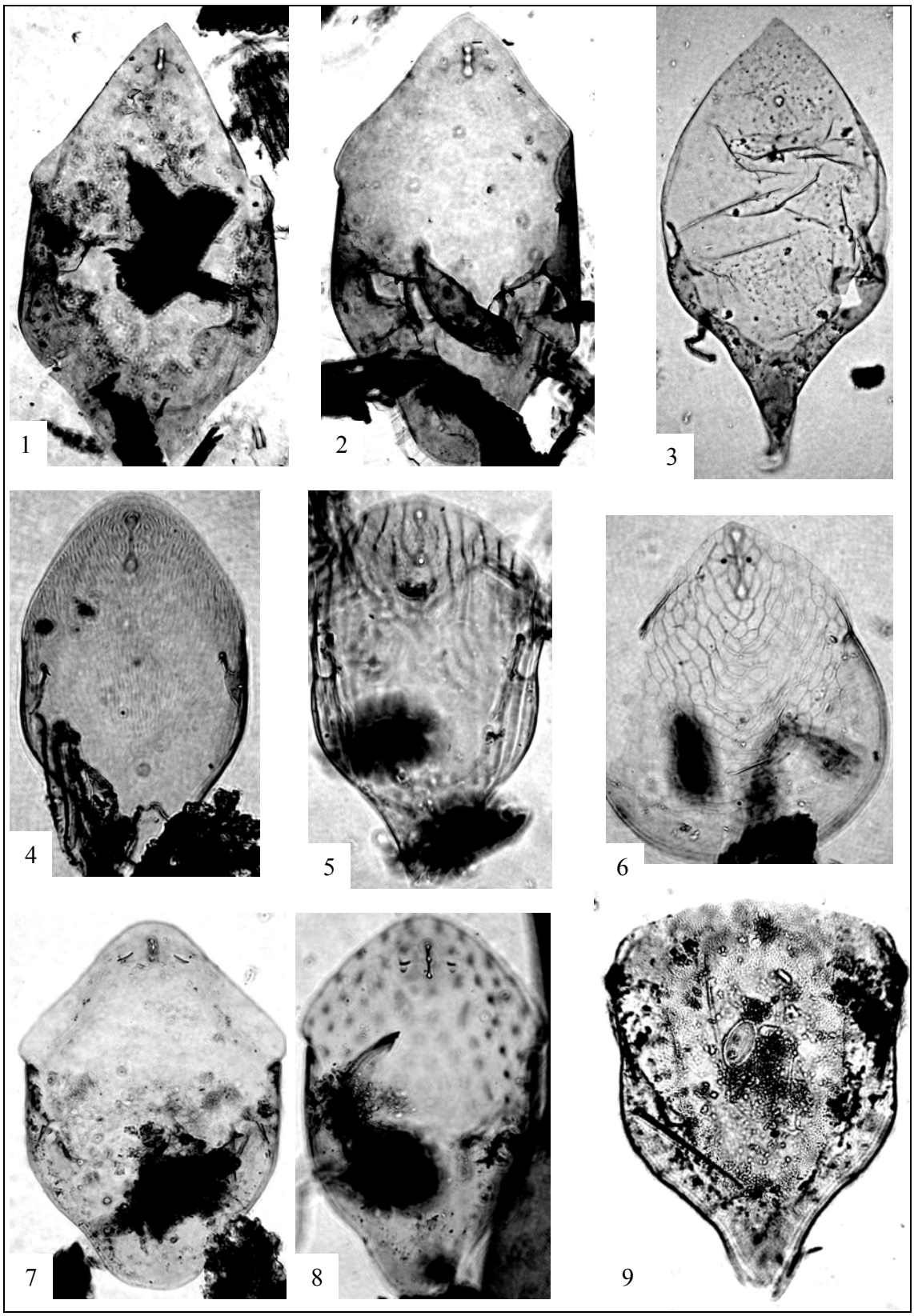

Figure 3.3 Headshields of some chydorids (1 *Alona affinis*, 2 *Alona quadrangularis*, 3 *Pleuroxus trigonellus*, 4 *Alonella excisa*, 5 *Alonella exigua*, 6 *Graptoleberis testudinaria*, 7 *Alona costata*, 8 *Alona rustica*, 9 *Monospilus dispar*).

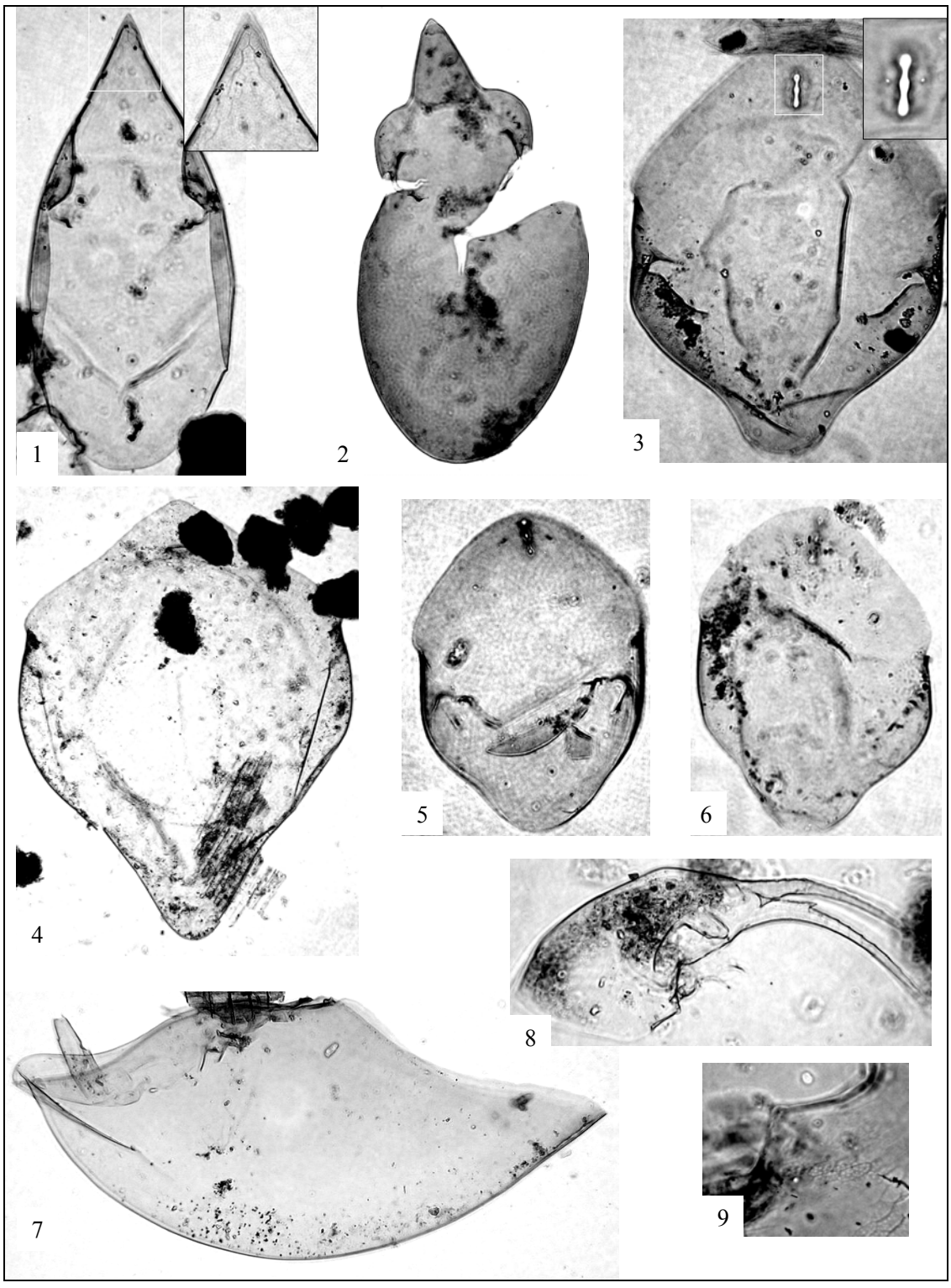

Figure 3.4 Headshields of some chydorids and bosnomids (1 *Chydorus Sphaericus* (tip part enlarged and inserted), 2 *Chydorus piger*, 3 *Leydigia leydigii* (head pores enlarged and inserted), 4 *Eurycercus lamellatus*, 5&6 *Alona guttata/rectangula*, 7 *Acroperus harpae*, 8 *Bosmina longirostris*, 9 *Bosmina longispina*).

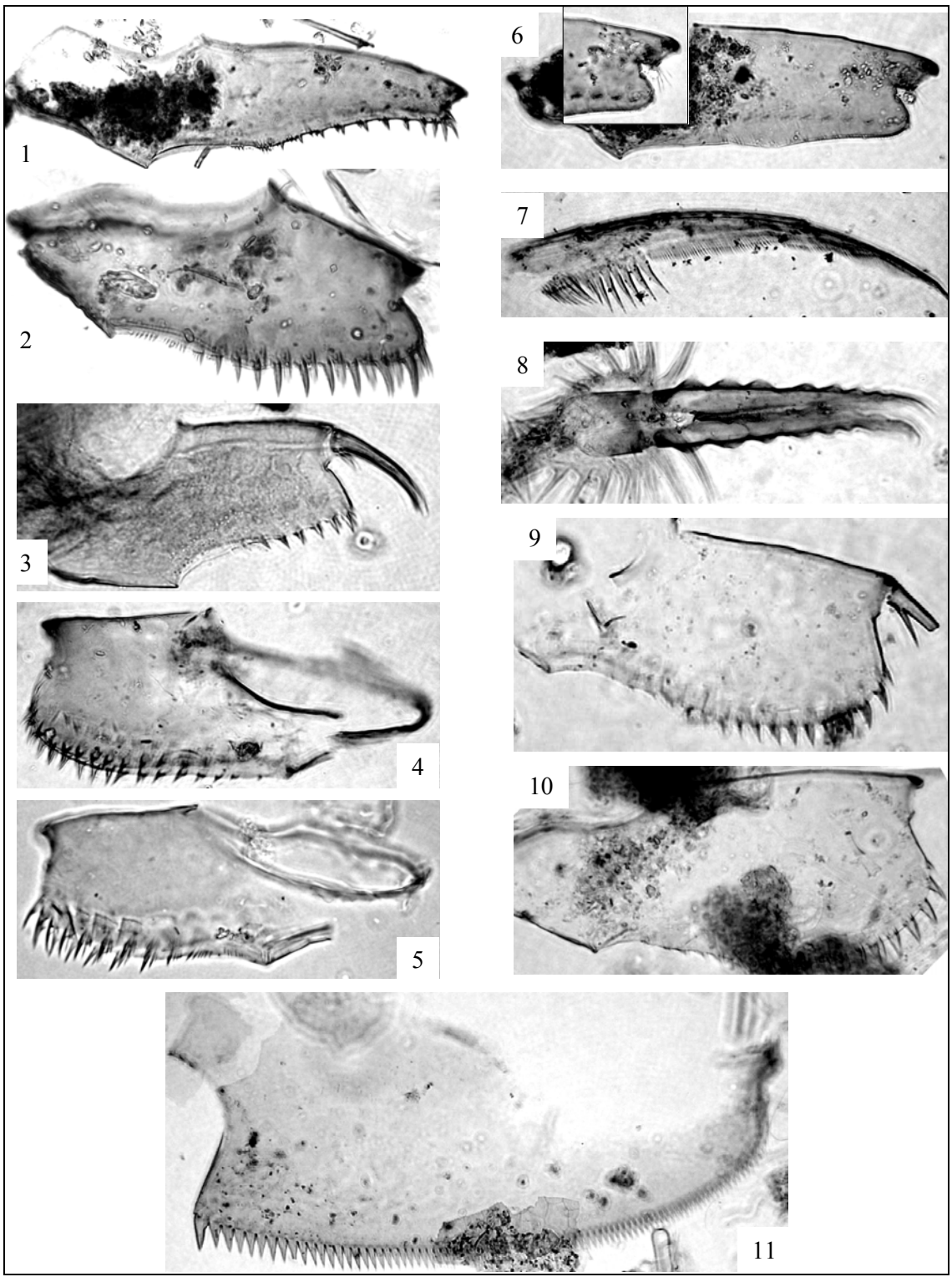

Figure 3.5 Postabdomens and claws of some cladocerans (1 *Pleuroxus laevis*, 2 *Alona rustica*, 3 *Alona guttata*, 4 *Alona affinis* (?), 5 *Alona rectangula*, 6 *Acroperus harpae* (part at the base of the claw enlarged and insersted), 7 *Daphnia pulex* group, 8 *Daphnia longispina* group, 9 *Alona affinis*, 10 *Alona quadrangularis*, 11 *Eurycercus lamellatus*).

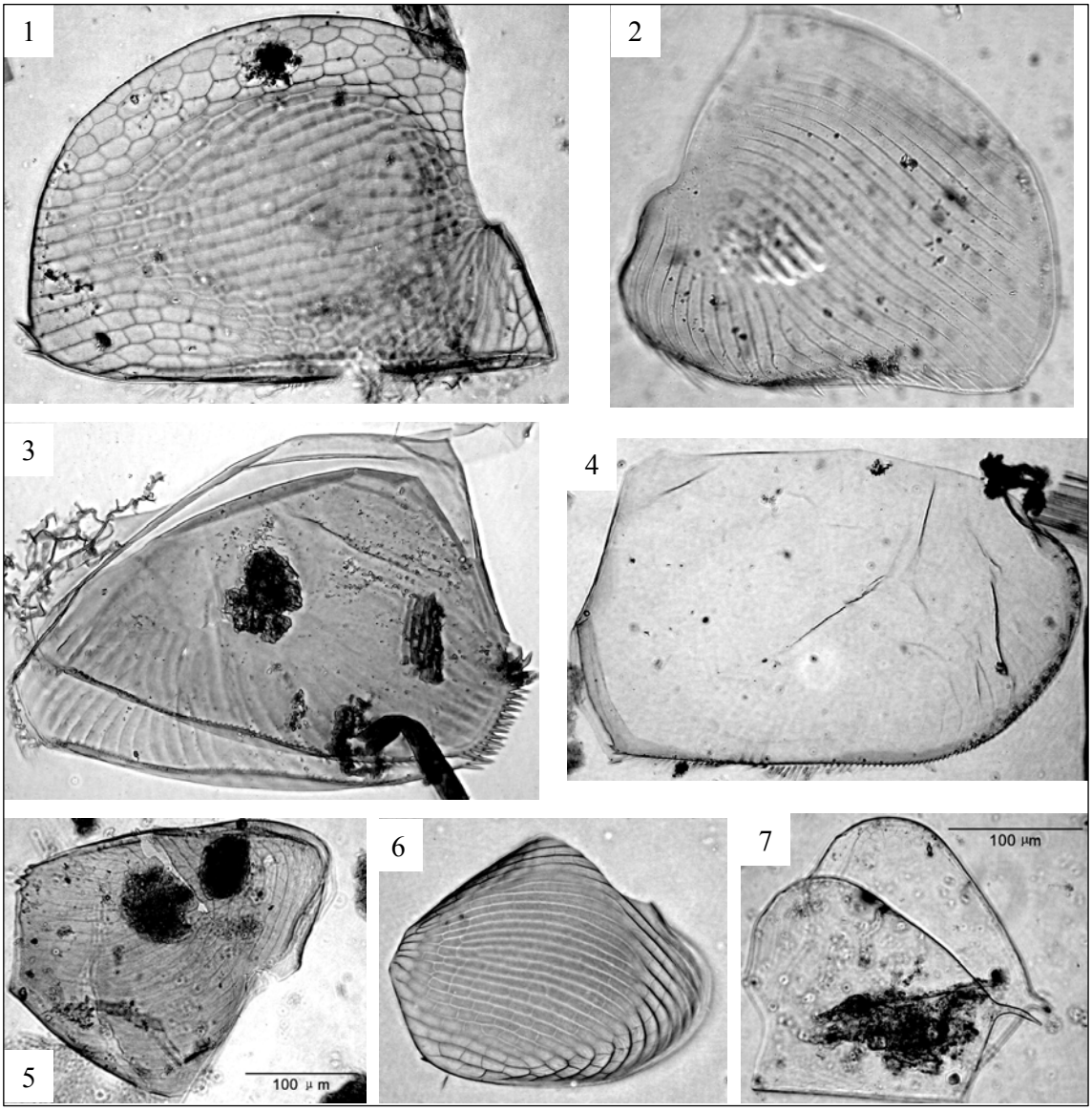

Figure 3.6 Carapaces of some cladocerans (1 *Graptoleberis testudinaria*, 2 *Acroperus harpae*, 3 *Pleuroxus truncatus*, 4 *Pleuroxus aduncus*, 5 *Alonella exigua*, 6 *Alonella nana*, 7 *Bosmina longirostris*).

Counts of remains were adjusted to represent the Cladocera individual with two shells, one headshield and one postabdomen. The most abundant, single remain was used in the analysis for each taxa or species (Frey, 1986). At least 100 individuals of Chydoridae were counted or at least 70 individuals with at least 20 Cladocera slides counted when the chydorid density was very low. Coverslips were enumerated completely in order to avoid any possible bias towards the relative abundance of the remains and to determine the absolute concentrations and biomass (Bredesen *et al.*, 2002).

All Chydoridae and Bosminidae remains were identified to species level (see Figure 3.4 to Figure 3.7), except when insufficient taxonomic feature could be observed. Species

of *Alona guttata* and *A. rctangula* were merged as *A. guttata*/*rectangula* group because their headshields (see 5&6 in Figure 3.5), often the most abundant fragments of both species in this study, were generally indistinguishable despite that their postabdomens (see 3&5 in Figure 3.6) and carapaces could be identified and distinguished. Headshields and carapaces are often the only fragments for the Bosminidae in this study (see 8&9 in Figure 3.5 and 7 in Figure 3.7). Position and shape of lateral pores on the headshield enabled the identification of *Bosmina* at species level (Goulden & Frey, 1963), but the features of carapaces alone could not be used to distinguish between *Bosmina longirostris* and *B. longispina*. However, number of *Bosmina* individuals indicated by headshields was higher than that indicated by related shells in this study and therefore there was no taxonomic problem for enumerating sedimentary *Bosmina* remains. *Daphnia* remains could only be identified to two species groups (*Daphnia* cf. *longispina* and *D.* cf. *pulex*) (Frey, 1958) mainly based on the postabdominal claws (as shown in Figure 3.6).

Ephippia were only found in surface sediments from Loughs Mullagh (1, the total number of ephippium), Ballyteige (1), Crans (1), Egish (21) and in the bottom sediment from Lough Crans (3). Therefore ephippia were not included in the subsequent data analysis due to their extremely low frequency of occurrence and abundance.

3.3 Numerical Analyses

The simplest data structure can be expressed as a single variable of an object (e.g. the maximum depth of lakes, count of one diatom species), often called univariate data. Multivariate data consist of many objects, with each object described by several variables. In this study the ecological data consists of percentage counts for the two fossil variables (diatom and Cladocera) in multiple lakes (objects). Ecological and environmental data were stored mainly in MS Excel spreadsheets as well as MS Access tables where the relational data sets were linked. Both diatom and cladoceran counts were expressed as percentage data of the total counts for the same sample respectively.

3.3.1 Exploratory Data Analysis (EDA)

For univariate data, the histogram was used in this study for displaying the distributional information. Quantile-quantile (Q-Q) plots are useful tools for determining if data are normally distributed (Venables & Ripley, 2002). They show the relationship between the distribution of a variable and a theoretical distribution. Quantile-Quantile (Q-Q) plots were employed in this study to test the normal probability of environmental data (but not displayed). For bivariate and multivariate data, the scatter plot was used to explore the data patterns. A scatter plot reveals relationships or association among variables. Such relationships manifest themselves by any non-random structure in the plot. They can provide information about the relationship and dependence among variables and any potential outliers. Scatter plots are a useful diagnostic tool for determining association and correlation. They can provide the key features of a set of data and an overall impression of the data distribution.

3.3.2 Multivariate Analysis

Any analysis that attempts to simultaneously examine the behaviour of more than one dependent variable is termed as multivariate analysis. Ordination and classification, the most commonly used multivariate methods, are applied in this study.

3.3.2.1 Missing Data and Normalizing Transformation of Data

Missing data is a feature of many physico-chemical data sets, including the current data set. Fourteen values of six physical variables and three values of two hydrochemical variables are missing (see Appendex B and D). Removing variables with missing values could eliminate critical physico-chemical factors determining the biological assemblages and valuable information can be lost. Instead the mean of the variable was substituted for the missing value prior to data transformation and multivariate analysis (Legendre & Legendre, 1998).

If the data have a highly skewed distribution e.g. with many small values and a few extremely large values, data transformation is recommended to improve the analysis (ter Braak, 1987b). Normalizing transformation of data for subsequent multivariate analysis

was employed to reduce the influence of extreme values and also to fit better to a normal distribution (also known as the Gaussian distribution) for further multivariate analysis. Square root and logarithm with bases of 10, natural and 2 were used to test the normality of each environment variable. Addition of 1 to the value equal to or below 0 was also used for square root and log- transformations. Transformation of each variable was judged and selected based on the Q-Q plot and frequency histogram analyses. Square root transformation was selected for ecological data in this study prior to multivariate analysis if not specified (Lepš & Šmilauer, 2003).

3.3.2.2 Gradient Analysis

Gradient analysis is the examination of species distribution along environmental gradients. Gradient analysis includes indirect gradient analysis (ordination), where community samples are displayed along axes of variation in composition that can subsequently be interpreted in terms of latent, not measured, environmental gradients and direct gradient analysis (constrained ordination), in which each species' abundance is described as a function of measured environmental variables (ter Braak & Prentice, 1988).

Models of species response to environmental gradients determine the techniques used for gradient analysis (see Table 3.3). Two types of species response models to an environmental gradient are used: linear and unimodal response models. The linear response model is the simplest approximation, and the unimodal model expects that the species have an optimum on an environmental gradient (Lepš & Šmilauer, 2003). Detrended Correspondence Analysis (DCA) is used to ascertain if the species data are suited to a linear or unimodal model (ter Braak & Prentice, 1988). DCA provided an estimate of the gradient length in relation to the underlying environmental gradients reported in standard deviation (SD) units for each axis (ter Braak & Prentice, 1988). If the gradient length is short (2 SD units or less) (see Table 3.3), taxa are generally behaving monotonically along the environment gradient and therefore linear methods are appropriate (Birks, 1995). If the gradient length is longer (2 SD or more), unimodal-based methods would be appropriate.

Table 3.3 Summary of basic ordination techniques used in this study.

Response model	Linear	Unimodal
Gradient length (SD)	≤ 2	> 2
Unconstrained	Principal Components Analysis (PCA)	Correspondence Analysis (CA)
Constrained	Redundancy Analysis (RDA)	Canonical Correspondence Analysis (CCA)

Linear response model

Species responses may seem to change linearly through short environmental gradients, so a linear response model may be a reasonable basis for exploring quantitative abundance data spanning a narrow range of environment variation. Linear methods (PCA and RDA) are generally used to explore the structure of the environmental data and are also used for the Cladocera data in this study.

Principal components analysis (PCA) transforms the original variables into a smaller set of linear combinations that account for most of the variance of the original data (Dillon & Goldstein, 1984). It is an indirect or unconstrained ordination technique that constructs the theoretical variable that minimizes the total residual sum of squares after fitting straight lines to the species data (ter Braak, 1987b). The purpose of PCA is to explain as much of the total variation in the data as possible with as few components as possible. The principal components are extracted so that the first principal component accounts for the largest amount of the total variation in the data. Total sum of squares of regressions is called eigenvalue (λ, a value for each eigenvector) and it is expressed as percentage of total variance in species data. The eigenvalue is actually equal to the maximized dispersion of the species scores on the ordination axis and is thus a measure of importance of the ordination axis (ter Braak, 1987b). The first of first PCA axis has the largest eigenvalue (λ_1), the second axis the second largest eigenvalue (λ_2), and so on. Broken stick model was used to examine which ordination axes should be retained for interpretation based on the percentage variance explained by each axis (Jackson, 1993) and this was performed in the program R.

In PCA without data transformation (also known as non-standardised or species-centred), each species is implicitly weighted by the variance of its abundance values and the abundant species with largest variances are emphasized within the covariance matrix. However, the species with low variance, often the rare ones, have only minor influence on the solution (ter Braak, 1987b). In standardised PCA all species receive equal weight including rare species in a correlation matrix. Standardisation of data is recommended when variables are measured in different and non-comparable units and the shape of the transformed data is left unchanged (Noy-Meir *et al.*, 1975; Shaw, 2003). Environmental data were therefore standardised before performing PCA as they were generally measured in different units in this study.

PCA is commonly displayed in a biplot where sites marked by points and species by arrows are represented jointly (Gabriel, 1971). Sites close together are inferred to resemble one another in species composition, which are tacitly assumed to have similar environments. Points near the origin have low magnitude of change from the average (the origin). Therefore, species on the edge of the diagram (far from the origin) are the most important for indicating site differences and species near the centre are of minor importance (ter Braak & Prentice, 1988). The arrow points in the direction of maximum variation in the species' abundance, and its length is proportional to this maximum rate of change (ter Braak & Prentice, 1988). Angles between vector arrows approximate their correlations: an angle of 0˚ means a completely positive correlation while an angle of 180˚ indicates a completely negative correlation.

Redundancy Analysis (RDA) is a constrained PCA method in which species are presumed to have linear relationships to environmental gradients. RDA selects linear combination of environmental variables that gives smallest total residual sum of squares. The results of RDA can be expressed in a triplot or biplot, where sites are indicated by points, and both species and environmental variables are indicated by arrows whose interpretation is similar to that of the arrows in the PCA biplot. The species scores in RDA are most accurately represented by arrows (the direction in which the species is increasing in abundance).

Unimodal response model

Correspondence Analysis (CA) is the technique that constructs the theoretical variable that best explains the species data using weighted averaging (ter Braak, 1987b). It is also known as Reciprocal Averaging, which means that sample scores are calculated as weighted averages of species scores, and species scores are calculated as a weighted average of sample scores. Species and samples are ordinated simultaneously in CA. The theoretical variable constructed by CA is termed the first CA axis and its values are the site scores on the first CA axis. A second and further CA axes are also constructed for maximizing the dispersion of the species scores, but the size of eigenvalues can be a guide for the number of axes worth interpreting. As in PCA broken stick model was used to decide which CA axes should be used for interpretation (Jackson, 1993).

Correspondence Analysis has a fault as the arch effect and this effect is caused by nonlinearity of species response curves (ter Braak, 1987b). Detrended Correspondence Analysis (DCA) was developed by Hill and Gauch (1980) as a heuristic modification and more sophisticated form of CA, and designed to correct the arch effect. In CA both sites and species are represented by points, and each site is located at the center of gravity of species that occur there (ter Braak, 1987b). Species points on the edge of the diagram are often rare species, either because they prefer extreme environmental conditions or because their few occurrences by chance happen to be at extreme sites.

Canonical Correspondence Analysis (CCA) is a modification of CA in which the ordination axes are restricted to be weighted sums of environmental variables. CCA aims to visualize a pattern of community variation, as in standard ordination, and also the main features of species' distributions along environmental variables. The species and sites are positioned as points in CCA and their joint interpretation is the same as that in CA. A site with a high abundance of a species tends to be close to the point for that species. Since species are assumed to have unimodal responses with respect to linear combinations of the environmental variables, the species are logically represented by points (corresponding to their approximate optima), and the environmental variables by arrows indicating their direction and rate of change (ter Braak & Prentice, 1988). The joint plot of species points and environmental arrows is actually a biplot that approximates the weighted averages of each of the species with respect to each of the environmental variables. The weighted averages are approximated as deviations from the mean of each environmental variable represented by the origin of the plot. The inferred weighted average is higher than the average if the projecting point lies on the

same side of the origin as the head of the arrow and is lower then average if the origin lies between the projecting point and the head of and arrow.

3.3.2.3 Minimum Adequate Model Selection

A significant feature of all statistical models is that they have two main parts: the systematic component, that describes the variability in the values of response variables explained by the explanatory variables, and the stochastic component, that refers to the residual variability in the response variables not described by the systematic part of the model and contributes to the noise in the data set (Lepš & Šmilauer, 2003). The systematic component is often used to judge the quality of the fitted model based on the amount of variance in response variables it can explain. The principal of parsimony is generally followed to include only those explanatory variables that significantly contribute to the model (Legendre & Legendre, 1998). However, strong collinearity among explanatory variables can affect the correct estimation of model parameters (Legendre & Legendre, 1998). To best describe the ordination models in this study, forward selection is used to select significant environmental variables and is run using functions of R package vegan (Oksanen, 2005b). This procedure uses a fairly new concept Akaike's information criterion (AIC) as the selection criterion (Godínez-Domínguez & Freire, 2003), rather than the maximum extra fit used in CANOCO (ter Braak & Šmilauer, 2002).

Forward selection, a type of stepwise selections, starts with a null model without any explanation variable (Legendre & Legendre, 1998). In automatic selection procedure a single variable that explains the largest amount of variability in the species data independently is selected (Lepš & Šmilauer, 2003). However, automatic selection can be misleading when several highly correlated variables are almost equally good as explanatory variables but small changes in the data can change the selection results (Oksanen, 2004; Birks, 2005a). Alternately an environmental variable (TP in this study) with a priori ecological interest can be manually selected in the forward selection as the starting variable. This variable is then used as a covariable and other variables are selected in the order of their additional effects (variance explained in addition to the selected variable(s)). Monte Carlo permutation tests are used to test the significance of

each variable and only those variables with $P \leq 0.05$ under 999 permutations are accepted (Lepš & Šmilauer, 2003; Birks, 2005a).

A major problem with forward selection is that all variables selected in previous steps are included in the model though some of them may contribute little to the multivariate regression model (Legendre & Legendre, 1998). The elimination of a variable does not mean that it is not ecological important and therefore the final selected model must be based both on ecological and statistical criteria (Oksanen, 2004; Birks, 2005a). Therefore the minimal adequate model selected by forward selection is only one possible model.

3.3.2.4 Cluster Analysis

Cluster analysis is used to identify homogeneous groups or clusters of objects (species, samples) (Lepš & Šmilauer, 2003). Hierarchical classification means that groups are nested within other groups. An agglomerative hierarchical method starts with small groups of few samples, and progressively groups them into larger and larger clusters, until the entire data set is sampled. Ward's minimum variance method is one type of hierarchical agglomerative clustering analysis. At the beginning of the procedure each object is in a cluster of its own and at each clustering step this method finds the pair of objects whose fusion increases the sum of squared distances between objects and cluster centroids as little as possible over all objects (Legendre & Legendre, 1998). In comparison with other hierarchical agglomerative methods like average linkage, complete linkage and single linkage methods, the minimum variance method is usually proven to most useful though it tends to produce clusters of fairly equal size (Birks, 2005a). Ward's minimum variance clustering method was used for classifying both the diatom and Cladocera data and selection of clusters were mainly determined by the homogeneity and separation of the clusters (van Tongeren, 1995). The resulting clusters were compared with the Irish EPA Lake Typology classes in this study. Transformed environmental and ecological data were employed for cluster analysis in this study.

3.3.2.5 Species Response Curves

There are many types of ecological response curves and a Gaussian unimodal model with symmetric curves is generally more appropriate as an ecological response curve than a monotonic one (ter Braak & Looman, 1987). As a type of GLM, Gaussian Logit Regression (GLR) is preferred as a practical method for summarizing species distributions along environmental gradients under a Gaussian-like species response curve (ter Braak & Looman, 1986). GLR is usually applied to presence-absence data, but it can also be used for proportional data as quasi-likelihood model (McCullagh & Nelder, 1989). Such GLR modelling for species with relative abundances and a fixed maximum abundance value have been applied in several diatom training sets used for modelling the diatom response to environmental gradients (ter Braak & van Dam, 1989; Birks *et al.*, 1990) and also other ecological studies (Odland *et al.*, 1995). GLR was also applied to model the response curves of Cladocera species in several Cladocera training sets (Davidson, 2005; Simpson, 2005a), but only using presence-absence data because of a high frequency of zero counts and also many rare species. In the current Cladocera training set proportional data was used as most species have relatively high number of occurrence with a relatively low frequency of zero counts. Therefore in this study GLR with a logit link function and a binomial distribution of response variables was used to model the response of diatom and Cladocera species with relative abundance data along the environmental gradients. The ecological and environmental data were not transformed for normalization or standardization in the GLR modelling analyses of this study.

3.3.3 Calibration Analysis

In comparison with regression analysis which is used to explore the response of species to environmental variables, calibration analysis expresses values of environmental variable as a function species data (ter Braak, 1987a). The regression function so constructed is the transfer function that is used for prediction of environmental variables based on species data. Calibration analysis considers only one environmental variable in each transfer function (ter Braak, 1995). A typical transfer function normally includes many taxa (e.g., 50-300) and 50-300 samples (Birks, 2005b). Ecological data are usually quantitative and expressed as relative abundance (%) of total sample count. While environmental data rarely contain zero values and often follow a log-normal distribution (Birks, 2005b).

The length of the first axis of Detrended CCA (DCCA) constrained by only one environmental variable of interest generally indicates the degree of species turnover, with a length of above 2 suggesting unimodal-based methods are appropriate and a length of less than 2 suggesting linear-based methods (Birks, 1995). Two powerful non-linear methods, weighted averaging (WA) and weighted averaging partial least squares (WA-PLS), and one linear method, partial least squares (PLS), are applied in this study. WA is also employed to calculate the TP optima of both diatom and Cladocera taxa. In many cases there may be no single optimal model and it is advised to use several models for reconstruction and to reach a consensus reconstruction based on different inference models (Birks, 1998). The use of multiple inference models for TP reconstruction was adopted in this study (see Chapter 7).

3.3.3.1 Weighted Averaging (WA)

When a species shows a unimodal response curve along an environment gradient, this species would most frequently and abundantly occur around the optimum of the curve. Weighted Averaging (WA) regression analysis estimates the species optima by taking the average of the values of the environmental variables over those sites where this species occurs (ter Braak & Looman, 1987). For calibration analysis WA is used to estimate the value of the environmental variable by taking a weighted average with weighting proportional to species abundance (ter Braak & Prentice, 1988). The species' tolerance (ecological amplitude) can be estimated as the weighted standard deviation of the environmental variable. Species with a narrow tolerance can be given greater weight than species with a wide tolerance if required (Birks *et al*., 1990). In WA the number of species in a site is not very important and a better indicator is the effective number, the Hill's N2 measure of diversity (ter Braak & Juggins, 1993).

In WA environmental reconstructions, averages are taken twice (in WA regression and calibration respectively) and this causes the shrinkage of the range of inferred values towards the mean of the environmental variable (Birks *et al.*, 1990; Birks, 1995). Two methods, inverse and classical deshrinking regressions, are available to correct this shrinkage effect (Birks *et al.*, 1990; ter Braak & Juggins, 1993; Birks, 1995). In inverse regression the observed values of environmental variable are regressed on the initial

inferred values. This method minimises the root mean squared error (RMSE) in the training set but is liable to produce bias at both high and low ends of the environment gradient (ter Braak & Juggins, 1993). For classical regression the observed values of an environmental variable, instead of the initial inferred values, are used for regression in the training set. Therefore classical regression deshrinks more than inverse regression and it uses the inferred values further away from the mean of the training set (Birks *et al*., 1990). For WA regression and calibration the choice of inverse or classical regression depends on the scale of environment gradient to be reconstructed. If higher accuracy is needed at high or low end of the gradient classical deshrinking is appropriate and if the middle part of the gradient is of interest the inverse method is a better choice (Birks, 1995).

WA has become a powerful method in palaeolimnology and palaeoecology because it can provide quantitative prediction and also prediction error (Birks, 2005b). It does not assume linear species-environment relationship and therefore is not hindered by multicollinearity between variables or by large number of taxa in training sets and is relatively less sensitive to outliers (ter Braak & Juggins, 1993; Birks, 2005b). WA functions best with noisy, species rich, compositional data, with species absent in many samples and a long ecological gradient (ter Braak & Juggins, 1993). Furthermore, WA performs well in no-analogue situations with relatively realistic inferences as long as there are reliable optima estimates for the fossil taxa of high numerical importance (Birks, 2005b). However, WA disregards species absence and therefore the reliability of its estimate depends on the distribution of the environmental variable in the samples (Braak & Looman, 1986). WA also disregards residual correlations in species data, namely correlations that remain after fitting the environmental variable of interest and that are often caused by environmental variables not considered in WA model (ter Braak & Juggins, 1993). WA can be modified by incorporating partial least squares (PLS) in the method WA-PLS, where residual correlations are taken into account to improve the optima estimates (ter Braak & Juggins, 1993).

3.3.3.2 Partial Least Square (PLS) and Weighted Averaging Partial Least Square (WA-PLS)

Partial least squares (PLS) is a linear-based modelling technique developed first in chemometrics (Birks, 1995). In PLS regression the original independent variables are replaced by a few orthogonalized components through employing the covariance matrix between the dependent variables and the independent variables (Nilsson *et al.*, 1996). A main feature of PLS is that components are selected not to maximize the variance of each component of environment variables but to maximize covariance between components that are linear combinations of the variables within environment and ecological data (Birks, 2005b). The optimal number of components that yields the model with highest prediction properties and lowest prediction errors is determined by cross-validation (ter Braak & Juggins, 1993; Nilsson *et al.*, 1996).

In the unimodal method, Weighted Averaging Partial Least Square (WA-PLS), components are chosen to maximize the covariance between the vector of weighted averages of environmental and ecological variables. Subsequent components are selected to implement the same task but are restricted to be orthogonal and therefore uncorrelated to earlier components (ter Braak *et al.*, 1993). It was found that WA is a form of PLS regression on transformed data when only the first component was used. The improvement in WA-PLS over WA lies in the use of further orthogonal components to improve predictive powers (ter Braak & Juggins, 1993). The additional components use the residual structure in species data to optimize the species parameters in the transfer function. In WA-PLS with each additional component the model fits the environmental variable better as measured by the root mean square error (RMSE) but with less predictability as the RMSE is not corrected for degrees of freedom (Birks, 1995). The estimated optimum component is often the one which gives the lowest prediction error (RMSEP, as discussed in the next subsection) in the training set tested using cross-validation. In WA-based models there are problems of edge effect that result in inevitable overestimation of optima at the low end of the environment gradient and underestimation at the high end of the gradient (ter Braak & Juggins, 1993). WA-PLS implicitly implements a weighted inverse deshrinking regression to pull the inferred values towards the mean of the training set. Therefore WA-PLS usually produce models with lower prediction errors and bias than WA (Birks, 1995).

3.3.3.3 Evaluation and Validation of Transfer Functions

Root mean square error (RMSE), coefficient of determination (r^2) and the bias between estimated and observed values are commonly calculated as measures of the performance of the inference model (Smol, 2002). These measures of the strength of the relationship between observed and inferred values allow comparison between transfer functions for different environmental variables. However, measures like the coefficient of determination (often referred to as apparent r^2) are generally an overly optimistic estimate of model performance because of their dual roles in both estimating the species parameters and testing the model performance (ter Braak & Juggins, 1993). A large training set can help to estimate the model performance as independent test but they are generally unavailable in reality. An alternate way is to employ computer-intensive cross-validation to evaluate the predictability and prediction errors of a training set.

The simplest cross-validation approach used in this study is jack-knifing (also known as leave-one-out) (Birks, 1995). A new subset of the original data set is created and each time one sample is left out in the new subset as an independent test sample for the transfer function based on the new subset. This sample is then returned in the subset and another sample is selected for running another independent test. This test is run repeatedly until all the samples in the original training sets have been used for independent tests. A second approach for cross-validation is bootstrapping which a more complex error estimation technique (Birks *et al.*, 1990). Like jack-knifing bootstrapping extracts a subset from the original training set each time, but it replaces those removed samples with samples from the subset to keep the size of the subset the same as that of the original training set (Birks *et al.*, 1990; Birks, 1995). This estimate is less prone to bias because it includes the full size of the training set. Some samples are selected twice because of replacement sampling and this procedure is run many times (1000 used in this study). Bootstrapping can also be used to estimate sample-specific RMSEP of fossil samples in reconstruction (Birks, 1998). The coefficient of determination from jack-knifing and bootstrapping are also termed as r^2_{jack} and r^2_{boot} respectively.

A model with a low prediction error in cross-validation (RMSEP), as well as low mean or maximum bias is preferred in model selection. These measures of model performance are independent of the range of observed environmental variables, unlike the coefficient of correlation (r^2), and they can be more directly interpretable than correlations (Gasse *et al.*, 1995). Coefficient of correlation (r^2) is a measure of relationship between the observed and predicted values of environmental variables and it is mainly used for

comparing different inference models (Birks, 1998). However, model selection often involves a compromise between the models with lower RMSEP and lower bias, mainly depending on the use of model in reconstruction (Birks, 1998). When the model is used for reconstruction near the end of the environment gradient, it may be appropriate to tolerate a high maximum bias at the other end of the gradient to obtain a low maximum bias for the predicted values of the environmental variables. In some cases it can be more appropriate to apply several models for reconstruction and to obtain concordant inferred values for fossil samples (Birks, 2005b; Lotter, 2005).

A potentially vigorous but rarely used method to validate inference models is to compare reconstruction data based on two or more biological indicators (Birks, 2005b). As two independent biological indicators (diatoms and Cladocera) are included in this study, the comparisons of diatom-based and Cladocera-based transfer functions could provide a strong and unique insight into their model performances as a form of multi-proxy approach (Ihaka & Gentleman, 1996). In addition inference models with bootstrapping were applied for TP reconstruction in this study as sample-specific errors could only be provided by bootstrapping (see Chapter 7).

3.3.4 Software and Packages

Most of the exploratory and multivariate analyses in this study were performed using the R program (version 2.2.1), a language and environment for statistical computing and graphics (Becker *et al.*, 1988; Fox, 2002; Venables & Ripley, 2002; Venables *et al.*, 2005). R is an integrated suite of software facilities for data manipulation, calculation and graphical display, and is available as open-source and free software (*http://www.r-project.org/*). It provides a wide variety of statistical and graphical techniques (Oksanen, 2004; Simpson, 2005b). It can be used in multivariate analysis of ecological and environmental data (Fox, 2002; Oksanen, 2005a). The basic functions provided by the standard packages in R are mainly used for the explanatory and cluster analyses in this study including the broken stick model test (Oksanen, 2005b, 2005a). Package vegan is employed in ordination analyses of both diatom and Cladocera training sets (including PCA/RDA, CA/DCA/CCA) and also for performing forward selection to help select significant environmental variables for both training sets in this study (Wood, 2006). Package mgcv is applied in the species response curve analysis (GLR) (Juggins).

C^2 (version 1.4.2), a software for ecological and paleoecological data analysis and visualization developed by Juggins (2003), is mainly used here for developing calibration models, TP reconstruction and plotting stratigraphic diagrams. It is also used for transforming and summarising the ecological and environmental data together with R.

Chapter 4: Relationships between Surface Sediment Diatom and Environment Variables of 72 Irish Lakes

This chapter will first illustrate the structure and distribution of related environmental and diatom data in a 72-lake diatom training set using numerical methods detailed in Chapter 3. Relationships between both diatoms and environmental variables are then explored using direct gradient analysis and species response curves. Classification of the 72 lakes based on the diatom assemblages is performed and compared with the Irish EPA lake typology classes.

4.1 Environment Variables

Seventy-two lakes across the Irish Ecoregion are included in the diatom training set and have been summarily described in Chapter 3. Twelve physico-chemical variables and five land cover variables are used in the data analysis of the diatom training set (see Appendex B and D). Summary statistics and frequency histograms of the 17 environmental variables are shown in Table 4.1 and Figure 4.1 respectively.

4.1.1 Explanatory Data Analysis

The majority of the 72 lakes in the diatom training set are located at altitudes of <100 m (see Figure 4.1). Most lakes have area of <200 ha with a median value of 50 ha and mean value of 177.3 ha (see Figure 4.1 and Table 4.1). Catchment areas of most lakes were smaller than 20,000 ha and their ratio to lake areas are generally smaller than 100:1. Most of the training set lakes have maximum depth between 5-25 m but they have a wide range of maximum depth between 1.1-45.7 m (see Figure 4.1). The mean depth of most lakes is below 10 m with mean and median values of 6.4 m and 5.5 m respectively.

Half of the training set lakes are surrounded by forests in the catchment, but most forest-covered lakes have forestry coverage of less than 10% (Figure 4.1 and Appendix C). Pasture land is much more common and over half of the lakes have pasture coverage of 22.4%. Other agricultural lands, including the permanent irrigated and non-irrigated

arable lands, are not common. Only half of the lakes have coverage of above 2.5% of other agriculture lands, with a mean average value of 7.9% for the whole training set (Table 4.1). Peat bogs occur frequently and half of the lakes had peat coverage of above 33.8% (Table 4.1). Urban land is quite uncommon in the catchments of the 72 lakes and most of the 72 lakes had very limited or no urban coverage (see Table 4.1 and Figure 4.1).

Table 4.1 Summary statistics of 17 environmental variables of 72 lakes.

Variables	Min	Max	Mean	Median	Standard deviation	N With data	N missing	N Non-zero
Physical								
Altitude (m)	7.0	378.0	79.9	57.0	72.7	70	2	70
Catchment Area (ha)	30.0	147874.0	6361.4	965.7	18628.1	69	3	69
Lake Area (ha)	1.2	1973.9	177.3	50.0	349.9	71	1	71
Catchment Area:Lake Area	2.8	2045.2	77.0	17.8	263.3	69	3	69
Maximum Depth (m)	1.1	45.7	16.3	14.0	10.2	70	2	70
Mean Depth (m)	0.7	19.8	6.4	5.5	4.4	70	2	70
Land Cover								
Agriculture (%)	0.0	84.7	7.9	2.5	14.3	72	0	39
Forestry (%)	0.0	43.9	5.7	0.8	9.2	72	0	38
Pasture (%)	0.0	100.0	33.3	22.4	34.7	72	0	47
Peat (%)	0.0	100.0	40.6	33.8	38.1	72	0	55
Urban (%)	0.0	10.9	0.6	0.0	1.8	72	0	15
Hydrochemical								
Alkalinity (mg l^{-1})	-1.0	208.6	55.2	23.0	58.9	72	0	69
Chlorophyll a (µg l^{-1})	0.4	62.7	8.7	3.7	12.2	72	0	72
Colour (mg l^{-1} PtCo/Hazen)	1.0	208.5	47.0	34.5	39.2	70	2	70
Conductivity (µS cm^{-1})	33.0	462.0	183.1	164.5	113.5	71	1	71
pH	5.1	8.5	7.3	7.4	0.9	72	0	72
TP (µg l^{-1})	0.0	142.3	25.9	10.0	31.3	72	0	71

Thirty-one lakes are categorised as low alkalinity lakes (<20 mg l^{-1} $CaCO_3$), 23 lakes as moderate alkalinity ones (20-100 mg l^{-1} $CaCO_3$) and the other 18 lakes as highly alkaline on the basis of the Irish Lake Typology scheme (e.g., Free *et al.*, 2005) (see Appendix D for hydrochemical data). Lake conductivity measurements range from 33 to 462 µS cm^{-1} (mean = 181.8 µS cm^{-1}, median = 157 µS cm^{-1}) (Table 4.1). More than 50 lakes have chlorophyll-*a* values of < 10 µg l^{-1} and also low colour values of <60 mg l^{-1} PtCo/Hazen (Figure 4.1 and Appendix D). The hydrochemical variable pH shows a wide range from 5.1 to 8.5 (mean = 7.3, median = 7.4). More than half of the 72 lakes in the diatom training set have a pH of above 7 (Table 4.1). The 72 lakes have a mean TP value of 25.9 µg l^{-1} and a median TP value of 10.0 µg l^{-1}. Seventy lakes had TP values

of <100 µg l^{-1} and the other two had TP values of 141 and 142.3 µg l^{-1} (see Appendix D). Based on the OECD classification scheme (OECD, 1982) (see Chapter 2 for details), 36 of the 70 lakes therefore are oligotrophic with 19 and 15 lakes categorised as mesotrophic and eutrophic respectively.

Frequency distribution plots of all six physical variables and five land cover variables display strong skewness to right with long tails (see Figure 4.1). Hydrochemical variables TP and chlorophyll-*a* are also strongly skewed to right with long tails. Conductivity, alkalinity and colour display log-normal distributions in the frequency histogram (see Figure 4.1).

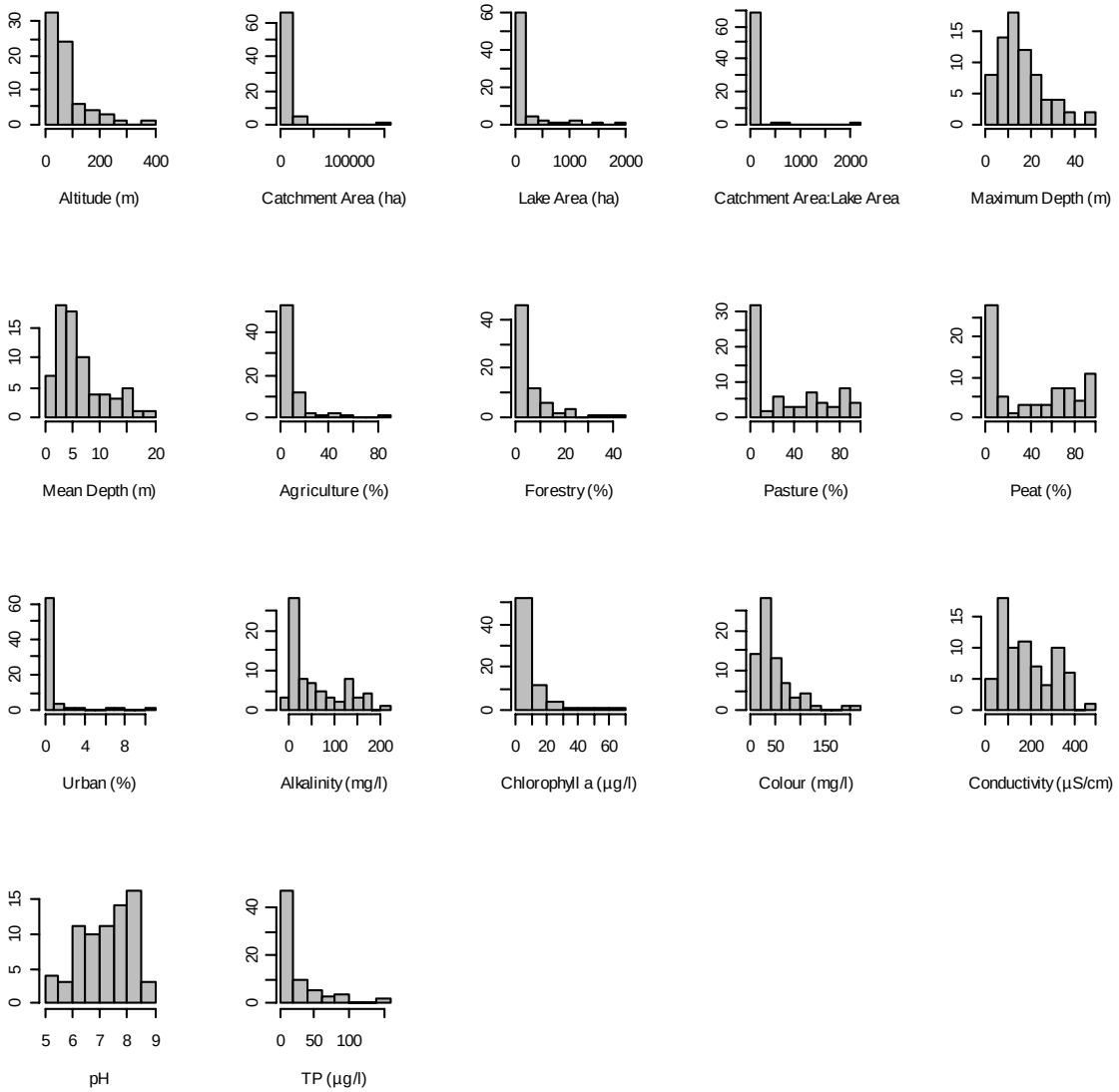

Figure 4.1 Frequency histograms of 17 environmental variables of 72 lakes (the frequency class set by fault in the statistical programme R).

All the twelve lake types categorised in the Irish lake typology scheme are covered in the 72-lake diatom training set, each containing 3-11 lakes (see Chapter 3 for details of the Irish lake typology scheme). Lake Type 4 (low alkalinity, deep and large) contains 11 lakes, the most populated class among the 12 lake types (see Appendex B and D for full physico-chemical data). Lake Types 2 (low alkalinity, shallow and small), 6 (moderate alkalinity, shallow and large) and 11 (high alkalinity, deep and small) contain only three lakes each.

4.1.2 Ordination of Environmental Data

Principal Component Analysis (PCA) was applied to explore the key physical, hydrochemical and land cover factors simultaneously in the training set lakes (see Chapter 3 on details of methods). In PCA all the environmental data are examined and can be best explained by a few latent axes based on a linear response model. Normalising transformation of environmental data was performed to reduce the influence of extreme data. Most physical data was $\log_{10}$-transformed while maximum depth and land cover variables were square-root transformed. Because of the occurrence of many zero values, $\log_{10}(1+)$ transformation was used for hydrochemical data except for pH. The normalised data were then scaled to zero mean and unit standard deviation for PCA analysis with sites unscaled for a correlation biplot to highlight the relationships between the environmental variables.

Two main gradients were identified along the first two axes of PCA (see Figure 4.2). Axis one with an eigenvalue (λ_1) of 0.338, explained most of the variance of the environmental data and accounted for 33.8% of the total variance, while the second axis captured 19.5% ($\lambda_2 = 0.195$) of the total variance (see Table 4.2). Cumulatively the first two axes explained 53.3% of the total variance and the first four axes explained 72.2% of the total variance in the environmental data (Table 4.2). The broken stick model indicated that only the first and second axes were significant components for further investigation (Jackson, 1993). The first axis is highly correlated with several closely clustered chemical variables (conductivity, alkalinity, pH and TP) and land cover variables (peat and pasture) (see Figure 4.2). Biplot scores for these variables were among the highest along the first principal component axis (see Table 4.2). Physical variables (catchment area, lake area and maximum depth) and colour were important on

the second axis. TP was correlated with both PCA axes 1 and 2, explaining variances in the data set along first two PCA axes of variation.

Angles between the environmental variables (vectors) approximate their correlations in PCA biplot and several groups of environmental variables are closely correlated (Figure 4.2). Obvious in the PCA biplot is the closely clustered chemical variables of pH, alkalinity and conductivity along the first principal component axis. This acidity and conductivity gradient is negatively correlated with peat land coverage, indicating that higher coverage of peat lands would produce more acidic and conductive lake waters. The very small angle between TP and chlorophyll-*a* in PCA plot indicates their close positive correlation. Scatter plots of both variables both in raw and log-transformed data are shown in Figure 4.3 where a strong linear relationship is evident between TP and chlorophyll-*a*. However, several sites were shown as outliers in Figure 4.3 (a), including Rosconnell [ROS], Ramor [RAM], Sillan [SIL] and Morgans [MOR].

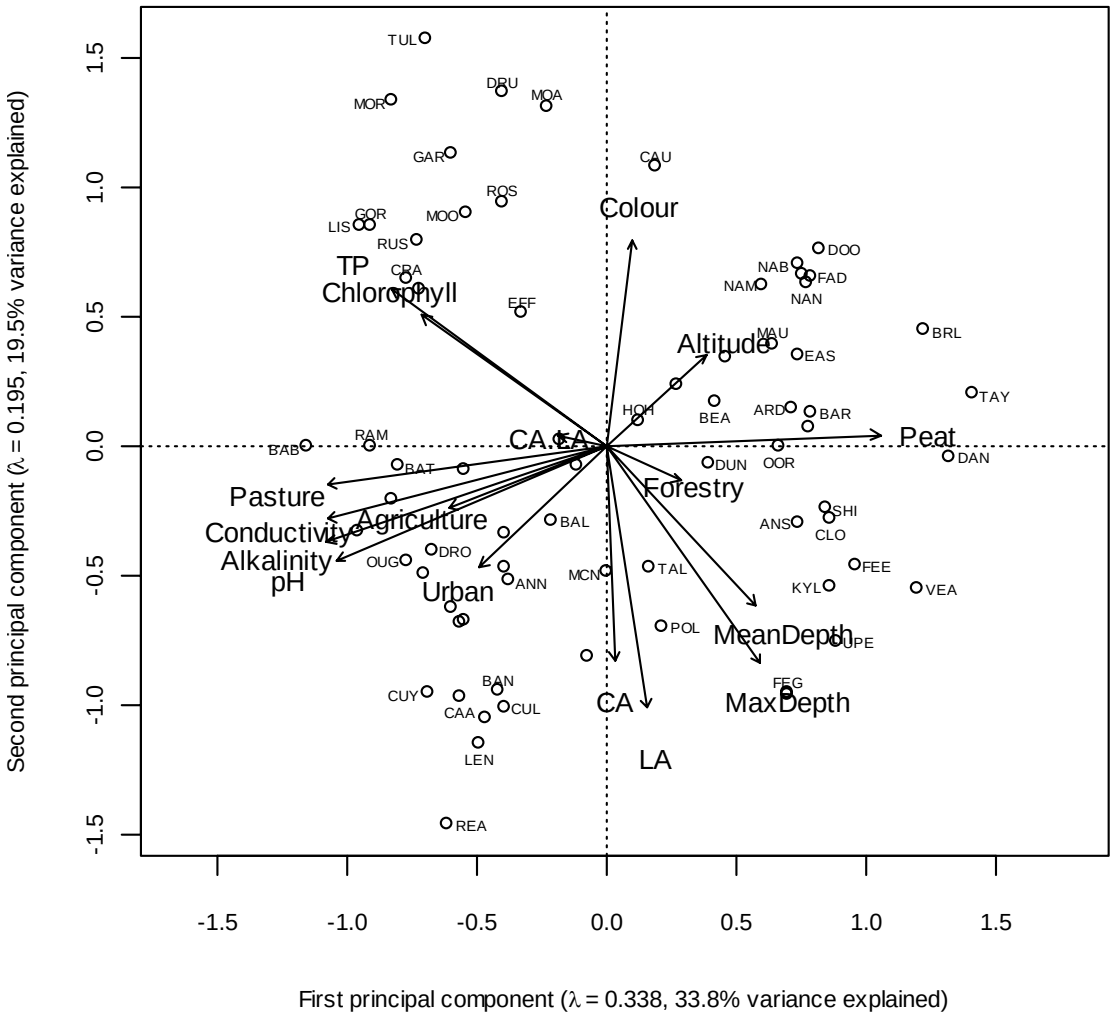

Figure 4.2 PCA Correlation biplot of 17 environmental variables at 72 lakes.

Table 4.2 Summary statistics of PCA of 17 environmental variables at 72 lakes.

PCA Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.338	0.195	0.109	0.079	1
Variance (%)	33.8	19.5	10.9	7.9	
Cumulative Variance (%)	33.8	53.3	64.2	72.2	
Total Unconstrained Eigenvalue					1
Biplot Scores for Environmental Variables					
Altitude	0.449	0.420	-0.632	-0.636	
Catchment Area (CA)	0.040	-0.974	0.907	-0.067	
Lake Area (LA)	0.206	-1.189	0.192	-0.273	
CA:LA	-0.221	0.042	1.153	0.245	
Maximum Depth	0.699	-0.982	-0.029	-0.353	
Mean Depth	0.680	-0.719	-0.294	-0.498	
Agriculture	-0.718	-0.277	0.037	0.182	
Forestry	0.334	-0.149	0.541	-0.589	
Pasture	-1.269	-0.176	-0.218	-0.260	
Peat	1.237	0.060	0.273	0.192	
Urban	-0.576	-0.546	-0.035	-0.651	
Alkalinity	-1.270	-0.436	-0.008	0.210	
Chlorophyll	-0.839	0.593	0.157	-0.668	
Colour	0.115	0.941	0.750	-0.382	
Conductivity	-1.267	-0.332	-0.033	0.263	
pH	-1.228	-0.518	-0.149	0.015	
TP	-0.981	0.717	0.300	-0.480	

Sites with projection points lying along the vector of a variable have higher than average values of the variable (the origin) and vice versa in the PCA bioplot (ter Braak, 1987b). In Figure 4.2 Loughs Tay [TAY, site code used in PCA plot], Dan [DAN] and Bray Lower [BRL] are characterised as being very acidic with extremely low conductivity and alkalinity values and with very high peatland coverage. Loughs Tullabrack [TUL], Morgans [MOR], Gortaganniv [GOR] and Lisnahan [LIS] have high values of colour, TP and chlorophyll-*a* as they lie near the end of these vectors. Sites on the lower bottom and right of the biplot (Figure 4.2), like Veagh [VEA] and Fee [FEE], are deep lakes with low TP values and high forestry coverage. Sites located at the edges of the biplot, like Tullabrack and Rea [REA], indicate big differences from other sites near the origin.

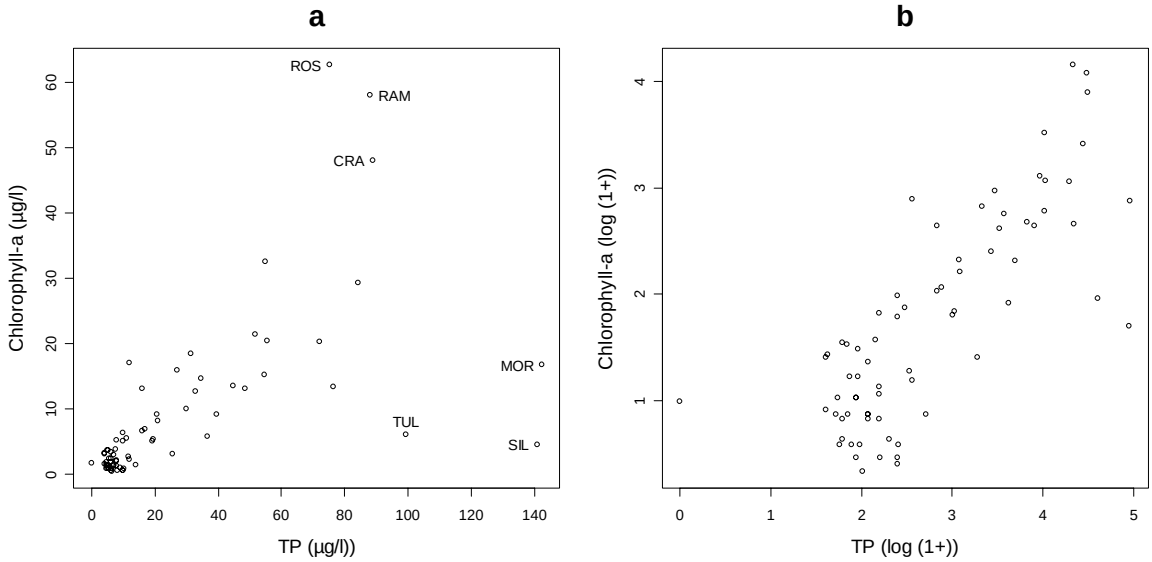

Figure 4.3 Scatter plots of TP and Chlorophyll-*a* with raw data in plot (a) and with data log(1+)-transformed in plot (b) (outlier sites are labelled with site code in plot (a)).

4.2 Surface Sediment Diatoms

In total 602 diatom taxa were counted in the surface sediments of 72 lakes. Of the total 602 taxa, 233 species occurred in least in 3 sites with a maximum abundance not less than 1%. To reduce the influence of rare species, only these 233 diatom taxa were used for multivariate analysis and model construction. Authorities of diatom taxa used in this study are provided in Appendix E.

Table 4.3 lists 79 common diatom taxa with a maximum relative abundance of above 5% and the basic statistics of these taxa are summarised. *Achnanthes minutissima* var. *minutissima* is the most common species and it occurred in 70 of the 72 lakes. It has the highest Hill's Number (N2 = 37.9), a measure of the effective number of occurrence in the samples (Hill, 1973). *Tabellaria flocculosa*, *Cyclotella radiosa*, *Cocconeis placentula*, *Fragilaria exigua*, *F.construens* f. *venter* and *Asterionella formosa* are also very common species occurring in more than half of the 72 lakes. They generally also have high N2, e.g. above 10 (see Table 4.3). Other common species with an occurrence in at least 30 lakes are *Eunotia incisa*, *Fragilaria brevistriata*, *F. pinnata*, *F. capucina var gracilis*, *Stephanodiscus parvus*, *Cymbella gracilis*, *Gomphonema parvulum* var. *parvulum*, *Navicula radiosa*, *Amphora pediculus*, *Aulacoseira subarctica* and *Cyclotella pseudostelligera* (see Table 4.3). Taxa like *Aulacoseira granulata* var. *angustissima*,

Mastogloia elliptica and *Stephanodiscus tenuis*, occurred in only three lakes. These taxa with very low frequency of occurrence also have low effective numbers, e.g. < 2 (see Table 4.3).

Table 4.3 Summary statistics of 79 common diatom taxa (≥ 5% at 3 sites) from surface sediments of 72 lakes (N =number of occurrence; N2= Hill's effective number of occurrence; max and mean refer to the maximum and mean values of relative abundance (%); SD= standard deviation)

Taxon Name	Code	N	N2	Max	Mean	SD
Achnanthes clevei	KCLE	11	4.8	5.2	0.2	0.7
Achnanthes helvetica	PHEL	13	2.2	34.3	0.7	4.2
Achnanthes laterostrata	KLAT	12	2.2	12.5	0.3	1.5
Achnanthes minutissima var. *minutissima*	ADMI	70	37.9	30.1	6.8	6.5
Achnanthes pusilla	ACNP	27	9.5	5.8	0.3	0.9
Achnanthes scotica	ADCA	21	10.5	6.2	0.4	0.9
Achnanthes sp cf saccula	PSAC	10	3.4	5.6	0.2	0.7
Achnanthes subatomoides	PSAT	20	8.2	6.8	0.4	1.0
Amphora inaeriensis	AINA	17	9.3	6.4	0.4	1.1
Amphora pediculus	APED	31	13.1	24.0	2.1	4.5
Anomoeoneis neoexilis	BEXI	27	17.5	6.2	0.8	1.5
Anomoeoneis vitrea	BVIT	10	2.2	15.3	0.3	1.8
Asterionella formosa	AFOR	41	17.0	37.7	4.6	8.2
Asterionella ralfsii	ARAL	11	3.1	64.7	1.8	8.5
Aulacoseira ambigua	AAMB	21	8.1	22.0	1.3	3.7
Aulacoseira granulata	AUGR	12	6.0	24.6	1.2	4.1
Aulacoseira granulata var. *angustissima*	AUGA	3	2.0	40.8	0.9	5.5
Aulacoseira islandica var. *islandica*	AUIS	12	5.4	18.4	0.8	2.7
Aulacoseira subarctica	AUSU	31	15.6	67.0	6.9	13.1
Cocconeis neothumensis	CNTH	8	3.2	10.4	0.3	1.3
Cocconeis placentula	CPLA	43	11.6	26.0	1.9	4.4
Cyclostephanos dubius	CDUB	18	6.0	19.9	1.0	3.2
Cyclostephanos invisitatus	CINV	11	4.6	9.3	0.4	1.3
Cyclotella atomus var. *gracilis*	CAGR	3	1.9	6.7	0.1	0.8
Cyclotella comensis	CCMS	27	13.6	35.7	3.3	6.9
Cyclotella comta	PUCO	12	3.0	16.8	0.5	2.2
Cyclotella distinguenda	CDTG	9	2.0	11.0	0.2	1.3
Cyclotella gordonensis	CGOR	6	2.1	28.4	0.6	3.5
Cyclotella kuetzingiana	CKRM	20	9.0	24.0	1.8	4.7
Cyclotella kuetzingiana cf striata	CSTR	3	2.1	7.3	0.2	1.0
Cyclotella meneghiniana	CMEN	19	5.7	10.7	0.4	1.4
Cyclotella ocellata	COCE	7	2.7	7.0	0.2	0.9
Cyclotella pseudostelligera	CPST	30	4.8	54.6	1.8	6.7
Cyclotella radiosa	PRAD	44	22.9	11.3	1.9	2.7
Cymbella gracilis	ENNG	33	15.8	9.6	0.9	1.8
Cymbella helvetica	CHEL	6	1.8	8.6	0.2	1.0
Cymbella laevis var. *capitata*	CLAE	6	4.7	5.7	0.3	1.0
Cymbella microcephala	ENCM	27	16.8	7.3	0.9	1.7
Denticula tenuis	DTEN	12	6.5	6.1	0.3	1.0
Eunotia incisa	EINC	36	14.6	11.5	1.3	2.5
Eunotia pectinalis var. *undulata*	EPUN	16	8.1	7.8	0.5	1.5
Fragilaria brevistriata	PSBR	35	12.2	13.7	1.3	2.9
Fragilaria brevistriata var. *binodis*	PBBI	6	1.7	6.2	0.1	0.7
Fragilaria capucina	FCAP	22	10.0	8.8	0.7	1.7
Fragilaria capucina var. *gracilis*	FGRA	33	15.8	9.8	1.0	1.8

Taxon Name	Code	N	N2	Max	Mean	SD
Fragilaria capucina var. *rumpens*	FCRP	10	4.9	6.2	0.2	0.9
Fragilaria construens f. *venter*	SCVE	41	8.1	60.9	2.8	7.9
Fragilaria crotonensis	FCRO	6	2.1	15.9	0.3	1.9
Fragilaria delicatissima	FDEL	4	1.2	6.8	0.1	0.8
Fragilaria elliptica	SELI	14	4.8	6.8	0.3	1.0
Fragilaria exigua	SEXG	43	17.8	25.8	3.0	5.2
Fragilaria leptostauron var. *martyi*	SMAT	4	2.2	8.4	0.2	1.1
Fragilaria nanana	FNAN	15	4.1	9.6	0.3	1.2
Fragilaria parasitica f. *subconstricta*	SDSU	3	1.2	6.0	0.1	0.7
Fragilaria pinnata	SPIN	35	13.8	14.2	1.0	2.1
Fragilaria ulna var. *angustissima*	FUAN	5	2.3	22.7	0.5	2.8
Fragilaria virescens	FVIR	7	1.7	23.1	0.4	2.8
Frustulia rhomboides	FRHO	29	12.7	8.3	0.7	1.5
Frustulia saxonica	FSAX	21	12.9	6.7	0.7	1.5
Gomphonema lateripunctatum	GLAT	16	9.3	10.6	0.8	2.0
Gomphonema parvulum var. *exilissimum*	GEXL	17	7.4	7.1	0.3	1.0
Gomphonema parvulum var. *parvulum*	GPAR	33	19.3	5.1	0.7	1.1
Gomphonema pumilum	GPUM	21	8.0	10.7	0.5	1.4
Mastogloia elliptica	MELL	3	1.2	9.8	0.1	1.2
Mastogloia lacustris	MLAC	8	5.0	13.3	0.8	2.8
Meridion circulare	MCIR	5	1.3	8.3	0.1	1.0
Navicula leptostriata	NLST	5	2.1	6.6	0.1	0.8
Navicula pseudoconstruens	FPCO	16	8.6	5.1	0.3	0.9
Navicula radiosa	NRAD	32	8.3	9.2	0.4	1.2
Nitzschia perminuta	NIPM	18	7.3	5.4	0.3	0.8
Peronia fibula	PFIB	12	4.1	6.4	0.3	1.1
Pinnularia subcapitata	PSCA	10	3.2	11.2	0.4	1.8
Stephanodiscus hantzschii	SHAN	19	8.2	7.1	0.5	1.3
Stephanodiscus minutulus	STMI	14	4.1	20.8	0.7	2.9
Stephanodiscus neoastreae	SNEO	13	4.8	13.1	0.5	2.0
Stephanodiscus parvus	SPAV	34	11.9	33.3	2.8	6.2
Stephanodiscus tenuis	SHTE	3	1.3	14.0	0.2	1.6
Tabellaria flocculosa	TFLO	48	21.8	25.4	2.9	4.4
Tabellaria quadriseptata	TQUA	6	1.8	11.3	0.2	1.3

Among the 11 species with maximum relative abundance of above 30%, most are centric diatoms like *Aulacoseira subarctica* (67.0%), *A. granulata* var. *angustissima* (40.8%), *Cyclotella pseudostelligera* (54.6%) and *C. comensis* (35.7%). Non-centric taxa include *Asterionella ralfsii* (64.7%), *Fragilaria construens* f. *venter* (60.9%) and *Asterionella ncise* (37.7%) (see Table 4.3). Taxa with low maximum relative abundance of less than 6% are all ncise diatoms among the 79 common diatom species and they have a varying frequency of occurrence between 6 and 33, like *Cymbella laevis* var. *capitata* (6, frequency of occurrence), *Achnanthes clevei* (11), *A pusilla* (27) and *Gomphonema parvulum* var. *parvulum* (33).

The distribution of abundant diatom taxa with Hill's N2 ≥ 5 and maximum relative abundance ≥ 5% are shown in Figure 4.4. The most common taxon, *Achnanthes minutissima* var. *minutissima*, is dominant at Upper [UPE], Anascaul [ANS], Gortaganniv [GOR], Moher [HOH] and Atedaun [ATE] in abundance of above 20% (see Figure 4.4). *A. minutissima* is often observed to prefer circumneutral and oligo- to eutrophic waters (Van Dam *et al.*, 1994). Other abundant *Achnanthes* species like *A. pusilla* and *A. scotica* are much less common or dominant than *A. minutissima* in this training set. The relative abundance of *Asterionella formosa* is above 20% in four lakes-Effernan [EFF], Mooghna [MOO], Pollaphuca [POL] and Morgans [MOR] (see Figure 4.4). *A. formosa* only occurs suspended in the open water of a lake with a narrow ecological amplitude (Round, 1981). This species was observed to be dominant and subdominant in mesotrophic and eutrophic lakes but not in oligotrophic and acid lakes in the Northwest of the former USSR (Trifonova, 1987). It is a major indicator species of long-term changes associated with increased nutrient inputs in four English lakes (Talling & Heaney, 1988).

Four *Aulacoseira* taxa are prevalent among the 72 lakes and *A. subarctica* is generally more common and dominant then the other three taxa (*A. ambigua*, *A. granulata* and *A. islandica* var. *islandica*). *A. subarctica* is a typically meroplanktonic diatom, which grows as part of the plankton but also spends an important part of its life in the lake bottom (Gibson *et al.*, 2003). It is considered to be a diatom of mesotrophic lakes and has been found to be common in surface sediments of meso-eutrophic lakes with contrasting depth (Anderson, 1997a; Gibson *et al.*, 2003). An increase in its population was believed to be part of the eutrophication pattern in English lakes (Lund, 1954, 1979). *A. subarctica* has a relative abundance of above 60% in Loughs Muckno [MUN] and also very high abundances in Caragh [CAR], Feeagh [FEG], Rushaun [RUS], Oughter [OUG] and Doo [DOC] (see Figure 4.4). All the other three *Aulacoseira* taxa mainly occur in lakes toward high end of the TP gradient as shown in Figure 4.4. *Cocconeis placentula* also occurred in lakes with high TP values, particularly in sites like Atedaun [ATE], Ballyallia [BAA], Ballyteige [BAT] and Gortaganniv [GOR] in abundances of above 10%. This species has its preference in eutrophic and alkaline lakes (Van Dam *et al*., 1994). Most of the *Cyclotella* taxa were abundant mainly in lakes with low TP values, like *C. comensis*, *C. kuetzingiana* and *C. radiosa*. *Cyclotella* is a common centric genus in freshwaters and live as planktonic organisms (Round *et al.*, 1990). This genus is generally noticed to be prominent in oligotrophic lakes (Reynolds,

1984). *C. comensis* is abundant at sites like Fad Inishowen [FAD], Owel [OWE] and Shindilla [SHI] and *C.kuetzingiana* at Veagh [VEA] and Shindilla [SHI] in abundances of above 20%.

Eunotia incisa is common at Dan [DAN] and Barra [BAR] in abundances of around 10% and *E. pectinalis* var. *undulate* is prominent at Dan [DAN] and Arderry [ARD] in abundance of between 5-10% (see Figure 4.4). *Eunotia* is a common genus in acidic freshwater habitats along with *Tabellaria* (Round, 1981; Lange-Bertalot & Metzeltin, 1996). Epiphytic *T. flocculosa* is more dominant at Doo [DOC], Dan [DAN] and Lickeen [LIC] in abundances of above 15%. Six *Fragilaria* taxa are abundant in this 72-lake diatom training set. Epiphytic *Fragilaria* are common in circumneutral and alkaline conditions but can have differential preference of trophic status of freshwaters (Van Dam *et al*., 1994): *F. exigua* has a preference for oligotrophic lakes, *F. capucina* var. *gracilis* for oligo-mesotrophic lakes, *F. capucina* for mesotrophic lakes, *F. contruens* var. *venter* prefer meso-eutrophic waters, while *F. brevistriata* and *F. pinnata* can tolerate a wide ecological amplitude from oligotrophic to eutrophic lakes. *F. exigua* has relative abundances of above 20% at Dunglow [DUN] and Easky [EAS]. *F. capucina* var. *gracilis* is most abundant at sites like Rosconnell [ROS] and Monamore [MOA] but with a maximum abundance below 10%. *F. contruens* var. *venter* has an extremely high abundance at Lisnahan [LIS] (60%) but is in much lower abundance at sites like Tullabrack [TUL] and Moanmore [MOA]. *F. brevistriata* occurs in abundances of above 10% at Muckanagh [MUC], Carra [CAA] and Bane [BAN] and *F. pinnata* is present only at Moanmore [MOA] with an abundance of above 10%.

Gomphonema is often a common component of the benthic community in freshwaters (Round *et al*., 1990). Four common *Gomphonema* taxa generally occur with low abundances and frequency of occurrence (see Figure 4.4). *G. parvulum* var. *parvulum* only occurs at Rosconnell [ROS] and *G. pumilum* at Atedaun [ATE] in abundances of above 5%. *Stephanodiscus* is a common diatom genus in eutrophic waters (Reynolds, 1984). *S. parvus* is much more frequent and dominant than *S. hantzschii* in this training set. *S. parvus* occurs in abundances of above 30% at Morgans [MOR] and Crans [CRA] and occurred in abundances greater than 5% in nine other sites, including Rushaun [RUS], Lene [LEN], Garvillaun [GAR] and Ramor [RAM]. Only three sites (Inchiquin [INQ], Ramor [RAM] and Caragh [CAR]) have *S. hantzschii* in abundances of above 5%.

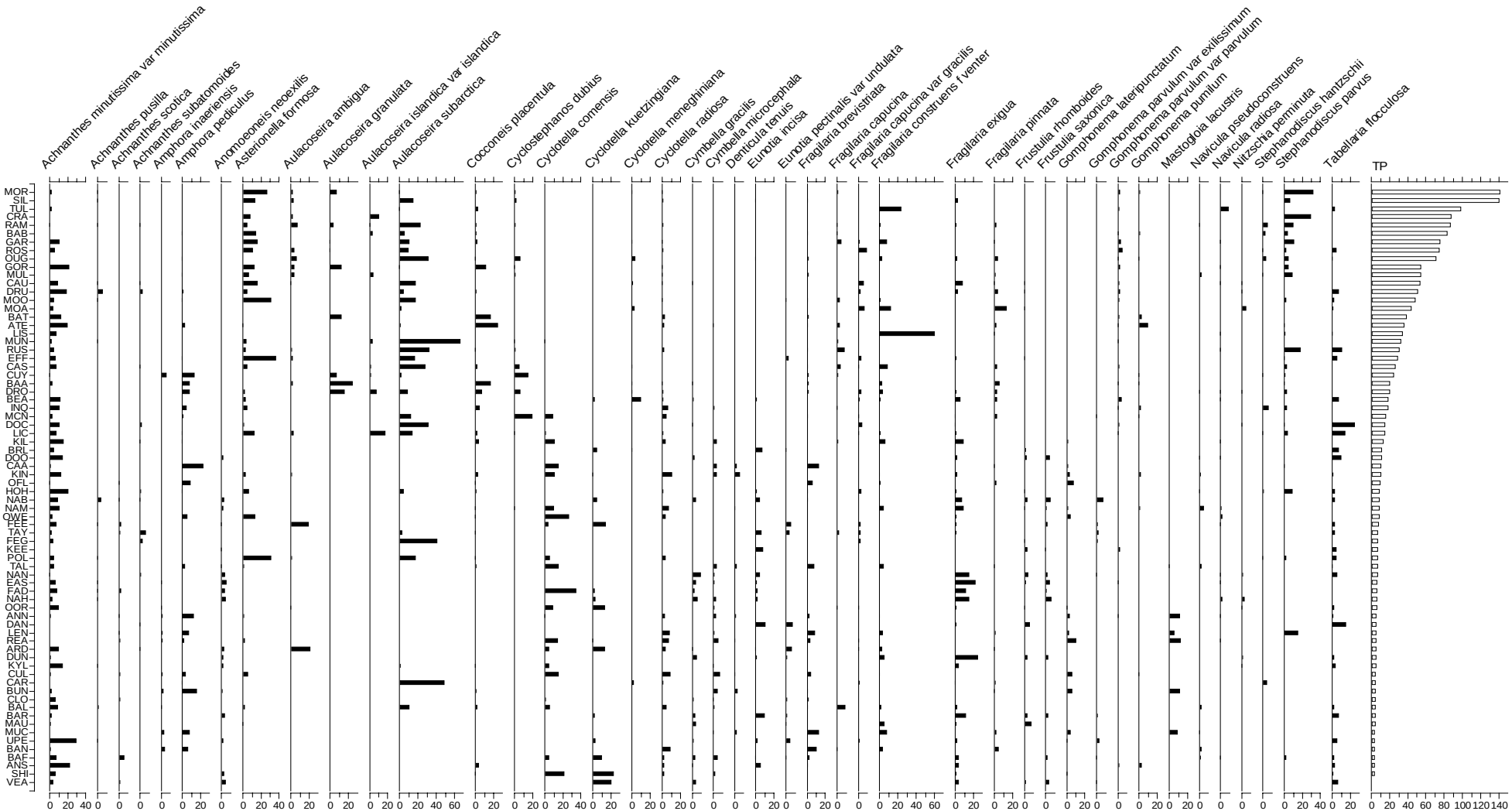

Figure 4.4 Distribution of 42 abundant diatom taxa (Hill's N2 ≥ 5 & maximum abundance ≥ 5%) (taxa are listed in alphabetic sequence and lakes are ordered according to their TP values from lowest at the bottom to highest on the top).

4.2.1 Indirect Gradient Analysis of Diatom Data

Before we explore the internal structure of surface sediment diatom data and their relationships with environmental variables, the gradient length of the diatom community is examined to determine the appropriate ordination methods. Detrended Correspondence Analysis (DCA) was used to estimate the heterogeneity in assemblage composition and the degree of species turnover. Rare taxa were downweighted to reduce unduly high influence in the pattern of diatom data. DCA analysis of 233 surface sediment diatom data produced a gradient length of 3.718 SD for the first axis. As the gradient length was in between 3 and 4, a unimodal response model was selected for further ordination and constrained ordination analysis as most taxa are not behaving monotonically over the underlying environmental gradients (ter Braak & Prentice, 1988). Therefore Correspondence Analysis (CA) was selected to explore the variation in diatom assemblages.

CA simultaneously ordinates both species and sites and such analysis aims to reveal the internal structure of species data. The results of CA is shown in Table 4.4 and two separate CA plots of diatoms and sites are shown in Figure 4.5 and Figure 4.6 respectively instead of one joint plot because of the large numbers of diatom species and sites. Eigenvalues for the first and second axes were 0.511 (λ_1) and 0.405 (λ_2), explaining 12.1% and 9.6% of the total variance of the diatom data. Only Axes 1 and 2 were significant in explaining the pattern of diatom data under the broken stick model and both axes together explained 21.7% of total variance in the diatom data.

Table 4.4 Summary statistics of CA on the diatom data of 72 lakes

CA Axis	1	2	3	4	Total Variance
Eigenvalue (λ)	0.511	0.405	0.182	0.166	4.222
Variance Explained (%)	12.1	9.6	4.3	3.9	
Cumulative Variance (%)	12.1	21.7	26.0	29.9	
Total Unconstrained Eigenvalue					4.222

Twenty diatom taxa with the highest species scores (both positive and negative) are listed for axes 1 and 2 in Table 4.5. Diatom taxa like *Eunotia elegans* [EELE, species code], *Gomphonema hebridense* [GHEB], *Tabellaria quadriseptata* [TQUA] and

Eunotia monodon var. *bidens* [EMBI] sit on the far right side of Figure 4.5. They show a closely positive relationship with axis 1 (Table 4.5). At the other end of the first axis, taxa like *Aulacoseira islandica* var. *islandica* [AUIS], *Stephanodiscus tenuis* [SHTE] and *Navicula cari* [NCAR] are dominant. While taxa including *Fragilaria ulna* var. *acus* [FUAC], *Stephanodiscus agassizensis* [SAGA], *Diatoma tenuis* [DITE] and *Stephanodiscus minutulus* [STMI] dominate at the bottom left of Figure 4.5 with a negative correlation with the axis 2. While several taxa like *Cymbella laevis* var. *capitata* [CLAE], *Mastogloia lacustris* [MLAC] and *Cymbella delicatula* [CDEL] with highest species scores are dominant along the other side of the axis 2 (the upper part of Figure 4.5).

Table 4.5 Twenty diatom taxa with highest species scores (both negative and positive) along the first two CA axes (taxa are listed in the order of their species scores (λ) along either axis)

Code	Taxon	λ_1	Code	Taxon	λ_1
EELE	*Eunotia elegans*	1.71	CLAE	*Cymbella laevis* var. *capitata*	2.63
GHEB	*Gomphonema hebridense*	1.56	MLAC	*Mastogloia lacustris*	2.42
NDPA	*Navicula parabryophila*	1.52	MELL	*Mastogloia elliptica*	2.41
TQUA	*Tabellaria quadriseptata*	1.51	MSMI	*Mastogloia smithii*	2.33
EMBI	*Eunotia monodon* var. *bidens*	1.50	CDEL	*Cymbella delicatula*	2.30
EUAL	*Achnanthes flexella* var. *alpestris*	1.47	GSUB	*Gomphonema subtile*	2.29
ERHY	*Eunotia rhynchocephala* v. *rhynchocephala*	1.45	EGOE	*Epithemia muelleri*	2.28
FERI	*Frustulia erifuga*	1.43	FPLA	*Fragilaria pinnata* var. *lancetttula*	2.24
ENPE	*Cymbella perpusilla*	1.43	FLAP	*Fragilaria lapponica*	2.17
PALT	*Achnanthes altaica*	1.43	NDEN	*Nitzschia denticula*	2.14
STMI	*Stephanodiscus minutulus*	-1.00	NESP	*Neidium sp.*	-0.84
GMCU	*Gomphonema minutum*	-1.07	SHTE	*Stephanodiscus tenuis*	-0.85
FUAN	*Fragilaria ulna* var. *angustissima*	-1.09	AUAL	*Aulacoseira alpigena*	-0.85
AUGR	*Aulacoseira granulata*	-1.09	FUAN	*Fragilaria ulna* var. *angustissima*	-0.86
CPED	*Cocconeis pediculus*	-1.10	STMI	*Stephanodiscus minutulus*	-0.89
AUGA	*Aulacoseira granulata* var. *angustissima*	-1.11	DITE	*Diatoma tenuis*	-0.90
AUIS	*Aulacoseira islandica* var. *islandica*	-1.13	FTEN	*Fragilaria tenera*	-0.90
NCAR	*Navicula cari*	-1.19	FUAC	*Fragilaria ulna* v. *acus*	-0.95
FUAC	*Fragilaria ulna* v. *acus*	-1.20	SAGA	*Stephanodiscus agassizensis*	-1.01
SHTE	*Stephanodiscus tenuis*	-1.35	AUIT	*Aulacoseira italica*	-1.09

In CA the species composition at a particular site is composed of those species sitting close to the site (ter Braak & Prentice, 1988). Loughs Veagh [VEA], Barra [BAR], Easky [EAS] and Doo [DOO], which are located to the far right side of axis 1 (Figure 4.6), have high abundances of *Frustulia saxonica* [FSAX], *Eunotia elegans* [EELE], *Eunotia monodon* var. *bidens* [EMBI], *Navicula parabryophila* [NDPA] (see Figure 4.5). While diatom assemblages at those sites near the other end of axis 1 including Mullagh [MUL], Dromore [DRO], Cullaunyheeda [CUY] and Inchiquin [INQ] consist of taxa

like *Cocconeis pediculus* [CPED], *Navicula cari* [NCAR], *Achnanthes clevei* [KCLE], *Stephanodiscus neoastreae* [SNEO] in high abundances. Similarly the sites located at the top end of CA axis 2 in Figure 4.6 including Bunny [BUN], Muckanagh [MUC] and Carra [CAA] are comprised of *Cymbella laevis* var. *capitata* [CLAE], *Mastogloia lacustris* [MLAC] and *Mastogloia smithii* [MSMI] with high positive species scores (see Figure 4.5 and Figure 4.6). At the lower end of axis 2 diatom assemblages at Muckno [MUN], Rushaun [RUS] and Rosconnell [ROS] are more influenced by taxa like *Aulacoseira italica* [AUIT], *Stephanodiscus agassizensis* [SAGA] and *Stephanodiscus minutulus* [STMI]. A distinctive feature in the CA plot of diatom species is the triangle-like configuration of the diatom taxa (see Figure 4.5). The diatom communities spread strictly along the first and second axes of CA. This is probably a reflection of the significance of the first two axes in controlling the diatom assemblages of the 72 lakes.

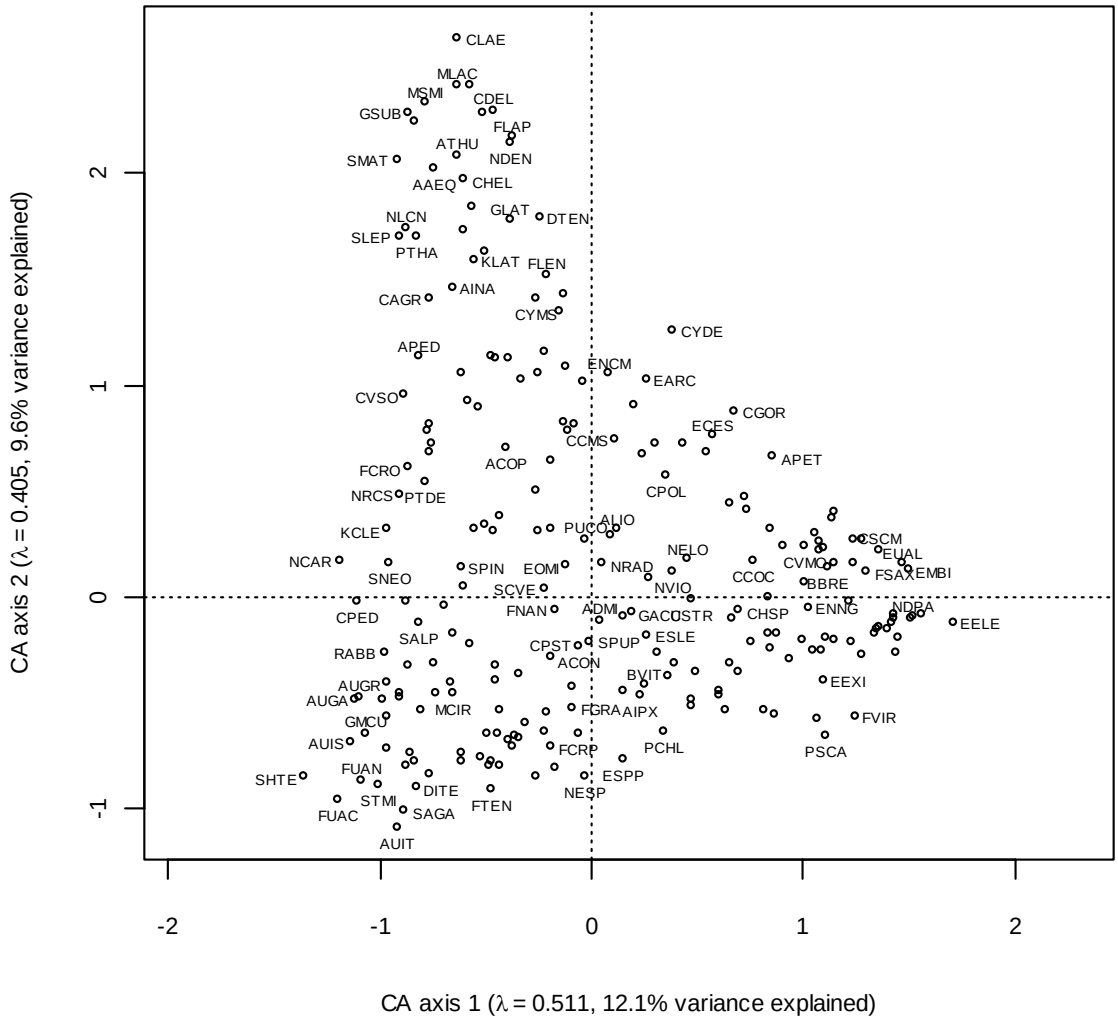

Figure 4.5 CA plot of species scores of diatom data from 72 lakes.

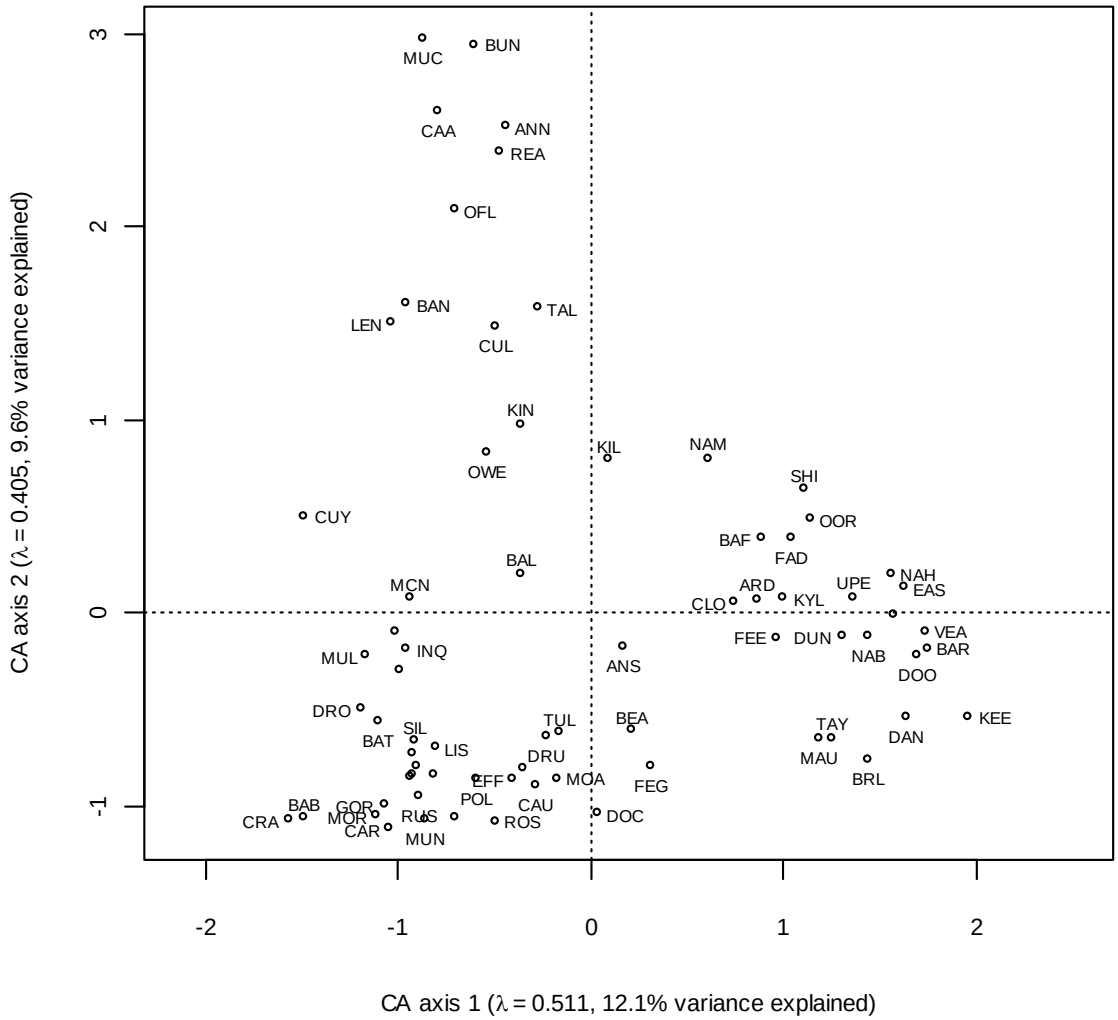

Figure 4.6 CA plot of site scores of diatom data from 72 lakes.

4.2.2 Classification of Diatom Assemblages

Alkalinity, mean depth and lake area were recognized as significant variables in characterising Irish lakes by the Irish EPA (Free *et al*., 2005). This physico-chemical categorization has only been verified by limited ecological data from around 60 lakes. The surface sediment diatom assemblages from 72 lakes are used here to see if these biological parameters reflect the physico-chemical classes of lakes. All twelve lake types are represented by the 72 lakes in the diatom training set and each lake type contains 3-11 lakes. Ward's minimum variance method was used to classify the surface sediment diatoms and seven clusters were identified based on the dissimilarity between diatom samples (see Figure 4.7). A summary of the seven clusters of diatom samples and their associated lake types are shown in Table 4.6. A comparison of diatom cluster analysis and the physico-chemical classification for the 72-lake diatom training-set

lakes is shown in Figure 4.10. Both Ward's seven diatom clusters and the 12 Irish lake types are superimposed on CA plots of diatom data.

Cluster 1 with 15 lakes is mainly dominated by planktonic diatoms like *Aulacoseira subarctica*, *Asterionella formosa* and *Cyclotella radiosa* (see Table 4.6). They are typical taxa of nutrient-rich and circumneutral to alkaline waters. Cluster 2 only contained seven lakes and is mainly composed of non-planktonic (including benthic, epiphytic and epilithic) taxa. The common taxa are acidophilous and circumneutral which suggests that this cluster has lower alkalinity than Cluster 1. Cluster 3 is dominated by planktonic taxa commonly found in circumneutral to alkaline lakes. It differs from Cluster 1 in that it has a greater abundance of *Stephanodiscus parvus* and a lesser abundance of *Aulacoseira subarctica* (see Table 4.6).

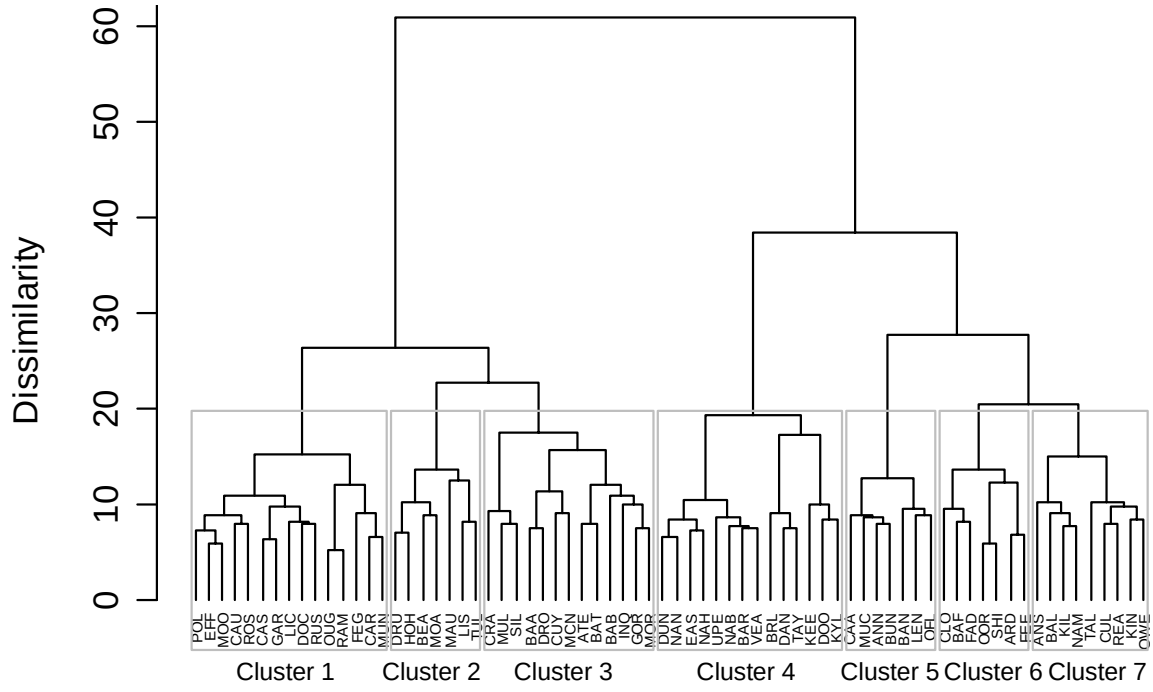

Figure 4.7 Dendrogram of surface sediment diatoms of 72 lakes according to Ward's minimum variance method (seven clusters are identified within grey frames and labelled under each frame).

Cluster 4 lakes are dominated by non-planktonic and acidophilous taxa, including acidophilous *Tabellaria flocculosa*, *Eunotia incisa* and *Frustulia saxonica*. In comparison with Cluster 4, Cluster 5 is also dominated by non-planktonic taxa but they are commonly found in alkaline lakes, like *Amphora pediculus*, *Fragilaria brevistriata* and *Fragilaria pinnata*. The seven Cluster 6 lakes are mainly composed of planktonic taxa commonly found in acidophilous to circumneutral waters, including *Cyclotella*

comensis and *C. kuetzingiana*. These taxa are also common component of nutrient-poor waters. In comparison with Cluster 4 the diatom assemblages implied higher alkalinity in Cluster 6. It also differs from Cluster 2 in its greater abundance of *C. comensis* and *C. kuetzingiana*. Cluster 7 is dominated by both non-planktonic and planktonic diatoms common found in circumneutral to alkaline waters. The most common taxa include circumneutral *Achnanthes minutissima* var. *minutissima* and *Cyclotella comensis*, as well as alkaliphilous *Fragilaria brevistriata* and *Cyclotella radiosa*. The greater abundance of non-planktonic taxa like benthic *Fragilaria* makes Cluster 7 distinguishable from Cluster 1 as Cluster 1 is dominated by planktonic taxa.

The diatom clusters show good agreement with the physico-chemical lake typology classification. Lakes with high alkalinity are represented by Clusters 3 and 5 in Figure 4.10 A and Types 9 to 12 in Figure 4.10 B. They are located in the left of both CA plots and are enclosed by the oval in the left of the plot in Figure B. The oval in the middle of Figure B is mainly composed of the Lake Types 5 to 8 with medium alkalinity, which corresponds well with the diatom Clusters 1, 7 and part of Cluster 2 in Figure A. While the oval to the right of Figure 4.10 B encloses most of the Lake Types 1 to 4 characteristic of low alkalinity and these are similar to diatom Clusters 4 and 6 and part of Cluster 2 in Figure A. This confirms that alkalinity used in the Irish Lake Typology is a significant factor controlling the diatom assemblage structure in these lakes. Lake depth and area, the other two variables employed in the Irish Lake Typology, also show influence on the diatom assemblages of the 72 lakes. In Figure A Clusters 3 and 5 contain lakes with high alkalinity but they lie at opposite ends of the CA plot (see Figure 4.10). This can be accounted for by their distinct lake areas as summarised in Table 4.6: lakes in Cluster 3 are mainly small and those in Cluster 5 are large lakes, corresponding to Lake Types 10 and 12 and 9 and 11 respectively in Figure B. Lake depth probably seprerates Cluster 2 from Cluster 6, both of which contain lakes mainly with low alkalinity (see Table 4.6).

Table 4.6 Summary of diatom clusters classified using Ward's method and related lake types.

Diatom Cluster	Cluster description	Common taxa	No. of lakes	Lake types
1	Planktonic dominant, circumneutral to alkaline taxa	*Aulacoseira subarctica* *Asterionella formosa* *Achnanthes minutissima* var. *minutissima* *Cyclotella radiosa* *Gomphonema parvulum* var. *parvulum*	15	Mainly medium alkalinity, Mainly shallow, Small and large
2	Non-planktonic dominant, acidophilous to circumneutral taxa	*Achnanthes minutissima* var. *minutissima* *Tabellaria flocculosa* *Fragilaria construens* f. *venter* *Cyclotella pseudostelligera* *Nitzschia palea*	7	Mainly low alkalinity, Mainly shallow, Small
3	Planktionic dominant, circumneutral to alkaline taxa (higher abundance of *S. parvus* and fewer *A. subarctica* than Cluster 1)	*Stephanodiscus parvus* *Asterionella formosa* *Cocconeis placentula* *Achnanthes minutissima* var. *minutissima* *Cyclotella radiosa*	13	High to medium alkalinity, Mainly deep, Mainly small
4	Non-planktonic dominant, acidophilous taxa (more acidophilous and fewer planktonic taxa than Cluster 2)	*Tabellaria flocculosa* *Achnanthes minutissima* var. *minutissima* *Eunotia incisa* *Frustulia saxonica* *Fragilaria exigua*	14	Low alkalinity, Deep and shallow, Small and large
5	Non-planktonic dominant, alkaline taxa	*Amphora pediculus* *Fragilaria brevistriata* *Fragilaria pinnata* *Gomphonema lateripunctatum* *Cymbella microcephala*	7	High alkalinity, Deep and shallow, Large
6	Planktonic dominant, acidophilous to circumneutral taxa (greater abundance of *C. comensis* and *C. kuetzingiana* than Clusters 2 and 4)	*Cyclotella comensis* *Cyclotella kuetzingiana* *Fragilaria exigua* *Achnanthes minutissima* var. *minutissima* *Tabellaria flocculosa*	7	Low alkalinity, Deep, Mainly large
7	Non-planktonic and planktonic co-dominant, with circumneutral to alkaline taxa (fewer planktonic taxa than Cluster 1)	*Achnanthes minutissima* var. *minutissima* *Cyclotella comensis* *Fragilaria brevistriata* *Cyclotella radiosa* *Fragilaria construens* f. *venter*	9	Mainly medium alkalinity, Deep, Small and large

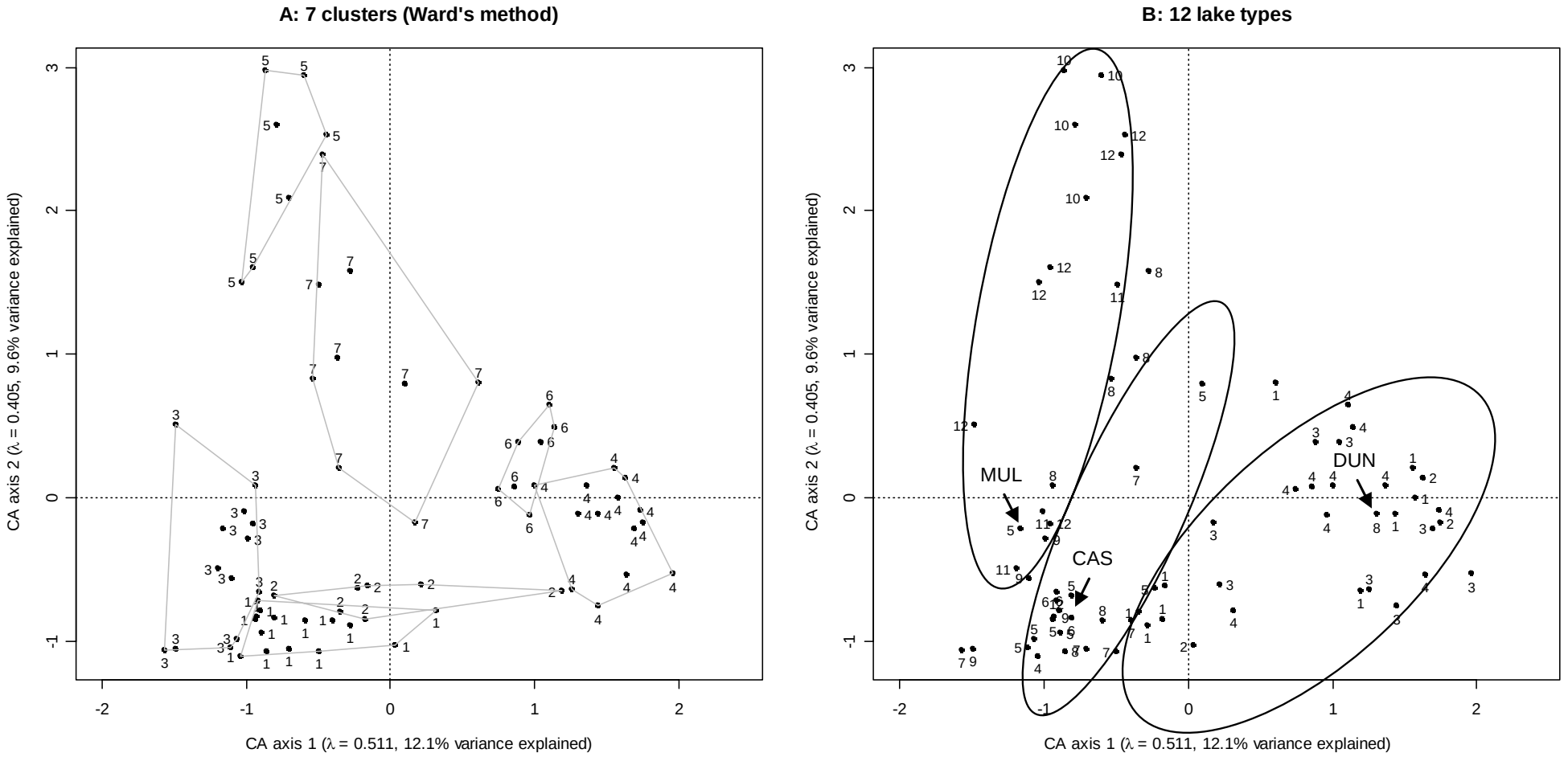

Figure 4.8 Comparison of seven Ward's minimum clusters (A) and 12 typology classes (B) of 72 lakes in the diatom training-set (both classifications are super- imposed on CA plots of diatom data; three ovals indicate three groups of lakes with high, medium and low alkalinities from the left to the right of Figure B; the arrows point to misclassfied sites based on diatom data).

A mismatching between the diatom clusters and the Irish Lake Typology classes is also observed in Figure 4.10. The arrows in Figure 4.10 (B) highlight three sites, Mullagh [MUL], Castle [CAS] and Dunglow [DUN] as the potentially displaced sites. Diatom assemblages at these three sites suggest high, medium and low alkalinities instead of the measured medium, high and medium alkalinity values respectively. The use of the typology classification assumes that there are discrete groups among lakes based on several physico-chemical variables. This may not occur in nature and some groups of lakes could overlap with other groups of lakes. In this 72-lake data set, the diatom Cluster 2 contains lakes both with low alkalinity and medium alkalinity (see Figure 4.10). However, most other diatom clusters do not overlap with each other in CA plot and this indicates strong dissimilarity between clusters and similarity within each diatom cluster. In general the diatom clusters correspond well with the EPA lake types and this provides biological verification of the classification scheme.

4.3 Direct Gradient Analysis

Internal patterns in the environmental data and the diatom data in the training set have been explored separately in the previous sections using the PCA and CA. Canonical Correspondence Analysis (CCA) is now used to constrain the diatom data with the environmental data based on the unimodal species response model. Study sites, diatom taxa and environmental variables are ordinated simultaneously in CCA. All the data analyses in this section can provide a vigorous basis to determine the degree of influence of environmental variables, particularly TP, on the diatom assemblages and to assess the viability of TP for further development in inference modelling and environmental reconstruction.

4.3.1 Constrained Correspondence Analysis (CCA)

CCA biplots of diatom data and sites constrained by all the 17 environmental variables are shown in Figure 4.9 and Figure 4.10, and the results are also summarised in Table 4.7. Eigenvalues constrained by the 17 environmental variables account for 35.1% (= 2.016/5.737) of the total variance in the diatom assemblages (see Table 4.7). Constrained eigenvalues of axes 1 to 4 were 0.488, 0.378, 0.150 and 0.131, explaining 8.5%, 6.6%, 2.6% and 2.3% of the total variance in the diatom data respectively. The

first two axes accounted for 15.1% of the total variance of the diatom assemblages while each of the additional axes explained a much smaller fraction of the total variance.

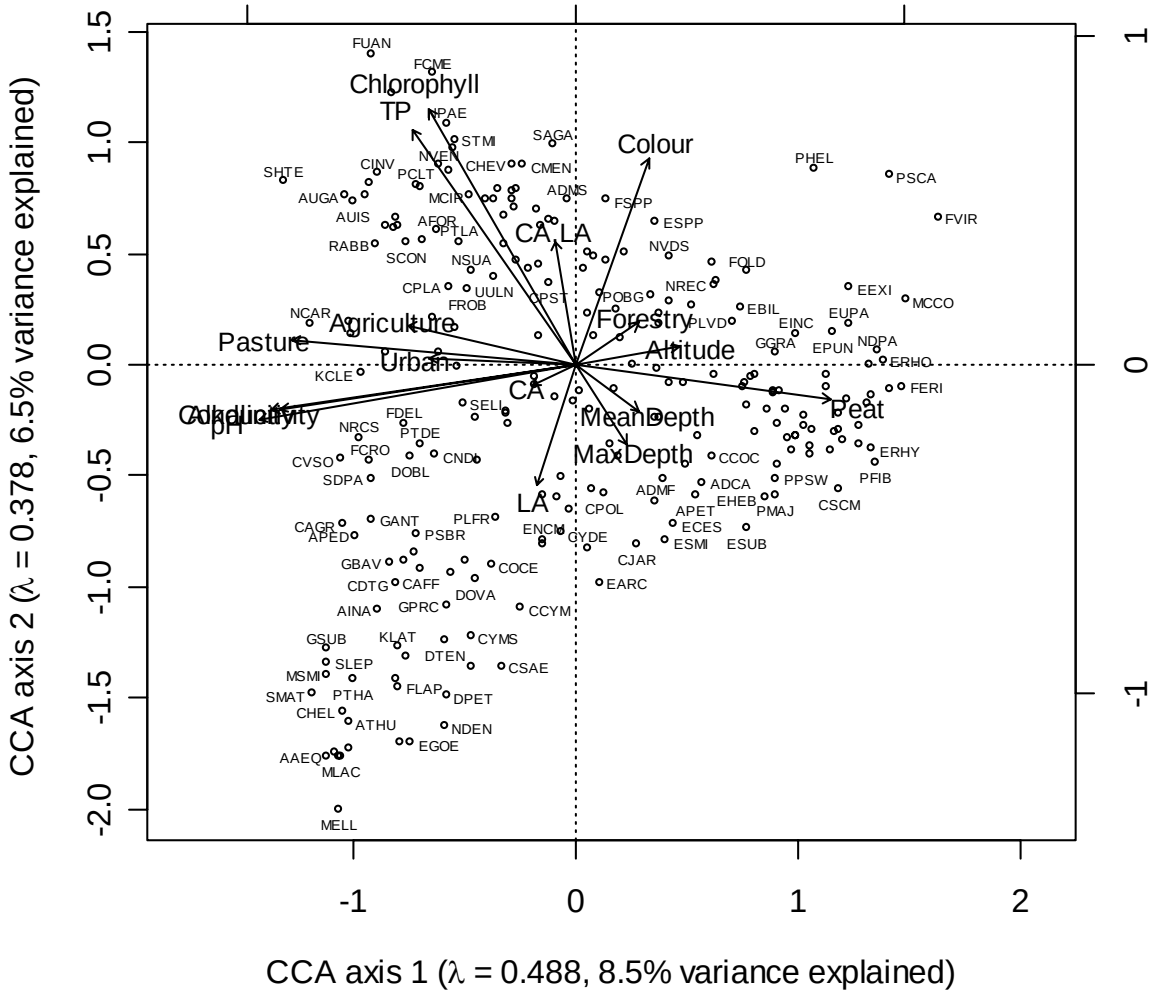

Figure 4.9 CCA biplot of species and 17 environment variables in 72 lakes in the diatom training set.

Compared with the unconstrained plots of species and site scores (seeFigure 4.5 and Figure 4.6), the location of most diatom taxa and configuration of the assemblages remains quite similar in the constrained plot. Many species like *Amphora aequalis* [AAEQ], *Amphora inaeriensis* [AINA], *Stephanodiscus tenuis* [SHTE], *Pinnularia subcapitata* [PSCA] and *Cymbella cesattii* [ECES] on CCA plot (Figure 4.9) remain in similar positions as in CA plot (Figure 4.5) or drift in a short distance, like *Navicula cari* [NCAR] which moves from above the first axis on the left of the CA plot to just below the first axis on the left of CCA plot (Figure 4.9). This implies that the inherent pattern of diatom assemblages unconstrained is captured when constrained by the environmental variables. This is also apparent when comparing the eigenvalues of unconstrained CA axes and constrained CCA axes (ter Braak, 1987b). The percentage

variances explained by the CCA first and second axes are close to those explained by axes 1 and 2 in the unconstrained CA (8.5% and 6.6% for CCA axes and 12.1% and 9.6% for CA axes respectively) (see Table 4.4 and Table 4.7). The variances explained by the constrained eigenvalues stabilise after the first two axes in both CCA and CA at a much smaller fraction of the total variance.

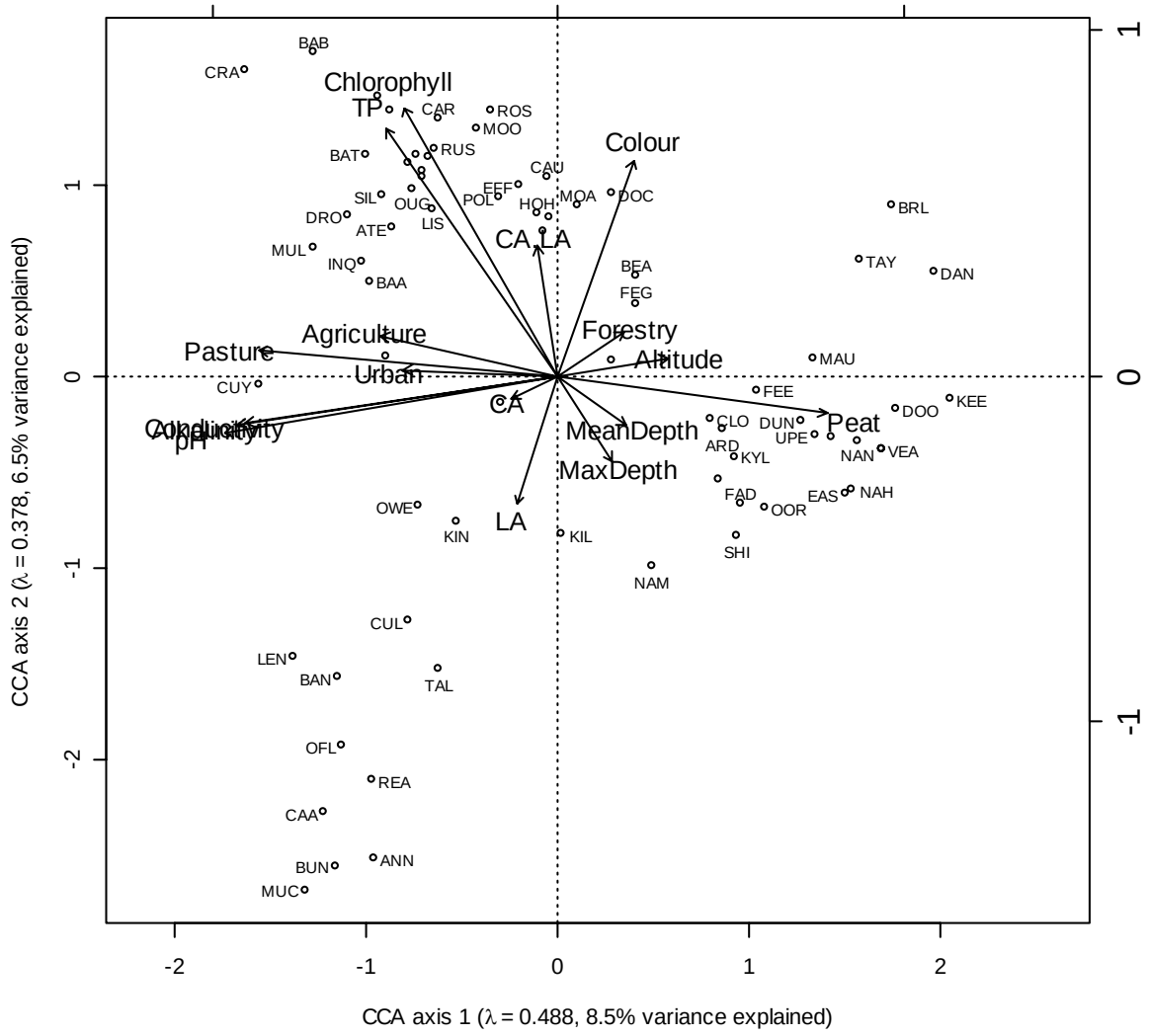

Figure 4.10 CCA biplot of sites and 17 environment variables in72 lakes in the diatom training set.

The first CCA axis is strongly correlated with pH, alkalinity, conductivity, pasture and peat while the axis 2 is dominated by chlorophyll a, TP and colour. The gradient compositions of the two CCA axes are very much the same as the structure in the first two PCA axes of environmental variables shown in Figure 4.2: the first axis is most correlated with acidity and conductivity gradient (pH, alkalinity, conductivity) and the second one with nutrient gradient (TP, chlorophyll-*a*). Differences in the structures of both PCA and CCA plots include the shortened vectors of all six physical variables in

CCA plot, indicating their reduced influences in constraining the diatom data (compare Figure 4.9 and Figure 4.2).

Table 4.7 Summary statistics of 17 environmental variables in the CCA of diatom data.

CCA Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.488	0.378	0.150	0.131	5.737
Speices-Environment Correlation	0.963	0.923	0.902	0.865	
Cumulative Variance (%)	8.5	15.1	17.7	20.0	
Constrained Eigenvalues					2.016
Total Unconstrained Eigenvalues					3.721
Biplot scores for constraining variables					
Altitude	0.316	-0.053	0.306	-0.160	
CA	-0.134	0.061	0.213	0.076	
LA	-0.122	0.365	0.452	0.269	
CA:LA	-0.057	-0.376	-0.225	-0.233	
MaxDepth	0.150	0.243	0.759	0.231	
MeanDepth	0.191	0.137	0.671	0.396	
Agriculture	-0.512	-0.123	0.014	0.032	
Forestry	0.193	-0.128	0.292	0.260	
Pasture	-0.867	-0.084	0.043	-0.284	
Peat	0.781	0.114	0.055	0.266	
Urban	-0.450	-0.025	0.303	0.000	
Alkalinity	-0.932	0.131	-0.153	-0.059	
Chlorophyll	-0.442	-0.781	-0.161	-0.004	
Colour	0.228	-0.624	-0.006	-0.155	
Conductivity	-0.905	0.128	-0.142	-0.212	
pH	-0.968	0.158	-0.072	0.097	
TP	-0.494	-0.720	-0.183	-0.113	

As weighted averages of species scores (Oksanen, 2005a) are used as site scores for CCA in Figure 4.9, the position of sites in CCA plot also show a similar pattern as those in CA plot (Figure 4.6). As in PCA, the origin of the CCA plot represents the weighted averages of the environmental variables and the sites with a projection on an environmental vector at the same side of the arrow have higher values of the environmental variable and vice versa. Sites located in the left side of CCA axis 1 are characteristic of being highly alkaline with high conductivity like Cullaunyheeda [CUY], Muchanagh [MUC] and Lene [LEN]. Highly acidic sites with very low conductivity including Keel [KEE], Doo [DOO] and Maumwee [MAU] are located in the opposite side of axis 1. Loughs with high nutrient levels (e.g. Crans [CRA] and Ballybeg [BAB]) could be found in the lower part along the axis 2, while those with low nutrient levels are mainly located in the other side of axis 2, including Muchanagh [MUC], Bunny [BUN] and Annaghmore [ANN] at the edge of the upper left of the plot (Figure 4.9). In

summary, sites in the CCA plot are located along the two major gradients, acidity-conductivity and nutrient.

4.3.2 Selecting Important Environment Variables

Two main axes identified in constrained ordination of the species and environment data are the acidity-conductivity and nutrient gradients. Each gradient is represented by several highly correlated variables. Along the first axis pH, alkalinity and conductivity are closely positively correlated. On the second axis TP and chlorophyll-*a* are closely correlated and have strongly negative relationships with lake depth (see Figure 4.9). Redundant information in the full CCA model using all the 17 environmental variables occurs as a result of collinearity among the environmental variables. The building of the CCA model often follows the principle of parsimony (also called 'Ockham's Razor'), which recommends that any parameter that does not significantly contribute to the model should be eliminated (Legendre & Legendre, 1998).

Forward selection was implemented in the R program to aid the selection a smaller set of significant environmental variables to explain the variation in the diatom training set. Variance Inflation Factors (VIF) of the 17 environmental variables reported in CCA output (see Table 4.8) enabled the exploration of co-relationship among the variables. A common rule of thumb is that variables with VIF of above 20 are strongly dependent on other variables and therefore could not provide unique information (ter Braak & Šmilauer, 2002). Catchment and lake areas are shown to be strongly correlated and they were the first candidates for exclusion from the ordination model due to their high VIFs (see Figure 4.9). Partial CCAs selecting just one environmental variable each time were also performed to constrain the diatom data and test for significance (see Table 4.8). Monto Carlo permutation tests (999) were performed for all 17 variables. Two variables, catchment area and forestry, are not significant at the P-level of 0.05 (Table 4.8). Variables including pH, alkalinity and conductivity independently explained the highest variances (8.2%, 7.7% and 7.4% respectively) in the diatom data and they are significant at the P-level of 0.001. But their VIF values suggest that moderate collinearity exists for these three variables (see Figure 4.9 and Figure 4.10). Each of the two nutrient variables, TP and chlorophyll-*a*, explains around 6% of the total variance significantly (P-level < 0.001).

Table 4.8 Summary of VIF and partial CCAs constrained by only one of the 17 environmental variables each time (999 Monto Carlo permutation tests were used; λ_1 and λ_2 refer to eigenvalues of the first two CCA axes; variables significant at the P-level of 0.001 are marked with *).

Variable	VIF	λ_1	λ_2	λ_1/λ_2	% variance explained	P-value
Altitude	1.6	0.126	0.498	0.253	2.2	0.010
Catchment Area (CA)	174.4	0.083	0.527	0.158	1.5	0.328
Lake Area (LA)	128.9	0.136	0.532	0.255	2.4	0.003
CA.LA	77.8	0.127	0.530	0.239	2.2	0.008
MaxDepth	5.1	0.148	0.524	0.282	2.6	0.002
MeanDepth	3.7	0.138	0.518	0.267	2.4	0.004
Agriculture	1.8	0.174	0.458	0.379	3.0	< 0.001*
Forestry	1.6	0.090	0.521	0.174	1.6	0.178
Pasture	9.2	0.392	0.458	0.855	6.8	< 0.001*
Peat	6.7	0.334	0.458	0.730	5.8	< 0.001*
Urban	1.7	0.168	0.464	0.362	2.9	< 0.001*
Alkalinity	14.2	0.439	0.464	0.948	7.7	< 0.001*
Chlorophyll-*a*	3.7	0.343	0.505	0.679	6.0	< 0.001*
Colour	3.2	0.223	0.530	0.420	3.9	< 0.001*
Conductivity	11.3	0.422	0.464	0.910	7.4	< 0.001*
pH	10.3	0.470	0.464	1.013	8.2	< 0.001*
TP	6.0	0.336	0.499	0.673	5.9	< 0.001*

Because TP has proven to be a significant variable controlling the diatom assemblages and is also the variable of primary interest in this study, it is manually included in forward selection before any other variable is added in the minimum model. Forward selection with Monte Carlo permutation test was performed in the R program. Two variables (pH and maximum depth) were automatically selected in addition to the manually selected TP. Table 4.9 summarises the minimum optimal CCA model for the 72-lake diatom training set constrained by pH, TP and maximum depth. This model accounted for more than 45% (0.768/2.016) of the total constrained variance explained by the 17 environmental variables (compare Table 4.9 with Table 4.7). Therefore, this minimal set containing three significant environmental variables could explain the species data nearly as well as the full model as it captured nearly half of the total variance in the full model constrained by all 17 environmental variables.

Table 4.9 Summary of CCA of diatom data constrained by TP, pH and maximum depth.

Axes	1	2	3	4	Total Variance
Eigenvalues for constrained Axes	0.4763	0.3056	0.1258		5.737
Speices-Environment Correlation	0.953	0.849	0.874		
Cumulative Variance (%) for Constrained Axes	8.3	13.6	15.8		
Eigenvalues for unconstrained Axes	0.2574	0.2458	0.2006	0.1833	
Cumulative Variance (%) for Unconstrained Axes	4.5	8.8	12.3	15.5	
Constrained Eigenvalue					0.908
Total Unconstrained Eigenvalue					4.829

4.3.3 Viability of TP for Constructing Transfer Functions

Both the full and the partial constrained ordinations of diatom data in the previous sections indicate that TP is one of the most significant environmental variables explaining the patterns of diatom assemblages in the 72 lakes. A good indicator for assessing the viability of environmental variable selected to construct transfer functions is the ratio of λ_1 to λ_2, the first two eigenvalues of partial CCA (ter Braak, 1988; Hall & Smol, 1992). The $\lambda_1/\ \lambda_2$ ratios for the 17 environmental variables are listed in Table 4.9. The three variables related with acidity-conductivity gradient have highest $\lambda_1/\ \lambda_2$ ratios above 0.7 among all the 17 variables. The $\lambda_1/\ \lambda_2$ ratios of pasture and peat are also quite high (0.855 and 0.730 respectively), but their high abundance of zero values implies that they are not viable for such modelling (Birks, 2005b). TP produces a $\lambda_1/\ \lambda_2$ ratio of 0.673. Variables with high λ_1:λ_2 ratios (e.g., >1) generally produce strong calibration models, but variables with lower ratios can still be used in transfer function development but care should be taken in ecological interpretation (ter Braak, 1987b; ter Braak, 1988). TP inference models have been developed with relative low $\lambda_1/\ \lambda_2$ ratios, e.g. 0.4 (Hall & Smol, 1992), 0.42 (Reavie *et al.*, 1995), 0.45 (Wunsam & Schmidt, 1995) 0.29 (Hall & Smol, 1996), 0.50 (Gregory-Eaves *et al.*, 1999), 0.83 (Kauppila *et al.*, 2002). This comparison of $\lambda_1/\ \lambda_2$ ratios indicates that TP can be used for developing a robust transfer function for this diatom training set.

4.4 Species Response Curves

In this section the distribution of individual diatom species along the TP gradient is explored using Gaussian Logit Regression (GLR) based on a Gaussian unimodal model (see Chapter 3 on methods). Nutrient gradient was identified as a significant factor influencing the diatom assemblage patterns in the training set and being viable for constructing robust transfer functions. The species response analysis would help understand the linear or non-linear response of some significant diatom species to TP before further development of TP transfer functions using modelling methods based on species response models.

Twenty-two common diatom taxa occurred with a maximum relative abundance above 5% and an effective number (N2) of above ten and their response curves to TP are shown in Figure 4.11. Of the 22 common diatom taxa, eight taxa showed a unimodal or unimodal-like response curve, 12 displayed a monotonic decreasing sigmoid curve and two showed a monotonic increasing sigmoid curve along the TP gradient of 0-142 μg l^{-1}. Abundances of *Achnanthes scotica* [ADCA], *Anomoeoneis neoexilis* [BEXI], *Cyclotella comensis* [CCMS], *C. radiosa* [PRAD], *Cymbella gracilis* [ENNG], *C. microcephala* [ENCM], *Eunotia incisa* [EINC], *Fragilaria brevistriata* [PSBR], *F. exigua* [SEXG], *Frustulia rhomboids* [FRHO], *F. saxonica* [FSAX] and *Tabellaria flocculosa* [TFLO] decreased rapidly with the increase of TP values in the 72-lake training set. The curves indicate their preference in oligotrophic lakes. In the mesotrophic (TP in the range of 10-35 μg^{-1}) and eutrophic (TP in the range of 35-100 μg^{-1}) waters, *Achnanthes minutissima* var. *minutissima* [ADMI], *Amphora pediculus* [APED], *Aulacoseira subarctica* [AUSU], *Cocconeis placentula* [CPLA], *Fragilaria capucina* [FCAP], *F. capucina* var. *gracilis* [FGRA], *F. pinnata* [SPIN] and *Gomphonema parvulum* var. *parvulum* [GPAR] displayed a maximum abundance in this TP range. While in more eutrophicated lakes plankontic species, *Asterionella formosa* [AFOR] and *Stephanodiscus parvus* [SPAV], showed a monotonically increased abundance with the increase of nutrient level and their optimums are assumed to lie beyond the TP range of 0-142 μg l^{-1}.

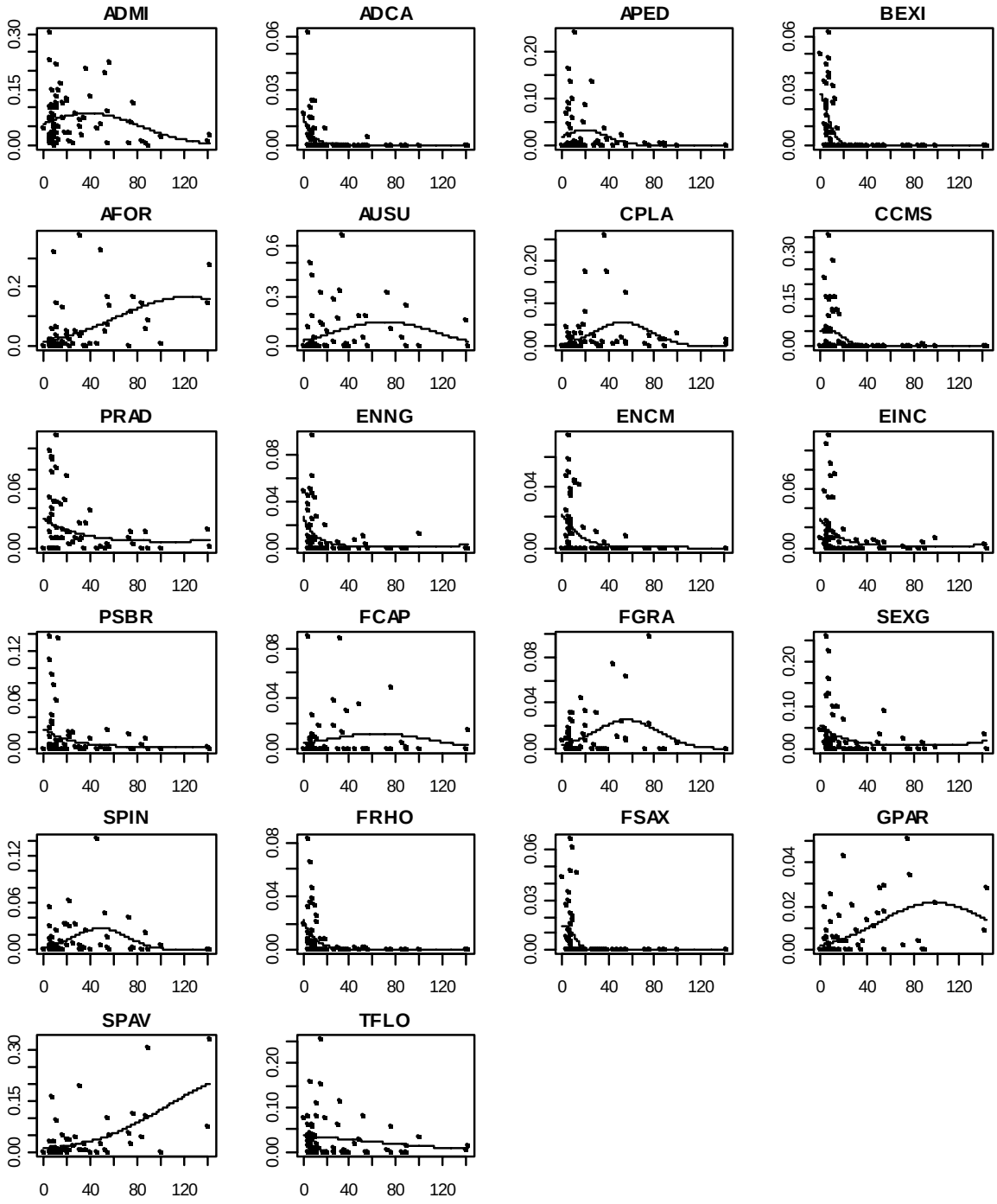

Figure 4.11 Response curves of 22 common diatom species (maximum ≥ 5%, N2 ≥ 10) to TP using Gaussian Logit Regression (GLR) (Horizontal axis represented TP (µg l^{-1}) and vertical axis represents the relative abundance for each taxon; ADMI *Achnanthes minutissima* var. *minutissima*, ADCA *Achnanthes scotica*, APED *Amphora pediculus*, BEXI *Anomoeoneis neoexilis*, AFOR *Asterionella formosa*, AUSU *Aulacoseira subarctica*, CPLA *Cocconeis placentula*, CCMS *Cyclotella comensis*, PRAD *Cyclotella radiosa*, ENNG *Cymbella gracilis*, ENCM *Cymbella microcephala*, EINC *Eunotia incisa*, PSBR *Fragilaria brevistriata*, FCAP *Fragilaria capucina*, FGRA *Fragilaria capucina* var. *gracilis*, SEXG *Fragilaria exigua*, SPIN *Fragilaria pinnata*, FRHO *Frustulia rhomboids*, FSAX *Frustulia saxonica*, GPAR *Gomphonema parvulum* var. *parvulum*, SPAV *Stephanodiscus parvus*, TFLO *Tabellaria flocculosa*).

4.5 Discussion and Conclusions

The main objectives of this chapter were to investigate the properties of environmental and diatom data in this diatom training set and to explore the quantitative relationships between both data sets for the 72 lakes. Several multivariate methods, including direct and indirect gradient analysis, classification, and species response analysis have been performed for achieving the above aims. The variable TP is highlighted as the primary variable of interest in this study and its relationship with the diatom community is the key to the development of a robust TP transfer function. This section aims to discuss the results of the numerical analyses and some key findings will also be concluded.

4.5.1 Characters of Environmental Variables

The ordination of the 17 environmental variables displayed a relatively even distribution of sites along the vectors of environmental variables for the diatom training set lakes. The most significant gradient associated with PCA axis 1 was composed of environmental variables related with acidity and conductivity and the second important gradient was mainly composed of nutrient-related variables along the PCA axis 2. When constrained to the diatom data in CCA, both gradients were found to be the most significant ones in explaining the structure of the diatom community. Therefore the measured environmental variables related with both gradients accounted for the main variation in diatom species composition.

The land cover variables peat and pasture are strongly correlated with the first PCA axis and remained a similar relationship with the first CCA axis. The other three land cover variables (urban, agriculture and forestry) were much less significant and also displayed little change in the length and direction of vectors relative to other variables in the PCA and CCA. The inverse relationship between forest coverage and nutrient gradients means that sites with low forest cover tend to have higher nutrient levels. Deforestation can make it easier for nutrient loads to be exported into the lakes through the catchment run-off (Jennings *et al.*, 2003) and this can account for the higher nutrient status for lakes with lower forestry coverage. Similarly the inverse relationships between peatland coverage and acidity and conductivity variables may be due to the high input of humic

materials transported from the surrounding catchment into the waters (e.g. Dalton, 1999).

The physical variables (e.g. catchment area, mean and maximum depth) generally show a reduced influence on the diatom community in comparison to their relative long vector length in PCA. This suggests that the physical variables have direct and significant influence on chemical composition of the lakes but generally exert indirect influence on the diatom assemblages. These physical variables may still play important roles in controlling the species structure indirectly through influencing the hydrochemical variables (e.g. TP and pH). Lakes at high altitudes tend to have low pH, alkalinity and conductivity values and this is also shown in other studies (e.g. Stevenson *et al.*, 1991). Two nutrient variables, TP and Chlorophyll-*a*, displayed an inverse relationship with lake depth (both mean and maximum) indicating that the nutrient level is generally low in deep lakes while shallow lakes are generally characterised of higher nutrient status. In shallow lakes, the intense sediment-water contact and low buffering capacity of nutrient loading make them more liable to nutrient enrichment in lake waters than deep lakes (Scheffer, 1998). This inverse relationship between nutrient level and lake depth was also indicated by many diatom-TP training sets (e.g. Dixit & Smol, 1994; Hall & Smol, 1996; Dixit *et al.*, 1999; Enache & Prairie, 2002).

TP and chlorophyll-*a* are strongly correlated and both are equal contributers to the total variance in the diatom assemblages. Chlorophyll-*a* is the most dominant chlorophyllous pigment of algae and cyanobacteria and is used to estimate algal biomass (Wetzel, 2001). It is generally an in-lake biological indicator and can be influenced by other variables like alkalinity and conductivity in addition to phosphorous concentration. In addition it is a surrogate for algal biomass and can reflect the combined influence of nutrient variables. Its relationship with diatom community is complex as diatoms have dual roles in constrained ordination analysis, both as the ecological variable and as part of the environment variable. Therefore TP is used as the primary nutrient variable in this study for further transfer function development to infer nutrient level.

4.5.2 Diatom-Environment Relationships

High variation in assemblage composition was observed in the indirection ordination of diatom data. This was reflected by the species response curves of the common diatom taxa along the TP gradient (see Figure 4.11). Eight of 22 common diatom taxa taxa showed a unimodal or unimodal-like response curve with 12 displaying a monotonic decreasing sigmoid curve and two showing a monotonic increasing sigmoid curve along the TP gradient of 0-142 μg l^{-1}. The relatively long TP gradient in this diatom training set is well comparable with other TP training sets (e.g. Fritz *et al.*, 1993; Gregory-Eaves *et al.*, 1999; Kauppila *et al.*, 2002; Miettinen, 2003; Ramstack *et al.*, 2003). The distribution of individual taxa along the TP gradient in this study generally correspond well with what has been published in many other studies (e.g. Wunsam & Schmidt, 1995; Lotter *et al.*, 1998; Bradshaw & Anderson, 2001; Miettinen, 2003).

The configuration of diatom taxa displayed the geometry of triangle in the CA plot of diatom data (see Figure 4.5) and a similar configuration of diatom distribution was also found in the CCA (see Figure 4.9). This shows that the internal structure of the diatom assemblages was captured by the 17 measured environmental variables. In addition the diatom assemblages are completely dominated by the first two axes of both CA and CCA plots, which were strongly correlated with acidity-conductivity and nutrient gradients respectively. The influences of other environmental gradients are too weak to pull the diatom distribution away from the first two axes in ordination space and therefore produce the triangle configuration in the ordination analysis. The relative small divergence between eigenvalues of the first two axes in the CCA and the CA of the 72 lakes means that a large portion of total variance in the diatom assemblages is accounted for by the measured environmental variables. But other unmeasured or unknown variables can also influence the diatom distribution. Several environmental variables unmeasured in this study including epilimnetic CO2 (e.g. Philibert & Prairie, 2002), water temperature (e.g. Bloom *et al.*, 2003), salinity (e.g. Fritz, 1990) can exert significant influence the diatom assemblages. In addition the biotic interactions may also be important regulators on diatom growth and community structure, like selective predation by zooplankton (Lampert & Sommer, 1997).

Despite the relatively long nutrient gradient in this diatom training set, the acidity and conductivity gradient stills plays a more important role in regulating the diatom assemblages. The acidity-conductivity gradient strongly correlated with the first CCA axis explained more of the total variance than the second axes which is mainly

composed of nutrient variables. This has also been found in several diatom training sets constructed for TP inference models (e.g., Reavie & Smol, 2001). The ratio of λ_1/λ_2 of pH was higher than that of TP in this study as in several other studies (e.g., Dixit & Smol, 1994; Dixit *et al.*, 1999; Enache & Prairie, 2002). Generally there is a strong relationship between diatoms and pH due to the direct physiological influence of pH on diatoms. A weaker diatom-TP relationship can be caused by the indirect diatom-TP relationship as other physical and hydrochemical factors (e.g. physical mixing, silica availability) are also involved (Reynolds, 1984). A weaker diatom-pH relationship than the diatom-TP one was also found in several diatom training sets and they generally included lakes with a relative narrow pH range (e.g. Hall & Smol, 1992; Bennion, 1994; Tibby, 2004).

In the CCA of diatoms, TP alone accounted for 5.9% of total variance in species composition and this is comparable with other studies on TP transfer functions, e.g., 5% (Wunsam & Schmidt, 1995), 8.4% (King *et al.*, 2000), 5.8% (Bradshaw & Anderson, 2001). Its relative high ratio of λ_1/λ_2 is well comparable with most other diatom-based TP transfer functions. The common diatom taxa also displayed high sensitivity to the TP gradient in both ordination and species response analyses. In spite of its secondary significance to the acidity-conductivity gradient in controlling the diatom assemblages, TP has been successfully used in constructing transfer functions in several other studies (e.g. Dixit *et al.*, 1999; Reavie & Smol, 2001). All these features point to that TP can be potentially used for developing robust transfer functions for this diatom training set.

4.5.3 Use of Diatoms in Verifying the Lake Typology

The diatom clusters using Ward's minimum clustering method display good correspondence with the lake typology classes mainly along the alkalinity gradient. Three groups of lakes with low, moderate and high alkalinities are clearly reflected in the diatom clusters for the 72 lakes. This strong influence of alkalinity on the biological community was also found by a study in the U.K. which employed the diatoms to validate the British Lake Typology of reference lakes (Bennion *et al.*, 2004a). Alkalinity is one of the key factors in determining the diatom community in lakes and the lake typology classification based on alkalinity has strong ecological response. A major difference between this study and the work by Bennion *et al*. (2004) is that both

reference and degraded lakes are included in this study. The influence of alkalinity on the diatom community is indicated by the strong association of the acidity-conductivity gradient with the first axis of CCA of diatom data despite that impacted lakes were selected mainly along the TP gradient in this study. This confirms the applicability and reliability of the lake typology classification in identifying the lake types of impacted lakes under strong human disturbance (e.g. eutrophication). The influence of lake area and depth used in the Irish Lake Typology is also demonstrated in the diatom clusters but they exert much weaker impact than alkalinity. This may indicate that lake area and depth are not as significant as alkalinity in affecting the ecological status of lakes. But lake depth did show significant influence in the diatom community of 219 samples from the U.K. (Bennion *et al.*, 2004a) and in comparison the relative small number of lakes (72) included in this study may provide insufficient data to evaluate the roles of both variables in affecting diatom community. Furthermore, more biological indictors should be included to fully assess the significance of the variables adopted in the Irish Lake Typology scheme in the overall ecological status.

The use of biological classification of lakes can help identify the potentially misclassified sites, like Mullagh, Castle and Dunglow highlighted in this study (see Figure 4.10). This mismatching of diatom and physico-chemical classification can be due to the seasonal fluctuation of alkalinity in lake waters and therefore insufficient measurement can cause the misclassification of lake types. Seasonal change of alkalinity has been found in Irish lakes (Dalton, 1999; Irvine *et al.*, 2001) and the use of fixed boundary of alkalinity in categorizing lakes can therefore be arbitrary. In addition insufficient lake bathymetry data can also influence the accuracy of mean depth used in lake type classification (Bennion *et al.*, 2004a). As the biological indicators often shown a gradual change in response to the physico-chemical gradients (Jeppesen *et al.*, 2000; Søndergaard *et al.*, 2005), the overlap between the neighbouring lake types also occurred in this study, e.g. at the borders of low to medium and medium to high alkalinities. Therefore the gradual biological response can be inconsistent with the fixed boundary physico-chemical variables adopted in lake classification and this can be a challenge in implementing the Water Framework Directive.

4.5.4 Conclusions

The use of surface sediment diatoms provided a vigorous tool in collecting integrated ecological information and they showed strong response to environment gradients in the multivariate analyses. This investigation showed that acidity and nutrient gradients were most significant in controlling the diatom assemblages. Peat and pasture coverage were strongly correlated with the acidity-conductivity gradient with physical variables being the least important environment variables for species composition in the Irish Ecoregion. Diatom assemblages displayed high heterogeneity and were distributed mainly along the acidity-conductivity and nutrient gradients. The diatom clusters displayed good correspondence with the lake groups based on alkalinity and less with the typology classes based on lake area or depth. This confirmed the importance of alkalinity in classifying the Irish lakes. TP can be used for developing a reliable transfer function as it explained a large portion of total variance in the diatom assemblages and some common taxa responded strictly along the TP gradient.

Chapter 5: Relationships between Surface Sediment Cladocera and Environment Variables of 33 Irish Lakes

This chapter aims to reveal the ecological responses of Cladocera to the physical, hydrochemical and catchment variables based on the surface sediment remains of 33 lakes in the Irish Ecoregion. The environmental characteristics of the 33 lakes and the surface sediment Cladocera assemblages are investigated and summarised separately. In addition the representability of Cladocera remains in the lake sediments is also evaluated via a comparison with contemporary Cladocera data collected for six Irish lakes. Cladocera data is then used to assess the lake typology classes of the 33 lakes based on cluster analysis. The relationship between Cladocera and environmental variables are then examined through constrained ordination analysis and species response curves are examined to check the ecological responses of individual Cladocera to significant environmental gradients.

5.1 Environmental Variables

Seventeen physical, chemical and land use variables (see Appendixes B, C and D) were included in this 33-lake Cladocera data set. Summary statistics of the physico-chemical data for the 33 lakes are shown in Table 5.1 and their frequency distributions are displayed in Figure 5.1. According to the Irish EPA Lake Typology classification scheme (see Chapter 3), eight lake types were included in this Cladocera training set. Lake Types 5 (medium alkalinity, shallow, small) and 12 (high alkalinity, deep, large) contained eight and six lakes respectively. Lake Types 1, 3, 7, 9 and 11 were composed of 3-5 lakes each and only one lake was included in Lake Type 10.

Most of the 33 Cladocera training-set lakes are located in lowlands with altitudes of <150 m except Tay [TAY] with an altitude of 250 m (see Figure 5.1 and Appendix B). Surface areas of the 33 lakes are between 1.2 and 416.3 ha and the majority of the lakes are relatively small with a median value of 21.4 ha (Table 5.1). Catchment area ranges from 30 ha to 29042 ha and the ratio of catchment area to lake area is less than 30:1 for more than half of the 33 lakes. Around half of the lakes have a maximum depth of greater than 12 m and a median depth of greater than 4 m (Table 5.1). More than half of the lakes have pasture coverage of >50% while 30 of the 33 lakes have an urban

coverage of not more than 2%. Most of the lakes have peat coverage of less than 20%. Catchment forest cover is generally low and more than 20 lakes have forested area of less than 5% (Figure 5.1). Thirty of the 33 lakes have agriculture coverage of not more than 20%.

Table 5.1 Summary of 17 environmental variables for 33 lakes.

Variables	Min	Max	Mean	Median	Standard deviation	N	N missing	N non-zero
Physical								
Altitude (m)	7.0	250.0	59.7	50.0	49.3	31	2	31
Catchment Area (ha)	30.0	29042.0	3779.2	390.5	7569.7	30	3	30
Lake Area (ha)	1.2	416.3	48.7	21.4	78.2	32	1	32
Catchment Area:Lake Area	2.8	2045.2	148.0	28.2	387.1	30	3	30
Maximum Depth (m)	1.1	32.8	12.8	12.0	8.4	31	2	31
Mean Depth (m)	0.7	17.1	5.7	4.4	4.1	32	1	32
Land Cover								
Agriculture (%)	0.0	52.1	8.6	3.6	12.8	33	0	19
Forestry (%)	0.0	23.1	4.4	0.0	6.8	33	0	14
Pasture (%)	0.0	100.0	51.2	54.5	31.5	33	0	30
Peat (%)	0.0	100.0	19.3	3.1	31.7	33	0	18
Urban (%)	0.0	10.9	0.7	0.0	2.2	33	0	8
Hydrochemical								
Alkalinity (mg l^{-1})	-0.3	208.6	87.5	78.0	62.0	33	0	32
Chlorophyll a (µg l^{-1})	0.6	62.7	13.1	9.2	13.4	33	0	33
Colour (mg l^{-1} PtCo/Hazen)	1.0	208.5	57.8	36.0	49.6	31	2	31
Conductivity (µS cm^{-1})	40.0	462.0	251.7	259.0	107.3	32	1	32
pH	5.1	8.5	7.7	7.9	0.7	33	0	33
TP (µg l^{-1})	4.0	142.3	43.3	34.7	35.7	33	0	33

Most lakes have pH values of above 7 in the range of 7-8.5 (see Figure 5.1). The TP gradient is in the range of 4-142.3 µg l^{-1} for the 33 lakes, but most lakes have TP values below 100 µg l^{-1} (see Table 5.1 and Figure 5.1). The median TP value of 34.7 µg l^{-1} indicates that there are nearly as many oligotrophic and mesotrophic lakes as the eutrophic and hypertrophic ones in the Cladocera training set. Alkalinity has a relatively wide range from -0.3 to 208.6 mg l^{-1} $CaCO_3$, with median and mean values of 78.6 and 87.8 mg l^{-1} $CaCO_3$ respectively (Table 5.1). Most lakes fall into the conductivity range of 100-400 µS cm^{-1} and the median conductivity value of 259 µS cm^{-1}. The majority of the environmental variables display strong skewness to right in the frequency histograms (Figure 5.1).

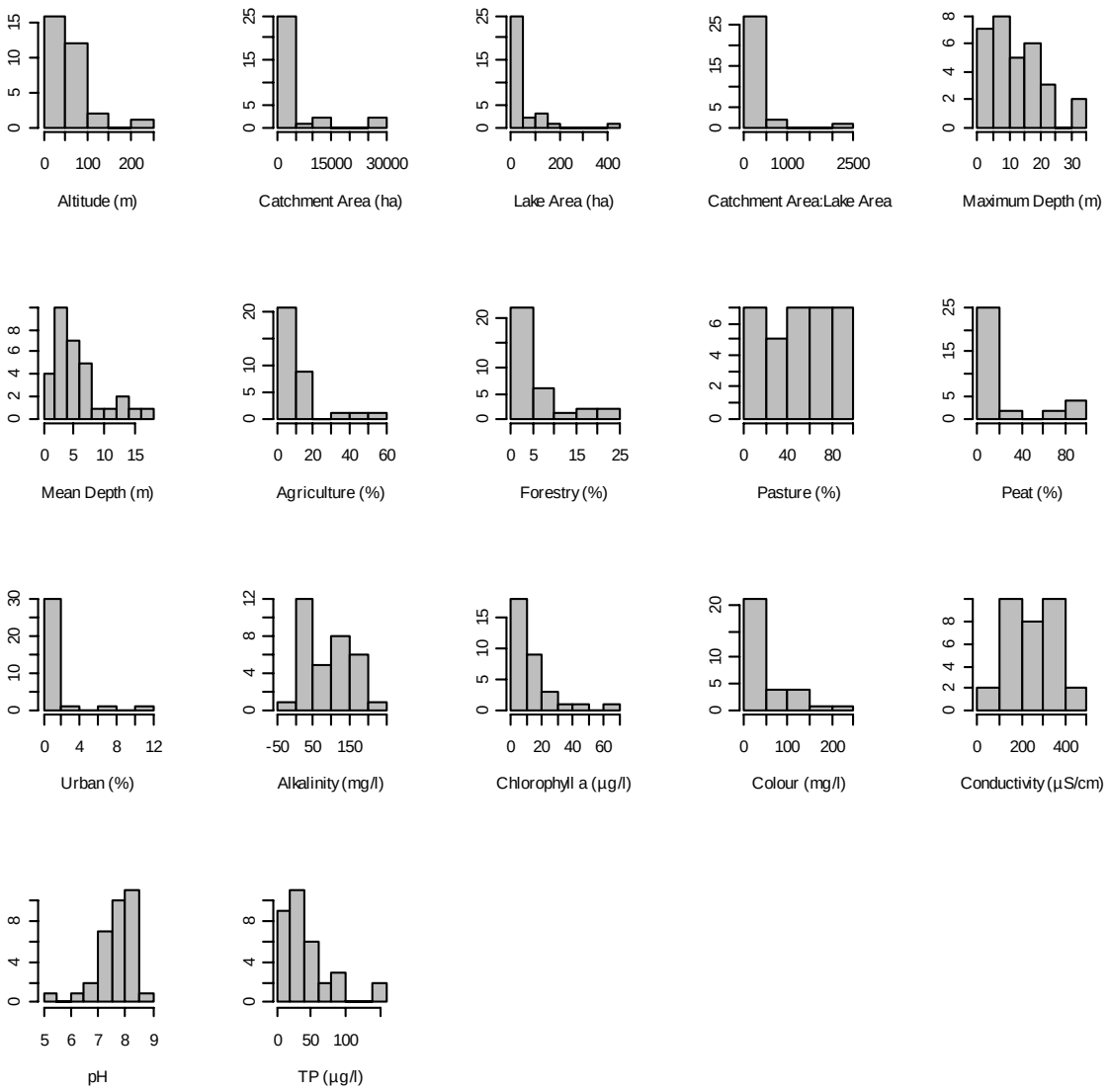

Figure 5.1 Histogram of 17 environmental variables in the Cladocera training set

5.1.1 Ordination of Environmental Variables

To further examine the data structure and the underlying gradients in the environmental and ecological variables, multivariate methods were applied in the Cladocera training set. Principal Component Analysis (PCA) was used to explore the physical, chemical and land use data. Normalising transformation of the environmental variables was performed to stabilize the variance prior to PCA. All the physical variables except maximum depth were $\log_{10}$-transformed, while TP, alkalinity, chlorophyll-*a* and peat coverage were $\log_{10}(1+)$-transformed as they contained values of either zero or between 0 and 1. All the other variables were square root transformed as they better approximate the normal distribution.

A correlation biplot of the first two PCA axes for all 17 environmental variables in the 33-lake Cladocera training set is shown in Figure 5.2 and the summary statistics of the first four PCA axes in Table 5.2. The broken stick model revealed that the first three PCA axes were significant enough for further exploration (Jackson, 1993). The first three principal components explained 66.8% of the total variance in the environmental variables while the first two axes accounted for 54.7% of the total variance (Table 5.2).

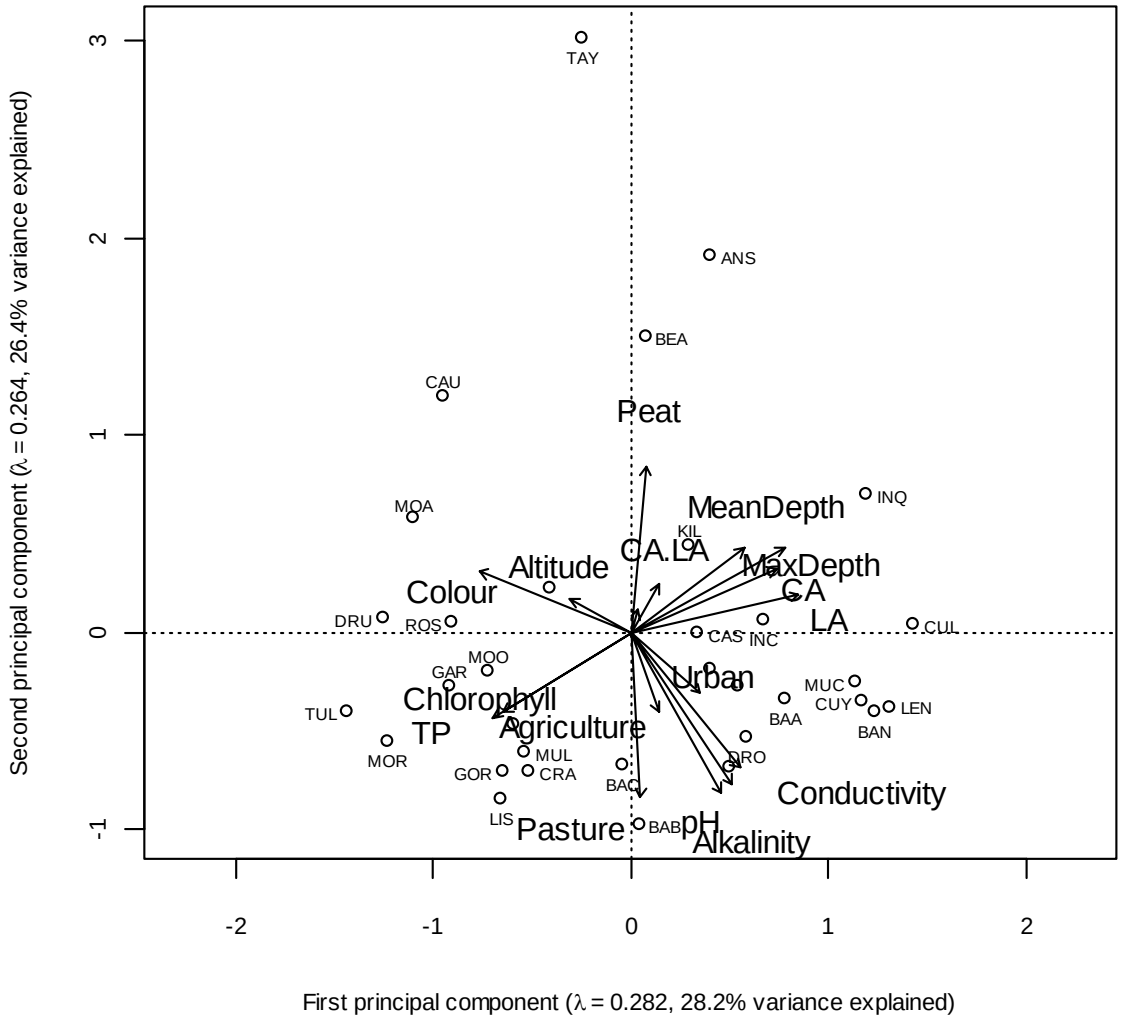

Figure 5.2 PCA correlation bilot for 17 environmental variables of 33 lakes.

The first principal component (axis one) accounted for 29.0% of the total variance in the environmental variables with an eigenvalue (λ_1) of 0.290. This axis is closely positively linked with lake area (0.989, the biplot score), maximum depth (0.915) and catchment area (0.893) and negatively with colour (-0.914) and TP (-0.792) (see Table 5.2). The angle between TP and chlorophyll-*a* was nearly zero indicating a significant

correlationship between both nutrient variables. The small angles among the variables related with lake depth and lake and catchment area also indicated their close covariance. These two clusters of nutrient and physical variables showed closely negative correlation: nutrient status (either TP or chlorophyll-*a*) decreases with the increase of lake depth or area and vice versa. Sites sitting along the vectors of depth and area like Inchiquin [INQ] are therefore deep and large lakes with large catchment areas but with low TP and chlorophyll-*a* values. While Tullabrack [TUL] and Morgans [MOR] sitting along the vectors of TP and chlorophyll-*a* are typically eutrophic and shallow lakes with very small lake and catchment areas.

Table 5.2 Summary statistics of PCA on the 17 environmental variables of 33 lakes

PCA Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.290	0.257	0.121	9.316	1
Variance (%)	29.0	25.7	12.1	9.3	
Cumulative Variance (%)	29.0	54.7	66.8	76.1	
Total Unconstrained Eigenvalue					1
Biplot scores for Environmental Variables					
Altitude	-0.395	0.177	-0.835	0.031	
Catchment Area (CA)	0.893	0.391	0.472	-0.192	
Lake Area (LA)	0.989	0.235	-0.150	0.263	
CA:LA	0.210	0.290	0.783	-0.521	
MaxDepth	0.915	0.507	-0.245	0.212	
MeanDepth	0.666	0.496	-0.470	0.316	
Agriculture	0.182	-0.463	0.052	-0.044	
Forestry	0.068	0.159	0.484	0.927	
Pasture	0.053	-0.984	-0.326	-0.019	
Peat	0.185	0.954	0.443	0.138	
Urban	0.417	-0.361	0.101	0.688	
Alkalinity	0.622	-0.897	0.245	-0.035	
Chlorophyll-*a*	-0.743	-0.472	0.273	0.437	
Colour	-0.914	0.447	0.318	0.091	
Conductivity	0.754	-0.719	0.286	-0.171	
pH	0.540	-0.950	0.042	0.038	
TP	-0.792	-0.511	0.350	0.253	

The eingenvalue (λ_2) for the second axis is 0.257 and this axis explained 25.7% of the total variance. The second axis was positively correlated with peat (0.954, biplot score) and negatively correlated with pasture, pH and alkalinity (-0.984, -0.950 and -0.897) (see Figure 5.2). Alkalinity and pH are highly correlated, while peat and pasture are strongly negatively correlated. Sites sitting along the peat vector like Tay [TAY] are

covered with high coverage of peatland and also have very limited coverage of pasture. Sites located along the pH and alkalinity vectors (e.g. Sillan [SIL]) are very alkaline, while sites at the opposite end of the same vectors (e.g. Caum [CAU]) are characteristic of being circumneutral in pH and having a very low alkalinity. The third axis has an eigenvalue (λ_3) of 0.121 and this explained 12.1% of total variance in the environmental variables. In the Table 5.2 we can see that the third axis is positively correlated with catchment area to lake area ratio (CA:LA, 0.783) and negatively correlated with altitude (-0.835).

5.2 Surface Sediment Cladocera Data

This section will first examine the reliability of surface sediment Cladocera samples in reflecting the modern Cladocera assemblage structure. The surface sediment Cladocera data from 33 lakes are then summarised and explored using numerical methods including indirect gradient analysis and cluster analysis.

5.2.1 Comparison of Modern and Surface Sediment Cladocera

Zooplankton often disarticulate into various fragments (e.g. carapaces, claws, postabdomens) after death and are then differentially preserved in lake sediments (Korhola & Rautio, 2001). Therefore the representation of live communities in sediment samples is a key to the use of sediment Cladocera in tracking the history of water quality and lake ecosystem change. A limited number of studies have been undertaken to verify the reliability of sediment Cladocera assemblages (e.g., Frey, 1960; Rautio *et al.*, 2000; Jeppesen *et al.*, 2003). In order to explore the data structure of surface sediment Cladocera assemblages, a comparison with modern Cladocera assemblages from the water columns is conducted to ascertain whether the fossil remains in the sediments are truly representative. A confirmation of close relationship between sediment and modern Claodcera assemblages from Irish lakes could provide a vigorous basis for constructing a surface sediment Cladocera-based TP transfer function. Littoral and planktonic Cladocera assemblages of six lakes are examined separately because of their differential preservation in the sediments and their distinctive habitat preferences across the whole lake area.

5.2.1.1 Littoral Cladocera

Modern water samples for six lakes (Cullaun, Dromore, Egish, Inchiquin, Lene and Mullagh) were taken for chydorid analysis by de Eyto (2000) and the data are used here to compare with the chydorid assemblages from the surface sediments of the same lakes (see Table 5.3). Chydorids are mainly littoral dwellers in diverse habitats and their abundance and community structure often display strong seasonality. Monthly sampling for modern chydorid analysis were conducted for Loughs Inchiquin and Lene and quarterly sampling for the other four lakes between July 1996 and September 1997 in three main littoral habitats (rocks, plants and littoral waters) (de Eyto, 2000). Surface sediment samples were collected in the summer of 2004 for these six lakes. Surface sediments were accumulated during the past several years and therefore fossil chydorid assemblages from the six lakes should be comparable with the modern samples.

Thirty chydorid species were observed in both surface sediment and multiple water samples of all the six lakes (see Table 5.3). The highest species richness (21) occurred in the surface sediment of Cullaun while the lowest (6) was found in the water samples of Egish. The average species richness in surface sediment and water samples for all six lakes were 17.3 and 14.7, with a median value of 16.5 and 16 correspondingly. In four of the six lakes there was higher species richness in the surface sediment compared to the water samples and only in Lough Dromore were there two more species found in water column than in sediments. Species richness in Dromore (19) was also the highest among the 26 lakes analysed by de Eyto (2000). Species richness in one-off sampling of surface sediments is generally higher than that in multiple water samples for the same lake (see Table 5.3). Surface sediments captured a minimum of two-thirds of total number of species in the water samples with an average value of 80% for the 6 lakes. For example in Lough Egish, all the six speices counted in the seven water samples were identified in the surface sediment and furthermore 10 additional chydorid species occurred in the surface sediment. Therefore the surface sediments captured the main chydorid assemblages from different littoral habitats in the six lakes.

Table 5.3 Comparison of the presence (represented by the symbol ●) of 30 chydorid species in surface sediments (*S*) and water column (*W*) for 6 lakes (water column data from de Eyto (2000))

Lakes	**Cullaun**		**Dromore**		**Egish**		**Inchiquin**		**Lene**		**Mullagh**	
Species/Sample	***S***	***W***	***S***	***W***	***S***	***W***	***S***	***W***	***S***	***W***	***S***	***W***
Acroperus harpae	●	●	●	●	●	●	●	●	●	●	●	●
Alona affinis	●	●	●	●	●	●	●	●	●	●	●	●
Alona costata	●	●	●	●			●	●	●	●	●	●
Alona guttata	●		●	●	●		●		●		●	●
Alona intermedia	●						●		●			
Alona quadrangularis	●		●		●		●	●	●	●	●	●
Alona rectangula	●	●	●	●	●		●	●	●	●	●	●
Alona rustica			●		●	●					●	
Alonella excisa	●	●	●		●		●	●	●			
Alonella exigua		●	●	●			●	●	●			
Alonella nana	●	●	●	●	●		●		●	●	●	●
Alonopsis elongata	●	●			●							
Anchistropus emarginatus	●	●		●								
Camptocercus rectirostris	●		●	●	●		●					●
Chydorus piger	●				●		●				●	
Chydorus sphaericus	●	●	●	●	●	●	●	●	●	●	●	●
Eurycercus lamellatus	●	●	●	●	●	●	●	●		●	●	
Graptoleberis testudinaria	●	●	●	●	●		●	●			●	●
Leydigia acanthocercoides									●			
Leydigia leydigiii					●						●	●
Monospilus dispar	●	●	●				●	●		●	●	
Oxyurella tenuicaudis		●			●	●						
Phrixura rostrata	●	●						●	●	●		●
Pleuroxus aduncus				●								
Pleuroxus laevis	●		●	●			●	●	●	●		
Pleuroxus trigonellus	●	●	●	●			●	●	●	●	●	
Pleuroxus truncatus		●		●			●	●		●		
Pleuroxus uncinatus	●			●				●			●	●
Pseudochydorus globosus				●				●		●		
Rhynchotalona falcata		●		●						●		
Species richness (No. captured in both samples)	**21 (14)**	**18 (14)**	**17 (13)**	**19 (13)**	**16 (6)**	**6 (6)**	**19 (14)**	**17 (14)**	**15 (10)**	**15 (10)**	**16 (11)**	**13 (11)**

5.2.1.2 Planktonic Cladocera

Planktonic Cladocera occurring in Irish lakes are mainly composed of taxa from two families, Daphniidae and Bosminidae (Irvine *et al.*, 2001). In comparison with Chydoridae the preservation of planktonic Cladocera is usually poor in lake sediments except for Bosminidae (Korhola & Rautio, 2001). Daphniidae, a key component of aquatic ecosystem, are poorly preserved because their exoskeleton is too fragile to

survive decomposition and attack by fungi. Post-abdominal claws are normally the most abundant Daphniidae remains in the sediments. As a result the Daphniidae species is difficult to distinguish based mainly on post-abdominal claws and can only enable the identification of two species groups, *Daphnia pulex* group and *D. longispina* group. The modern zooplankton data were sampled 4-8 times in the Summer of 1996 in open waters of the same 6 lakes as used in the previous subsection (Irvine *et al.*, 2001).

Table 5.4 Comparison of the presence of planktonic Cladocera taxa (represented by the symbol ●) in surface sediment (*S*) and water samples (*W*) for six lakes (double symbols (●●) indicate the dominant taxa; data of water samples from Irvine *et al.* (2001)).

Lake	Cullaun		Dromore		Egish		Inchiquin		Lene		Mullagh	
Taxa/Sample	*S*	*W*	*S*	*W*	*S*	*W*	*S*	*W*	*S*	*W*	*S*	*W*
Bosmina longirostris	●	●	●	●	●	●	●		●	●	●●	●●
Bosmina longispina	●						●		●		●	
Ceriodaphnia quadrangula								●				
Ceriodaphnia pulchella												●
Daphnia longispina group	●●	●	●●	●●	●●	●●	●●	●●	●●	●●	●●	●●
Daphnia pulex group			●				●					
Diaphanosoma brachyurum		●										
Bythotrephes longimanus		●						●		●		
Leptodora kindti	●		●	●	●					●	●	●
Sida crystalline			●		●						●	
Latona setifera	●											

In total nine species and two species groups occurred in all samples of the six lakes and *Daphnia longispina* group and *Bosmina longirostris* occurred in most of the samples (see Table 5.4). The planktonic Cladocera assemblages are generally composed of 1-4 species in addition to one or two species groups and their species richness are relatively low in comparison with the species richness of 15-20 for chydorids in the six lakes. *D. longispina* group was dominant in both surface sediments and water samples of all six lakes except in water samples from Lough Cullaun. *B. longirostris* was sub-dominant in sediment and water samples of Mullagh. The dominant taxa in water samples were well captured in the surface sediment samples in all lakes except for Culluan where there were no clearly dominant taxa in water samples (Table 5.4). However, few other species or groups corresponded well with each other in water and sediment samples. Only *Leptodora kindti* was recorded in both types of samples in two lakes (Dromore and Mullagh). This may be due to not only the occurrence of differential preservation in the

sediment samples but also the spatial heterogeneity of the modern planktonic Cladocera distribution in the water column.

5.2.2 Surface Sediment Cladocera

In total 39 Cladocera taxa and groups were counted in the surface sediments of 33 Irish lakes and their summary statistics are shown in Table 5.5. Of the 39 taxa and groups only eight were planktonic and all the other 31 taxa or species group were littoral chydorids. *Acroperus harpae*, *Alona affinis*, *A. guttata/rectangula* group, *A. quadrangularis* and *Alonella nana* occurred in all 33 lakes while *Latona setifera*, *Pleuroxus denticulatus* and *Leydigia acanthocercoides* only occurred in one or two lakes (see Table 5.5). Planktonic Cladocera *Bosmina longirostris* and *Daphnia longispina* group were the most dominant taxa in most of the lakes with the highest maximum abundances (87% and 77% respectively), highest mean values (17% and 30%) and also highest standard deviations (21 for both). Among the littoral Cladocera (Chydoridae) *Alona guttata/rectangula* group, *Chydorus sphaericus*, *Alona affinis*, *A. quadrangularis* and *Alonella nana* are the most dominant taxa and they generally have mean values of relative abundance in the range of 4-8% for all the lakes (Table 5.5). Some common species like *Acroperus harpae*, *Alonella exigua* and *Camptocercus rectirostris* had relatively low values of maximum relative abundance (generally <6%) while some uncommon species like *Alona rustica* occurred with a maximum relative abundance of above 22%.

The distribution of 31 common Cladocera taxa with occurrence in at least two sites and maximum relative abundances of above 1% are shown in Figure 5.3. *Bosmina longirostris*, a small planktonic cladoceran, has the highest abundance of over 80% at Caum [CAU] in this Cladocera training set and is also dominant in abundances of over 40% at Rosconnell [ROS], Ballybeg [BAB], Rushaun [RUS], Mullagh [MUL] and Effernan [EFF]. The other *Bosmina* species in this training set, *B. longispina* is much less common and is most abundant (over 30%) at Tay [TAY]. The species replacement of *B. longispina* by *B. longirostris* has been well documented as a consequence of nutrient enrichment (e.g. Deevey & Deevey, 1971; Boucherle & Züllig, 1983). High abundance of bosminids can occur when fish predation is intense due to their small size (Dodson & Frey, 2001). *Daphnia* is the other common planktonic Cladocera genus in

this study and are composed of often abundant *D. longispina* group and much rare *D. pulex* group. Large-sized *D. pulex* group only occurs at Crans [CRA] in abundance of above 10%. In comparison *D. longispina* group has abundances of above 20% at 21 of the 33 sites and is dominant at Crans, Inchiquin [INQ], Morgans [MOR], Garvillaun [GAR], Mooghna [MOO], Dromore [DRO], Gortaganniv [GOR] and Bane [BAN] in abundances of over 50%. Large-sized *Daphnia* is sensitive to selective predation of fishes and high abundance of *Daphnia* is often found in lakes lack of or with low density of planktivorous fishes (Brooks & Dodson, 1965; Jeppesen *et al.*, 1996; Lampert & Sommer, 1997). It is often rare or absent in highly acid lakes (Nilssen & Sandoy, 1990; Caroni, 2001)

Littoral *Acroperus harpae* is often found in lakes with low production and alkalinity (Whiteside, 1970; Lotter *et al.*, 1998) and associated with macrophytes (Freyer, 1968). It has abundances of above 5% only at sites Drumanure [DRU] and Atedaun [ATE]. *Alona affinis* and *A. quadrangularis* are common taxa in Ireland and were observed to have similar distribution (Duigan, 1992). *A. affinis* and *A. quadrangularis* are both cosmopolitan species inhabiting the bottom muds of lakes (Freyer, 1968). *A. affinis* is abundant at Muckanagh [MUC], Beaghcauneen [BEA] and Kiltooris [KIL] in this study. While *A. quadrangularis* occurs more often in lakes with high production (Whiteside, 1970) and it is most abundant at Castle [CAS] and Tullabrack [TUL] in abundances of above 10%. Both *Alona guttata* and *A. rectangula* are very common and widespread chydorids in Ireland (Duigan, 1992). *A. rectangula* is often found in lakes with high production while *A. guttata* is found in nutrient-poor and acidic waters (Whiteside, 1970; Brodersen *et al.*, 1998), but *A. guttata* was also found to survive nutrient enrichment (Duigan, 1992). The often indistinguishable headshields of both species in sediments lead to a combined species group and it is most abundant at sites Lisnahan [LIS], Tullabrack [TUL] and Lene [LEN] in abundances of over 20%. *Alona intermedia*, often found in acid-oligotrophic waters (Whiteside, 1970; Duigan, 1992), is most abundant at Lene [LEN] in abundance of ca. 15% but much less abundant at other sites like Beaghcauneen [BEA] and Moanmore [MOA]. *Alona rustica* is most abundant at sites Anascaul [ANS], Drumanure [DRU] and Tay [TAY] in abundances of above 10%. This is a typical species of oligotrophic lakes with low alkalinity in Europe (Whiteside, 1970; Duigan, 1992; de Eyto *et al.*, 2003). Among the three *Alonella* species all of which occurred in most of the sites, *A. excisa* and *A. nana* are generally more abundant than *A. exigua*. *A. excisa* often has high abundance in acidic and nutrient-poor lakes (de Eyto *et*

al., 2003) and is most abundant at sites Anascaul [ANS], Atedaun [ATE] and Tullabrack [TUL] in abundances of above 5%. *A. nana* was found to have similar ecological preferences as A. *excisa* except that *A. nana* prefers higher nutrient levels (Whiteside, 1970; Duigan, 1992).

Table 5.5 Summary statistics of 39 Cladocera taxa or species groups counted in surface sediments of 33 lakes (N = number of occurrence; N2 = Hill's effective number of occurrence; max, mean and median refer to maximum, mean and median relative abundance (%); SD = standard deviation).

Taxa	N	N2	Max	Mean	Median	SD
Acroperus harpae	33	19.91	5.80	1.87	1.25	1.52
Alona affinis	33	18.38	18.24	4.10	3.44	3.65
Alona costata	28	16.08	4.48	1.15	0.71	1.17
Alona guttata/rectangula group	33	22.04	26.79	8.47	7.69	5.97
Alona intermedia	21	5.88	14.53	1.18	0.63	2.54
Alona quadrangularis	33	22.40	15.03	4.70	3.74	3.23
Alona rustica	14	5.08	22.38	2.23	0.00	5.22
Alonella excisa	27	14.23	9.79	2.25	0.90	2.58
Alonella exigua	28	16.18	4.31	1.11	0.79	1.13
Alonella nana	33	21.71	11.63	4.21	3.91	3.03
Alonopsis elongata	8	6.39	1.39	0.15	0.00	0.30
Anchistropus emarginatus	5	4.44	0.86	0.09	0.00	0.23
Camptocercus rectirostris	23	13.59	2.78	0.59	0.40	0.70
Chydorus piger	26	9.63	12.50	1.97	0.60	3.08
Chydorus sphaericus	32	19.26	21.24	6.13	4.27	5.18
Eurycercus lamellatus	26	15.14	4.42	1.00	0.63	1.08
Graptoleberis testudinaria	31	18.06	9.03	2.35	1.65	2.14
Leydigia acanthocercoides	2	1.98	0.70	0.04	0.00	0.15
Leydigia leydigii	16	7.97	4.17	0.50	0.00	0.88
Monospilus dispar	18	10.23	2.80	0.44	0.16	0.65
Oxyurella tenuicaudis	4	2.45	2.49	0.13	0.00	0.47
Phrixura rostrata	16	11.28	2.18	0.35	0.00	0.49
Pleuroxus aduncus	7	6.42	0.88	0.12	0.00	0.24
Pleuroxus denticulatus	1	1.00	0.41	0.01	0.00	0.07
Pleuroxus laevis	17	10.51	2.00	0.35	0.06	0.52
Pleuroxus sp.	4	3.22	0.88	0.06	0.00	0.18
Pleuroxus trigonellus	22	14.15	2.65	0.60	0.41	0.70
Pleuroxus truncatus	15	13.11	0.72	0.19	0.00	0.23
Pleuroxus uncinatus	16	9.39	2.89	0.37	0.00	0.59
Rhynchotalona falcata	8	5.78	2.10	0.22	0.00	0.49
Bosmina longirostris	28	12.74	86.77	16.94	5.07	21.37
Bosmina longispina	18	8.86	33.73	4.37	1.03	7.22
Daphnia longispina group	32	22.00	77.35	29.71	26.67	21.01
Daphnia pulex group	13	4.94	15.16	1.16	0.00	2.76
Ilyocryptus silvaeducensis	4	3.12	0.67	0.05	0.00	0.16
Leptodora kindti	10	5.78	3.47	0.34	0.00	0.74
Latona setifera	1	1.00	0.44	0.01	0.00	0.07
Sida crystalline	15	9.39	1.88	0.26	0.00	0.42

.

Two *Chydorus* species are common in most of the 33 lakes but *C. sphaericus* is generally more abundant and common than *C. piger*. Despite its cosmopolitan distribution in different lake types, *C. sphaericus* often has the highest abundance in productive lakes with high alkalinity (Whiteside, 1970; Brodersen *et al.*, 1998; de Eyto *et al.*, 2003). It is a littoral dweller mainly associated with macrophytes and can occur in open water though its structure and adaption are not those of a planktonic animal (Freyer, 1968; de Eyto, 2000). This species has the highest abundance (around 20%) at Ballyallia [BAA] and is also abundant at Ballyteige [BAT], Ballybeg [BAB], Atedaun [ATE], Tullabrack [TUL] and Lisnahan [LIS] in abundances of above 10%. In comparison *C. piger* occurs most frequently on sand-rock substrates (Duigan & Kovach, 1991) and is often found in lakes with low productivity and alkalinity (Whiteside, 1970; Duigan, 1992). It has high abundance of above 10% at Beaghcauneen [BEA] and Moanmore [MOA].

Eurycercus lamellatus is a large-sized littoral cladoceran mainly associated with macrophytes (Duigan, 1992) and is often found in lakes with low productivity and alkalinity (Whiteside, 1970). This species occurred in abundances of less than 5% in this study and is most abundant at Ballyallia [BAA] and Kiltooris [KIL]. *Graptoleberis testudinaria* occurres frequently on macrophytes and was usually found in nutrient-rich waters (Duigan, 1992) but it has also occurred in less productive lakes (Brodersen *et al.*, 1998). It was abundant at sites like Kiltooris [KIL], Tullabrack [TUL], Ballyteige [BAT], Lisnahan [LIS] and Ballycar [BAC] in abundances of above 5%. *Leydigia leydigii*, a benthic-dwelling species, is typically found in nutrient-rich waters with high alkalinity (Whiteside, 1970; Duigan, 1992). It is not abundant in this study and was found at Lisnahan [LIS] and Tullabrack [TUL] in abundances of less than 5%. Six species of the littoral *Pleuroxus* genus were identified from the 33 sites with the numbers of occurrence ranging from one to 22. They also occurred in very low abundances of less than 3% in all samples.

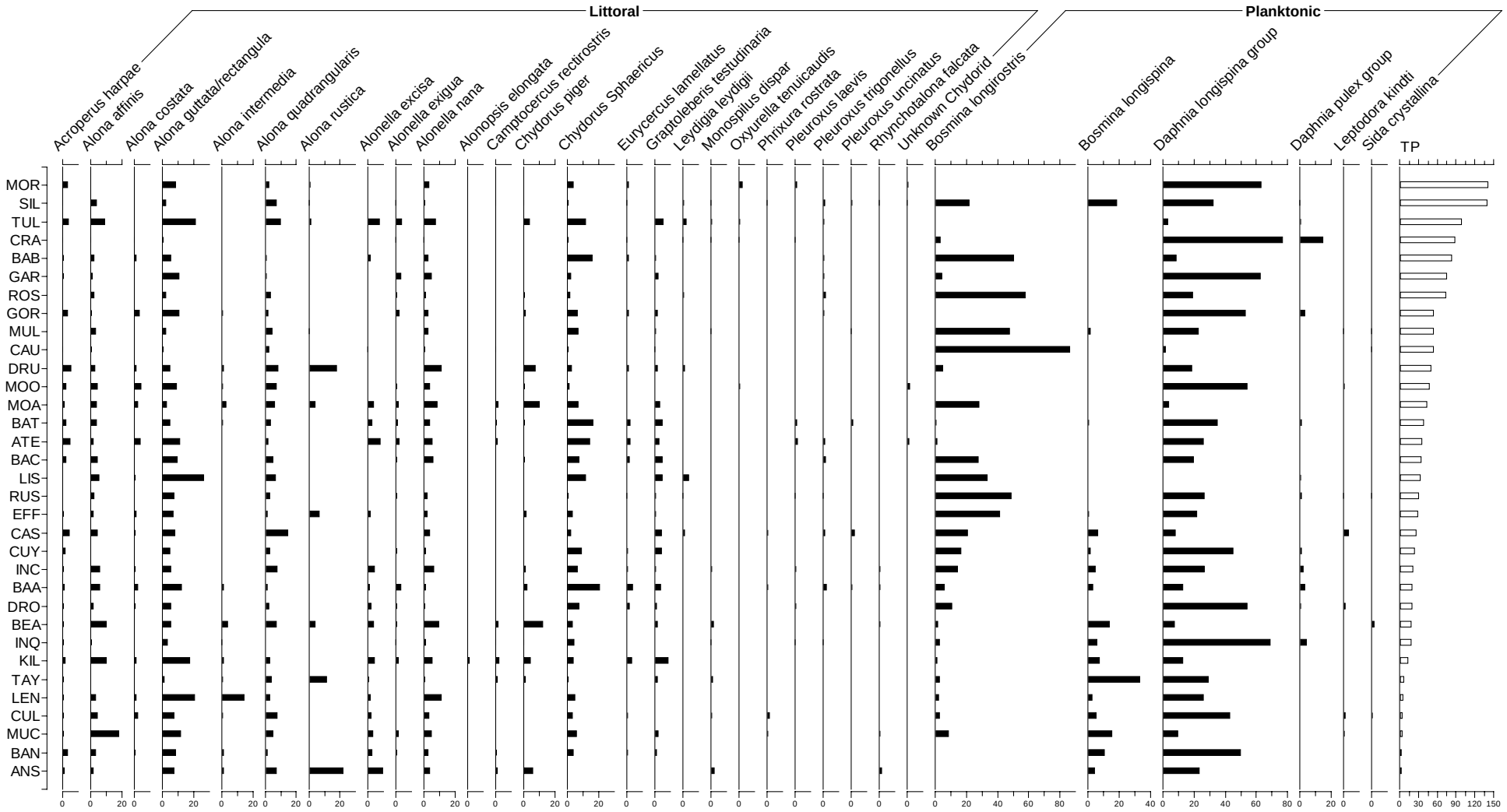

Figure 5.3 Distribution of 31 common Cladocera taxa (maximum abundance at least 1%, 2 sites) along the TP (µg l^{-1}) gradient of 33 lakes (taxa are listed in alphabetic sequence in either planktonic or littoral group; lakes are ordered according to their TP values with the lowest TP at the bottom).

5.2.3 Indirect Gradient Analysis of Cladocera Data

Gradient analysis techniques depend on the species response to environmental gradients. Detrended Correspondence Analysis (DCA) was used to determine the gradient length of the Cladocera data before performing the analysis. The length of the first axis was 1.83 SD units with rare species downweighted. This implied that the Cladocera taxa are generally behaving monotonically along the gradient and statistical techniques based on linear response model is more appropriate as the gradient length is less than 2 SD units (Birks, 1995). Therefore Principal Component Analysis (PCA) is used here to examine the structure of Cladocera data. Summary statistics of PCA on Cladocera data is shown in Table 5.6. The first five axes are statistically significant under the broken stick model. The first four axes explained 51.2% of the total variance of the Cladocera data in the 33-lake training set and the first and second axes explained 18.2% and 13.3% respectively. PCA plots of species and site scores are displayed in Figure 5.4 and Figure 5.5.

Table 5.6 Summary statistics of PCA on Cladocera data in the 33-lake Cladocera training set.

PCA Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.182	0.133	0.101	0.095	1
Variance (%)	18.2	13.3	10.1	9.6	
Cumulative Variance (%)	18.2	31.5	41.6	51.2	
Total Unconstrained Eigenvalue					1

Selected Cladocera taxa with high species scores are listed along the first five axes in Table 5.7 to aid the interpretation of the PCA plot of species in Figure 5.4. More Cladocera taxa are located on the right side of the plot than the left side and also the lengths of vectors on the right side are generally longer than those on the left indicating stronger species turnover between sites. Most taxa close to the second axis are located in the lower part of the plot with long vectors. The first axis is positively correlated with Chydordiae taxa like *Alonella excisa* (0.735, species score), *Camptocercus rectirostris* (0.705) and *Chydorus piger* (0.700), and negatively related to planktonic taxa like *Daphnia pulex* group (-0.489), *D. longispina* group (-0.423) and *Bosmina longirostris* (-0.356) (see Table 5.7 and Figure 5.4). Several planktonic taxa are also important components along other PCA axes, including *Leptodora kindti* (-0.751) in the third axis and *D. longispina* group (-0.696) and *B. longirostris* (0.407) in the fourth axis (Table 5.7). Some important chydorid taxa for these axes are *Chydorus Sphaericus*, *Alona*

rustica, *Monospilus dispar*, *Eurycercus lamellatus*, *Alonella exigua*, *Leydigia leydigii* and *Oxyurella tenuicaudis*.

Table 5.7 Species scores of selected Cladocera taxa along the first five PCA axes (taxa are ordered acording to their species cores)

Code	Taxon	Score	Code	Taxon	Score
	PCA axis 1			PCA axis 3	
ALOEXC	*Alonella excisa*	0.735	ALOEXI	*Alonella exigua*	0.397
CAMREC	*Camptocercus rectirostris*	0.705	ALORUS	*Alona rustica*	0.330
CHYPIG	*Chydorus piger*	0.700	CHYPIG	*Chydorus piger*	0.324
ALONAN	*Alonella nana*	0.688	PHRROS	*Phrixura rostrata*	-0.640
ALOAFF	*Alona affinis*	0.682	PLEUNC	*Pleuroxus uncinatus*	-0.693
OXYTEN	*Oxyurella tenuicaudis*	-0.160	LEPKIN	*Leptodora kindti*	-0.751
LEYLEY	*Leydigia leydigii*	-0.202			
BOSLOR	*Bosmina longirostris*	-0.356		PCA axis 4	
DAPLOG	*Daphnia longispina* group	-0.423	LEYLEY	*Leydigia leydigii*	0.694
DAPPUG	*Daphnia pulex* group	-0.489	BOSLOR	*Bosmina longirostris*	0.407
			GRATES	*Graptoleberis testudinaria*	0.341
	PCA axis 2		PLELAE	*Pleuroxus laevis*	-0.428
ALORUS	*Alona rustica*	0.645	UNKCH	Unknown Chydorid	-0.439
MONDIS	*Monospilus dispar*	0.611	DAPLOG	*Daphnia longispina* group	-0.696
RHYFAL	*Rhynchotalona falcata*	0.465			
BOSLOS	*Bosmina longispina*	0.351		PCA axis 5	
CHYPIG	*Chydorus piger*	0.220	OXYTEN	*Oxyurella tenuicaudis*	0.719
PLELAE	*Pleuroxus laevis*	-0.512	UNKCH	Unknown Chydorid	0.507
ALOGR	*Alona guttata/rectangula* group	-0.549	LEYLEY	*Leydigia leydigii*	0.407
ACRHAR	*Acroperus harpae*	-0.549	ALOEXI	*Alonella exigua*	-0.327
EURLAM	*Eurycercus lamellatus*	-0.609	GRATES	*Graptoleberis testudinaria*	-0.330
CHYSPH	*Chydorus Sphaericus*	-0.737	CHYSPH	*Chydorus Sphaericus*	-0.353

In PCA plot sites which are located close to each other have similar species compositions and therefore surface sediment Cladocera assemblages at sites like Garvillaun [GAR] and Lisnahan [LIS] have quite similar structure. Comparison of the site (Figure 5.5) and species vectors (Figure 5.4) is helpful with disclosing the assemblage structure. Sites close to a particular species vector are composed of that species with a higher than average relative abundance, while sites located in the opposite direction only have low abundances of that species. Loughs Anascaul [ANS] and Tay [TAY] on the top right of Figure 5.5 are composed of taxa with relative abundances much higher than averages, like *Alona rustica*, *Monospilus dispar* and *Rhynchotalona falcata* sitting at the top right of Figure 5.4. Sites close to the origin of the plot like Bane [BAN] and Lene [LEN] are composed of species in average abundances in Figure 5.5.

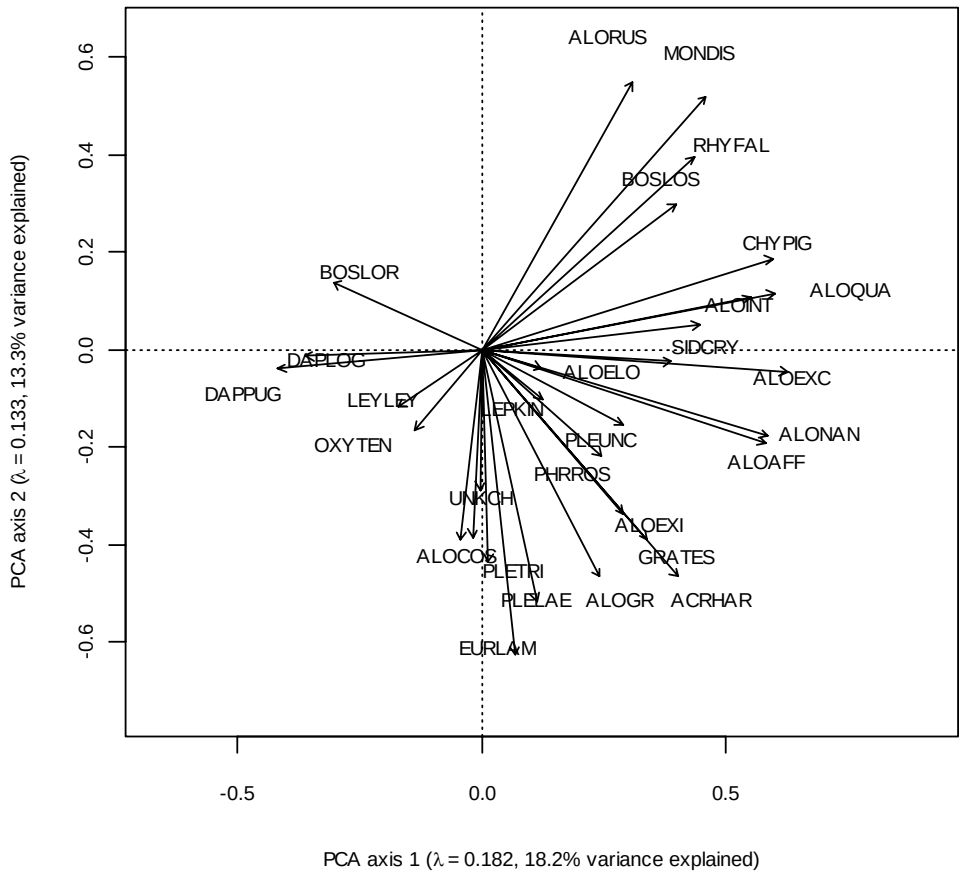

Figure 5.4 PCA of Cladocera data in the 33-lake Cladocera training set (only species are shown).

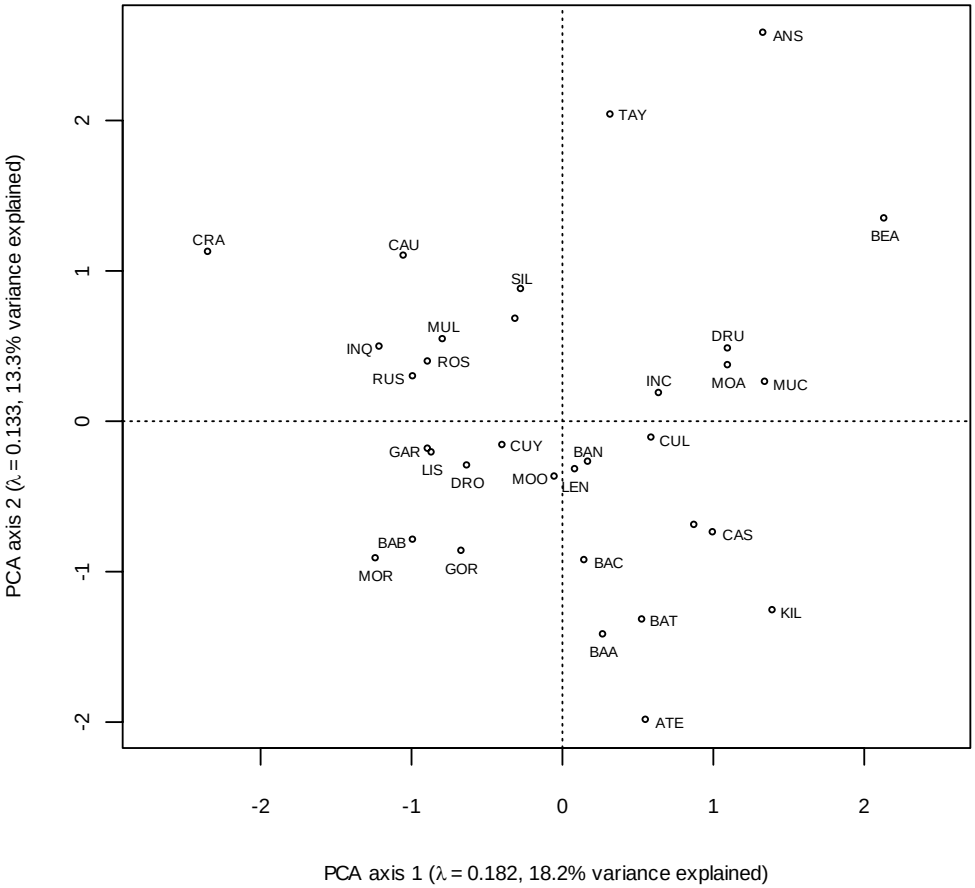

Figure 5.5 PCA of Cladocera data in the 33-lake Cladocera training set (only sites are shown).

5.3 Classification of Cladocera Assemblages

Surface sediment chydorid remains were first explored in conjunction with lake types by Whiteside (1970). Surface sediment Cladocera have also been used to examine Cladocera distribution and ecology (Duigan & Kovach, 1991; de Eyto & Irvine, 2002; de Eyto *et al.*, 2003; Simpson, 2005a) and their relationship with macrophyte communities (Duigan & Kovach, 1994). Cladocera are used here to evaluate lake typology classification. This section aims to classify the Cladocera data from 33 lakes and compare the biological clustering with the Irish Lake Typology classes. Planktonic Cladocera are not included in this cluster analysis as they are liable to factors like fish predation (Brooks & Dodson, 1965; Jeppesen *et al.*, 1996; Lampert & Sommer, 1997). Ward's minimum variance method, a kind of agglomerative cluster analysis, was used to extract the cluster structure in the chydorid data for the 33 lakes and four clusters of sites are produced in Figure 5.6.

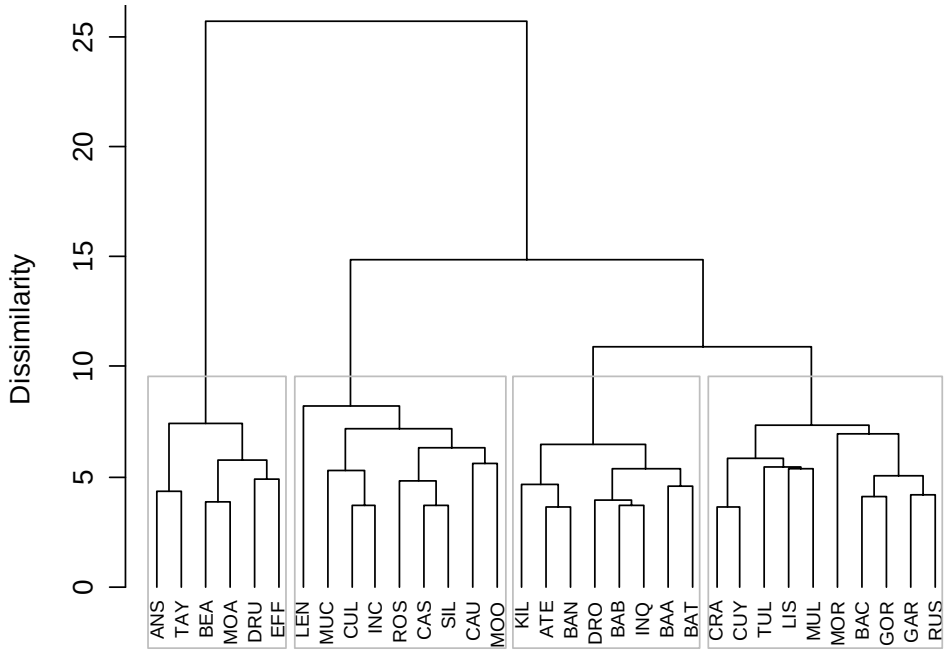

Figure 5.6 Dendrogram of surface sediment chydorid of 33 lakes using Ward's minimum variance method (Clusters 1 to 4 are framed from the left to the right and see Appendix A on site names).

A summary of the four chydorid clusters and related lake types are detailed in Table 5.8. *Alona guttata*/*rectangula* group and *A. quadrangularis* are predominant in the first two clusters. *A. guttata* and *A. rectangula* generally have distinct nutrient preferences, the former preferring nutrient-poor and acidic lakes and the latter preferring more

productive lakes (Brodersen *et al.*, 1998). *A. quadrangularis* was often found in a variety of waterbodies in Ireland (Duigan, 1992). The subdominant taxa in Clusters 1 and 2 represent two different habitats (see Table 5.8): *Alona rustica*, *A. intermedia*, *Chydorus piger* and *Alonella excisa* in Cluster 1 are indicative of an oligotrophic environment with low alkalinity; *Alona affinis*, *A. nana* and *Phrixura rostrata* in Cluster 2 are indicative of mesotrophic environment with higher alkalinity.

Table 5.8 Summary of chydorid clusters classified using Ward's minimum method and related lake types (B = benthic habitat, P = macrophyte-associated and S = semi-planktonic).

Chydorid cluster	Cluster description	Common taxa	Habitat	No. of lakes	Lake types
1	Acidophilous and oligotrophic taxa sub-dominant	*Alona guttata/rectangula*	B	6	Mainly low alkalinity, Mainly deep, Small
		Alona quadrangularis	B		
		Alona rustica	M		
		Chydorus piger	B		
		Alonella excisa	B		
		Alona intermedia	B		
2	Mesotrophic taxa subdominant, no acidophilous taxa	*Alona guttata/rectangula*	B	9	Mainly high alkalinity, Mainly deep, Mainly large
		Alona quadrangularis	B		
		Alona affinis	B		
		Alonella nana	B		
		Acroperus harpae	M		
		Phrixura rostrata	B		
3	Alkaline and meso-eutrophic taxa dominant with oligotrophic taxa subdominant (more macrophyte-associated taxa than Cluster 2)	*Chydorus sphaericus*	S & M	8	High alkalinity, Deep and shallow, Mainly small
		Alona guttata/rectangula	B		
		Graptoleberis testudinaria	M		
		Alona affinis	B		
		Alonella excisa	B		
		Eurycercus lamellatus	M		
4	Alkaline and meso-eutrophic taxa dominant, with eutrophic taxa subdominant	*Chydorus sphaericus*	S & M	10	Mainly medium alkalinity, Mainly shallow, Small
		Alona guttata/rectangula	B		
		Alonella nana	B		
		Graptoleberis testudinaria	M		
		Alona quadrangularis	B		
		Leydigia leydigi	B		

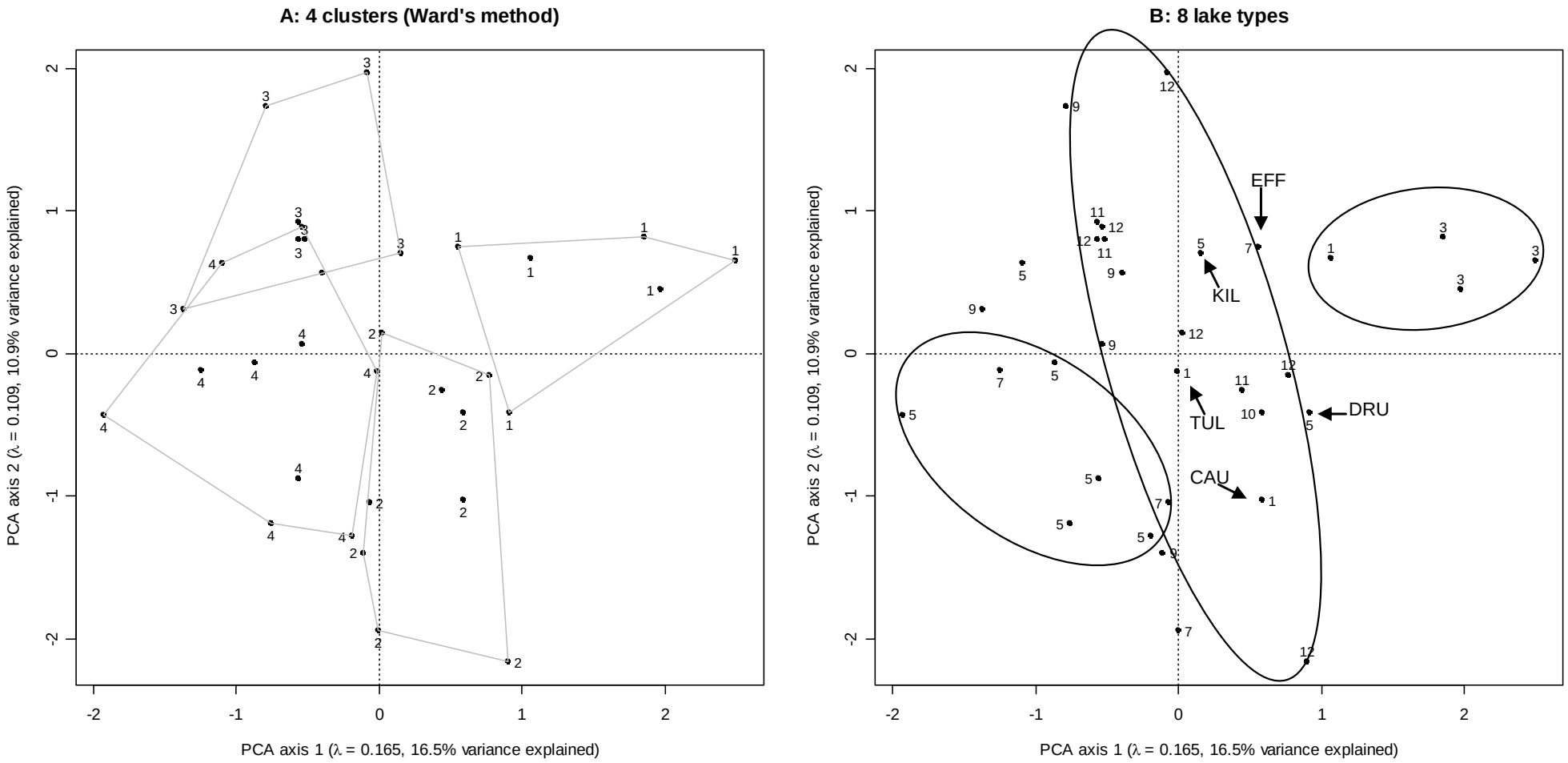

Figure 5.7 Comparison of four Ward's minimum clusters of littoral Cladocera (chydorids) (Figure A) and 8 EPA lake typology classes (Figure B) of 33 lakes in the Cladocera training set (both classification are superimposed on PCA plots of chydorid data; three ovals indicate three groups of lakes with medium, high and low alkalinities respectively from the left to the right of Figure B; the arrows point to misclassified sites based on chydorid data).

Both Cluster 3 and Cluster 4 are dominated by *Chydorus sphaericus*. A high abundance of this species often occurs in lakes with high productivity and alkalinity (Brodersen *et al.*, 1998) but it can be found in different lake types of Ireland (de Eyto & Irvine, 2002). *Alona guttata*/*rectangula* group and *Graptoleberis testudinaria* are also subdominant in both clusters and *G. testudinaria* was found to be abundant in nutrient-rich waters in Ireland (Duigan, 1992). The subdominant *A. nana* and *A. quadrangularis* in Cluster 4 are generally found in more productive lakes than the subdominant *A. excisa* and *A. affinis* in Cluster 3 respectively (Whiteside, 1970; Duigan, 1992). The subdominance of the eutrophic and alkaline *Leydigia leydigii* in Cluster 4 also indicates a productive lake environment while *Eurycercus lamellatus* in Cluster 3 is generally not abundant in productive waters (Whiteside, 1970; Brodersen *et al.*, 1998).

A comparison of the four chydorid clusters with the eight lake typology classes is superimposed on PCA plots of chydorid data (see Figure 5.7). The four chydorid clusters generally display distinct clustering from each other with slight overlap between Clusters 3 and 4 (Figure 5.7 A). Three lake groups with low, high and medium alkalinities are outlined in three ovals and they also show good homogeneity within each group (see Figure 5.7 B). Lakes with low alkalinity (Lake Types 1 and 3) correspond well with the chydorid Cluster 1, which is reflected by the sub-dominance of acidophilous and oligotrophic taxa (see Table 5.8). Lakes with medium alkalinity (Lake Types 5 and 7) are also consistent with the chydorid Cluster 4 (compare plots A and B of Figure 5.7). The other lake group with high alkalinity (Lake Types 9-12) corresponds well with the chydorid Clusters 2 and 3. Therefore the lake groups with different alkalinity are well reflected by the chydorid clusters and they show good concordance in the clustering configuration. However, several sites with low or medium alkalinity (e.g. Tullabrack [TUL] and Drumanure [DRU] as labelled in Figure 5.7 B) are located within the high alkalinity lake group and they show strong dissimilarity between the chydorid assemblage and related lake type. In addition the lake groups outlined in plot B is not ordered along the alkalinity gradient. The lake group in the middle of the PCA plot has high alkalinity while the lake groups with low and medium alkalinities display the biggest discrepancy between the three lake groups. This may indicate that other physico-chemical variables exert significant influence on the chydorid assemblages of lakes with medium or high alkalinity. Noticeably the high abundance of eutrophic taxa (e.g., *Chydorus sphaericus* and *Leydigia leydigii*) in chydorid Cluster 4 may indicate

that nutrient level may account for the considerable dissimilarity between chydorid Clusters 4 and 1 in addition to alkalinity (see Table 5.8).

5.4 Direct Gradient Analysis

In the previous section PCA was used to investigate the internal pattern of surface sediment Cladocera assemblages of 33 lakes. This section aims to examine the relationship between Cladocera assemblages and environmental variables and also select the most significant variables influencing the Cladocera assemblages. Redundancy Analysis (RDA) is used to constrain the ordination analysis of Cladocera data with 17 environmental variables.

5.4.1 Redundancy Analysis (RDA)

Summary RDA information for the first four axes is shown in Table 5.9 and RDA biplots of species and sites are displayed in Figure 5.8 and Figure 5.9 respectively. The eigenvalues for the first four axes were 0.149, 0.115, 0.086 and 0.057 and each axis explained 14.9%, 11.5%, 8.6% and 5.7% of the total variance in the Cladocera data. The four axes explained 40.7% of total variance of Cladocera assemblages in comparison with 51.2% explained by the first four PCA axes of Cladocera data indicating that the measured variables captured a large portion of the total variance in the Cladocera data (see Table 5.9).

Nutrient variables like chlorophyll-*a* and TP were strongly positively correlated with the first axis as indicated by their high scores of 0.648 and 0.556 (see Table 5.9). Alkalinity, pH and conductivity are strongly negatively correlated with the second axis and they have highly negative scores along this axis as shown in Table 5.9. *Leydigia leydigii*, *Oxyurella tenuicaudis*, *Bosmina longirostris*, *Daphnia pulex* group and *D. longispina* group are positively related with nutrient variables (see Figure 5.8) and they are generally in high abundances at sites close to the nutrient vectors including Cullaunyheeda [CUY], Garvillaun [GAR] and Dromore [DRO]. *Alonella excisa*, *Alona intermedia* and *Camptocercus rectirostris* are strongly negatively correlated (Figure 5.8) and they are abundant at sites Beaghcauneen [BEA], Anascaul [ANS] and Tay [TAY] (Figure 5.9). Several species like *Alona rustica*, *Monospilus dispar*, *Rhynchotalona*

alcate, *Chydorus piger* and *Bosmina longispina* are closely negatively correlated with both nutrient and acidity variables at sites like Ballybeg [BAB] and Morgans [MOR]. *Pleuroxus trigonellus* and *C. sphaericus* are positively linked with both conductivity and acidity gradients and they have high abundances at sites like Ballycar [BAC], Mooghna [MOO] and Ballyallia [BAA] (Figure 5.9). The high species-environment correlation scores also display strong association between species and the environment gradients for the first four RDA axes (see Table 5.9).

Table 5.9 Summary of RDA of Cladocera data constrained by 17 environmental variables.

RDA Axes	**1**	**2**	**3**	**4**	**Total Variance**
Eigenvalue (λ)	0.149	0.115	0.086	0.057	1
Species-Environment Correlation	0.925	0.938	0.954	0.842	
Cumulative Variance (%)	14.9	26.4	34.9	40.7	
Constrained Eigenvalue					0.611
Total Unconstrained Eigenvalue					0.389
Biplot scores for constraining variables					
Altitude	0.428	0.512	-0.126	0.423	
Catchment Area (CA)	-0.512	0.024	0.607	-0.139	
Lake Area (LA)	-0.315	0.223	0.616	-0.125	
CA.LA	-0.372	-0.191	0.199	-0.063	
Maximum Depth	-0.151	0.344	0.736	0.216	
Mean Depth	0.071	0.506	0.666	0.328	
Agriculture	0.257	-0.313	0.088	-0.084	
Forestry	0.419	0.050	0.253	-0.332	
Pasture	0.254	-0.679	0.022	0.115	
Peat	-0.400	0.529	0.187	-0.268	
Urban	0.146	-0.230	0.077	-0.054	
Alkalinity	0.172	-0.698	0.421	-0.115	
Chlorophyll-*a*	0.648	-0.199	-0.366	-0.233	
Colour	0.142	0.287	-0.386	-0.091	
Conductivity	-0.035	-0.583	0.410	-0.218	
pH	0.288	-0.604	0.349	-0.104	
TP	0.556	-0.330	-0.393	-0.062	

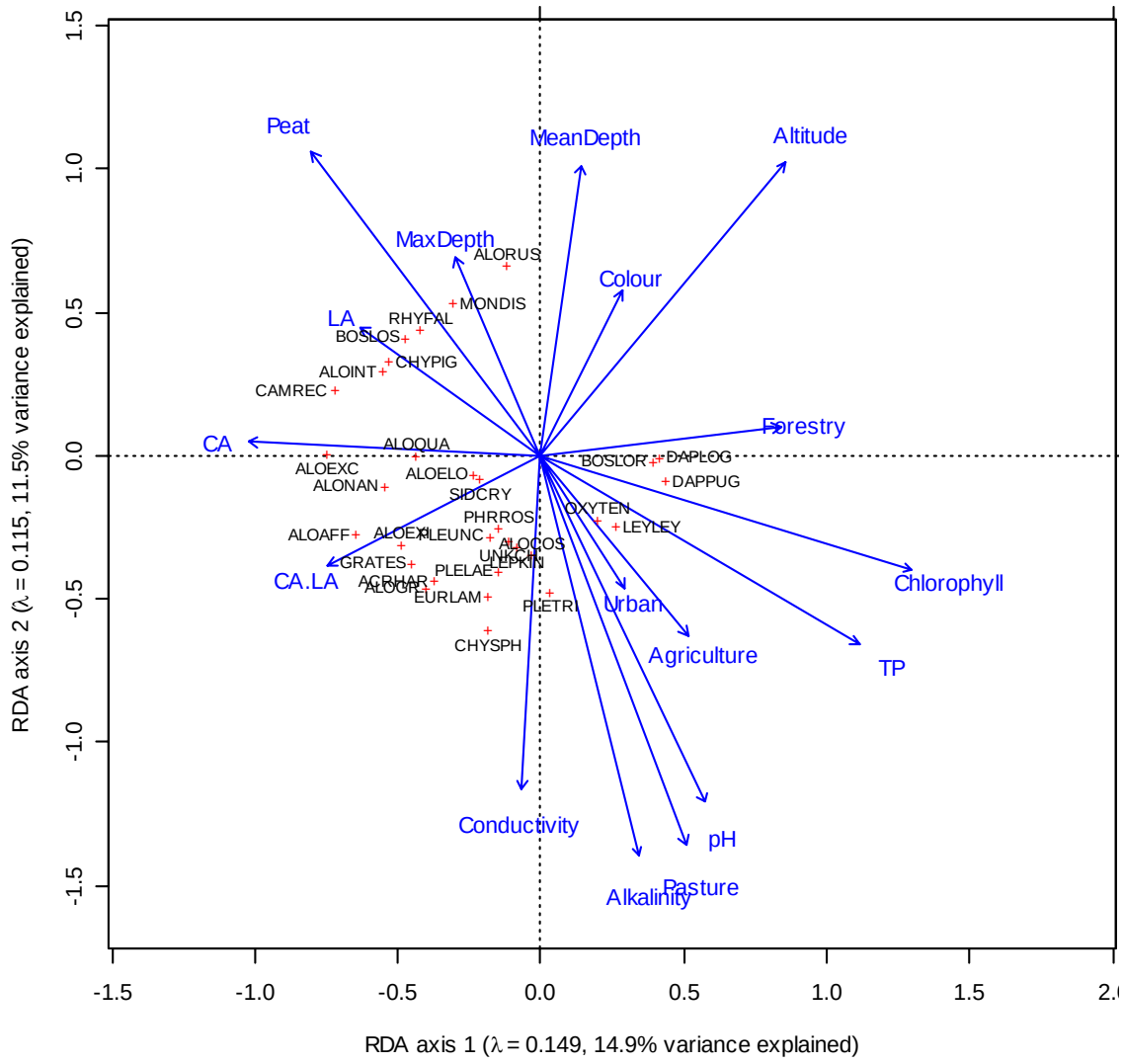

Figure 5.8 RDA biplot of Cladocera taxa and 17 environmental variables for the 33-lake Cladocera training set.

An obvious feature, when comparing the unconstrained PCA plot of species (Figure 5.4) and RDA plot of species and environmental variables (Figure 5.8), is that the configuration of Cladocera taxa remained very similar except the exchange of their positions along the second axis. Species located to the left side of the second axis in PCA have moved to the right side in RDA and vice versa. Also when we compare the configuration of environmental variables in RDA with PCA plot of environmental variables in Figure 5.2, the main pattern remains the same except that most environmental variables have moved to the opposite side along the second axis (Figure 5.2). This implies that the main pattern of Cladocera taxa revealed in PCA is captured in the RDA constrained by the measured environmental variables.

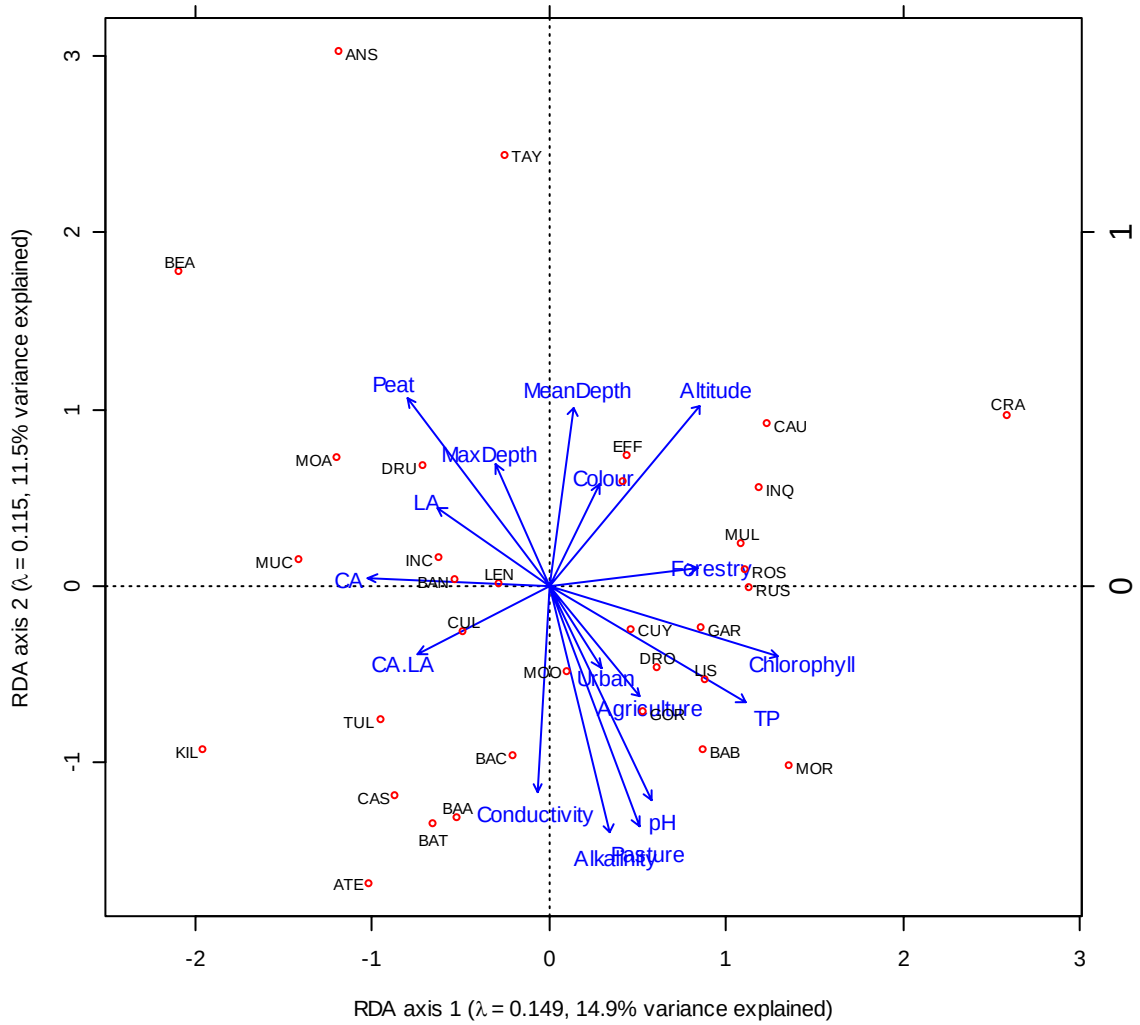

Figure 5.9 RDA biplot of sites and 17 environmental variables for the 33-lake Cladocera training set.

5.4.2 Selecting Significant Environmental Variables

This section aims to identify significant environmental variables influencing the Cladocera data among the 17 environmental variables. Partial RDA constrained by only one environment variable can provide information on the independent contribution of each variable to the total variance of the biological community. Summary information on partial RDAs of Cladocera data constrained by each of the 17 environmental variables is provided in Table 5.10 as well as the results of significance tests. Five variables, agriculture, catchment area/lake area ratio (CA:LA), colour, forestry and urban are not statistically significant at the P-level of 0.05 (see Table 5.10). Variance inflation factors (VIF) for the other 12 environmental variables (not shown here) are all below 20 indicating that each of the 12 variables have unique contribution to the total

variance. Altitude, catchment area, mean depth, pasture coverage, alkalinity, Chlorophyll-*a* and TP are statistically significant at the 99.9% level (see Table 5.10). Among the 17 environmental variables Chlorophyll-*a* makes the largest contribution (8.9%) to the total variance in the Cladocera data followed by alkalinity and TP (8.4% and 8.1% respectively).

Table 5.10 Summary of partial RDAs constrained by only one of the 17 environmental variables (Ratio of eigenvalue of the first axis (λ_1) to that of the second one (λ_2) and significance levels are also shown; variables were ordered according to their λ_1/λ_2 ratios; variables significant at the P-value of 0.001 are marked with * using 999 Monte Carlo permutation tests).

Variable	λ_1	λ_1/λ_2	% variance explained	P-value
Altitude	2.390	0.446	7.7	0.001*
Catchment Area (CA)	2.41	0.491	7.8	< 0.001*
Lake Area (LA)	2.01	0.372	6.5	0.006*
CA:LA	1.28	0.250	4.1	0.166
Maximum Depth	2.26	0.405	7.3	0.006
Mean Depth	2.4	0.426	7.7	< 0.001*
Agriculture	0.99	0.186	3.2	0.407
Forestry	1.51	0.300	4.9	0.068
Pasture	2.24	0.413	7.2	0.001*
Peat	2.21	0.426	7.1	0.003
Urban	0.75	0.137	2.4	0.734
Alkalinity	2.6	0.471	8.4	< 0.001*
Chlorophyll-*a*	2.75	0.608	8.9	< 0.001*
Colour	1.46	0.260	4.7	0.078
Conductivity	2.14	0.380	6.9	0.002
pH	2.31	0.431	7.5	0.002
TP	2.5	0.519	8.1	< 0.001*

Forward selection can help select significant environmental variables together with evaluation of their ecological significance (see Chapter 3 on methods). The nutrient gradient (Chlorophyll-*a* and TP) proves to be the most significant variables influencing the Cladocera assemblage along the first axis of RDA. Chlorophyll-*a* reflects the combined influence of nutrient variables on algae biomass because of its close and complex relationship with nutrient variables, particular phosphorus and nitrogen (Dillon & Rigler, 1974; Wetzel, 2001). In this 33-lake Cladocera training set a close positive relationship occurs between TP and Chlorophyll-*a* ($r = 0.751$ (Pearson), $P < 0.001$). However, Chlorophyll-*a* can also be strongly influenced by other factors like light availability (turbidity), resuspension, flushing and lake depth (Scheffer, 1998). This may account for a slightly bigger contribution of Chlorophyll-*a* (8.9%) to the total variance of Cladocera data than that of TP (8.1%) (see Table 5.10). Automatic forward selection

often omits those variables that are closely correlated with other variables even though they are ecologically significant. Therefore TP is manually selected as the first variable in the forward selection due to its ecological significance. Both manual and automatic forward selections were performed in the R program. Chlorophyll-*a*, maximum depth, alkalinity and altitude are the automatically selected variables and all four variables explained around 29.1% of the total variance in Cladocera data. The manually selected variables TP, maximum depth, alkalinity and catchment area explain 28.2% of the total variance. Both groups of variables account for nearly half of the total constrained variance (61.1%) by all 17 environmental variables in RDA (see Table 5.9). Both selections are acceptable as the minimum adequate models for this Cladocera training set in consideration of less than 1% difference in variance explained by both models.

The forward selected variables have a significant relationship with the Cladocera assemblages and can be used for calibration analysis for transfer function development (ter Braak, 1987b). The ratios of λ_1 to λ_2 for all variables are displayed in Table 5.10. Only Chlorophyll-*a* and TP produced λ_1:λ_2 ratios of above 0.5. Transfer functions with relatively low λ_1:λ_2 ratios of TP have been constructed, e.g. 0.42-0.44 for diatom-based TP transfer functions in Canada (Hall & Smol, 1992; Reavie *et al.*, 1995) and 0.48 (Amsinck *et al.*, 2005) and 0.72 (Lotter *et al.*, 1998) for Cladocera-based TP models. TP in the diatom training set of the current study (see Chapter 4) has a λ_1:λ_2 ratio of 0.673 and this is slightly higher than that (0.519) in this Cladocera training set. Cladocera-based training set can produce lower λ_1:λ_2 ratios in constrained ordination analysis than diatom-based one as diatoms are the primary producers in lake systems and can be more sensitive to and affected by the lake conditions than cladocerans, the secondary producers in lakes. The variance explained by TP (8.1%) is comparable with other Cladocera training sets, e.g. 8.0% (Lotter *et al.*, 1998), 5.8% (Amsinck *et al.*, 2005). Therefore TP is viable for further calibration analysis for developing transfer functions although other factors also influence the Cladocera assemblages.

5.5 Species Response Curves of Cladocera

This section explores the species response curves to highlight the response of individual Cladocera taxa to the environment gradients. This method can also aid the understanding of species response in direct gradient analysis. In comparison with the diatom responses explored in many studies (e.g., Birks *et al.*, 1990; Smilauer, 1995;

Anderson, 1997b), few such analyses have been conducted for Cladocera species (Davidson, 2005; Simpson, 2005a). Gaussian logit regression (GLR) method was applied to 20 common Cladocera taxa (maximum relative abundance above 3%, Hill's number above 10). Their distributions along TP, Chlorophyll-*a* and alkalinity are shown in Figure 5.10, Figure 5.11 and Figure 5.12 respectively.

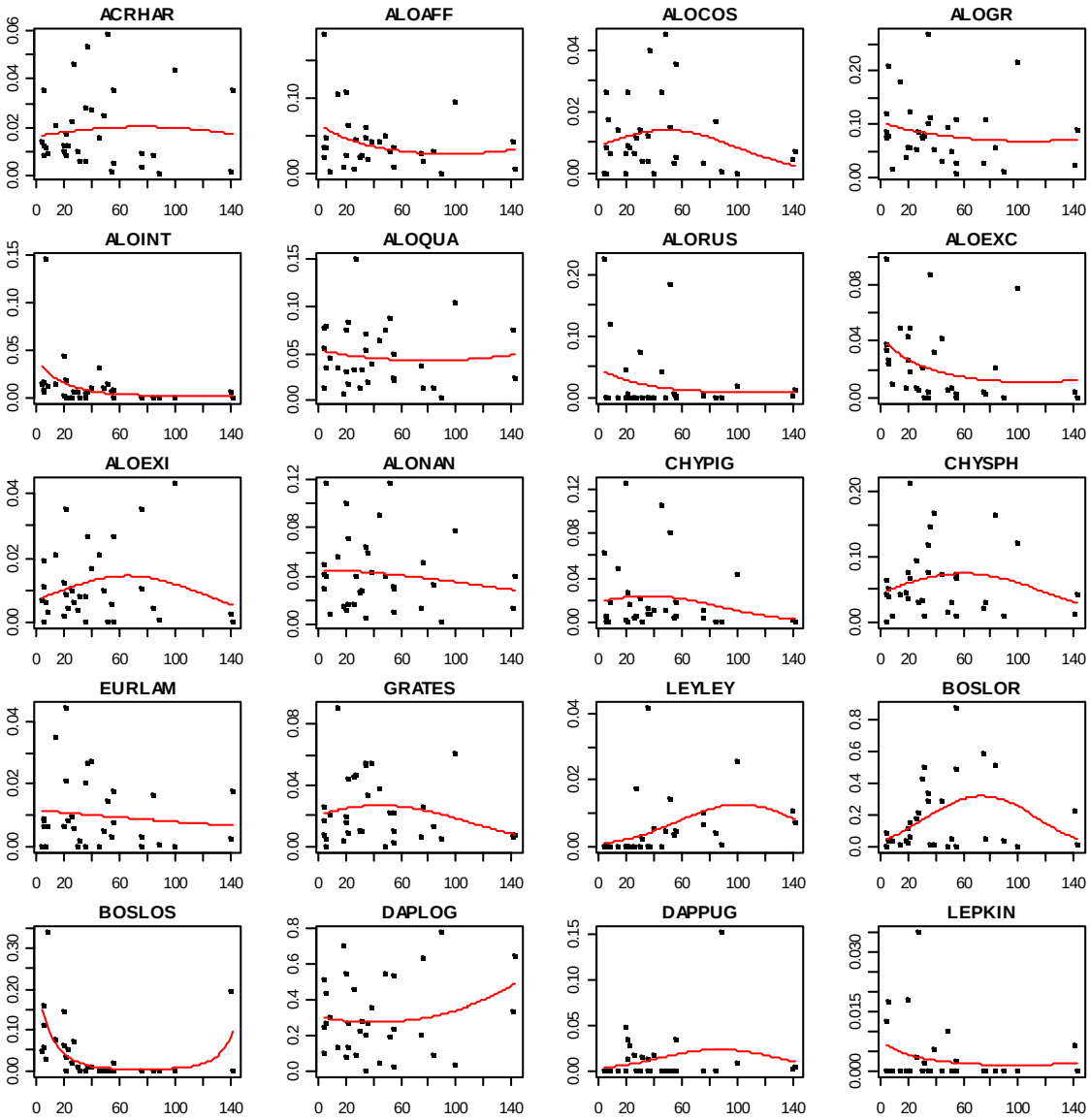

Figure 5.10 TP response curves of 20 common Cladocera taxa using Gaussian Logit Regression (GLR) (Horizontal axes represent TP (µg l^{-1}) and vertical axes represent the relative abundances (%); ACRHAR *Acroperus harpae*, ALOAFF *Alona affinis*, ALOCOS *Alona costata*, ALOGR *Alona guttata/rectangula* group, ALOINT *Alona intermedia*, ALOQUA *Alona quadrangularis*, ALORUS *Alona rustica*, ALOEXC *Alonella excisa*, ALOEXI *Alonella exigua*, ALONAN *Alonella nana*, CHYPIG *Chydorus piger*, CHYSPH *Chydorus Sphaericus*, EURLAM *Eurycercus lamellatus*, GRATES *Graptoleberis testudinaria*, LEYLEY *Leydigia leydigii*, BOSLOR *Bosmina longirostris*, BOSLOS *Bosmina longispina*, DAPLOG *Daphnia longispina* group, DAPPUG *Daphnia pulex* group, LEPKIN *Leptodora kindti*).

Littoral species like *Alona costata* [ALOCOS], *A. intermedia* [ALOINT], *Chydorus piger* [CHYPIG], *C. sphaericus* [CHYSPH] and *Leydigia leydigii* [LEYLEY] display strong responses to TP and chlorophyll-*a*. *A. intermedia*, *C. piger* and *Alonella excisa* [ALOEXC] responded to an increase in both nutrient variables with monotonically decreased abundance and showed a strong preference for waters poor in nutrients (Figure 5.10 and Figure 5.11). In comparison species like *C. sphaericus* and *A. costata* showed unimodal responses along the nutrient gradient: the abundance increases with the elevated nutrient level, maximises at a certain point and then decreases with a higher nutrient level. Planktonic *Bosmina longirostris* [BOSLOR] and *B. longispina* [BOSLOS] also display strong responses along the TP and chlorophyll-*a* gradients.

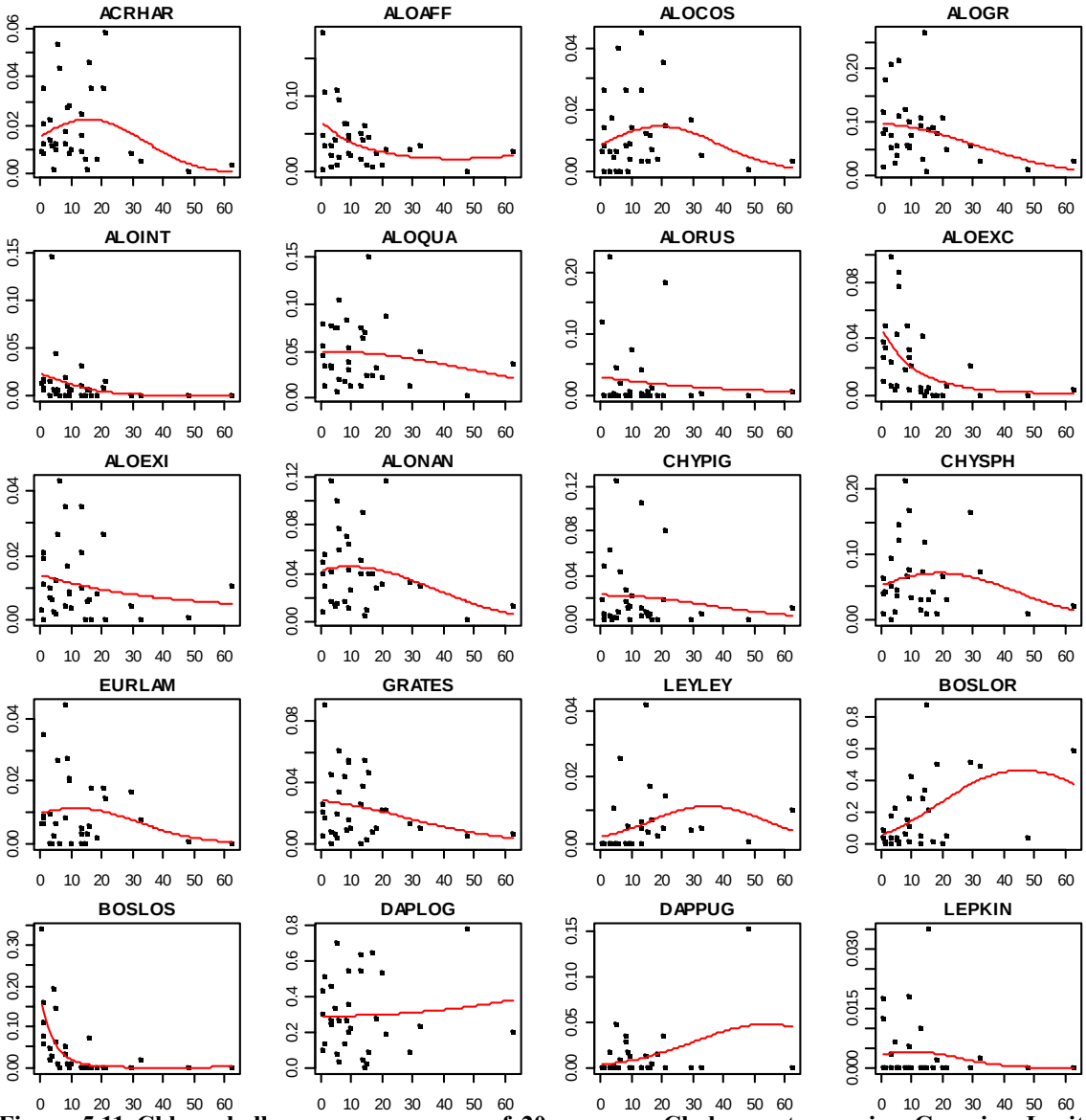

Figure 5.11 Chlorophyll-*a* response curves of 20 common Cladocera taxa using Gaussian Logit Regression (GLR) (Horizontal axes represent Chlorophyll-*a* (µg l^{-1}) and vertical axes represent the relative abundances (%); see Figure 5.10 for taxa names).

Acroperus harpae [ACRHAR], *Alonella nana* [ALONAN] and *Eurycercus lamellatus* [EURLAM] showed stronger responses to Chlorophyll-*a* than to TP. Abundances of *A. harpae* changed little with increased TP but showed a clear unimodal response to chlorophyll-*a*. *A. nana* and *E. lamellatus* display a truncated unimodal model along the chlorophyll-*a* gradient and a monotonical decline with increase in TP. In contrast *Alonella exigua* [ALOEXI] and *Graptoleberis testudinaria* [GRATES] display a unimodal response to TP but a monotonic decrease with an increase in chlorophyll-*a* (see Figure 5.10 and Figure 5.11). The species group [ALOGR] composed of *Alona guttata* and *A. rectangula* showed no clear response to TP but their abundance was relatively high in waters with low chlorophyll-*a* and decreased rapidly with increases in chlorophyll-*a*.

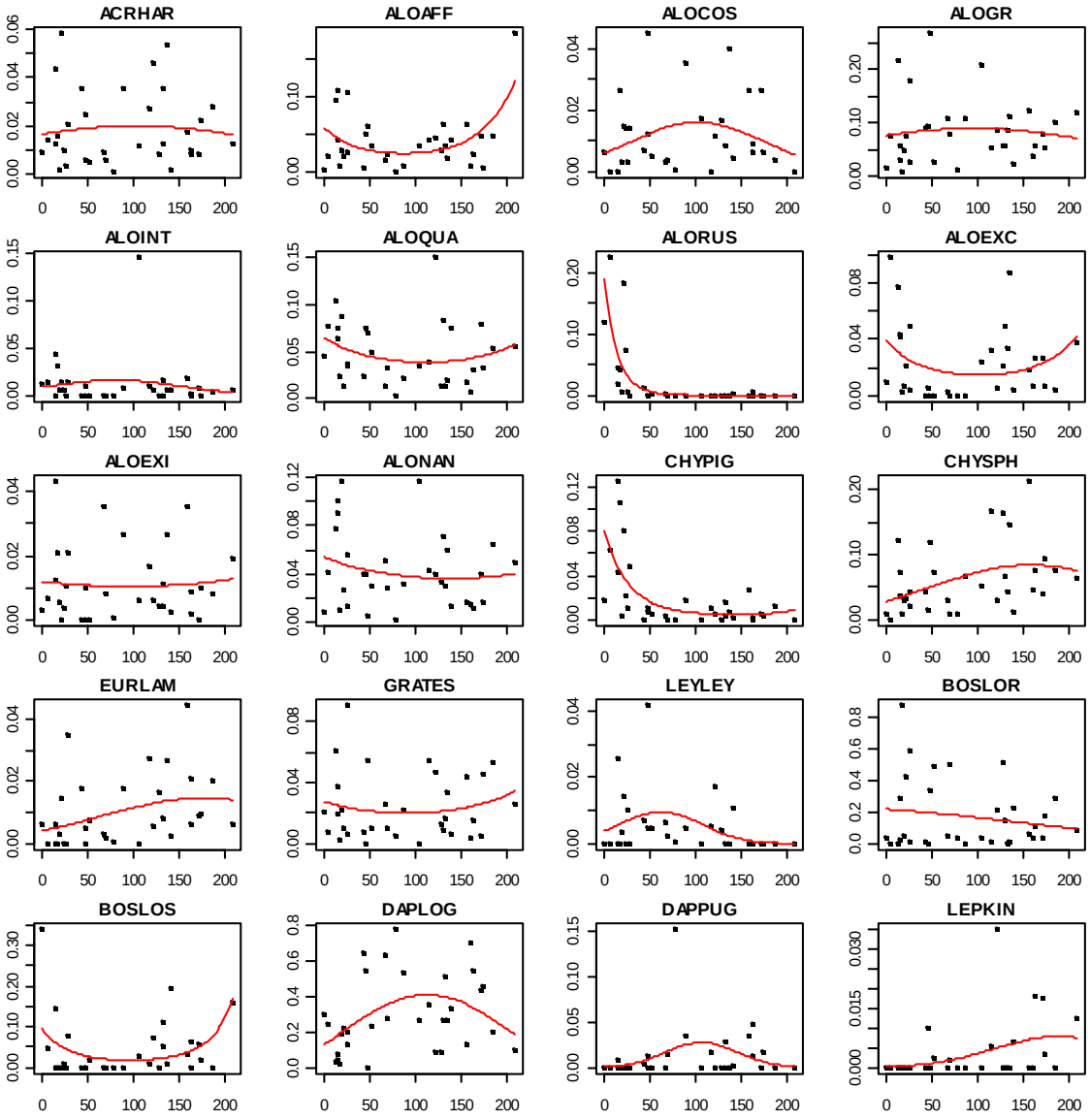

Figure 5.12 Alkalinity response curves of 20 common Cladocera taxa using Gaussian Logit Regression (GLR) (Horizontal axes represent alkalinity (mg l^{-1}) and vertical axes represent the relative abundances (%); see Figure 5.10 for taxa names).

Several species like *Alona rustica* [ALORUS] and *C. piger* were found to be more sensitive to alkalinity than nutrient levels. Both species show strong preferences for low alkalinity and their abundances decrease abruptly with an increase in alkalinity (see Figure 5.12). *L. leydigii* displays unimodal responses to both nutrient and alkalinity gradients: its abundance increases with elevated nutrients and alkalinity but decreases after certain threshold levels of nutrients and alkalinity are passed. A similar trend also occurs with the planktonic *Daphnia pulex* group [DAPPUG] and *Chydorus Sphaericus*. The response models of *D. longispina* group [DAPLOG] along TP and chlorophyll-*a* are much less clear when compared with the unimodal response to alkalinity. Many littoral species like *A. harpae*, *A. exigua* and *Alona quadrangularis* [ALOQUA] show little sensitivity to alkalinity and therefore their response curves are more or less flat along the alkalinity gradient (see Figure 5.12).

5.6 Discussion and Conclusions

5.6.1 Representation of Sediment Cladocera

The examination of similarity between sediment and live Cladocera assemblages can provide a strong basis for further exploration. Comparison of modern and surface sediment chydorid assemblages in six lakes revealed that relatively high species richness (in the range of 15-20) occurred in most of the surface sediment and all-year water samples of six Irish lakes. Surface sediments captured the main assemblage structure of live chydorids observed in water samples during the all-year investigation. In addition the more intensive sampling of water samples generally provided less species than the surface sediment sampling. This could be due to seasonal changes in live communities and spatial heterogeneity of chydorid habitats. Strong seasonal variations in chydorid species have been observed in the monthly and seasonal sampling of 29 Irish lakes (Irvine *et al.*, 2001). Different littoral species can have distinct habitat preferences, e.g. vegetation, rock, sand and mud (Freyer, 1968; Hofmann, 1987b; Hann, 1989). In contrast the surface sediments from the profoundal lake area may integrate the chydorids from different habitats of the whole lake. The exoskeleton fragments of littoral groups often occur abundantly in the deep-water sediments of lakes and therefore are highly representative of modern chydorid assemblages (Frey, 1960; Frey, 1988a). Planktonic *Daphina longispina* group were dominant in most of the sediment

and water samples of the six lakes. This dominant species group and the common species *Bosmina longirostris* in water samples are also recovered from the surface sediment samples. This corresponds well with findings from other studies on the good agreement between surface sediment records and contemporary data (Rautio *et al.*, 2000; Jeppesen *et al.*, 2003). Apparent dissimilarity occurred between the sediment and water samples for the uncommon planktonic taxa. The poor preservation of the planktonic Cladocera remains in the sediments may be partly responsible for the lack of correspondence of rare taxa between the sediment and water samples (Korhola & Rautio, 2001).

5.6.2 Cladocera Distribution and Classification

Thirty-nine Cladocera taxa and species groups were counted in this 33-lake Cladocera training set including 28 littoral species and one species group of chydords. This is comparable with contemporatory chydorid investigations of Irish waters. Thirty-one species of chydorids were found from 29 lakes of West and Central Ireland (de Eyto & Irvine, 2002) and 41 chydorid taxa were collected from 287 sampling sites across Ireland (Duigan & Kovach, 1991). This implies that the chydorid remains in the surface sediments of lakes can generally reflect the biogeography of the contemporary lake community. *Acroperus harpae*, *Alona affinis*, *A. quadrangularis* and *Alonella nana* were found to be the common taxa in Ireland during the two contemporary surveys and all of these species occurred in all of the 33 lakes in this study. *Chydorus sphaericus* is probably the most ubiquitous chydorid species in Ireland (Duigan & Kovach, 1991) and it was found in 32 of the 33 lakes in this study. It is abundant in lakes with high nutrient content and a high abundance of this species has indicated nutrient enrichment in Ireland (de Eyto & Irvine, 2002) and also other European ecoregions (e.g. Whiteside, 1970; de Eyto *et al.*, 2003). *C. sphaericus* is only absent in Lough Anascaul, the site with the lowest TP level (4 μg l^{-1}) in the Cladocera training set. This lake is mainly surrounded with peatland, has low pH and alkalinity. *Alona rustica* has the highest abundance in this lake. Duigan (1992) found that *A. rustica* often occurs in lakes poor in nutrients or with high coverage of peatland. Rare species in the modern investigations are generally also uncommon in the surface sediments, including *Leydigia Leydigii*, *Oxyurella tenuicaudis* and *Anchistropus emarginatus*. However, several species like *Alona intermedia*, *Alonella excisa* and *Graptoleberis testudinaria* were infrequent in the

modern surveys (e.g., in 5-9 of the 29 lakes by de Eyto & Irvine (2002)) but they were collected in 21-31 of the 33 lakes in this study. This suggests that the use of chydorid assemblages in surface sediments can be advantageous over the contemporary surveys of live communities. The results of modern surveys are often limited by the sampling time and sites while chydorids from different habitats can be continuously accumulated in surface sediment of deep waters.

Some disadvantages of surface sediment sampling compared to contemporary survey are apparent in the planktonic Cladocera assemblages. Six planktonic species and two species groups were collected from surface sediments the 33 lakes in this study. In comparison a 31-lake zooplankton survey (Irvine *et al.*, 2001) collected 16 species of planktonic Cladocera in Ireland. The low species diversity in the sediments can result from that the majority of the planktonic Cladocera have poor preservation in sediments (Rautio *et al.*, 2000). Postabdominal claws are the most abundant fragments of Daphniidae in surface sediments and only two *Daphina* species groups (*D. longispina* and *D. pulex* groups) can be identified for the Daphniidae family. In comparison six species of the *Daphina* genus and three species of the *Ceriodaphina* genus were identified for the Daphniidae family in the contemporary survey (Irvine *et al.*, 2001). Poor preservation of planktonic Cladocera in lake sediments and their low taxonomic resolution are the main clogs for them to be reliable indication of the original live community. Predation by fishes can also selectively eliminate the planktonic taxa like large-sized *Sida crystalline* and *Daphnia pulex* group and therefore reduce their preservation in sediments (Lampert & Sommer, 1997; Jeppesen *et al.*, 2002). The lack of fish density data in this study prevented the assessment of the predation impacts on planktonic Cladocera.

Thirty-three lakes were classified into four clusters in terms of chydorid assemblages. All the four clusters conformed well to the physico-chemical lake typology classes of low, medium and high alkalinities. However, in the clusters dominated by *Chydorus sphaericus*, an indicator species of alkaline and meso-eutrophic lakes, the subdominant species were different with one group containing eutrophic species like *Leydigia leydigii* and the other including mesotrophic species like *Alonella excisa*. Differences in species nutrient preference may account for the inconsistence of the lake groups along the alkalinity gradient as the group with high alkalinity was placed in the middle of the groups with low and medium alkalinities in the PCA of chydorid data. The high

abundance of eutrophic species in lakes of medium alkalinity influenced its position in PCA away from the group with low alkalinity. This result was expected as lakes were selected along the TP gradient for this 33-lake training set and lakes with high TP were included. In summary the chydorid clustering is comparable with the lake typology classes mainly based on alkalinity. Therefore alkalinity had strong ecological response in the surface sediment chydorids. However, only 33 lakes were included in this study covering eight of the 13 Irish lake types and sampling additional sites with a full coverage of the Irish lake types would help fully assess the chydorid clustering and lake typology. In addition, the exclusion of degraded lakes can improve the assessment as chydorids are strongly influenced by the human disturbance, like nutrient enrichment. Biological responses in different lake of 709 Danish lakes along the TP gradient were evaluated and most biological indicators were found to respond strongly to the increase in TP for different lake types (Søndergaard *et al.*, 2005). Therefore TP can potentially be used to categorize the lake types in implementing the Water Framework Directive.

5.6.3 Cladocera-Environment Relationships

The relatively short DCA gradient length of 1.8 SD for this 33-lake Cladocera training set is comparable with several Cladocera training sets, e.g. 0.9 for Cladocera in 83 lakes (Simpson, 2005a). 1.5 for littoral Cladocera in 69 lakes (Lotter *et al.*, 1998) and 2.0 for littoral Cladocera in 32 lakes (Brodersen *et al.*, 1998). This implies that the Cladocera structure is relatively different between sites and species turnover along underlying environmental gradients is relative strong. But this gradient length of less than 2 SD indicates that most species respond linearly to underlying ecological gradients (ter Braak, 1987b). Longer ecological gradients were found in other studies, e.g. 3.1 for zooplankton (mainly composed of Cladocera) in 36 lakes (Amsinck *et al.*, 2005), 3.0 for Cladocera in 28 lakes (Sweetman & Smol, 2006). All the measured variables explained 61.1% of the total variance of Cladocera data in constrained ordination and therefore the main pattern of community structure is captured. TP, Chlorophyll-*a*, alkalinity, lake depth, altitude and catchment area played important roles in explaining the total variance of the Cladocera data.

TP was also found to be one of the most significant variables in determining the surface sediment Cladocera community in other studies (e.g. Brodersen *et al.*, 1998; Bos &

Cumming, 2003). The strong correlation between TP and Chlorophyll-*a* in this study has long been observed in many studies (e.g. Dillon & Rigler, 1974; OECD, 1982). Phosphorus can influence the food quality (e.g. algae) of Cladocera and also directly affect the body growth of planktonic Cladocera (Sterner & Hessen, 1994; Urabe *et al.*, 1997). A change in nutrient levels can cause the loss of macrophyte habitat, increased fish predation and deficiency of oxygen at the sediment-water interface and all these can affect the Cladocera community directly (Jeppesen *et al.*, 2001; Vadeboncouer *et al.*, 2003). In this study a large portion of the total variance in Cladocera data was explained by Chlorophyll-*a* and TP. Around seven to nine of the 20 common Cladocera taxa displayed unimodal or unimodal-like responses along the TP or Chlorophyll-*a* gradient. All these confirmed that the nutrient variables had significant influences on the Cladocera assemblages in this study. The importance of alkalinity (or other acidity variables) in Cladocera distribution was reported in many studies (e.g., Whiteside, 1970; Krause-Dellin & Steinberg, 1986; Uimonensimola & Tolonen, 1987). The acid-tolerant species *Bosmina longirostris* displayed a decreased abundance with increase in alkalinity and *Alona rustica* and *Chydorus piger* were typically acidophilous taxa in this study. However, only around five of the 20 common taxa displayed unimodal- or unimodal-like response along the alkalinity gradient and this implied that alkalinity had a weaker effect on the species than the nutrient variables. It was found that changes in species abundance and community structure may be indirectly affected by acidity (Nilssen & Sandoy, 1990). These changes can be directly influenced by acidity-related factors like predation and macrophyte habitat (Steinberg *et al.*, 1988; Korhola & Rautio, 2001).

Three physical variables, altitude, lake depth and catchment area, were identified as important factors determining the Cladocera assemblages. The altitudinal distribution of Cladocera has been observed in modern surveys and the influence of altitude was often related to air temperature (e.g. Green, 1995). The strong temperature-Cladocera relationship has enabled the construction of two temperature transfer functions based on sediment Cladocera (Lotter *et al.*, 1997; Korhola, 1999). However, temperature may not account for all the variation in Cladocera community explained by altitude. A decrease in pasture land coverage was found with the increase in altitude in this training set and therefore altitude may influence the Cladocera community through related change in land use type. The change in macrophyte cover and predator-prey relationships along an altitude gradient may also indirectly affect the Cladocera assemblages. Therefore more

work is needed before a clear altitude-Cladocera relationship can be built. As littoral Cladocera dwell mainly in the shallow part of the lake and planktonic Cladocera in the open water, it is not unexpected that a high abundance of chydorids can be found in shallow lakes and abundant planktonic Cladocera in deep lakes. Lake depth has been recognized as an important factor for surface sediment Cladocera distribution in several studies (Bos & Cumming, 2003; Simpson, 2005a; Sweetman & Smol, 2006). A quantitative inference model has developed for lake depth using the Cladocera assemblages (Korhola *et al.*, 2000). However, transfer functions based on variables of intrinsic morphometric features of each lake like water depth can be problematic as the variable may not be appropriate for a space-for-time substitution in environmental reconstruction (Birks, 1998). Therefore lake depth is not considered for further development of transfer functions in this study. The influence of catchment area on Cladocera data may be exerted through the hydro-chemical variables and land cover variables. A large catchment area can provide more nutrients for a lake which is mainly surrounded by pasture lands or more humic materials if peatland is the main land cover type. In either way the catchment area can change the hydro-chemical conditions and therefore affect the Cladocera assemblages.

5.6.4 Conclusions

Cladocera (particularly Chydoridae) remains in surface sediments are a faithful indication of the modern community of the whole lake and can therefore be reliably used in tracking the history of lake environment. This can have important implications in water quality monitoring as the one-off surface sediment sampling can provide a more thorough investigation than multiple water sampling methods in recovering Cladocera community. Nutrient variables, alkalinity, altitude, catchment area and lake depth all had significant influences on the Cladocera assemblages. Strong ecological response of chydorids to alkalinity was also evident in the good agreement between the chydorid clusters and the physico-chemical lake typology classification. TP accounted for a large portion of the total variance of Cladocera data and both community structure and individual species displayed strong responses along the TP gradient. TP therefore can be used for constructing vigorous transfer functions for this Cladocera training set. In all Cladocera can serve as an independent and reliable indicator for TP like diatoms as illustrated in the previous chapter.

Chapter 6: Construction and Evaluation of TP Transfer Functions

This chapter employs calibration methods in constructing TP inference models for diatoms and Cladocera in the Irish Ecoregion by quantifying the responses of both indicators along a TP gradient. A useful guide for deciding whether to use linear or unimodal-based modelling methods is based on the length of the first axis of Detrended Canonical Correspondence Analysis (DCCA) constrained by TP only. A short gradient length of less than 2 SD suggests that linear modelling methods are appropriate while a gradient length longer than 2 SD suggests unimodal-based methods. This criterion has been widely used in transfer function development (e.g. Philibert & Prairie, 2002; Miettinen, 2003; Werner & Smol, 2005). Unimodal-based methods were also found to perform well in the case of a short gradient length (<2 SD) by some studies (e.g. Lotter *et al.*, 1998; Dalton, 1999). It is recommended that the comparative studies between the linear and unimodal-based methods be taken while applying this criterion (Birks, 1995).

The transformation of environment and ecological data can critically influence model performances (Birks, 1995), but selection of optimal transformation has been rarely considered in model development (e.g. Cumming & Smol, 1993; Koster *et al.*, 2004). Both untransformed and log-transformed TP and untransformed and square root transformed ecological data will be explored in the model development. Also both jack-knifing and bootstrapping are practiced for model cross-validation. Performances of all the data transformation and model cross-validation are assessed before an optimal format is selected for model development. In addition, the removal of outlier sites in the training set can also improve the model performance (e.g. Hall & Smol, 1992; Gasse *et al.*, 1995). This chapter will present a wide range of transfer functions in detail as there has been insufficient presentation of transfer function performance results in model selection (H.J. Birks, personal communication). The constructed TP models are then evaluated through comparing model performances and TP optima of the same taxa with those from other published models. In addition diatom- and Cladocera-inferred TP for the same 29 sites are compared to assess both TP models. All the inference models are developed using the software C2 (version 1.4.2) (Juggins, 2003).

6.1 Diatom-based TP Transfer Functions

The diatom training set is composed of 72 lakes across the Irish Ecoregion with a TP range of 0-142 μg l^{-1}. Surface sediments of these sites consist of 233 common diatom taxa with maximum relative abundance of ≥ 1% and occurrence in at least three sites. As shown in Chapter 4, TP has been identified as one of the most significant environmental variables influencing the assemblage structure of surface sediment diatoms. A relatively high λ_1/λ_2 ratio (0.673) in a partial CCA constrained by TP indicates that TP is statistically powerful for the development of inference models. DCCA constrained by TP gives a long gradient length of 3.435 for the diatom data, indicating that non-linear modelling methods, Weighted Averaging (WA) and Weighted Averaging Partial Least Square (WA-PLS), are appropriate for developing TP transfer functions for the diatom training set.

Data manipulation is also employed in model construction as it can affect the model performance depending on different calibration methods (Koster *et al.*, 2004). Both log-transformed and untransformed TP data are included in the WA and WA-PLS modelling and the inference models with transformed TP data show better performances (not illustrated here). Models based on log-transformed TP and either untransformed or transformed diatom data are used. Two cross-validation methods, jack-knifing and bootstrapping, were performed respectively for each model and jack-knifing generally showed a better performance than bootstrapping (n=1000). Therefore only models cross-validated with jack-knifing are reported here.

6.1.1 Weighted Averaging (WA) Modelling

A summary of 16 WA modelling results for the diatom training set is shown in Table 6.1. Eight WA models are produced for 72 lakes using classical or inverse deshrinking methods, with or without tolerance downweighted and based on untransformed or square root transformed diatom data as shown in Table 6.1. WA models based on the untransformed ecological data outperform the models based on square root transformed data when diatom species with wide tolerance (ecological amplitude) are downweighted (see Table 6.1). For example, the model WA_{TOL_Cla} (tolerance downweighted, classical deshrinking) based on raw diatom data has lower root mean square error (RMSE) and

RMSE of prediction (RMSEP) and higher coefficient of determination (r^2) and jack-knifed r^2 (r^2_{jack}) than the same WA_{TOL_Cla} but based on the square-root transformed diatom data. The former model also has lower average and maximum bias (both apparent and jack-knifed). But when tolerance downweighting is not applied, the WA models based on square-root transformed data outperform the same WA models based on untransformed diatom data.

Table 6.1 Summary of WA models for the diatom training set (TP data are $log_{10}(1+)$-transformed; both raw and square-root transformed taxa data are used; WA_{Inv} = inverse deshrinking, WA_{Cla} = classical deshrinking, WA_{TOL_Inv} = tolerance downweighted and inverse deshrinking, WA_{TOL_Cla} = tolerance downweighted and classical deshrinking; jack-knifing is used for cross-validation; the optimal model is highlighted in bold).

Sites	Taxa data	Model	r^2	RMSE	r^2_{Jack}	RMSEP	Average $Bias_{Jack}$	Max. $Bias_{Jack}$
72	raw	WA_{Inv}	0.734	0.226	0.576	0.285	-0.009	0.800
72	raw	WA_{Cla}	0.734	0.263	0.584	0.317	-0.011	0.653
72	**raw**	**WA_{TOL_Inv}**	**0.804**	**0.193**	**0.644**	**0.263**	**0.013**	**0.848**
72	raw	WA_{TOL_Cla}	0.804	0.216	0.649	0.288	0.017	0.772
72	sqrt	WA_{Inv}	0.722	0.230	0.608	0.274	-0.003	0.795
72	sqrt	WA_{Cla}	0.722	0.271	0.616	0.305	-0.004	0.640
72	sqrt	WA_{TOL_Inv}	0.798	0.196	0.621	0.272	0.018	0.868
72	sqrt	WA_{TOL_Cla}	0.798	0.220	0.624	0.301	0.022	0.797
70	raw	WA_{Inv}	0.775	0.197	0.646	0.248	-0.008	0.451
70	raw	WA_{Cla}	0.775	0.224	0.650	0.268	-0.010	0.308
70	**raw**	**WA_{TOL_Inv}**	**0.866**	**0.152**	**0.743**	**0.213**	**0.018**	**0.411**
70	raw	WA_{TOL_Cla}	0.866	0.163	0.745	0.224	0.021	0.328
70	sqrt	WA_{Inv}	0.772	0.199	0.682	0.235	-0.003	0.511
70	sqrt	WA_{Cla}	0.772	0.226	0.686	0.253	-0.003	0.384
70	sqrt	WA_{TOL_Inv}	0.866	0.152	0.720	0.223	0.026	0.529
70	sqrt	WA_{TOL_Cla}	0.866	0.164	0.721	0.236	0.030	0.467

Among the eight WA models, model WA_{TOL_Inv} based on untransformed diatom data gives the highest r^2 (0.804) with the lowest RMSE (0.193). These measures of model performance (e.g. r^2, RMSE) are generally over-optimistic due to the lack of independent test of model performance, therefore jack-knifed measures are used to produce more realistic models through cross-validation. The model WA_{TOL_Inv} based on untransformed data still shows a good performance because of its lowest RMSEP (0.263), second highest r^2_{jack} (0.644), a relatively low jack-knifed average bias (0.013) but it still has a second highest jack-knifed maximum bias (0.848). None of the eight WA models consistently give the best performance when all the measures of model performances are considered. However, the WA_{TOL_Inv} based on raw diatom data gives the best performance if maximum bias is ignored.

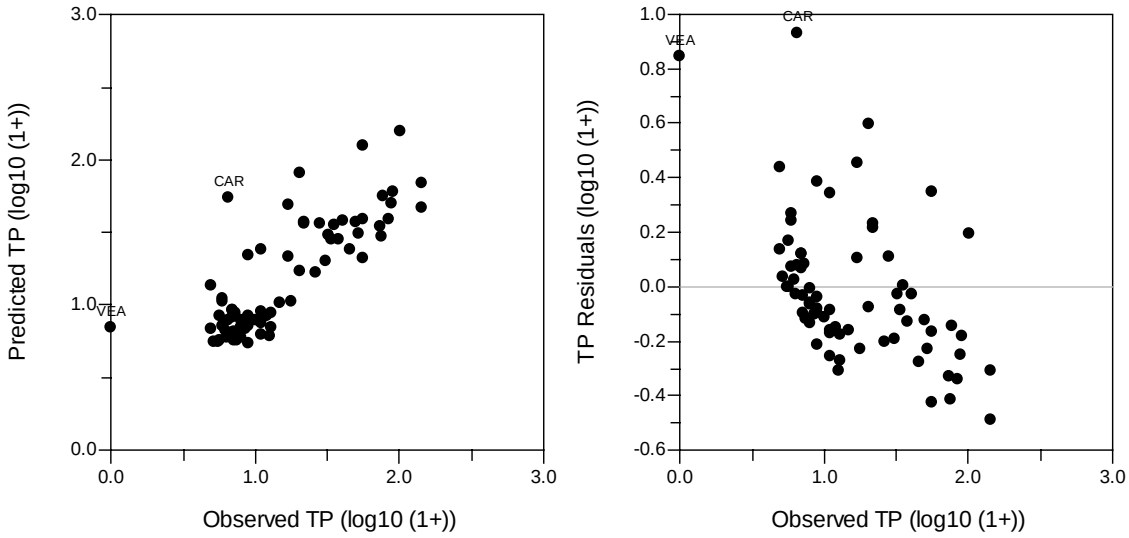

Figure 6.1 Relationship between observed TP and predicted TP and TP residuals in a jack-knifed WA_{TOL_Inv} model of 72 lakes (two outlier lakes are labelled with site codes; TP are log_{10}-transformed and diatom data are untransformed).

Figure 6.1 shows the observed and WA_{TOL_Inv} predicted TP values (log-transformed) as well as the TP residual values (observed minus predicted TP). A positive relationship between observed and predicted TP values of the model WA_{TOL_Inv} is apparent. It is also evidenced in the TP residuals that predicted TP is underestimated at the lower end of TP gradient and overestimated at the higher end of the TP gradient. TP residuals also show the highest discrepancy between observed and predicted TP values for two sites (Veagh [VEA] and Caragh [CAR]) as labeled in Figure 6.1. Accordingly sites VEA and CAR are removed as outliers and WA inference models based on 70 lakes are developed and summarized in Table 6.1.

In comparison with the WA models based on square root transformed ecological data, the WA models based on untransformed data from 70 lakes perform better when tolerance is downweighted. However, among the eight WA models of 70 lakes, the WA_{TOL_Inv} based on untransformed diatom data performs the best with the lowest RMSEP (0.213) and RMSE (0.152), highest r^2 (0.866) and almost highest r^2_{jack} (0.743, second to 0.745 of WA_{TOL_Cla}) and relatively low bias values (see Table 6.1). Also after two outlier sites are removed, the performance of WA_{TOL_Inv} of 70 lakes is clearly improved in comparison with the full WA_{TOL_Inv} model, e.g. the RMSEP is reduced from 0.263 to 0.213, and the r^2_{jack} value is raised from 0.644 to 0.743. Therefore the WA_{TOL_Inv} based on the untransformed diatom data of 70 lakes shows the best

performance among all the 16 WA models in consideration of the coefficients of determination (r^2) and their associated errors with a moderate maximum bias.

6.1.2 Weighted Averaging Partial Least Square (WA-PLS) Modelling

WA-PLS is another unimodal modelling method for developing transfer functions and differs from WA in that it exploits the residual structure in species data to optimise the species parameters in transfer functions (see Section 3.4.3 in methods). Performances of twenty variations of the WA-PLS model are summarised in Table 6.2. The measures of model performance with cross-validation, e.g. the jack-knifed coefficient of correlation (r^2) and RMSEP, are used for selecting the optimum WA-PLS model (Birks, 1998).

Table 6.2 Summary of the first five component WA-PLS models for the diatom training set (TP data are $log_{10}(1+)$-transformed; both raw and square-root transformed taxa data are used; jack-knifing is used for cross-validation; the optimal model is highlighted in bold).

Sites	Taxa data	Component	r^2	RMSE	r^2_{Jack}	RMSEP	Average $Bias_{Jack}$	Max. $Bias_{Jack}$
72	raw	1	0.734	0.226	0.576	0.285	-0.010	0.800
72	raw	2	0.878	0.152	0.544	0.307	-0.009	0.813
72	raw	3	0.932	0.114	0.558	0.307	-0.015	0.737
72	raw	4	0.955	0.092	0.580	0.301	-0.011	0.749
72	raw	5	0.975	0.069	0.575	0.308	0.007	0.711
72	**sqrt**	**1**	**0.722**	**0.230**	**0.608**	**0.274**	**-0.008**	**0.793**
72	sqrt	2	0.871	0.157	0.581	0.288	-0.010	0.884
72	sqrt	3	0.934	0.112	0.574	0.293	-0.012	0.775
72	sqrt	4	0.975	0.069	0.585	0.291	-0.012	0.754
72	sqrt	5	0.990	0.043	0.592	0.291	-0.006	0.791
70	raw	1	0.775	0.197	0.646	0.248	-0.009	0.451
70	raw	2	0.895	0.135	0.617	0.267	-0.007	0.526
70	raw	3	0.935	0.106	0.631	0.267	-0.012	0.479
70	raw	4	0.958	0.085	0.650	0.264	-0.008	0.450
70	raw	5	0.974	0.067	0.636	0.277	0.009	0.478
70	**sqrt**	**1**	**0.772**	**0.199**	**0.682**	**0.235**	**-0.005**	**0.504**
70	sqrt	2	0.892	0.137	0.652	0.248	-0.007	0.538
70	sqrt	3	0.948	0.095	0.649	0.251	-0.008	0.536
70	sqrt	4	0.980	0.058	0.673	0.243	-0.007	0.496
70	sqrt	5	0.992	0.038	0.674	0.245	-0.003	0.451

Two WA-PLS models based on both untransformed and square-root transformed diatom data from 72 lakes are developed and the first five components of both models are shown in Table 6.2. The first components of both models generally outperform other components with lowest RMSEP and highest or almost highest jack-knifed coefficients

of correlation (r^2_{jack}) among all the five components. For example, the first component of WA-PLS model based on raw diatom data produces the lowest RMSEP of 0.285 and a high r^2_{jack} of 0.576 (second to 0.580 of the fourth component of the same model). In comparison with the WA-PLS model based on raw diatom data of 72 lakes, the model based on square root transformed diatom data outperforms with a higher RMSEP and r^2_{jack} (see Table 6.2). The first component WA-PLS model (WA-PLS-1) with transformed diatom data produces a RMSEP of 0.274 and r^2_{jack} of 0.608. The scatter plots between observed TP values and predicted TP and TP residuals for this model are shown in Figure 6.2. The same trend of predicted TP values along the TP gradient observed in the WA $_{TOL\_Inv}$ also occurs in the WA-PLS-1 model, with overestimation at the low end of the TP gradient and underestimation at the upper end. VEA and CAR are again identified as outliers similar to the model WA $_{TOL\_Inv}$ (see Figure 6.1). Therefore, these two sites are removed and the WA-PLS models were re-run with 70 lakes (Table 6.2). After the removal of two sites the model performance is improved in comparison with the 72-lake model. The WA-PLS models based on square-root transformed diatom data performs best and the first component performs better than the other four components. Therefore, the optimal WA-PLS model is based on the squared root transformed diatom data of 70 lakes and its first component produces the lowest value of RMSEP (0.235) and the highest r^2_{jack} (0.682). The WA-PLS-1 predicted TP, TP residuals and observed values for 70 lakes are shown in Figure 6.2. A good relationship between observed and WA-PLS-1 predicted TP is apparent. In addition predicted TP is clearly underestimated at the lower end of TP gradient and overestimated at the higher end of the TP gradient.

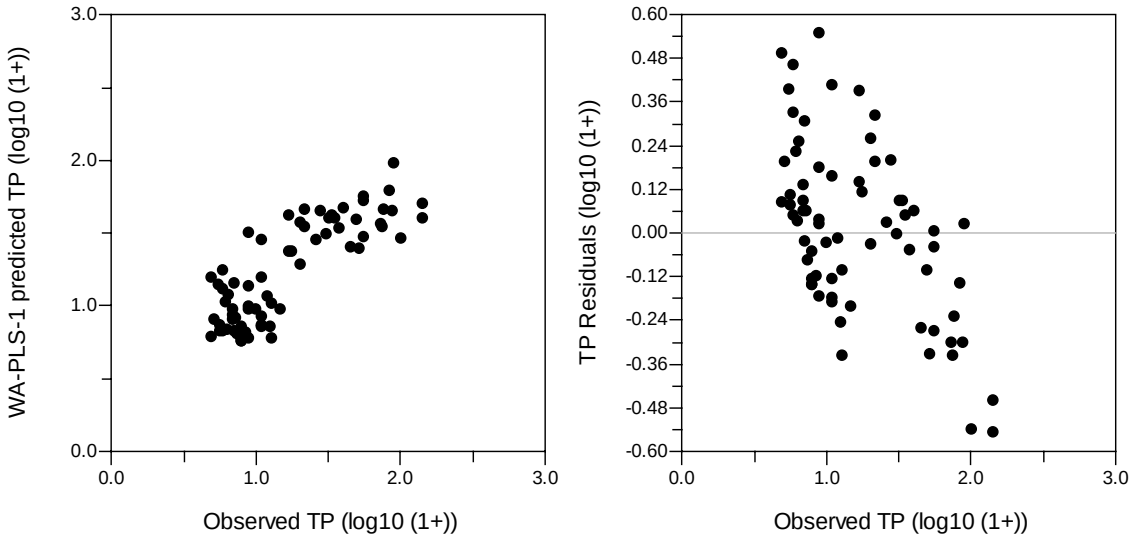

Figure 6.2 Relationship between observed TP and predicted TP and TP residuals of WA-PLS-1 model of 70 lakes (TP are log_{10}-transformed and diatom data are square root transformed).

6.1.3 Model Comparison and Evaluation

Two unimodal modelling methods, WA and WA-PLS, have been applied based on both untransformed and transformed diatom data and cross-validated with jack-knifing in the previous sections. The removal of two outlier sites (CAR and VEA) from the full 72-lake diatom training set improves the performance of both WA and WA-PLS models. In WA modelling of 70 lakes, the model WA_{TOL_Inv} based on untransformed diatom data gives the best performance with a RMSEP of 0.213 and r^2_{jack} of 0.743 (see Table 6.1). The first component of the WA-PLS model based on square root transformed diatom data from 70 lakes outperforms the other components of the same model with a RMSEP of 0.235 and r^2_{jack} of 0.682 (see Table 6.2). Therefore the optimum unimodal model for the 70-lake diatom training set is the WA_{TOL_Inv} based on the untransformed diatom data as it outperforms the first component of WA-PLS based on the transformed species data in reducing the RMSEP by 0.022 and improving r^2_{jack} by 0.061 (compare Table 6.1 and Table 6.2). This model also has relatively low bias.

Twenty-five diatom TP training sets including the current study and their performances are summarised in Table 6.3. The optimum WA model for the Irish Ecoregion is comparable with most other diatom TP transfer functions in terms of lake size, TP range and model performances. Only five of the 25 training sets have training set lakes (see Table 6.3). A Finish TP transfer function based on 78 lakes with a TP gradient (3-125 µg l^{-1}) has a performance with r^2_{jack} of 0.73 and RMSEP of 0.19 (Miettinen, 2003). Several TP transfer functions based on a smaller number of lakes or a shorter TP gradient often display weaker predictability and/or higher prediction error, including the training set composed of 59 lakes from British Columbia of Canada with a TP range of 6-42 µg l^{-1} (Reavie *et al.*, 1995), an Alaskan training set of 51 lakes with a TP gradient of 3-83 µg l^{-1} (Gregory-Eaves *et al.*, 1999), a 43-lake Swedish transfer function (Bradshaw & Anderson, 2001) and an Australian training set composed of 33 lakes and reservoirs (Tibby, 2004). In all the performance of the inference model developed for 70 lakes from the Irish Ecoregion shows strong predictability and low prediction error and is comparable with most other TP transfer functions based on diatoms.

Table 6.3 Summary of diatom-TP transfer functions and their performances (each training set is ordered in the name of its country and region; - = no data).

Study Area	Sites	Range of TP (µg/l)	TP Data	TP data Transfor-mation	Model	r^2	RMSE	r^2 jack/boot	RMSEP	References
Europe (Northwest)	152	5-1190	Annual mean	log	WA-PLS 2	0.91	0.15	0.81	0.21	(Bennion *et al.*, 1996)
Europe (Central)	86	2-266	Not mentioned	log	WA_{Tol}	0.57	0.32	-	0.35	(Wunsam & Schmidt, 1995)
Denmark	29	24-1145	Annual mean	log	WA-PLS 2	0.86	-	0.37	0.28	(Bradshaw *et al.*, 2002)
England (Southeast)	30	25-646	Annual mean	log	WA	0.79	0.16	-	0.28	(Bennion, 1994)
Finland (south)	61	3-89	Autumn	log	WA-PLS	-	-	0.76	0.16	(Kauppila *et al.*, 2002)
Finland (southeast)	78	3-125	Autumn	log	WA	0.81	0.16	0.73	0.19	(Miettinen, 2003)
Ireland	70	4-142	Annual mean	Log(1+)	WA_{Tol}	0.87	0.15	0.74	0.21	Current study
Northern Ireland	49	15-800	Annual mean	log	WA	0.80	0.73	0.19	0.24	(Anderson & Rippey, 1994)
Northern Ireland	43	25-800	Annual mean	log	WA	0.75	0.17	-	-	(Anderson *et al.*, 1993)
Sweden	43	7-369	Annual mean	log	WA_{Tol}	0.79	0.15	0.36	0.27	(Bradshaw & Anderson, 2001)
Switzerland (Alps)	68	6-520	Early spring	log	WA-PLS 2	0.93	0.11	0.79	0.19	(Lotter *et al.*, 1998)
Canada (BC)	59	6-42	Spring-autumn	none	WA	0.73	-	0.46	0.48	(Reavie *et al.*, 1995)
Canada (BC)	46	5-28	May-October	ln (1+)	WA	0.73	0.21	-	-	(Hall & Smol, 1992)
Canada (Ontario)	54	3-24	Spring	none	WA	0.62	3.50	-	-	(Hall & Smol, 1996)
Canada (Ontario)	64	4-54	Spring	none	WA	0.64	7.00	0.47	10.00	(Reavie & Smol, 2001)
Canada (Ontario)	30	6-49	Spring	log	WA	0.57	-	0.44	0.20	(Werner & Smol, 2005)
Canada (Quebec)	76	9-1687	Mainly summer	none	WA-PLS 2	0.89	-	0.51	3.20	(Philibert & Prairie, 2002)
Canada (Quebec)	41	3-30	June-August	none	WA-PLS	0.89	-	-	2.37	(Enache & Prairie, 2002)
USA (Alaska)	51	3-83	Summer	log	WA	0.77	0.16	0.52	0.23	(Gregory-Eaves *et al.*, 1999)
USA (Michigan)	41	1-51	July	ln (1+)	WA	0.73	0.41	-	-	(Fritz *et al.*, 1993)
USA (Minnesota)	55	7-139	May-October	log	WA	0.68	0.19	-	0.25	(Ramstack *et al.*, 2003)
USA (Northeast)	257	3-48	Not mentioned	ln	WA	0.55	-	-	0.79	(Dixit *et al.*, 1999)
USA (Northeast)	64	1-155	Annual mean	ln (1+)	WA	0.66	0.62	-	-	(Dixit & Smol, 1994)
Australia (Southeast)	33	7-451	Annual mean	log	WA-PLS 2	0.94	0.11	0.69	0.25	(Tibby, 2004)
China (East)	43	30-515	Not mentioned	log	WA	-	-	0.82	0.12	(Dong *et al.*, 2006)

A TP transfer function based on diatoms from Northern Ireland was developed by Anderson *et al*. (1993). The model is composed of 49 lakes from Northern Ireland with a TP range of 15-800 μg l^{-1} and it also has high predictability with an r^2_{boot} of 0.73 and an RMSEP of 0.244 (Anderson & Rippey, 1994). The optimum WA model developed in the current study outperforms the transfer function from North Ireland with a stronger predictability (the cross-validated r^2 improved by 0.013) and a lower prediction error (RMSEP reduced by 0.031). However, the TP gradient covered is much shorter but more lakes are included in the current study than the one from Northern Ireland.

6.1.4 TP Optima of Diatom Taxa

TP optima and tolerance of 233 common diatom taxa inferred by the WA model selected are listed in Appendix G. TP optima values for the 233 common taxa are all below 100 μg l^{-1} and only 19 taxa have TP optima of above 40 μg l^{-1}. Furthermore, nearly half (112 taxa) of the common taxa, have TP optima of less than 10 μg l^{-1} (see Appendix G). All these features reflect the predominance of lakes located at the low end of TP gradient in this training set. The estimated TP tolerance for most taxa lies within 3 μg l^{-1} and only 19 taxa have TP tolerances greater than 3 μg l^{-1}, indicating relatively narrow ecological amplitudes for most taxa in the study lakes. The tolerances of diatom taxa in this study are relatively small in comparison with other studies with reports on tolerances of diatom taxa. Nearly half of the 164 diatom taxa in a Canadian training set have TP tolerance of above 5 μg l^{-1} though it has a short TP gradient of 3-30 μg l^{-1} (Enache & Prairie, 2002).

An epilithic algae investigation of 32 Irish lakes by DeNicola *et al*. (2004) with a TP gradient of 3.6-90.5 μg l^{-1} also provided information on TP optima for some diatom taxa. TP optima of some common taxa in the Irish Eco-region indicated are shown in Table 6.4 as well as those from other Ecoregions. Eight of the ten diatom taxa in both studies with TP optima values are mainly benthic or littoral dwellers in lakes, including *Fragilaria pinnata*, *F. construens* f. *venter* and *Gomphonema parvulum*. Their TP optima generally show good correspondence with each other and also a similar sequence of species succession along the TP gradient is evident in both studies, e.g. 13.3 and 19.4 μg l^{-1} for *Tabellaria flocculosa*, 23.1 and 23.6 μg l^{-1} for *F. pinnata* and 25.6 and 25.3 μg l^{-1} for *F. capucina* var. *gracilis* respectively. However, *Aulacoseira ambigua* and

Asterionella formosa, display relative big divergence in both studies with TP optima of 22.7 and 36.0 μg l^{-1} for both species in the current study in comparison with 50.3 and 21.9 μg l^{-1} respectively in the study by DeNicola *et al*. (2004). Both species are commonly found in open waters and their dissimilarity in TP optima could be due to the fact that the epilithic algae were sampled only once in the littoral area at a water depth of 0.5-1 m and planktonic taxa are not well represented (DeNicola *et al.*, 2004).

Table 6.4 Comparison of WA-inferred TP (μg l^{-1}) optima of selected diatom taxa of the current study (Ireland[1]) with those from the same Irish Ecoregion (Ireland[2]) (DeNicola *et al.*, 2004), England (Bennion, 1994), Sweden (Bradshaw & Anderson, 2001), Finland (Miettinen, 2003) and the Alps (Wunsam & Schmidt, 1995) (Taxa are listed according to their TP optima in Ireland[1]; - = no available data).

Taxon	Ireland[1]	Ireland[2]	England	Sweden	Finland	the Alps
Cyclotella comensis	7.9	-	-	26	-	10.4
Fragilaria brevistriata	8.9	-	94.8	-	28.0	13.4
Cyclotella radiosa	10.9	-	70.8	-	12.5	17.8
Cyclotella pseudostelligera	12.7	-	158.1	55	18.5	-
Tabellaria flocculosa	13.3	19.4	50.2	-	17.5	-
Achnanthes minutissima	14.6	21.4	66.1	34	15.8	13.3
Stephanodiscus alpinus	16.4	-	-	46	11.2	10.4
Aulacoseira ambigua	22.7	50.3	95.7	57	23.4	19.0
Fragilaria pinnata	23.1	23.6	93.8	-	19.1	20.6
Navicula radiosa	23.4	-	60.0	-	20.0	-
Fragilaria construens f. *venter*	24.9	21.9	71.1	-	21.4	17.2
Fragilaria capucina var. *gracilis*	25.6	25.3	-	-	21.0	16.4
Cocconeis placentula	27.3	20.9	89.9	-	14.3	-
Aulacoseira subarctica	29.1	-	-	72	23.8/17.1	-
Cyclostephanos dubius	29.3	-	214.8	64	29.1	-
Aulacoseira islandica	33.1	-	-	44	16.0	13.2
Nitzschia palea	33.9	23.0	129.1	-	26.3	12.8
Gomphonema parvulum	34.9	29.9	138.4	-	22.6	-
Aulacoseira granulata	35.1	-	-	51	30.6	52.4
Asterionella formosa	36.0	21.9	152.8	61	16.4	-
Stephanodiscus hantzschii	43.1	-	288.4	74	51.5	111.2
Diatoma tenuis	43.7	-	-	66	29.0	-
Stephanodiscus parvus	46.7	-	200.9	125	-	26.8

Five species with TP optima published from North Ireland (Anderson, 1997b) show higher TP optima values than those in the current study but to a varying degree: slightly higher for *Cyclotella radiosa* and *A. subarctica* with TP optima at the low end of TP gradient and much more higher for taxa with high TP optima, like *Stephanodiscus parvus* and *S. hantzschii*. This is a result of the high numbers of eutrophic and

hypertrophic lakes and much longer TP gradient of 15-800 μg l^{-1} in the North Irish training set, e.g. more than half of the 54 lakes have TP values of above 50 μg l^{-1}. In contrast only 13 of 72 lakes in the current study have TP values of above 50 μg l^{-1}.

A comparison of TP optima of some common diatom taxa from the Irish Ecoregion with those from the European Ecoregions is also shown in Table 6.4. As expected diatom taxa from the Irish Ecoregion generally display a similar sequence of TP preferences as those from other European Ecoregions, e.g. *Cyclotella comensis* and *C. radiosa* preferring nutrient-poor waters, *Aulacoseira ambigua* and *Fragilaria construens* f. *venter* favouring more nutrient-enriched lakes and *Stephanodiscus hantzschii* and *S. parvus* as typical taxa in eutrophic and hypertrophic waters (see Table 6.4). However, some taxa display very variable TP preferences among training sets and European Ecoregions: *C. pseudostelligera* prefers lower end of TP gradient in Ireland and Finland but are found at the higher end in England. An obvious feature in Table 6.4 is the low TP optima values for diatom taxa from Ireland, Finland and the Alps, in comparison with those from England and to a less degree Sweden. For the same taxa from Southeast England TP optimum is generally 3-7 times higher than Ireland, like the relative ratios of 4.2 (152.8 μg l^{-1} (England) /36.0 μg l^{-1} (Ireland)) for *Asterionella formosa* and 6.7 (288.4/43.1) for *Stephanodiscus hantzschii*. Certain species from England have TP optima more than ten times higher than those from Ireland, like *Cyclotella meneghiniana*, *C. pseudostelligera* and *Fragilaria brevistriata*. This is due to the much longer TP gradient (25-646 μg l^{-1}) in the English training set (see Table 6.3). This English training set is mainly composed of shallow lakes and ponds (< 3m depth) with a limited range of lake types from a small geographical area in comparison to the Irish Ecoregion training set. Diatom species from Sweden also shows higher TP optima, approximately 1.3-3 times those from Ireland.

6.2 Cladocera-based TP Transfer Functions

The Cladocera training set consists of 31 common taxa (maximum relative abundance of ≥ 1% and in at least two sites) in the surface sediments of 33 Irish lakes with a TP range of 4-142.3 μg l^{-1}. DCCA constrained by TP only gives a gradient length of 0.886 for the Cladocera training set, indicating that a linear-based method is appropriate for developing a Cladocera-based TP transfer function (Birks, 1995). However, non-linear

methods are also recommended for model development as they can still outperform the linear method for compositional data with short gradients (ter Braak *et al.*, 1993; Birks, 1998)guangjie.chen. Therefore Partial Least Squares (PLS), WA-PLS and WA are all practiced for model comparison. Log_{10}-transformed TP and square root transformed Cladocera data are used in all three models as they provide models with better performances than those based on untransformed data (the latter are not illustrated here). Jack-knifing is used in all models for cross-validation. Summaries of the 28 variations of TP inference models are shown in Table 6.5.

Table 6.5 Summary of Cladocera-based TP inference models (TP and Cladocera data are log10- and square root transformed respectively in all models; jack-knifing is used for cross-validation; the optimal model is highlighted in bold).

Sites	Model	RMSE	r^2	r^2_{Jack}	RMSEP	Average $Bias_{Jack}$	Max. $Bias_{Jack}$
33	PLS 1	0.308	0.409	0.182	0.369	0.001	0.776
33	PLS 2	0.268	0.554	0.223	0.364	0.004	0.779
33	PLS 3	0.236	0.655	0.274	0.353	-0.001	0.866
33	PLS 4	0.190	0.775	0.286	0.353	-0.008	0.767
33	PLS 5	0.168	0.824	0.355	0.332	-0.007	0.728
31	PLS 1	0.268	0.541	0.310	0.331	0.003	0.694
31	PLS 2	0.198	0.750	0.477	0.288	0.010	0.416
31	**PLS 3**	**0.173**	**0.808**	**0.551**	**0.267**	**0.004**	**0.487**
31	PLS 4	0.139	0.876	0.569	0.264	0.006	0.419
31	PLS 5	0.121	0.907	0.601	0.252	0.003	0.414
33	WA-PLS 1	0.254	0.601	0.375	0.318	-0.006	0.654
33	WA-PLS 2	0.211	0.722	0.384	0.321	-0.016	0.584
33	WA-PLS 3	0.160	0.842	0.384	0.331	-0.024	0.435
33	WA-PLS 4	0.147	0.866	0.398	0.327	-0.027	0.485
33	WA-PLS 5	0.136	0.885	0.460	0.306	-0.020	0.480
31	WA-PLS 1	0.186	0.780	0.620	0.244	-0.001	0.369
31	**WA-PLS 2**	**0.131**	**0.891**	**0.729**	**0.206**	**-0.001**	**0.325**
31	WA-PLS 3	0.104	0.931	0.742	0.201	0.004	0.288
31	WA-PLS 4	0.094	0.943	0.742	0.201	0.002	0.264
31	WA-PLS 5	0.083	0.956	0.736	0.205	0.005	0.229
33	$WA_{\_Inv}$	0.254	0.601	0.374	0.318	0.003	0.665
33	$WA_{\_Cla}$	0.327	0.601	0.424	0.354	0.003	0.595
33	WA_{TOL_Inv}	0.227	0.679	0.434	0.305	-0.011	0.466
33	WA_{TOL_Cla}	0.276	0.679	0.448	0.358	-0.024	0.326
31	$WA_{\_Inv}$	0.185	0.780	0.622	0.244	0.007	0.362
31	$WA_{\_Cla}$	0.210	0.780	0.642	0.243	0.007	0.256
31	$\mathbf{WA_{TOL\_Inv}}$	**0.161**	**0.835**	**0.687**	**0.221**	**-0.007**	**0.328**
31	WA_{TOL_Cla}	0.176	0.835	0.676	0.239	-0.015	0.247

6.2.1 Partial Least Square (PLS) Modelling

Performances of the first five components of the PLS models based on square-root transformed Cladocera data from 33 lakes are listed in Table 6.5. The performance of each PLS model is based on cross-validated RMSE and r^2 (ter Braak & Juggins, 1993; Birks, 1995). A practical method to select the optimal number of components in a PLS model is a 5% or more reduction in RMSEP with each additional component (Birks, 1995). The addition of the second component of the PLS model based on all the 33 lakes shows a little improvement in the RMSEP in comparison with the first component and therefore PLS 1 (first component) model is selected as the optima PLS model for the 33 lakes (see Table 6.5). However, the performance of PLS 1 model is relatively poor with a moderate RMSEP of 0.396 and a low r^2_{jack} of 0.182 and the removal of two outlier sites (Sillan [SIL] and Lisnahan [LIS]) with high TP residuals (not shown here) help to improve the model performance evidently (Table 6.5). The optima PLS model based on 31 lakes is the third component (PLS 3) with an RMSEP of 0.267 and r^2_{jack} of 0.551 as the fourth component gives a less than 5% reduction of RMSEP than the PLS 3. The scatter plots of observed TP and predicted TP and TP residuals are shown in Figure 6.3 for the PLS 3 model based on 31 lakes. A good correlation between observed and predicted TP values is shown and overestimation at the low end of observed TP gradient and underestimation at the high end are evidenced in the TP residual plot.

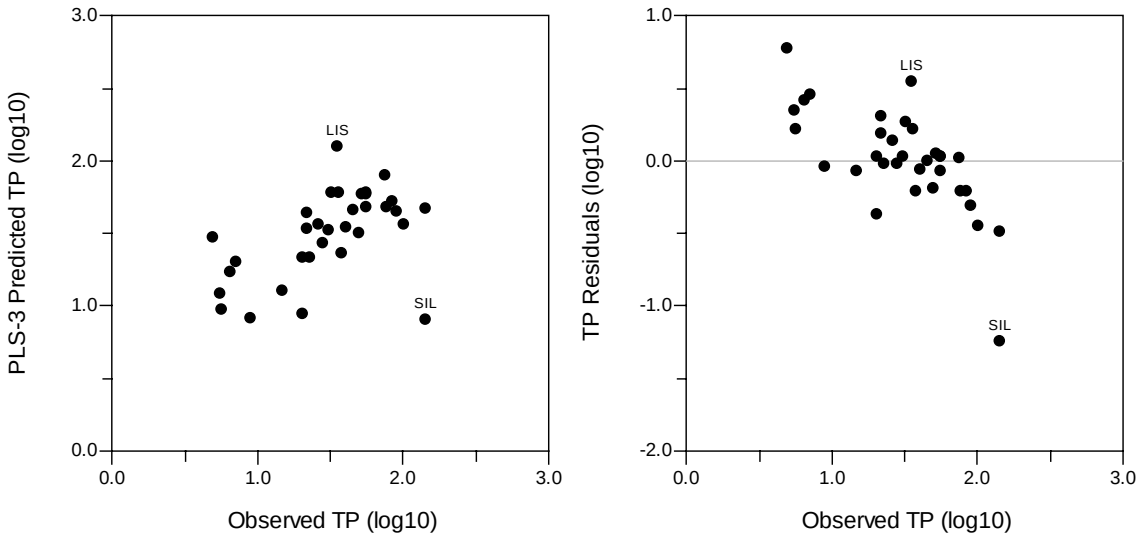

Figure 6.3 Scatter plots of observed and predicted TP and TP residuals for the PLS 3 model of 31 lakes (two lakes are labelled with site codes; TP data are log_{10}-transformed and Cladocera data are square root transformed).

6.2.2 Weighted Averaging Partial Least Square (WA-PLS) Modelling

Performances of the first five components of WA-PLS models are summarised in Table 6.5. Each of the five WA-PLS components outperforms the PLS model with lower RMSEP and higher jack-knifed r^2. After identifying and removing two sites (SIL and LIS, the same sites removed in the PLS model) with high TP residuals, the performance of each of the five WA-PLS components is improved (see Table 6.5). The second component of the WA-PLS model based on 31 lakes gives the best performance as no additional component can significant improve its RMSEP of 0.206 and r^2_{jack} of 0.729. The correlation between observed TP and TP values predicted by the WA-PLS 2 model is shown in Figure 6.4. Underestimation at the high end of observed TP occurs for this model but with a narrow range (-0.2-0 $\log_{10}$TP) of residuals, while overestimation at the low end occurs with a wider range (0-0.6 $\log_{10}$TP) (see Figure 6.4).

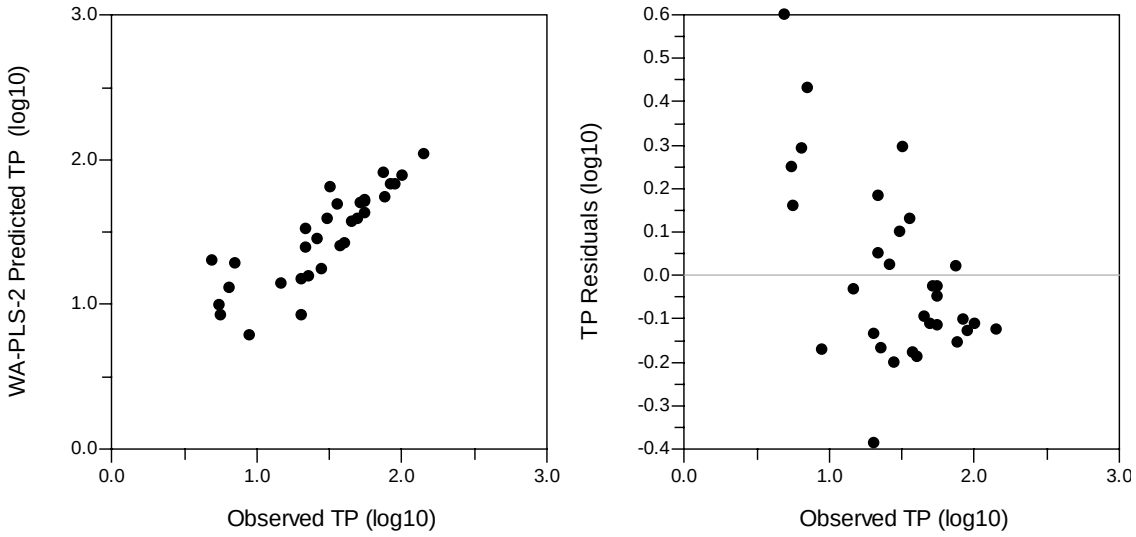

Figure 6.4 Scatter plots of observed TP and predicted TP and TP residuals of the WA-PLS 2 model of 31 lakes (TP data are $\log_{10}$-transformed and Cladocera data are square root transformed).

6.2.3 Weighted Averaging (WA) Modelling

Both inverse and classical deshrinking with and without tolerance downweighted were performed for WA modelling. Among the four WA models of 33 lakes the model with inverse deshrinking and tolerance downweighted gives the lowest RMSEP of 0.305 and second highest r^2_{jack} of 0.434. After the removal of sites SIL and LIS as in PLS and WA-PLS modelling, performance of the same model is improved significantly with an

RMSEP of 0.221 and r^2_{jack} of 0.687 (see Table 6.5). Observed and predicted TP values by this model for 31 lakes display strong correlation as shown in Figure 6.5. Both overestimation and underestimation occur along the low-middle part of the observed TP gradient and underestimation occurs for the high end of the TP gradient (see Figure 6.5).

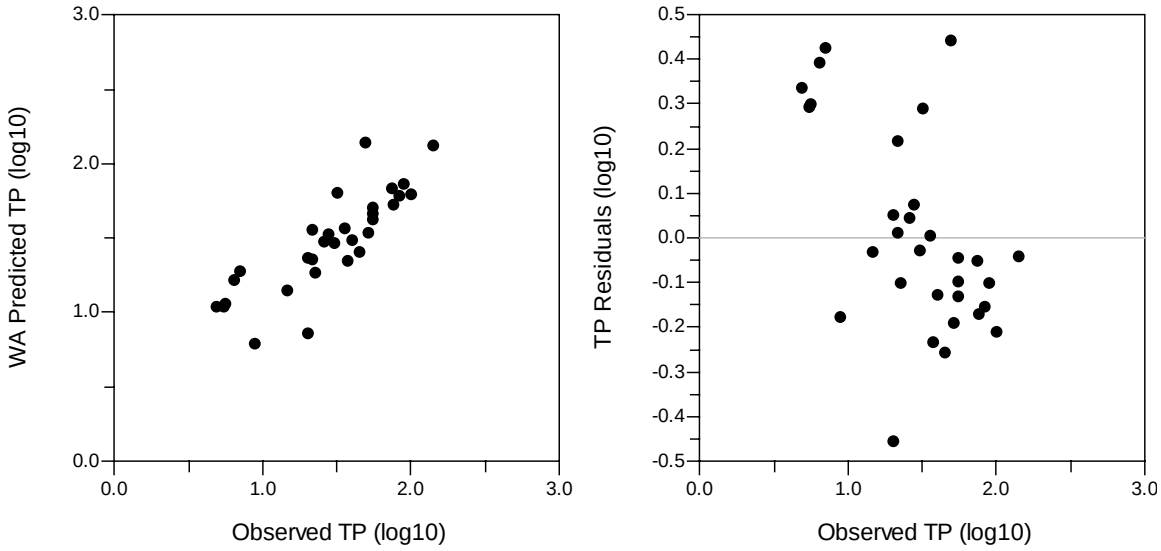

Figure 6.5 Relationship between observed and predicted TP and TP residuals of WA $_{TOL\_Inv}$ of 31 lakes (TP data are log_{10}-transformed and Cladocera data are square root transformed).

6.2.4 Model Comparison and Evaluation

Both linear- and unimodal-based modelling methods have been performed based on log_{10}-transformed TP data and square-root transformed Cladocera data of 33 lakes (see Table 6.5). The unimodal-based modelling method WA-PLS generally outperforms the linear-based method PLS for the same set of data. After removal of outlier sites Sillan [SIL] and Lisnahan [LIS], performances of all three models are substantially improved in both RMSEP and jack-knifed coefficient of determination (r^2). The best models are the third component PLS model (RMSEP = 0.267, r^2_{jack} = 0.551), the second component WA-PLS model (RMSEP = 0.206, r^2_{jack} = 0.729) and the WA $_{TOL\_Inv}$ model (RMSEP = 0.221, r^2_{jack} = 0.687). Of these three the optimal model with best predictability and least prediction error for the 31 lakes is the WA-PLS-2 model.

Table 6.6 Comparison of Cladocera-inferred TP transfer functions from Denmark[1] (Brodersen *et al.*, 1998), Denmark[2] (Amsinck *et al.*, 2005), Switzerland (Lotter *et al.*, 1998), Canada (Bos & Cumming, 2003) and Ireland (current study) (TP data are log-transformed in all models).

Study area	Sites	TP range (µ g/l)	Indicator	Model	r^2	RMSE	$r^2_{jack/boot}$	$RMSEP_{jack/boot}$
Denmark[1]	32	16-765	Chydorid	WA	-	-	0.79	0.24
Denmark[2]	36	27-327	Cladocera	WA	0.53	0.11	0.32	0.29
Switzerland	68	6-520	Chydorid	WA-PLS-1	0.63	0.24	0.49	0.28
Canada (BC)	49	2-146	Cladocera	WA	0.73	0.25	0.61	0.30
Ireland	31	4-142	Cladocera	WA-PLS-2	0.89	0.18	0.73	0.21

In comparison with four other published Cladocera-based TP transfer functions, the Irish training set has a similar TP gradient (4-142.6 µg l^{-1}) as that of Canada (British Columbia), and is shorter than the two Danish and the Swiss training sets (see Table 6.6). The Irish Ecoregion training set is of similar size as those from the Denmark, but is smaller than the Canadian and the Swiss data sets which have lake numbers of 49 and 68 lakes respectively. The Irish, Canadian and Danish models are based on the whole Cladocera assemblages, while the other two inference models only use the littoral Cladocera assemblages (see Table 6.6). The performance of the Irish Ecoregion TP transfer function is comparable with the other four TP transfer functions and gives the lowest RMSEP (0.21). The Danish training set containing 32 lakes with a long TP gradient (16-765 µg l^{-1}) has stronger predictability (r^2_{boot} = 0.79) but higher prediction error (RMSEP = 0.24). The WA-PLS-2 model for the Irish Ecoregion with an r^2_{jack} of 0.73 outperforms all the other transfer functions in both predictability and prediction error (see Table 6.6).

6.2.5 TP Optima of Cladocera Taxa

The Weighted Averaging (WA) method is used to produce the TP optima of common Cladocera taxa in the 31-lake training set as many taxa display a unimodal-like response along the TP gradient as shown in Figure 5.8. TP optima and tolerance of 30 common Cladocera taxa are listed in Table 6.7. *Rhynchotalona falcata* and *Bosmina longispina* have the lowest TP optima of 10.4 and 13.2 µg l^{-1} respectively; while *Oxyurella tenuicaudis* and *Leydigia leydigii* have the highest TP optima of 96.1 and 61.9 µg l^{-1} among all the 30 Cladocera taxa (see Table 6.7). All the other Cladocera taxa have TP optima in the range of 16-41 µg l^{-1}. Most Cladocera taxa have estimated TP tolerances

of less than 3 μg l^{-1} with the highest value of 3.05 μg l^{-1} for *Alona rustica* (see Table 6.7), reflecting relatively narrow ecological amplitude for most taxa in the study lakes.

Table 6.7 Weighted averaging TP optima and tolerance of 30 common Cladocera taxa (≥1 % at two sites) for 31 Irish lakes (TP optimum and tolerance are back-transformed to μg l^{-1} units; N2 data are based on square root transformed data; littoral and planktonic taxa are separated and listed in alphabetic order).

Taxon	Count	Max (%)	N2	Optimum (μg l^{-1})	Tolerance (μg l^{-1})
Acroperus harpae	31	5.80	26.56	29.92	2.51
Alona affinis	31	18.24	25.51	25.69	2.55
Alona costata	26	4.48	21.71	29.92	2.30
Alona guttata/rectangula group	31	21.55	27.93	27.44	2.62
Alona intermedia	20	14.53	14.79	17.82	2.44
Alona quadrangularis	31	15.03	27.12	27.54	2.53
Alona rustica	13	22.38	8.52	24.59	3.05
Alonella excisa	26	9.79	20.73	21.70	2.62
Alonella exigua	27	4.31	22.68	29.67	2.42
Alonella nana	31	11.63	27.36	28.35	2.55
Alonopsis elongate	7	1.39	6.68	16.87	2.19
Camptocercus rectirostris	22	2.78	18.82	20.48	2.27
Chydorus piger	24	12.50	17.07	28.13	2.32
Chydorus sphaericus	30	21.24	25.56	31.59	2.35
Eurycercus lamellatus	25	4.42	21.15	29.50	2.28
Graptoleberis testudinaria	29	9.03	24.20	30.10	2.33
Leydigia leydigii	14	2.59	11.89	61.86	1.61
Monospilus dispar	17	2.80	14.38	20.42	2.72
Oxyurella tenuicaudis	4	2.49	3.20	96.05	1.68
Phrixura rostrata	15	2.18	13.94	23.63	2.63
Pleuroxus laevis	17	2.00	14.59	29.93	2.48
Pleuroxus trigonellus	21	2.65	18.30	34.80	2.25
Pleuroxus uncinatus	15	2.89	13.18	27.44	2.02
Rhynchotalona falcate	7	2.10	6.50	10.44	2.06
Bosmina longirostris	26	86.77	18.11	36.30	2.04
Bosmina longispina	17	33.73	13.44	13.20	2.08
Daphnia longispina group	31	77.35	26.98	29.32	2.60
Daphnia pulex group	11	15.17	8.65	41.06	1.98
Leptodora kindtii	9	3.47	7.48	20.44	2.32
Sida crystalline	15	1.88	13.27	23.86	2.07

A comparison of TP optima for some common taxa with those from other Ecoregions is summarised in Table 6.8. *Rhynchotalona falcata* and *Alona intermedia* have the lowest TP optima in Ireland similar to those from Denmark. In meso-eutrophic lakes taxa like *Monospilus dispar*, *Camptocercus rectirostris*, *Alona affinis*, amd *Eurycercus lamellatus* are then commonly found across the Ecoregions. Among species with high TP optima three species, *Chydorus sphaericus*, *Pleuroxus trigonellus* and *Leydigia leydigii*, are generally dominant in Cladocera assemblages both in Ireland and Denmark. TP optima for these taxa are generally higher in Denmark than those in Ireland mainly due to the

larger number of eutrophic and hypertrophic lakes included the Danish training set (half of the 32 lakes have TP of above 100 μg l^{-1} in comparison to only two of the 33 lakes in the current study). Some common taxa were also comparable between Ireland and Canada in TP optima, like *Camptocercus rectirostris*, *Eurycercus lamellatus* and *Daphnia pulex* group. However, many species in Canada show morphological differences from those in Europe (Chengalath, 1987) and therefore it is difficult to compare some taxa like *C. sphaericus*, a common species across Europe (e.g. de Eyto *et al.*, 2003).

Table 6.8 Comparison of WA-inferred TP (μg l^{-1}) optima of selected Cladocera taxa from Ireland (current study), Denmark[1] (Brodersen *et al.*, 1998), Denmark[2] (Amsinck *et al.*, 2005) and British Columbia of Canada (Bos & Cumming, 2003) (taxa are ordered according to their TP values estimated by this study).

Taxa	Ireland	Denmark[1]	Denmark[2]	BC, Canada
Rhynchotalona falcata	10.4	26	-	-
Alona intermedia	17.8	19	-	-
Monospilus dispar	20.4	40	42	-
Leptodora kindtii	20.4	-	-	33
Camptocercus rectirostris	20.5	50	38	21
Alonella excisa	21.7	29	-	16
Sida crystallina	23.9	-	48	26
Alona rustica	24.6	24	-	-
Alona affinis	25.7	50	50	16
Alona guttata/ alcate lar group	27.4	25/109	75	24
Alona quadrangularis	27.5	78	46	24
Chydorus piger	28.1	27	-	13
Alonella nana	28.3	28	70	15
Eurycercus lamellatus	29.5	46	60	23
Alona costata	29.9	31	-	-
Acroperus harpae	29.9	44	50	23
Graptoleberis testudinaria	30.1	32	62	43
Chydorus sphaericus	31.6	123	76	-
Pleuroxus trigonellus	34.8	99	59	21
Daphnia pulex group	41.1	-	-	53
Leydigia leydigii	61.9	102	51	27

6.3 Comparison of Diatom- and Cladocera-inferred TP for 29 Sites

TP inference models have been developed based on diatoms and Cladocera respectively in the previous sections. The optimum TP inference models are the WA_{TOL_Inv} for the 70-lake diatom training set and the second component of WA-PLS from the 31-lake cladocera training set. Twenty-nine lakes are common to both training sets, therefore a comparison of TP values predicted for these sites can provide a unique insight into the

performances of models based on two different biological indicators, diatoms and Cladocera (see Figure 6.6). Both TP models were based on log_{10}-tranformed TP. In lake monitoring and management the measurement units (µg l^{-1}) are used for TP data. Therefore TP predicted by both models are compared in both transformed unit (log_{10}-tranformed TP) and measurement unit (µg l^{-1} TP) for theoretical and practical performance interpretation.

Diatom- and Cladocera-inferred TP for the 29 sites display good correlations (r = 0.816 and 0.854) with the observed data in Figure 6.6 (a) and (b). There are several outlier sites with slight deviance from the 1:1 line for both the diatom-inferred TP and the Cladocera-inferred TP, like Anascaul [ANS], Inchiquin [INQ] and Lene [LEN]. In the plot (c) good agreement between diatom- and Cladocera-predicted TP is evident with an r of 0.682 with several sites displaying moderate deviance from 1:1 line, e.g. Inchiquin, Lene and Mullagh [MUL]. This confirms that the diatom and Cladocera inference models are performing well and they show close correspondence in log_{10}-transformed TP for most of the 29 sites.

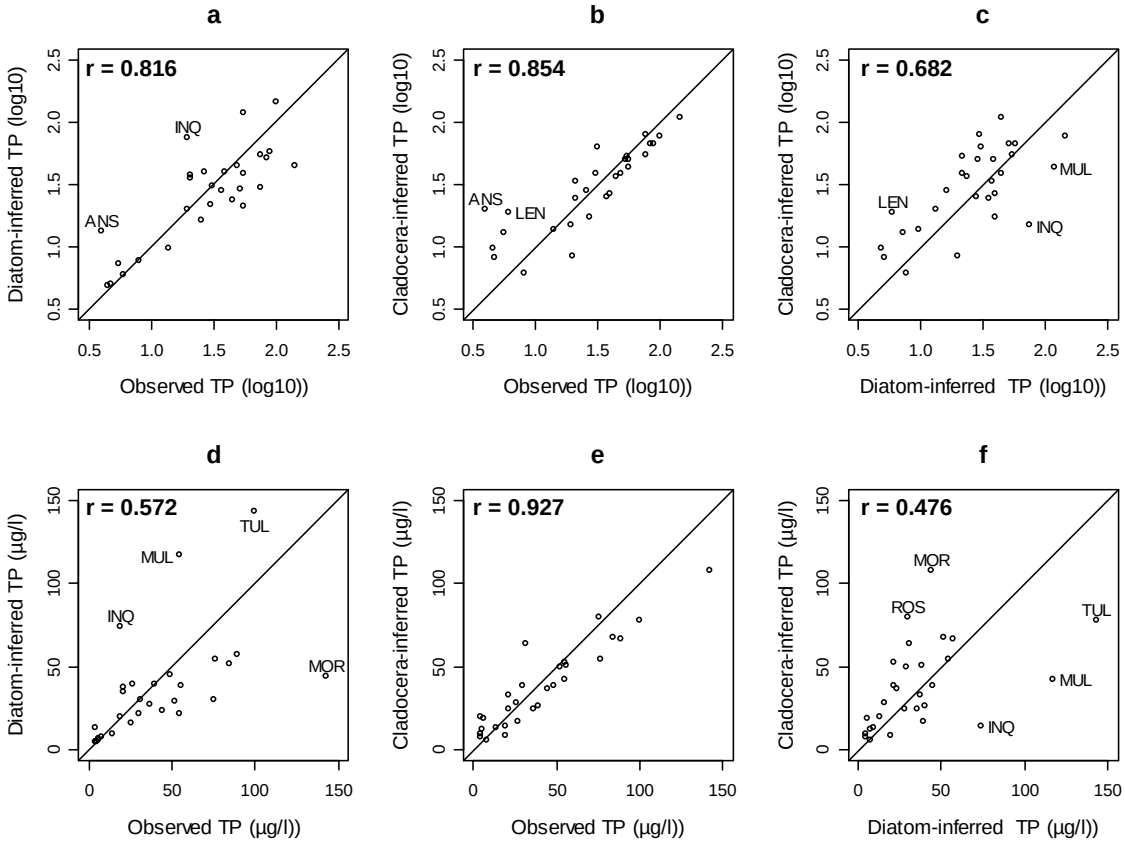

Figure 6.6 Comparison of observed TP, diatom- and Cladocera-inferred TP for 29 lakes (TP are log_{10}-transformed in plots a-c and back-transformed in plots d-f; Pearson's correlation coefficient is highlighted in each plot; sites with strong deviance from the 1:1 line are labelled with site codes).

However, when TP data are back-transformed to the measurement unit (μg l^{-1}), the relationship between the diatom-predicted and observed data decreases with an r of 0.572 (Figure 6.6 (d)). Similarly there is a decreased correlation between diatom- and Cladocera-inferred TP with an r of 0.476 (Figure 6.6 (f)). In plot (d) with diatom-inferred and observed TP several sites display strong divergence from the 1:1 line, including Inchiquin [INQ], Mullagh [MUL], Tullabrack [TUL] and Morgans [MOR] and these sites also displayed strong deviation from the 1:1 line in plot (f) with diatom- and Cladocera-predicted TP (Figure 6.6). These outlier sites labelled in plot (f) show stronger deviance from the 1:1 line than sites labelled in plot (c) with $\log_{10}$-transformed TP. In plot (f) most sites with big deviance from the 1:1 line are also obvious outliers in plot (d) with diatom-inferred and observed TP (Figure 6.6). In contrast the relationship between the Cladocera-inferred TP and the observed data is improved with an r of 0.927. This is also confirmed by the absence of obvious outlier sites in plot (e). This indicates that Cladocera-based model performs better than the diatom-based model when TP is back-transformed for the 29 sites.

6.4 Discussion and Conclusions

6.4.1 Data Transformation

Transformation of both measured TP and biological data has been shown to influence the model performances in the model development for both diatom and Cladocera training sets. Log-transformation of TP data improves the performances of all the models for both training sets. However, insignificant improvement of model performance was observed after log-transformation of TP data in an 82-lake diatom training set in Northeast America (Koster *et al.*, 2004). It was suggested that weaker sensitivity of diatom assemblages to TP than to pH/alkalinity in the American training set could account for the minor influence of TP data transformation. The current study in contrast indicates that model performance is improved in both WA, WA-PLS and PLS models after the data transformation even though TP is of secondary importance after pH/alkalinity in explaining the diatom assemblage variation (see Chapter 4). The improved influence of log-transformed TP data on diatom and Cladocera model performance could be due to the more normalized distribution of TP in the current

training set, which would strengthen the unimodal response of diatom and Cladocera species to TP (Koster *et al*., 2004). Untransformed biological data only outperform the square root transformed data in the WA model with tolerance downweighted in the diatom training set, but this model gives the best performance with the lowest RMSEP and highest or almost highest r^2_{jack} among all the WA and WA-PLS models for diatom data. This implies that in some cases untransformed ecological data can provide valuable information on the strength of relationship between biological assemblages and environmental variables. This phenomenon was also observed by Koster *et al.* (2004), where untransformed species data provided a model which predicted TP values in better accordance with the measured data. Therefore lack of data manipulation may result in lower model performance and therefore valuable information for further model evaluation, application and ecological interpretation may be lost. It is therefore recommended to test all possible inference models with both transformed and untransformed ecological and environmental data, before model selection is made on the basis of model performances.

6.4.2 Diatom and Cladocera TP Models

Performances of WA modelling and WA-PLS modelling generally show no great differences in regard to RMSEP and bias for the diatom training set. The WA model generally outperforms the WA-PLS model in regard of jack-knifed r^2 for the diatom training set. The short gradient length (0.886 SD) in the Cladocera data indicated that a linear response model was appropriate, however, the linear PLS models were generally outperformed by the unimodal-based models, WA and WA-PLS in terms of RMSEP and jack-knifed r^2. In particular the WA-PLS model displayed a significantly improved performance in comparison to the PLS model for the same set of data, indicating that non-linear ecological responses played a significant role. This is also evidenced by the response curves of Cladocera taxa along the TP gradient as shown in Figure 5.8 where almost half of the 20 common Cladocera taxa showed unimodal-like response curves. A better performance of unimodal-based WA-PLS in comparison to linear-based PLS for datasets with short ecological gradients has also been found by Birks (1998) for the first component of both models. Therefore the gradient length determined by the DCCA constrained by the environmental variable of interest (like TP in this study) may not be a good guide for selecting either linear or non-linear modelling methods. After suggesting

this criteria for method selection, Birks (1995) also highlighted that this guide may not work for specific data sets due to the lack of knowledge on model development, including the statistical properties of the reconstruction methods and the amount of noise within the data. Therefore non-linear modelling methods are recommended for model development when the ecological responses show linearity due to the short gradient length as determined by DCCA.

The strong performances of both TP models were evidenced by the close correlation between the $\log_{10}$-transformed TP predicted by diatom and Cladocera models at 29 sites. However, the much reduced correlation between the back-transformed TP given by both models implies that the optimal models based on transformed TP will produce stronger prediction errors when their reconstructed TP is back-transformed and used in the ecological assessment of lakes. This could pose a challenge for the application of TP inference models in identifying the reference conditions and setting targets for lake restoration. The relationship between TP residuals and observed values shows the same trend with overestimation at the low end of TP gradient and underestimation at the high end in the diatom and Cladocera models as in other studies (e.g. Bennion (1994), Lotter *et al*. (1998)). Therefore care is needed to be taken when these models are applied to fossil samples for environmental reconstruction as more errors can be created when the reconstructed values are near either end of the training set environment gradient. Other model methods, including WA with classical deshrinking and cross-validation with bootstrapping for sample-specific errors, can be used in reconstruction to get reliable and consensus results (Birks *et al.*, 1990; Birks, 1998).

In the absence of sampling additional sites the removal of outlier sites was found to improve the performance of inference model (Gasse *et al.*, 1995; Tibby, 2004). The removal of two oligotrophic sites, Veagh [VEA] and Caragh [CAR], significantly improved the performance of both WA and WA-PLS models for the diatom training set. Lough Veagh is identified as outlier site mainly due to its measured TP value of zero. Zero values used in model development indicate no statistical relationship with the ecological data and this explains the big divergence between the observed and predicted TP and TP residual. The surface sediment diatom assemblage in Caragh is dominated by *Aulacoseira subarctica* (49.5%), a meso-eutrophic species while the measured TP value is only 5.5 μg l^{-1}. Also in the 33-lake Cladocera training set, Lough Sillan [SIL] is a hypertrophic lake with a medium to large lake area (140 ha) with an observed TP of 141

μg l^{-1} but a high abundance (19.2%) of oligo-mesotrophic *Bosmina longispina* in the surface sediments (see Figure 5.3). Lough Lisnahan [LIS] is small (5.9 ha) and shallow (mean depth 1.44 m) and it also serves as source of water supply for local town Kilkee (Wemaëre, 2005). A high ratio of volume of daily abstraction in comparison to the lake volume can disturb the biological communities. Therefore the deviation between observed and predicted TP values is significant. Performance of Cladocera-inferred TP models were also evidently improved after both sites were excluded (see Table 6.5).

Littoral diatom taxa in the current study show analogous TP optima for the same taxa as inferred by a contemporary littoral algae survey of 32 Irish lakes (DeNicola *et al*., 2004). But TP optima of some taxa, particularly planktonic ones, display dissimilarity between both studies. This is a result of different sampling methods used. Only epilithic algae were sampled in the contemporary survey. Surface sediments used in this study generally contain diatom remains from a wide range of habitats in lakes. Therefore the surface sediment samples can provide more reliable information on the diatom responses to the nutrient level of lake waters. The same diatom taxa generally have similar TP optima in Finland (Miettinen, 2003) and the Alps (Wunsam & Schmidt, 1995) and both training sets have a similar TP gradient as the current study. Much higher TP optima were found in English lakes which contained more eutrophic lakes and covered a longer TP gradient (Bennion, 1994). Similarly a higher number of eutrophic lakes in the Danish training set (Brodersen *et al*., 1998) is a probable cause of higher TP optima for some Cladocera taxa.. Therefore TP optimum of the same Cladocera or diatom species can vary between training sets depending on the features of training set, including the length of TP gradient, Ecoregion feature, number of sites, geographical area and lake types.

6.4.3 Sources of Uncertainty

Although the Irish Ecoregion TP inference models based on diatoms and Cladocera are robust with high predictability and low prediction errors and bias, there are still many possible sources of error as highlighted by Anderson (1995a). These errors include taxonomic harmonization, spatial variability of biological assemblages in surface sediments and the influence of unmeasured variables. As the top 2-3 cm of 17 samples were sampled using an Echman Grab, rather than top 0.5-1 cm sampled by the gravity

corer for all other lakes, the inconsistent sampling methods can contribute to the errors of inference model. Furthermore, another potential error source is the water chemistry variability, both temporal and spatial variability, in this training set. Many of the training set lakes were only sampled for hydrochemistry on one occasion, generally in the summer season and water sampling and hydrochemical analysis were conducted by several projects in Ireland between 1996 and 2001. The lack of internal consistency and multiplicity of water sampling and analysis may reduce the reliability of hydrochemical data used in this study (Birks, 1998). Even so, the strong predictability and low errors of the developed TP inference models indicate that these factors may increase the model errors and reduce the accuracy of inference models but these influences are still relatively insignificant for the diatom and Cladocera models developed.

6.4.4 Conclusions

A diatom-based TP model (r^2_{jack} = 0.743, RMSEP = 0.213, n = 70) using WA with tolerance downweighted and inverse deshrinking, and a Cladocera-based TP model (r^2_{jack} = 0.729, RMSEP = 0.206, n = 31) using the second component of WA-PLS were developed for the Irish Ecoregion. Data transformation of both TP and biological data and removal of outlier sites significantly improved the performances of inference models. It is therefore strongly recommended that data manipulation of both environment and ecological data be tested before an optimal model can be decided. A close correlation between diatom and Cladocera-inferred TP for the same 29 sites also validated the performance of both diatom and Cladocera inference models developed. The construction of optimal inference models has not been easy due to the insufficient knowledge on related statistical methods and data properties. The selection of linear or non-linear modelling methods determined by the gradient length of DCCA constrained by the environment variable of interest can be misleading as shown in the development of Cladocera-TP models in this study. In the case of a short ecological gradient both linear and non-linear methods (particularly WA-PLS) should be performed for model comparison. Comparison of TP optima of some common taxa with those from other training sets show that TP optima for diatom or Cladocera taxa can vary between training sets depending on the gradient length, lake types, number of sites, Ecoregion feature etc.

Chapter 7: Identification of Reference Status of Seven Irish Lakes

Determination of reference status is a key factor for lake restoration and management under the Water Framework Directive (WFD) (European Union, 2000). Examination of the degree of deviation from the reference conditions is required to assess water quality status. However, very limited lake water monitoring data can provide long-term information on water quality in the absence of significant human influence. Palaeolimnological methods have been successfully applied providing the information on pre-impact conditions for polluted lakes (e.g. Smol, 2002; Cohen, 2003). These methods have included the 'top and bottom' approach which assumes that sediment core bottom samples represent reference conditions of study lakes (e.g. Dixit *et al.*, 1999; Bennion *et al.*, 2004a).

The establishment of pre-impact conditions in lakes has been a priority task to help implement the WFD for Ireland (Irvine *et al.*, 2002), as well as other EU member states like UK (Bennion *et al.*, 2004a) and Denmark (Søndergaard *et al.*, 2005). Eutrophication has long been the principal pressure on lake water quality in Ireland (Jennings *et al.*, 2003; Toner *et al.*, 2005). This Chapter aims to reconstruct TP levels under the pre-impact conditions using the TP inference models developed in Chapter 6. As single palaeolimnological proxies have their own strengths and weaknesses, multi-proxy analysis on the other hand can be advantageous in reaching consensus results and identifying the disagreement among proxies (Lotter, 2005). Both diatoms and Cladocera are combined and employed to infer nutrient status via the examination of the top and bottom of sediment cores for seven impacted lakes. TP values inferred by diatom- and Cladocera-based transfer functions provide independent reconstructions of nutrient status and comparison of both indicators can help cross-validate the reference conditions of the seven study lakes.

7.1 Study Sites

Seven sites were selected for top-bottom analysis of sediment cores and they are mainly located in the west and North of Ireland (see Figure 7.2), with three lakes (Atedaun, Ballybeg and Inchiquin) from Co. Clare, two from Co. Cavan (Loughs Sillan and

Mullagh) and the other two from Counties Tyrone (Lough Crans) and Monaghan (Lough Egish). They are all located at altitudes below 150 m and Loughs Egish, Inchiquin and Sillan have lake areas of greater than 100 ha while the other four have lake areas of less than 50 ha. Catchment areas of the seven lakes vary widely between 60-28,250 ha and are mainly covered by pasture (> 50%) (Table 7.1), except for Atedaun and Inchiquin, whose pasture cover are surpassed by the mixed semi-natural areas (Wemaëre, 2005). The main bedrock geology for these seven lakes includes Carboniferous Limestone and Silurian Quartzite (Taylor *et al.*, 2006). All the seven lakes have medium to high alkalinity values between 52.0-161.8 mg l^{-1} with high pH values in the range of 7.8-8.5 (see Table 7.1). Five lake types are represented by these seven lakes as shown in Table 7.1: two lakes in Lake Type 9 (small, shallow, high alkalinity), two in Lake Type 12 (large, deep, high alkalinity), one in each of Lake Types 5 (small, shallow, medium alkalinity), 7 (small, deep, medium alkalinity) and 8 (large, deep, medium alkalinity) respectively.

Figure 7.1 Site location map of 7 study lakes.

Table 7.1 Summary of main physico-chemical and catchment features of seven lakes.

Lake Name	Irish Grid Reference	Alt. (m)	Lake Area (ha)	Catch. Area (ha)	Max Depth (m)	Mean Depth (m)	Alkal. (mg l^{-1})	pH	TP (µg l^{-1})	Pasture Land (%)	Lake Type
Atedaun	R 295 885	22	38.0	28250.0	13.0	1.4	135.4	8.0	36.7	38.4	9
Ballybeg	R 330 739	10	19.7	414.0	5.7	2.7	128.0	7.9	84.3	53.7	9
Crans	H 711 568	95	8.5	59.5	12.0	6.7	78.0	8.5	89.0	85.0	7
Egish	H 795 132	162	121.7	784.3	12.0	5.0	69.0	8.1	344.0	86.4	8
Inchiquin	R 268 897	35	106.9	14714.0	31.0	12.2	161.8	8.2	19.3	32.0	12
Sillan	H 700 070	94	140.0	-	12.0	6.0	140.0	8.3	141.0	79.8	12
Mullagh	N 677 855	120	35.1	114.2	8.1	2.3	52.0	7.8	55.0	95.7	5

Six of the seven lakes are eutrophic and hypertrophic and only Lough Inchiquin is mesotrophic with the lowest TP value of 19.3 µg l^{-1} based on the trophic classification of OECD (1982). Generally all these seven lakes are impacted not only by agricultural activities as evidenced by the dominant land cover of pasture lands for all seven lakes, but also influenced by urban and industrial development surrounding these lakes. For example, the total area farmed relative to the whole drainage area is 68%, 73% and 64% for Loughs Atedaun, Ballybeg and Inchiquin respectively based on an agricultural census in 2000 (Wemaëre, 2005). Cattle density for Loughs Mullagh and Egish was surveyed and found to be 2.85 and 1.68/ha in 1990 (Irvine *et al.*, 2001). There is a creamery and meat-processing factory close to Loughs Egish and Sillan respectively (Taylor *et al.*, 2006). Treated sewage is channelled to Lough Sillan and there is a caravan site close to the lake. Lough Inchiquin also serves as the source of water supply for the local town of Corrofin (Wemaëre, 2005).

All seven lakes were cored in August and September 2004 using a Renberg gravity corer. The lengths of lake cores vary between 31-41cm. The dating based on ^{210}Pb, ^{241}Am and ^{137}Cs measurements and sedimentation rates for six of the seven lakes are summarised in Table 7.2. An irregular ^{210}Pb concentration profile was measured for the sediment core of Atedaun and this precluded the use of other lake models to estimate chronology and sedimentation rate (Dalton *et al.*, 2006). Estimated ages for core bottoms of the other six lakes ranged from late 18th Century (Egish) to the middle of 20th Century (Mullagh). The author conducted Cladocera analysis and Manel Leira provided diatom data for the same samples. Diatom and Cladocera assemblages are summarised for each of the seven lakes in the following section before the application of TP inference models for quantitative reconstruction of TP.

Table 7.2 Depth and estimated ages of bottom samples of seven lakes (Taylor *et al.*, 2006).

Lake	Depth of Bottom Sample (cm)	Estimated Sedimentation Rate (g cm^{-2} yr^{-1})	Estimated Age (±1 SD)
Atedaun	39-40	-	-
Ballybeg	30-31	0.026 ± 0.002	ca. 1889 AD ± 11 at 29-30cm
Crans	39-40	0.028 ± 0.007	ca. 1825 AD ± 37 at 38-39cm
Egish	31-32	0.017 ± 0.004	ca. 1781 AD ± 50 at 30-31cm
Inchiquin	40-41	0.12 ± 0.02	ca. 1931 AD ± 8 at 38-39cm
Mullagh	38-39	0.04 ± 0.01	Pre-1950 AD ± 8 at 37-38cm
Sillan	38-39	0.053 ± 0.012	ca. 1905 AD ± 13 at 37-38cm

7.2 Diatom and Cladocera Assemblages of Top and Bottom Samples

In the core top and bottom samples of the seven study lakes 114 diatom taxa and 34 Cladocera taxa occur respectively. Common diatom and Cladocera assemblages for the core top and bottom of the seven lakes are shown in Figure 7.2 and Figure 7.3. All the diatom and Cladocera species or species group are expressed in relative abundance (%).

7.2.1 Lough Atedaun

Forty-three and 46 diatom taxa occur in the sediment core bottom and top from Lough Atedaun respectively and diatom assemblages in both samples are mainly composed of non-planktonic taxa. The dominant taxa are meso-eutrophic taxa, *Cocconeis placentula*, *Achnanthes minutissima* and *Amphora pediculus* and only slight differences are found for these dominant taxa between the core top and bottom (see Figure 7.2). Changes in diatom assemblages include a moderate increase in the subdominant eutrophic taxa *Cyclostephanos invisitatus* and meso-eutrophic *Gomphonema pumilum*, as well as an decrease of oligo-mesotrophic *Cymbella microcephala*, *Navicula cryptotenelloides* and *Fragilaria brevistriata* in the core top. Lough Atedaun is currently a eutrophic lake with a mean TP value of 36.7 μg l^{-1} (see Table 7.1) and the dominance of meso-eutrophic diatom taxa correspond well with the current nutrient status. The dominance of non-planktonic diatoms also reflects the shallow lake basin of Atedaun (mean depth = 1.2 m).

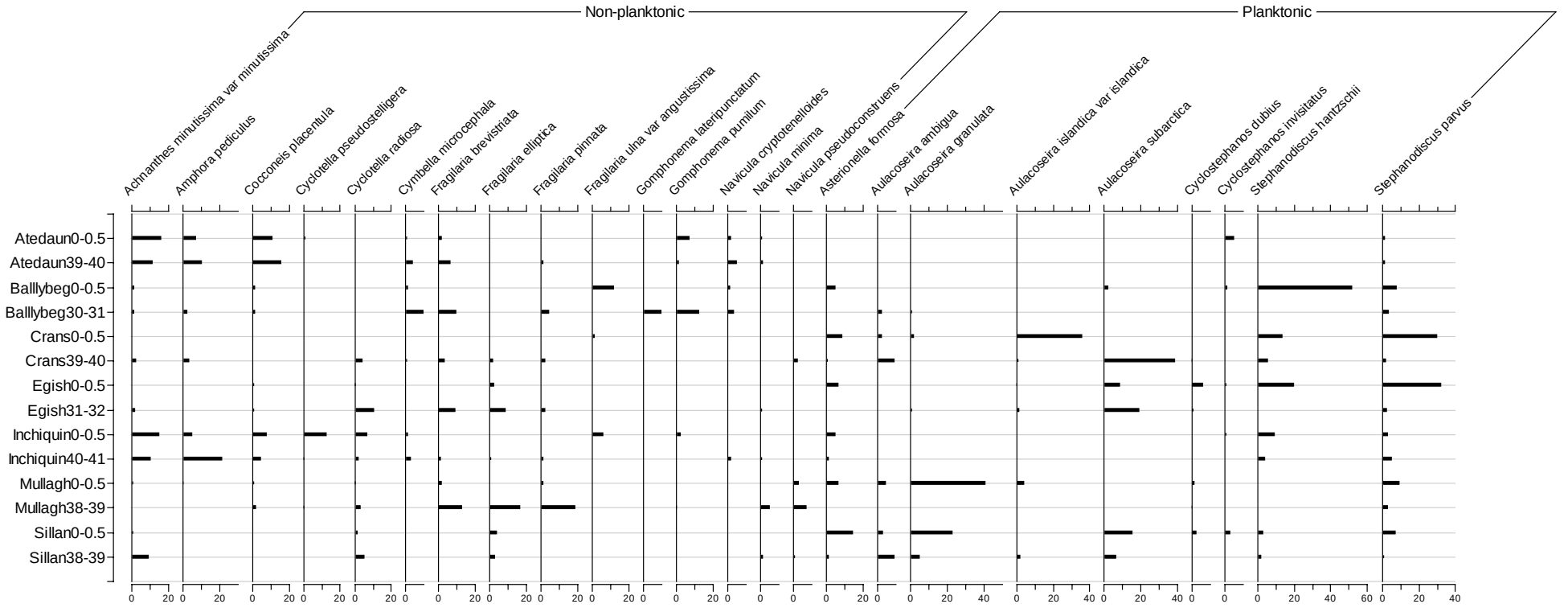

Figure 7.2 Comparison of dominant diatoms (above 5%) in surface and bottom sediments of seven study lakes (data provided by Manel Leira).

Cladocera species richness displayed little change between the core bottom (23) and top (21). Chydoridae taxa, which prefer the littoral areas of lakes, are dominant among the Cladocera assemblage for both surface and bottom sediment samples (see Figure 7.3). Dominant littoral Cladocera taxa, including *Chydorus sphaericus* and *Alona guttata/rectangula* group, show relatively stable abundance between core bottom and top. *C. sphaericus* is often found to tolerate a wide range of environments but prefers eutrophic conditions (de Eyto *et al*., 2002). Subdominant taxa in the core bottom like *Alonella exigua* and *Graptoleberis testudinaria*, typically found in mesotrophic lakes with abundant macrophytes (Duigan, 1992), display a decrease in relative abundance between the core bottom and core top. A reduction in macrophyte cover was found to support less plant-associated chydorids (Jeppesen, *et al*., 2001). A sharp increase in planktonic *Daphnia longispina* group occurred between the core bottom to core top with the disappearance of oligo-mesotrophic *Bosmina longispina* group. These changes correspond well with reduced abundances of plant-associated taxa like *A. exigua* and *G. testudinaria*, indicating an increased nutrient status in association with a decrease in macrophyte cover (see Figure 7.3).

7.2.2 Lough Ballybeg

The abundance of diatom taxa decreased by 25% from 32 taxa in the core bottom to 24 in core top from Ballybeg. Non-planktonic diatoms dominate in the core bottom, including *Gomphonema pumilum*, *G. lateripunctatum*, *Fragilaria brevistriata* and *Cymbella microcephala* (see Figure 7.2). These oligo-mesotrophic taxa are replaced by planktonic diatoms in the core top sample, particularly *Stephanodiscus hantzschii* which has a relative abundance of over 50%. Other subdominant planktonic taxa include *S. parvus*, *Fragilaria ulna var. angustissima* and *Asterionella formosa*. These species are good indicators of lake eutrophication. The complete shift from a benthic-dominated to a planktonic-dominated assemblage points to a change from a nutrient-poor to a nutrient-enriched lake.

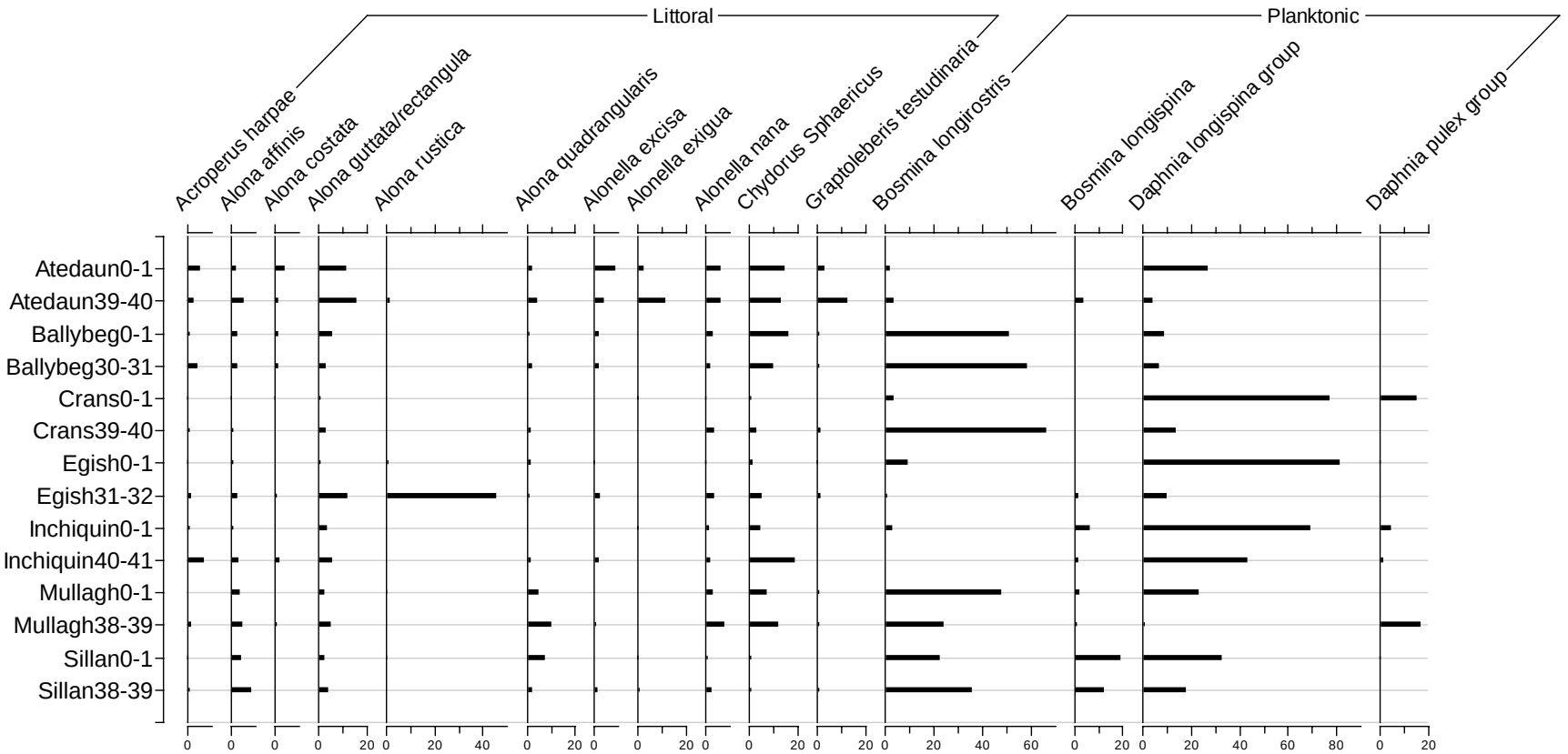

Figure 7.3 Comparison of dominant Cladocera (over 5%) in core top and bottom sediments of seven lakes.

The Cladocera assemblages show little variation in species richness unlike the diatom assemblages. The planktonic cladoceran assemblage was more abundant than chydorid assemblage in both top and bottom samples and the abundance and structure of planktonic Cladocera changed only slightly (see Figure 7.3). The littoral species *C. sphaericus* was dominant in both top and bottom samples and the main changes of note are an increase in *C. sphaericus* in the surface sample, and a concomitant decrease in plant-associated species *Acroperus harpae*. These small changes in the Cladocera assemblages may indicate an increase in nutrient status but no major change is implied based on the minor assemblage shift between the core bottom and top. The dominance of meso-eutrophic taxa, mainly including littoral *C. sphaericus* and planktonic *Bosmina longirostris*, suggest a relative stable meso-eutrophic nutrient level at Ballybeg between the core bottom and top.

7.2.3 Lough Crans

Meroplanktonic and meso-eutrophic *Aulacoseira subarctica* is dominant in the core bottom from Crans in relative abundance of nearly 40% with subdominant taxa *A. ambigua* and *Stephanodiscus hantzschii* in much lower abundances (see Figure 7.2). In comparison, both *A. islandica* and *S. parvus* are co-dominant in the core top with relative abundance of above 30%, while *S. hantzschii* and *Asterionella ormosa* are subdominant in abundances of around 10%. Diatom species richness decreased dramatically from 35 in the bottom sample to 14 in the core top. The shift from diatom assemblage dominated by meso-eutrophic taxa in the bottom sample to the one dominated by eutrophic taxa implies an increase in nutrient level from the reference period to the current day. This may explain the sharp decrease in diatom species abundance between the core bottom and the top samples.

In comparison to the large decrease in diatom species richness, only a slight decrease in species richness was observed for Cladocera assemblages with 20 taxa in the core bottom and 17 at the top. Planktonic taxa were much more dominant in the surface compared to the bottom sediment (see Figure 7.3). This may indicate a shift away from littoral species coincident with a decrease in suitable littoral habitat. The dominant Cladocera shifted from *Bosmina longirostris* in the bottom sample to the *Daphnia*

longispina group in the surface sediment with increased abundance in the *D. pulex* group. Littoral taxa display reduced abundance in the core bottom compared to the top. The high abundance of planktonic Cladocera in the surface sample may reduce the relative weight of littoral taxa among the total Cladocera assemblage and obscure a direct comparison with the littoral assemblage in the bottom sample.

7.2.4 Lough Egish

The dominant diatom species in the core bottom from Lough Egish is *Aulacoseira subarctica*, which was also dominant in the bottom sediment of Lough Crans, but in lower abundance of around 20%. Subdominant taxa are mainly oligo-mesotrophic, including both planktonic *Cyclotella radiosa* and periphytic *Fragilaria brevistriata* and *F. elliptica*. In the core top planktonic and eutrophic taxa *Stephanodiscus parvus* and *S. hantzschii* are dominant in abundances of ca. 30% and 20% respectively. Planktonic *Aulacoseira subarctica*, *Asterionella frmosa* and *Cyclostephanos dubius* are subdominant in abundances of less than 10%. The shift from a diatom assemblage dominated by a mixture of planktonic and epiphytic diatoms to one dominated by planktonic diatoms indicates a change from a meso-eutrophic state in the core bottom to an eutrophic state at the top. However, only a minor decrease of species abundance from 43 to 37 was observed between the core bottom and the top.

There is little change in species richness in the Cladocera assemblages from 21 in the bottom to 20 in the core top corresponding to the relative stable species richness of diatom community. Nevertheless, Cladocera assemblage structure changes completely from being chydorid-dominated in the bottom sample to being planktonic-dominated in the surface sample (see Figure 7.3). This shift in the Chydoridae community over the core sedimentation period includes a decline in relative abundance from nearly 50% to less than 10% for *Alona rustica*, a common species in dystrophic water bodies and generally absent in enriched environments (Duigan, 1992). The surface assemblage is dominated by the planktonic *Daphnia longispina* group in abundance of around 80% and sub-dominated by the planktonic *Bosmina longirostris* in abundance of 10%. The complete shift in assemblage structure reflects an unambiguous increase in the nutrient status of Lough Egish between the sedimentation periods.

7.2.5 Lough Inchiquin

Benthic diatom taxa are the main components in the bottom sample of Inchiquin, including the dominant oligo-mesotrophic *Amphora pediculus* and subdominant ubiquitous *Achnanthes minutissima* (see Figure 7.2). The latter taxa becomes dominant in the surface assemblage and planktonic taxa are the main subdominant taxa, e.g. mesotrophic *Cyclotella pseudostelligera* and eutrophic *Stephanodiscus hantzschii*. Alhough the abundance of planktonic diatoms increases in the surface assemblage compared to that in the bottom assemblage, benthic taxa are still common in the surface sediment in constrast to the predominance of planktonic taxa in surface sediments of Loughs Ballybeg, Crans and Egish mentioned above. Therefore, a moderate shift from a benthic assemblage in the core bottom to a planktonic-benthic assemblage in the core top may indicate a moderate change in water quality in Lough Inchiquin. Diatom species richness decreases by 22% from 37 in the core bottom to 29 at the core top.

The planktonic *Daphnia longispina* group are dominant in both top and bottom samples and its abundance increases from around 40% in the bottom to 70% in the recent sediments (see Figure 7.3). No obvious shift in dominant taxa can be observed within the Chydoridae assemblage between the core bottom and the top, but most of common chydorids display decreased abundance, e.g. *C. sphaericus* decreased from 20% in the bottom to less than 10% in the top. Generally, no significant change was observed in the Cladocera assemblage structure except the moderately increased abundance of planktonic taxa between the core bottom and the top. At the same time, a slight increase in Cladocera species richness from 22 in the core bottom to 24 in the core top was found in Lough Inchiquin.

7.2.6 Lough Mullagh

Three epiphytic diatom taxa *Fragilaria pinnata*, *F. elliptica* and *F. brevistriata* co-dominate the bottom assemblage in the sediment core from Mullagh, with benthic *Navicula pseudoconstruens* and *N. minima* sub-dominant (see Figure 7.2). However, these common taxa in bottom sample diminish dramatically in the surface sediment sample and are all replaced by planktonic taxa. *Aulacoseira granulata* is dominant in abundance of around 40% with *Stephanodiscus parvus* and *Asterionella formosa* sub-

dominant. The obvious diatom succession from benthic taxa in the core bottom to the planktonic and eutrophic taxa indicates a changed water quality with an increased nutrient level for Lough Mullagh. This was also reflected by the decrease in diatom species richness from 34 at the core bottom to 28 at the core top.

Both bottom and surface Cladocera assemblages are dominated by planktonic *Bosmina longirostris* but with increased abundance from a basal 20% to over 40% in the surface sediments (see Figure 7.3). The subdominant taxa changed from the *Daphnia pulex* group in the core bottom to the *D. longispina* group in the core top. All the common littoral taxa in the bottom sample show decreased abundance in the core top including *C. sphaericus*, a dominant chydorid species. Therefore, the increased proportation of planktonic taxa in the recent sediments may reflect a change in water quality for Lough Mullagh but no clear trend is indicated by the Cladocera assemblage. Species richness in Cladocera assemblage decreases from 25 in the core bottom to 20 in the core top concurrent with the decrease in diatom species richness.

7.2.7 Lough Sillan

Diatom species richness was reduced by one-quarter from 48 in the core bottom to 36 in the core top at Lough Sillan. Planktonic *Aulacoseira ambigua* and epiphytic *Achnanthes minutissima* co-dominate the bottom assemblage in abundances of around 10%, with mainly planktonic taxa sub-dominating including *Aulacoseira subarctica*, *A. granulata* and *Cyclotella radiosa* (see Figure 7.2). The domimance of these mesotrophic and eutrophic taxa, as well as the high diversity of diatom taxa, reflect a meso-eutrophic lake environment. In the core top dominant taxa are the planktonic *A. granulata* in abundance of around 20%, with *A. subarctica*, *Asterionella formosa* and *Stephanodiscus parvus* sub-dominating. The clear shift in dominant diatom taxa and concurrent reduction in species richness point to an increased nutrient level.

Dominant Cladocera taxa shifts from *Bosmina longirostris* in the bottom sample to the *Daphnia longispina* group in the surface sediment, and both are typical taxa in open waters (see Figure 7.3). Oligo-mesotrophic *Bosmina longispina* sub-dominates in both samples with only slight change in abundance. No clear change in littoral Cladocera was observed, however, they only represent a proportion (25-34%) of the Cladocera

assemblage (see Table 7.3). Cladocera species richness increases from 24 in the core bottom to 28 in the core top in contrast to the clear reduction in diatom species richness. The Cladocera taxa show moderate change in assemblage structure between the core bottom and top and may therefore indicate insignificant changes in aquatic environment in Lough Sillan.

7.2.8 Assessment of the Roles of Planktonic Cladocera in Indicating Nutrient Level

Planktonic Cladocera, particularly the *Daphnia longispina* group and *Bosmina longirostris*, are dominant in most of the top and bottom samples of the seven lakes (Figure 7.3). Detailed ratios of the planktonic and littoral Cladocera are summarised in Table 7.3. Six of the seven lakes show clear increases in the ratio of planktonic to littoral Cladocera in the core bottom compared to the top except for Lough Ballybeg. The most distinct increase in planktonic/littoral ratio is found in Loughs Egish (0.16 in the core bottom to 11.02 at core top) and Crans (4.12 to 25.27). The increased planktonic/littoral ratio corresponds with the decrease in mesotrophic *Alona rustica* in Egish but no clear change in chydorid assemblages can be found for Lough Crans. The proportions of littoral and planktonic Cladocera are interdependent when both groups are included in the Cladocera assemblage as indicators of lake environments (Hofmann, 1987b). Therefore, the higher proportion of planktonic Cladocera may obscure the change in the littoral Cladocera assemblage in indicating nutrient level. For example, the whole Cladocera assemblages from Lough Sillan (see Figure 7.3) show little change between the core bottom and the top, however, most of the mesotrophic chydorid taxa show decreased abundances (e.g. *Alona intermedia*, *Alonella excisa*, *Chydorus piger* and *Monospilus dispar*) (see Figure 7.4). Therefore, high abundances of planktonic Cladocera can reduce the strength of chydorid assemblage in reflecting the actual nutrient level when both groups are included.

Table 7.3 Summary of planktonic and littoral Cladocera for the seven lakes.

Lake	Atedaun		Ballybeg		Crans		Egish		Inchiquin		Mullagh		Sillan	
Depth (cm)	0-1	39-40	0-1	30-31	0-1	39-40	0-1	31-32	0-1	40-41	0-1	38-39	0-1	38-39
% Planktonic	30	12	60	66	96	80	92	14	84	46	74	43	75	66
% Littoral	70	88	40	34	4	20	8	86	16	54	26	57	25	34
Planktonic/ Littoral	0.43	0.14	1.48	1.94	25.27	4.12	11.02	0.16	5.07	0.86	2.8	0.75	3.03	1.94

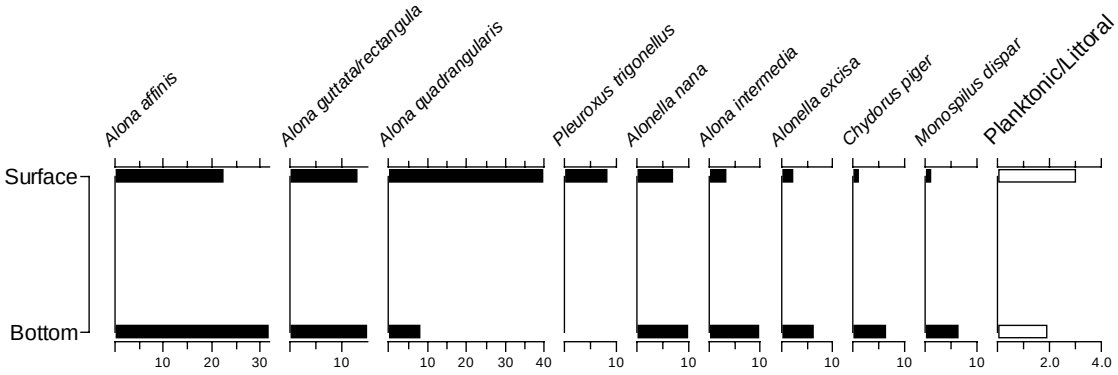

Figure 7.4 Comparison of the dominant taxa (5%) of chydorid assemblage in surface and bottom samples of Lough Sillan.

Planktonic Cladocera mainly occupy open water habitats and they are important components of the zooplankton community in lakes (Wetzel, 2001). Some planktonic taxa have been well documented as good indicators of nutrient level, e.g. a shift from *Bosmina longispina* to *B. longirostris* as a consequence of eutrophication (Frey, 1988a). However, the population and structure of planktonic Cladocera are liable to selective predation by invertebrate and vertebrate predators, particularly fish (e.g. Brooks & Dodson, 1965; de Bernardi *et al.*, 1987; Jeppesen *et al.*, 2001). The abundance and structure of planktonic Cladocera in lake sediments have been observed to correspond well with fish introductions (Leavitt *et al.*, 1994; Jacques *et al.*, 2005) and change in fish population structure and stocking (Leavitt *et al.*, 1989). The use of planktonic Cladocera remains in lake sediments can also be problematic as the remains of some planktonic Cladocera are often not as well represented as littoral chydorids in lake sediments (Frey, 1988a; Rautio *et al.*, 2000). Care is needed be taken when the full Cladocera assemblage in lake sediments are used for inferring the history of water quality. Both the inclusion and exclusion of planktonic Cladocera should be explored before a consensus conclusion is made on water quality.

7.3 Comparison and Evaluation of Diatom- and Cladocera-inferred TP

The TP inference models based on diatoms and Cladocera developed in Chapter 6 can be applied in surface and bottom samples of the seven lakes to reconstruct the nutrient levels. A TP transfer function based on only littoral Cladocera (chydorids) (not illustrated here) was also developed to exclude the influence of planktonic Cladocera. A

third component Partial Least Square (PLS-3) model gave the best performance with an r^2_{jack} of 0.664 and RMSEP of 0.226. This model was based on square root transformed chydorid data and log_{10}-transformed TP from 32 lakes. Prediction errors can vary between samples from different lakes. Bootstrapping is used in the reconstruction as this method can provide sample-specific errors (Birks *et al.*, 1990). Four diatom- and Cladocera-based TP models, which were cross-validated with bootstrapping and gave the best performances, are selected for TP reconstruction for the seven lakes. All the four models are based on log_{10}-transformed TP data. Reconstructed and back-transformed TP (μg l^{-1}) with errors for the top and bottom samples of the seven study lakes inferred by these four models are shown in Figure 7.6. In Weighted Averaging (WA) models, overestimation at low end of TP gradient and underestimation at high end observed can exaggerate the errors of predicted TP values. Therefore, diatom-based WA models based on 70 lakes are applied with both classical and inverse deshrinking. The WA-PLS-2 model is based on the full Cladocera assemblages from 31 lakes and PLS-3 model based on chydorids from 32 lakes. The results for each of the seven lakes are summarized below.

For Lough Atedaun, all the predicted TP values of the top sample given by the four models are close to the observed TP value of 36.7 μg l^{-1}. The TP estimated by the PLS-3 model are higher in both bottom and top samples than the other three models (see Figure 7.6). For Lough Ballybeg, PLS-3 and WA-PLS-2 models give higher TP values for both core top and bottom than the two diatom-based WA models. All the models show the same trend of a decreased TP in the bottom than in the core top with an increase of 30-40 μg l^{-1} TP by the diatom models compared to ca. 10 μg l^{-1} by the two Cladocera models. For the surface sediment of Lough Crans, PLS-3, WA-PLS-2 and WA_{Inv} models all give an inferred TP similar to the observed TP of 89 μg l^{-1}. Little change in TP estimated by the Cladocera-based models can be observed between the core bottom and top (see Figure 7.6). However, the diatom models infer a significant increase in TP values by ca. 60-80 μg l^{-1}. For the surface sediment from Lough Egish, predicted TP by all the four models show good correspondence in the range of 45-70 μg l^{-1}. The inferred values, however, deviate substantially from the observed TP value of 344 μg l^{-1} (see Table 7.1). TP predicted for the core bottom by the four models are all consistent in the TP range of 20-30 μg l^{-1}. All the four models shows an increase in inferred TP by around 20-40 μg l^{-1} between the core bottom and top indicating a clear trend of eutrophication (see Figure 7.6).

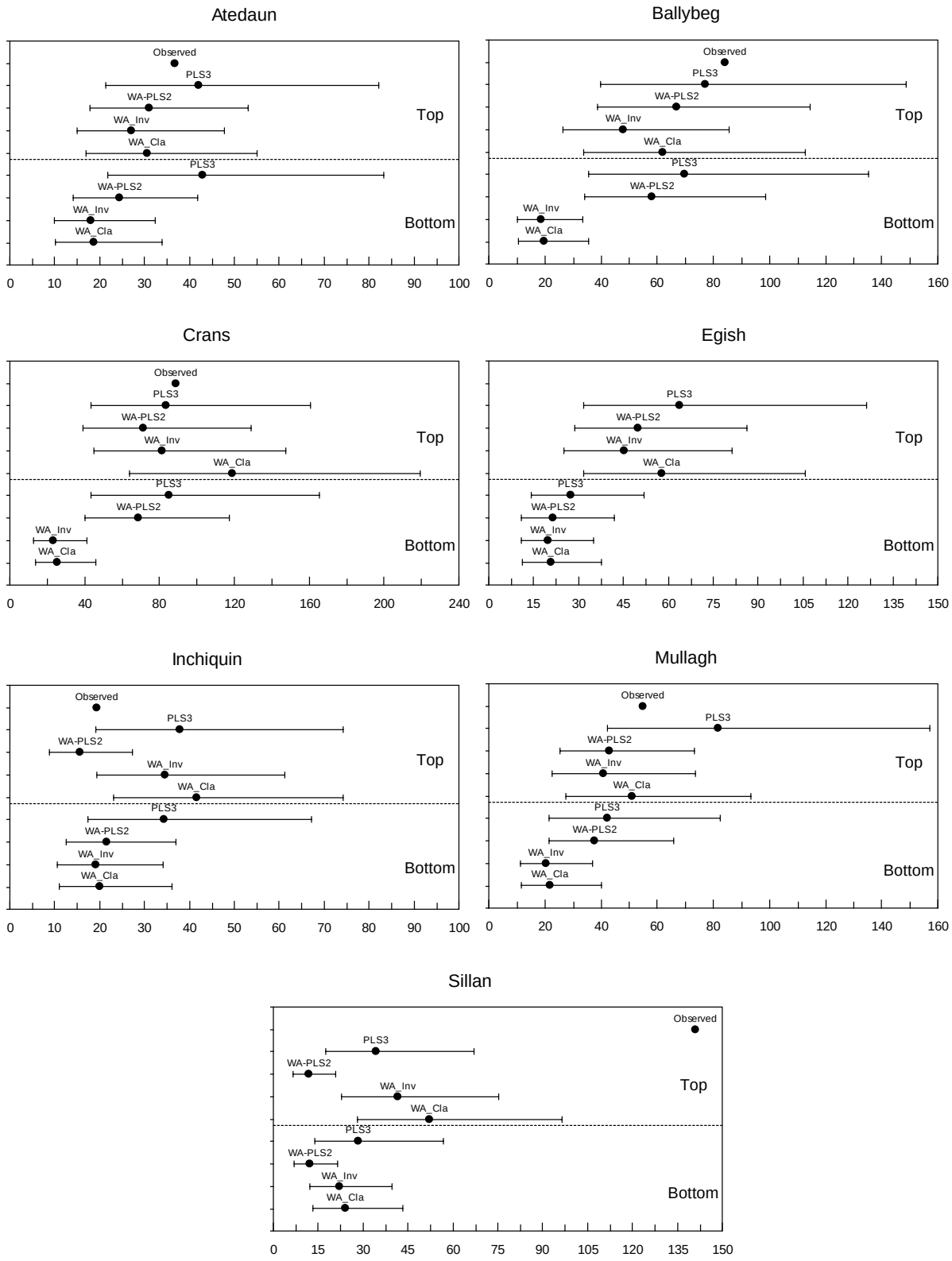

Figure 7.5 Comparison of TP values predicted for the core bottom and top for seven lakes by four inference models (Horizontal axis shows the TP values (µg l^{-1}); see text on details of PLS-3, WA-PLS-2, WA_{Inv} and WA_{Cla} models; error bar represents prediction error; observed TP for Egish (344 µg l^{-1}) is not shown).

The PLS-3, WA_{Inv} and WA_{Cla} models all gave predicted TP of around 35-40 μg l^{-1} for the surface sediment from Lough Inchiquin and this is higher than the observed TP of 19.3 μg l^{-1}. The chydorid-based PLS-3 model estimated a slight increase in TP between the core bottom and top, while the diatom models inferred a more pronounced increase. A higher TP (ca. 80 μg l^{-1}) is estimated by the PLS-3 model compared to the other three models (ca. 40-50 μg l^{-1}) for the surface sediment from Lough Mullagh (see Figure 7.6). For the bottom sample, all the four models gave predicted TP in the range of ca. 20-40 μg l^{-1}. Predicted TP for the surface sediment of Lough Sillan is in the range of around 30-50 μg l^{-1} based on the PLS-3 and the two WA models compared to ca. 10 μg l^{-1} by the WA-PLS-2 model. All the results are in strong contrast with the observed TP of 141 μg l^{-1}. Little change was found in estimated TP based on the WA-PLS-2 model between the core bottom and top while the PLS-3 model indicated an increase of less than 10 μg l^{-1} TP. Both diatom models suggested an increase of 20-30 μg l^{-1} in TP between the core bottom and top.

The quantitative TP inference models show obviously increased nutrient level in Loughs Ballybeg, Egish and Mullagh between core bottom and core top based on both diatoms and Cladocera. For the other four lakes (Atedaun, Crans, Inchiquin and Sillan), Cladocera-based models display little changes in estimated TP between the core bottom and the top in comparison to the clearly increased TP suggested by the diatom models. Therefore predicted TP for the bottom samples based on the two diatom-based WA models are ca. 20 μg l^{-1} for all the seven lakes, lower than those estimated for the core top of all seven lakes and also lower than observed TP values except for Lough Inchiquin (see Figure 7.6). The WA-PLS-2 model based on the full Cladocera gives similar TP values for the bottom samples of four lakes (Atedaun, Egish, Inchiquin and Sillan) but higher TP by 20-40 μg l^{-1} for the other three lakes. TP estimated by the PLS-3 model based on littoral Cladocera are higher than those by the other three models for the bottom samples of all seven lakes. This model also produces higher inferred TP values for the surface sediments of the seven lakes than all the other models, except for the TP predicted by the WA_{Cla} model for Loughs Crans and Inchiquin and the two WA models for Lough Sillan.

The two diatom-based WA models with inverse and classical deshrinking show good agreement in predicting TP for the core surface and bottom samples. TP values estimated by the WA_{Cla} model are slightly higher than those by the WA_{Inv} for both the

top and bottom samples of all seven lakes. Such deviation between $WA_{\_Cla}$- and $WA_{\_Inv}$-inferred TP is stronger in the core top than that for the core bottom, particulary at Lough Crans. Generally, both methods show comparable predictability and no major divergence in predicted TP indicating model consensus in the reconstruction of the nutrient levels. For the core bottom samples from all the seven lakes, the predicted TP by the two diatom WA models based on diatoms are ca. 20 μg l^{-1}, lower than the TP estimated by the two Cladocera-based models except for Loughs Egish and Sillan, where nearly identical TP are reconstructed by both diatom- and Cladocera-based TP models.

7.4 Discussion and Conclusions

7.4.1 Discussion

The weighted averaging method disregards species abundance and therefore the accuracy of its prediction is dependent on the distribution of the environmental variable in the training set (Braak & Looman, 1986). WA reliably estimates the species optima only in the special case of an even or uniform distribution of sites over the whole range of the environmental gradient. The TP distribution in the 70-lake diatom training set is strongly skewed towards the low end of the TP gradient as nearly half of the study sites have TP of less than 10 μg l^{-1} in the context of a TP range of 0-142 μg l^{-1} (see Table 4.1). While the TP distribution of the 31- or 32-lake Cladocera training set is less skewed and the median TP value for the data set is around 35 μg l^{-1} (see Table 5.1). The WA-inferred TP based on the diatom taxa is under-estimated because of this uneven TP distribution. This underestimation may account for the generally lower diatom-inferred TP than those predicted by the Cladocera-based models for the bottom samples and the lower inferred TP compared to the observed measurements for Loughs Ballybeg, Egish and Sillan. Therefore, the actual nutrient level for the reference state of the seven study lakes should be higher than the TP predicted by the diatom-based WA models.

TP values inferred by the chydorid-based model were higher than TP inferred by the other three models for Loughs Atedaun, Ballybeg and Mullagh. These three lakes are all shallow with mean depths of 1.4, 2.7 and 2.3 m respectively (see Table 7.1). A higher TP inferred by the chydorid-based model could be due to the synchrony of abundant

chydorids with a TP peak during summer in shallow lakes. In a 14-month sampling period of chydorids in 11 Irish lakes (de Eyto, 2000), high abundance of chydorids occurred mainly between June and September and very low chydorid abundance was observed mainly between October and April. Hydrochemical data from Loughs Mullagh, Ballybeg and Atedaun generally displayed a peak in TP values during the summer season (Free, 2002; Wemaëre, 2005). In shallow lakes the intensive sediment-water interaction and relatively high temperature of sediments in the summer lead to the release of phosphorus from the sediments to the water column and increase the total phosphorus concentration during the summer (Scheffer, 1998). In contrast, seasonal nutrient dynamics in deep lakes follows an opposite direction, where a continuous loss of nutrients from the epilimnion to the hypolimnion occurs in the summer before the autumn turnover when the well-mixed water enables the return of nutrients into the water column and therefore increases the nutrient level. This feature in deep lakes could account for a higher or at least similar TP inferred by the diatom- and full Cladocera-based inference models in comparison to that predicted by the chydorid-based models in lakes of Crans, Inchiquin and Sillan, particularly for the surface sediments (see Figure 7.6). Higher estimated TP by the chydorid-based model than the model based on full-species Cladocera may also be due to the influence of planktonic Cladocera. High abundance of planktonic Cladocera and factors other than nutrient levels (like predation) can reduce the accuracy of the whole Cladocera model.

The comparison of diatom-, Cladocera- and Chydorid-inferred TP for the core top with the measured TP is complicated due to the strong inter-annual and intra-annual variability of TP measurements. The mean and range of TP for the seven study lakes during 1996-2003 are shown in Table 7.4 based on published data. Low-frequency water sampling was conducted for Loughs Atedaun, Ballybeg, Crans and Egish in comparison to more frequent water sampling for 4-7 years for the other three lakes. These four lakes still show strong inter-annual and intra-annual fluctuation in TP. TP estimated by all the four inference models for the surface sediments of Atedaun, Ballybeg and Crans generally display good comparability with the annual mean of observed TP. This indicates that the diatom and Cladocera assemblages are reliable indicators for the average nutrient status of these lakes and the high TP variability exerts limited influence on the biological assemblages. TP predicted by the four inference models for Lough Egish show big divergence from the observed mean TP of 334 μg l^{-1} with an intra-annual range of 226-421 μg l^{-1} in 1997 (Free, 2002). Good correspondence

in inferred TP, however, exists between the models based on diatoms, Cladocera and littoral Cladocera. This poor comparability between observed and inferred TP may be due to the strong variability in the measured TP as the annual range can be greater in eutrophic lakes (e.g. Gibson *et al.*, 1996; Bennion & Smith, 2000). High-frequency water sampling in Lough Sillan between 1998 and 2003 shows a wide range of annual mean TP between 63.7 and 85.3 μg l^{-1} (see Table 7.4). Similarly, the annual mean TP fluctuated between 44.3 and 61.6 μg l^{-1} and the intra-annual fluctuation displays a wider range between 29 and 80 μg l^{-1} in 1996 in Lough Mullagh. This strong inter-annual variability in nutrient levels of Sillan and Mullagh can account for the dissimilarity between the measured and estimated TP by the inference models.

Table 7.4 Summary of recent TP measurements and TP ranges for the seven study lakes.

Lake	Year of Sampling	Sampling frequency	Mean TP (μg l^{-1})	TP range (μg l^{-1})	Reference
Atedaun	2000	4	36.7	17.3-83.7	(Wemaëre, 2005)
	1997	1	23.0	-	(Irvine *et al.*, 2001)
Ballybeg	2001	7	84.3	49.3-123.7	(Wemaëre, 2005)
	2000	8	74.5	28.7-166.1	(Wemaëre, 2005)
Crans	1987	1	241.0	-	(Gibson, 1991)
	1989/90	-	89.0	-	DARD (1991)
Egish	1997	6	336.2	226-421	(Free, 2002)
	1996	3	358.3	279-429	(Free, 2002)
Inchiquin	2001	10	16.8	-	(Toner *et al.*, 2005)
	2001	8	19.3	12.4-30.4	(Wemaëre, 2005)
	2000	8	20.2	10.4-35.1	(Wemaëre, 2005)
	1997	10	26.0	6.0-41.0	(Irvine *et al.*, 2001)
	1996	7	16.0	5.0-28.0	(Irvine *et al.*, 2001)
Mullagh	2003	11	56.7	-	(Toner *et al.*, 2005)
	2002	11	61.6	-	(Toner *et al.*, 2005)
	2001	10	44.3	-	(Toner *et al.*, 2005)
	2000	9	47.9	-	(McGarrigle *et al.*, 2002)
	1996	4	53.0	40.0-80.0	(Free, 2002)
	1997	6	56.3	29-69	(Free, 2002)
Sillan	2003	12	71.2	-	(Toner *et al.*, 2005)
	2002	12	63.7	-	(Toner *et al.*, 2005)
	2001	14	78.6	-	(Toner *et al.*, 2005)
	2000	10	72.0	-	(McGarrigle *et al.*, 2002)
	1999	12	66.4	-	(McGarrigle *et al.*, 2002)
	1998	12	85.3	-	(McGarrigle *et al.*, 2002)
	1996/7	1	141.0	-	(Irvine *et al.*, 2001)

Strong temporal (both intra-annual and inter-annual) and spatial variability of nutrient levels in lake waters imply that it is almost impossible to determine the nutrient level in fixed values as pre-impact status for lake restoration. Therefore, a range of nutrient levels would be more viable and practical as a reference for setting lake restoration targets. The minimum of this TP range could be the estimated TP by the two diatom-based models due to the skewed TP distribution of training set lakes towards the low end of the TP gradient. The upper limit of reference status could be the TP value predicted by the chydorid-based model as this model excludes the possible influence of selective predation on the plaktonic Cladocera and the TP distribution of the Cladocera training set lakes are more even along the TP gradient. In addition, this chydorid-based model may reflect the highest nutrient level due to the release of phosphorus from the sediments during the summer in shallow lakes. Therefore, 20-30 μg l^{-1} of TP is estimated as the reference range for Loughs Egish and Sillan, 20-40 μg l^{-1} for Loughs Atedaun, Inchiquin and Mullagh, and 20-70 μg l^{-1} for Loughs Ballybeg and Crans based on the diatom and Cladocera inference models (see Figure 7.6).

The response of chydorids to nutrient level may not be clear in all cases. Chydorids in the littoral zone may not respond to changes associated with increasing productivity until the littoral habitats (e.g. macrophytes) have been affected (Hofmann, 1987a). Submerged macrophytes are a significant habitat for chydorids and play an important role in the trophic structure and nutrient dynamics of shallow lakes (Scheffer *et al.*, 1993). The relative abundance of chydorid remains in surface sediments of 38 Australian billabongs showed a strong correlation with aquatic macrophyte cover (Thoms *et al.*, 1999). In contrast a large-scale investigation of 51 freshwater Scottish lochs found little correlation between the littoral microcrustacea (dominated by chydorids) and aquatic macrophyte communities (Duigan & Kovach, 1994). Therefore, the response of chydorids to nutrient levels of lake waters can be complicated by the macrophyte-nutrient relationship. A systematic investigation of aquatic macrophytes was not undertaken in this study. Such a study would improve our knowledge of the chydorid-macrophyte relationship for the training-set lakes and potentially reduce the errors in the application of a chydorid-based TP inference model. Also in shallow lakes, the dominance of non-planktonic diatoms, particularly small *Fragilaria*, have been observed in several studies as they can tolerate a wide range of water quality (Anderson *et al.*, 1993; Bennion *et al.*, 1995; Sayer, 2001). Though this phenomenon is not observed in the seven test lakes for top-bottom analysis, care should be taken when

additional application of the diatom training set is taken. The distinct features associated with seasonal nutrient dynamics in shallow and deep lakes can affect the accuracy in TP reconstruction using Cladocera-based models. A division of diatom- or chydorid-based training sets into two lake groups, deep and shallow lakes, can help distinguish the differential responses of biological indicators to the nutrient level in these two distinct habitat environment and improve the model performances.

7.4.2 Conclusions

The general trends in nutrient status for seven lakes were indicated by TP inferred by diatom and Cladocera inference models as well as the biological assemblages. The diatom-based models with inverse and classical deshrinking displayed good consensus in the reconstructed TP. They generally produced lower inferred TP than the chydorid- and Cladocera-based models due to the large number of oligotrophic lakes included in the diatom training set, which subsequently lowered the WA-inferred TP optima for diatom taxa. A TP range is recommended as the reference status for each of the seven lakes in consideration of the strong inter- and intra-annual variability of nutrient levels of lake waters. Diatom- and chydorid-inferred TP are used as the lower and upper limits respectively for the TP reference range. A TP range of 20-30 μg l^{-1} is estimated for Loughs Egish and Sillan, 20-40 μg l^{-1} for Loughs Atedaun, Inchiquin and Mullagh, and 20-70 μg l^{-1} for Loughs Ballybeg and Crans.

Chapter 8: Summary and Conclusions

The primary aim of this study was to investigate the relationships between the measured environment variables and biological community (diatoms and Cladocera) in Irish lakes. Total Phosphorus (TP) was highlighted in this study and its relationships with diatom and Cladocera communities were the focus in this study. A range of palaeolimnological techniques and numerical methods have been employed to achieve this aim. This chapter presents the summary of research results and findings and then concludes with suggestions for future research.

8.1 Summary of Results

The main research work and results were introduced and discussed in the previous chapters: Chapter 4 examined the surface sediment diatoms and environmental variables from 72 lakes, Chapter 5 investigated the surface sediment Cladocera and environmental data from 33 lakes, Chapter 6 developed TP transfer functions based on diatoms and Cladocera and finally the inference models were applied in TP reconstructions of seven lakes in Chapter 7. The main parts of this study are summarised in sequence below.

8.1.1 The Diatom Training Set

This diatom training set is composed of 72 lakes from the Irish Ecoregion along a TP gradient of 0-142.3 μg l^{-1}. The majority of lakes were located at low elevation and were relatively deep with variable lake areas. Peatland and pasture were the dominant land cover types. This training set includes a majority of the lake types recognised in the Irish EPA Lake Typology classification scheme. However, lakes with low TP were over-represented as half of the 72 lakes were oligotrophic and only two were hypertrophic. Land use and lake and catchment morphology displayed strong influences on the hydrochemical nature of the lake waters. Ordination of environmental variables highlighted that pH, alkalinity and conductivity were strongly correlated and influenced by catchment peat and pasture cover. The highly correlated nutrient variables, TP and chlorophyll-*a*, displayed an inverse relationship with lake depth.

High diatom species diversity was found in this 72-lake training set with 233 diatom taxa occurring in at least three sites and with a maximum abundance of greater than 1%. *Achnanthes minutissima*, a species commonly found in a wide range of lake types, was the most common species with presence in 70 lakes. A long ecological gradient indicated strong species turnover and a high degree of heterogeneity in the diatom assemblage composition. Eleven species occurred with a maximum abundance of over 30%, including *Aulacoseira subarctica*, *Asterionella formosa* and *Cyclotella comensis*. Diatom distribution along the TP gradient displayed a clear shift in assemblage structure, with species like *Cyclotella kuetzingiana*, *C. comensis* and *Frustulia saxonica* abundant at the low end of the TP gradient. A high abundance of taxa like *Tabellaria flocculosa* and *A. subarctica* occurred with an increase in TP while planktonic diatoms were dominant at the high end of the TP gradient, including *Asterionella formosa* and *Stephnanodiscus parvus*. The relatively even distribution of species along the first two axes of Correspondence Analysis (CA) indicated relatively continuous variation of species composition in the data set. This strong ecological response was also reflected in the good agreement between the diatom clusters and the physico-chemical lake typology classes. Alkalinity and to a lesser extent, lake area and size, were significant in controlling the diatom assemblages.

A large portion of the total variance in the diatom community was accounted for by the environment variables included in the Constrained Correspondence Analysis (CCA). Acidity and nutrient gradients were most significant in influencing the diatom distribution. This is also implied in the configuration of diatom species in CA plot where the triangle-like geometry of species along the first two axes indicated two dominant underlying environmental gradients. This triangle shape of species in CCA similar to that in CA confirmed that the main pattern in the species data were captured by the acidity and nutrient gradients. The forward selection exercise also identified pH and TP as the most significant variables controlling the diatom community. The examination of species response curves for some common diatoms showed a range of unimodal, sigmoidal increasing and decreasing curves to TP using Gaussian Logit Regression (GLR). This implied that a long TP gradient was covered in the training set and most of the common taxa displayed strong responses along this TP gradient. In addition the relative high λ_1/λ_2 ratio (0.673) in partial CCA indicated the importance of

TP in controlling the pattern of diatom assemblages. This implied that TP was appropriate for developing a robust transfer function for the diatom training set.

8.1.2 The Cladocera Training Set

Thirty-three lakes included in this Cladocera training set were mainly small and alkaline with various basin depths along a relatively long TP gradient (4-142.3 μg l^{-1}) similar to the diatom training set. Most of the lakes were strongly influenced by extensive agriculture activities, e.g. pastureland. Eight lake types were included in this training set according to the Irish Lake Typology classes and Lake Types 2, 4, 6 and 8 (all large lakes) were not represented. PCA of the environmental data revealed similar data structure as in the diatom training set: nutrient variables displayed an inverse relationship with lake depth while the highly correlated acidity-conductivity gradient was significantly influenced by catchment peat and pasture cover.

The comparison of Cladocera assemblages from surface sediments and contemporary water samples from six lakes was an attempt to check how well the sediment samples represented the live community. For littoral Cladocera (chydorids) higher species richness was often observed in sediments than in water samples with the main pattern of species composition captured in the sediments. The dominant planktonic Cladocera were also reflected in the sediments, however, clear dissimilarity occurred for the less common taxa. In addition low taxonomic resolution was achieved for the remains of some planktonic taxa. Cladocera assemblages in lake sediments are generally representative of original live community despite their differential disarticulation and preservation, but care needs be taken in interpreting the planktonic Cladocera remains. Thirty-nine Cladocera taxa and species groups were collected from surface sediments of 33 lakes with relatively high species diversity. However, a relatively short gradient length within the Cladocera data indicated an insignificant species shift along underlying environment gradients, as evidenced by the occurrence of five species or species group in all of the 33 lakes. Planktonic *Daphnina longispina* group and *Bosmina longirostris* were the most dominant taxa in this training set and littoral *Chydorus sphaericus* and *Alona guttata*/*rectangula* group were also common. The classification of chydorid data revealed four chydorid clusters with little overlap. They corresponded

well with low, medium and high alkalinity lake types indicating a strong ecological response of littoral Cladocera to alkalinity.

The first axis in a PCA of environment data was mainly correlated with catchment and physical variables but in the Redundancy Analysis (RDA) of Cladocera data the first axis was most strongly related to nutrient variables. This indicates that nutrient gradient is the most significant gradient in determining the Cladocera community while catchment and physical variables were less important. Alkalinity was strongly correlated with the second axes of both PCA and RDA. The species response curves also suggested strong ecological responses to the nutrients and alkalinity. In addition lake depth, altitude and catchment area were also identified as important variables influencing Cladocera distribution. Large portions of the total variance in the Cladocera data were explained by the strongly correlated nutrient variables, Chlorophyll-*a* (8.9%) and TP (8.1%), respectively in partial RDAs. A moderate $\lambda_1/\ \lambda_2$ ratio of 0.519 implied that TP can be potentially used for developing Cladocera-based transfer functions.

8.1.3 TP Inference Models and Their Application

Three transfer function methods, Weighted Averaging (WA), Partial Least Square (PLS) and Weighted Averaging Partial Least Square (WA-PLS) were used in developing TP transfer functions for the diatom and Cladocera training sets. Data manipulation was explored and multiple models were developed in this study. Model performances and prediction errors were presented to highlight the selection process to derive the optimal models. Log-transformed TP was used in all models as they provided models outperforming those based on untransformed TP. Jack-knifing was used for cross-validating each model as it outperformed bootstrapping. Both raw and square root transformed ecological data were included in the model development. The optimal diatom-TP transfer function was a WA model with tolerance downweighted and inverse deshrinking and it produced a jack-knifed r^2 of 0.743 and RMSEP of 0.213 based on untransformed diatom data from 70 lakes. The optimal Cladocera-TP transfer function utilised the second component of WA-PLS and had a jack-knifed r^2 of 0.729 and RMSEP of 0.206 based on square root transformed Cladocera data from 31 lakes.

The diatom- and Cladocera TP transfer functions produced were also evaluated and compared with published diatom and Cladocera TP models. Both models developed in this study covered relative long TP gradients and displayed strong performances in terms of cross-validated r^2 and RMSEP. TP optima of common diatom and Cladocera taxa were also produced using the Weighted Averaging method. Common taxa of diatoms and Cladocera in the Irish Ecoregion displayed good comparability in TP optima with those from other European Ecoregions. However, the under-representation of lakes with high TP in this diatom training set generally provided lower TP optima than those training sets composed of more eutrophic lakes. As 29 lakes were common in the diatom and Cladocera training sets, a comparison between the diatom- and Cladocera-inferred TP provided information on the performances and agreement of both transfer functions. A good correlation ($r = 0.685$) was found between the log-transformed TP inferred by both models for the 29 sites. Back-transformed TP ($\mu g\ l^{-1}$) in contrast had a lower correlation ($r = 0.476$).

Finally the diatom and Cladocera TP transfer functions including a littoral Cladocera TP model were applied to the top and bottom samples from seven lakes to evaluate the developed TP models and also help identify the pre-impact reference conditions. TP models cross-validated with bootstrapping were used as they provide sample-specific error information and the diatom WA models with both classical and inverse deshrinking were applied. An increase in the reconstructed TP was observed between the bottom and top samples for most of the seven lakes based on all the TP models and a similar trend was also revealed by the diatom and Cladocera assemblage changes. However, certain dissimilarities were also found between Cladocera- and diatom-inferred TP and also between Cladocera- and littoral Cladocera-inferred TP. Strong divergence between the observed and predicted TP by the diatom and Cladocera models occurred in the two currently hypertrophic lakes (Loughs Egish and Sillan). Based on the TP predicted for the bottom samples, a range of reference TP conditions were identified as 20-30 $\mu g\ l^{-1}$ for Loughs Egish and Sillan, 20-40 $\mu g\ l^{-1}$ for Atedaun, Inchiquin and Mullagh, and 20-70 $\mu g\ l^{-1}$ for Ballybeg and Crans. The adoption of a TP range for the reference conditions also reflected the nature of inter- and intra-annual variability of nutrient levels in lake waters.

8.2 Conclusions

Surface sediment diatoms from 72 lakes displayed strong response to environment gradients. Acidity and nutrient gradients were the most important in controlling the diatom distribution. The reliability of the surface sediment Cladocera in representing the original live community was confirmed in this study but the planktonic species were less well represented and lower taxonomic resolution was a contributory factor. A suite of environmental variables demonstrated strong influence on the Cladocera assemblages from 33 lakes and the nutrient gradient was shown to be the most significant. Alkalinity, altitude, lake depth and catchment area were also important in explaining the variation in the Cladocera data. TP was identified as one of the most significant variables in influencing both the diatom and the Cladocera communities using a range of multivariate and ordination methods. The classification of both diatom and Cladocera assemblages generally showed good agreement with the physico-chemical lake typology classes. Alkalinity displayed the strongest ecological response and lake depth and lake area to a lesser extent.

Both unimodal-based and linear methods were employed for modelling the quantitative relationship between TP and the biological assemblages. Data manipulation was found to be important in affecting the model performance. The good correlation between the diatom- and Cladocera-inferred TP for the same lakes also confirmed the performances of the constructed TP models. Similar TP trends were observed between the top and bottom samples of seven lakes using the diatom and Cladocera models including a littoral Cladocera-TP model. The over-representation of oligotrophic lakes in the diatom training set often produced lower TP in reconstruction than the Cladocera models. A TP range was proposed as the pre-impact reference condition for the seven degraded lakes taking account of the seasonal and inter-annual variation of TP in lake waters. Refinement of the TP transfer functions in the future can help improve the accuracy in TP reconstruction and defining lake restoration targets.

8.3 Future Directions

The performances of both inference models can potentially be improved by incorporating more lakes in the diatom and Cladocera training sets. Collecting more

samples and related environmental data can help to expand the training set and improve the accuracy of inference models. A higher frequency of water sampling and a more systematic method of measurement are required. In addition the collection of macrophyte and fish data would improve our knowledge on the Cladocera-environment relationships as both macrophyte cover and fish density were found to have strong influences on Cladocera assemblges (e.g. Jeppesen *et al.*, 1996; Davidson, 2005). Alternatively the merging of regional datasets can be used to generate new transfer functions with wider applicability (Gasse *et al.*, 1995). A combined diatom-inferred TP transfer function was constructed from 152 lakes after incorporating six regional training sets across Northwest Europe (Bennion *et al.*, 1996). This combined training set improved the model performance with better accuracy and strength. The prediction errors did not increase with the expanded heterogeneity in the training set. Therefore an expanded TP gradient and a larger training set can improve the performance of inference models. This method could be employed to combine the two existing diatom-based TP transfer functions created for the Irish Ecoregion. The current diatom training set is mainly composed of oligotrophic and mesotrophic lakes covering the majority of Irish lake types while a data set from Northern Ireland includes fewer lake types and is dominated by more eutrophic lakes (Anderson & Rippey, 1994). A combined training set with a more even distribution of TP, an expanded TP gradient and increased species diversity would improve the performance of the inference model for the Irish Ecoregion and this remains to be explored in the future.

The influence of uneven distribution of sites along the TP gradient on model reconstruction can be reduced through several methods to refine the inference models. Removal of some oligotrophic lakes in the diatom training set could reduce the skewness of site distribution along the TP gradient and improve the accuracy of TP optima of diatom taxa. Also the trimming of some diatom taxa irrelevant to the environmental variables of interest has improved model performance significantly (Racca *et al.*, 2004). The application of Modern Analogue Technique (MAT) can be performed for TP reconstruction to verify the performances of inference models and also to achieve consensus results in TP reconstructions. MAT assumes that current aquatic communities in reference lakes closely resemble those that formerly existed in the currently polluted lakes (Flower *et al.*, 1997). Dissimilarity coefficients measure the difference between fossil and modern assemblages and this method has been applied successfully in several studies (e.g. Flower *et al.*, 1997; Bennion *et al.*, 2004a).

Another opportunity for the current study, as suggested by Jeppensen *et al.* (2001), will be the development of inference models that use Cladocera taxa (e.g. *Daphnia*, *Bosmina*) to explain the variation in diatom assemblages as both organisms are important components in the food webs of lakes with diatoms as the primary producers and planktonic Cladocera the predatory zooplankton (e.g. Reynolds, 1984; Lampert & Sommer, 1997). The construction of such models would enable exploration of the Cladocera-diatom relationship and help understand their biological interaction and their influence on community structure.

References

Allott, N., Free, G., Irvine, K., Mills, P., Mullins, T. E., Bowman, J. J., Champ, W. S. T., Clabby, K. J., & McGarrigle, M. L. (1998). Land use and aquatic systems in the Republic of Ireland. In P. Giller (Ed.), *Studies in Irish Limnology*. Dublin: Marine Institute.

Allott, N. A. (1986). Temperature, oxygen and heat-budgets of six small western Irish lakes. *Freshwater Biology, 16*(2), 145-154.

Alonso, M. (1996). *Crustacea, Branchiopoda* (Vol. 7). Madrid: Museo Nacional de Ciencias Naturales, CSIC.

Amoros, C. (1984). Introduction pratique a la systematique des organismes des eaux continentales francaises-Crustacés Cladocères. *Bulletin de la Societe Linnéenne de Lyon, 53*, 72-145.

Amsinck, S. L., Jeppesen, E., & Landkildehus, F. (2005). Relationships between environmental variables and zooplankton subfossils in the surface sediments of 36 shallow coastal brackish lakes with special emphasis on the role of fish. *Journal of Paleolimnology, 33*(1), 39 - 51.

Anderson, N. J. (1990). Variability of sediment diatom assemblages in an upland, wind-stressed lake (Loch Fleet, Galloway, S.W. Scotland). *Journal of Paleolimnology, 4*(1), 43 - 59.

Anderson, N. J. (1995a). Diatom-based phosphorus transfer functions - errors and validation. In S. T. Patrick & N. J. Anderson (Eds.), *Ecology and Paleoecology of lake eutrophication* (Vol. Service report, pp. 39-40). Copenhagen: Geological Survey of Denmark.

Anderson, N. J. (1995b). Using the past to predict the future: lake sediments and the modelling of limnological disturbance. *Ecological Modelling, 78*, 149-172.

Anderson, N. J. (1997a). Historical changes in epilimnetic phosphorus concentrations in six rural lakes in Northern Ireland. *Freshwater Biology, 38*, 427-440.

Anderson, N. J. (1997b). Reconstructing historical phosphorus concentrations in rural lakes using diatom models. In H. Tunney, O. T. Carton, P. C. Brookes & A. E. Johnston (Eds.), *Phosphorus Loss from Soil to Water* (pp. 95-118). Wallingford, U.K.: Cab International.

Anderson, N. J., & Rippey, B. (1994). Monitoring lake recovery from point-source eutrophication: The use of diatom-inferred epilimnetic total phosphorus and sediment chemistry. *Freshwater Biology, 32*(3), 625-639.

Anderson, N. J., Rippey, B., & Gibson, C. E. (1993). A comparison of sedimentary and diatom-inferred phosphorus profiles: implications for defining pre-disturbance nutrient conditions. *Hydrobiologia, 253*(1 - 3), 357 - 366.

Anderson, N. J., Rippey, B., & Stevenson, A. C. (1990). Change to a Diatom Assemblage in a Eutrophic Lake Following Point Source Nutrient Re-Direction: A Palaeolimnological Approach. *Freshwater Biology, 23*, 205-217.

Barber, H. G., & Haworth, E. Y. (1981). *A Guide to the Morphology of the Diatom Frustule, with a Key to the British Freshwater Genera*: Freshwater Biological Association.

Barker, P. A., Pates, J. M., Payne, R. J., & Healey, R. M. (2005). Changing nutrient levels in Grasmere, English Lake District, during recent centuries. *Freshwater Biology, 50*(12), 1971-1981.

Battarbee, R. W. (1986). Diatom Analysis. In B. E. Berglund (Ed.), *Handbook of Holocene Palaeoecology and Palaeohydrology* (pp. 527-570). UK: John Wiley & Sons.

Battarbee, R. W., Anderson, N. J., Jeppesen, E., & Leavitt, P. R. (2005). Combining palaeolimnological and limnological approaches in assessing lake ecosystem response to nutrient reduction. *Freshwater Biology, 50*, 1365-2427.

Battarbee, R. W., Jones, V. J., Flower, R. J., Cameron, N. G., Bennion, H., Carvalho, L., & Juggins, S. (2001). Diatoms. In *Terrestrial, Algal, and Siliceous Indicators* (Vol. 3, pp. 155-202). Dordrecht: Kluwer Academic Publishers.

Becker, R. A., Chambers, J., M., & Wilks, A. R. (1988). *The new S language*. Pacific Grove, Ca: Wadsworth & Brooks.

Beeton, A. M. (1965). Eutrophication of the St. Lawrence Great Lakes. *Limnology and Oceanography, 10*(2), 240-254.

Bennion, H. (1994). A diatom-phosphorus transfer function for shallow, eutrophic ponds in southeast England. *Hydrobiologia, 275/276*, 391-410.

Bennion, H., Appleby, P. G., & Phillips, G. L. (2001). Reconstructing nutrient histories in the Norfolk Broads, UK: implications for the role of diatom-total phosphorus transfer functions in shallow lake management. *Journal of Paleolimnology, 26*(2), 181 -204.

Bennion, H., Davidson, T. A., Simpson, G. L., Solovieva, N., Rose, N., Theophile, S., Yang, H. D., Anderson, N. J., Brooks, S. J., & Peglar, S. (2004a). *Identification of reference lakes and evaluation of palaeoecological approaches to define reference conditions for UK (England, Wales, Scotland & North Ireland) ecotypes*. Edinburgh: SNIFFER.

Bennion, H., Fluin, J., & Simpson, G. L. (2004b). Assessing eutrophication and reference conditions for Scottish freshwater lochs using subfossil diatoms. *Journal of Applied Ecology, 41*, 124-138.

Bennion, H., Johnes, P., Ferrier, R., Phillips, G., & Haworth, E. (2005). A comparison of diatom phosphorus transfer functions and export coefficient models as tools for reconstructing lake nutrient histories. *Freshwater Biology, 50*, 1651-1670.

Bennion, H., Juggins, S., & Anderson, N. J. (1996). Predicting Epilimnetic Phosphorus Concentrations Using an Improved Diatom-Based Transfer Function and Its Application to Lake Eutrophication Management. *Environment Science and Technology, 30*(6), 2004 - 2007.

Bennion, H., & Smith, M. A. (2000). Variability in the water chemistry of shallow ponds in southeast England, with special reference to the seasonality of nutrients and implications for modelling trophic status. *Hydrobiologia, 436*(1-3), 145 - 158.

Bennion, H., Wunsam, S., & Schmidt, R. (1995). The validation of diatom-phosphorus transfer functions: An example from Mondsee, Austria. *Freshwater biology, 34*(22), 271-283.

Birks, H. J. B. (1995). Quantitative palaeoenvironmental reconstructions. In D. Maddy & J. S. Brew (Eds.), *Statistical modelling of Quaternary science data* (pp. 161-254). Cambridge: Quaternary Research Association.

Birks, H. J. B. (1998). Numerical tools in quantitative paleolimnology - progress, potentialities and problems. *Journal of paleolimnology, 20*, 301-332.

Birks, H. J. B. (2005a). *Numerical Analysis of Biological and Environmental Data*.Unpublished manuscript, University College London.

Birks, H. J. B. (2005b). Quantitative Palaeoenvironmental Reconstructions from Holocene Biological Data. In A. W. Mackay, R. W. Battarbee, H. J. B. Birks & F. Oldfield (Eds.), *Global Change in the Holocene* (pp. 107-123): Hodder Arnold.

Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., & ter Braak, C. J. F. (1990). Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society, London B, 327*, 263-278.

Bloom, A. M., Moser, K. A., Porinchu, D. F., & MacDonald, G. M. (2003). Diatom-inference models for surface-water temperature and salinity developed from a 57-lake calibration set from the Sierra Nevada, California, USA. *Journal of Paleolimnology, 29*(2), 235 - 255.

Bos, D. G. (2001). *Sedimentary Cladoceran remains, a key to interpreting past changes in nutrients and trophic interactions.* Unpublished Ph.D. thesis, Queen's University, Kingston, Ontario.

Bos, D. G., & Cumming, B. F. (2003). Sedimentary Cladoceran remains and their relationship to nutrients and other limnological variables in 53 lakes from British Columbia, Canada. *Canadian Journal of Fisheries and Aquatic Sciences, 60*(10), 1177-1189.

Boucherle, M. M., & Züllig, H. (1983). Cladoceran remains as evidence of change in trophic state in three Swiss lakes. *Hydrobiologia, 103*(1), 141 -146.

Bowman, J. J., & Clabby, K. J. (1998). Water quality of rivers and lakes in the Republic of Ireland. In J. G. Wilson (Ed.), *Eutrophication in Irish Waters* (pp. 55-63). Dublin: Royal Irish Academy.

Bowman, J. J., Clabby, K. J., Lucey, J., McGarrigle, M. L., & Toner, P. F. (1996). *Water Quality in Ireland: 1991-1994*. Wexford: Environmental Protection Agency.

Braak, C. J. F., & Looman, C. W. N. (1986). Weighted averaging, logistic regression and the Gaussian response model. *Plant Ecology, 65*(1), 3-11.

Bradshaw, E. G., & Anderson, N. J. (2001). Validation of a diatom-phosphorus calibration set for Sweden. *Freshwater Biology, 46*, 1035-1048.

Bradshaw, E. G., Anderson, N. J., Jensen, J. P., & Jeppesen, E. (2002). Phosphorus dynamics in Danish lakes and the implications for diatom ecology and palaeoecology. *Freshwater Biology, 47*, 1963-1975.

Bredesen, E. L., Bos, D. G., Laird, K. R., & Cumming, B. F. (2002). A cladoceran-based paleolimnological assessment of the impact of forest harvesting on four lakes from the central interior of British Columbia, Canada. *Journal of Paleolimnology, 28*(4), 389 - 402.

Brodersen, K. P., Whiteside, M. C., & Lindegaard, C. (1998). Reconstruction of trophic state in Danish lakes using subfossil chydorid(Cladocera) assemblages. *Canadian Journal of Fisheries and Aquatic Sciences, 55*(5), 1093-1103.

Brönmark, C., & Hansson, L.-A. (2005). *The Biology of Lakes and Ponds*. Oxford: Oxford University Press.

Brooks, J. L., & Dodson, S. I. (1965). Predation, body size, and composition of plankton. *Science, 150*, 28-35.

Brooks, S. J., Bennion, H., & Birks, H. J. B. (2001). Tracing lake trophic history with a chironomid-total phosphorus inference model. *Freshwater Biology, 46*(4), 513-533.

Brugam, R. B. (1983). The relationship between fossil diatom assemblages and limnological conditions. *Hydrobiologia, 98*(3), 223 - 235.

Brugam, R. B., & Speziale, B. J. (1983). Human Disturbance and the Paleolimnological Record of Change in the Zooplankton Community of Lake Harriet, Minnesota. *Ecology, 64*(3), 578-591.

Cameron, N. G. (1995). The representation of diatom communities by fossil assemblages in a small acid lake. *Journal of Paleolimnology, 14*(2), 185 - 223.

Caroni, R. (2001). *Freshwater zooplankton and pH: a study of natural patterns and experimental responses.* Unpublished Ph.D. thesis, Trinity College Dublin (University of Dublin).

Chengalath, R. (1987). The Distribution Of Chydorid Cladocera In Canada. *Hydrobiologia, 145*, 151-157.

Clabby, K. J., Lucey, J., McGarrigle, M. L., Bowman, J. J., Flanagan, P. J., & Toner, P. F. (1992). *Water Quality in Ireland, 1987-1990. Part One: General Assessment.* Dublin: Environmental Research Unit.

Cohen, A. S. (2003). *Paleolimnology-the History and Evolution of Lake Systems*. New York: Oxford University Press.

Cox, E. J. (1996). *Identification of Freshwater Diatoms from Live Material*. London: Chapman & Hall.

Cumming, B. F., & Smol, J. P. (1993). Scaled chrysophytes and pH inference models: the effects of converting scale counts to cell counts and other species data transformations. *Journal of Paleolimnology, 9*(2), 147 - 153.

Cumming, B. F., Smol, J. P., Kingston, J. C., Charles, D. F., Birks, H. J. B., Camburn, K. E., Dixit, S. S., Uutala, A. J., & Selle, A. R. (1992). How much acidification has occurred in Adirondack region lakes (New York, USA) since preindustrial times? *Canadian Journal of Fisheries and Aquatic Sciences, 49*(1), 128-141.

Dalton, C. (1999). *A palaeolimnological investigation of acidity in humic lakewaters in Connemara, W. Ireland.* Unpublished Ph.D. thesis, University College London.

Dalton, C., Leira, M., Taylor, D., Jordan, P., Chen, G., Bennion, H., León-Vintró, L., Irvine, K., & Nolan, T. (2006). *IN-SIGHT EPA/ERTDI Project # 2002-W-LS/7 Work Package (WP) 3 (months 21-36)*: Department of Geography, University of Limerick (Mary Immaculate College).

Davidson, T. A. (2005). *Zooplankton ecology and palaeoecology in nutrient enriched shallow lakes.* Unpublished Ph.D., University College London, London.

de Bernardi, R., Giussani, G., & Manca, M. (1987). Cladocera - Predators And Prey. *Hydrobiologia, 145*, 225-243.

de Eyto, E. (2000). *The ecology of the family Chydoridae (Branchiopoda, Anomopoda) and its application to lake monitoring.* Unpublished Ph. D. thesis, Trinity College Dublin (University of Dublin), Dublin.

de Eyto, E., & Irvine, K. (2002). The use of the family Chydoridae (Anomopoda, Branchiopoda) as indicators of lake ecological quality in Ireland. *Biology and Environment, 102B*, 81-91.

de Eyto, E., Irvine, K., Garcia-Criado, F., Gyllstrom, M., Jeppensen, E., Kornijow, R., Miracle, M. R., Nykanen, M., Bareiss, C., Cerbin, S., Salujoe, J., Franken, R., Stephens, D., & Moss, B. (2003). The distribution of chydorids (Branchiopoda, Anomopoda) in European shallow lakes and its application to ecological quality monitoring. *Archiv Fur Hydrobiologie, 156*, 181-202.

Deevey, E. S., Jr., & Deevey, G. B. (1971). The American species of Eubosmina Seligo (Crustacea, Cladocera). *Limnology and Oceanography, 16*, 201-218.

DeMott, W. R., Ramesh, D. G., & Donk, E. V. (2001). Effects of dietary phosphorus deficiency on the abundance, phosphorus balance, and growth of *Daphnia cucullata* in three hypereutrophic Dutch lakes. *Limnology and Oceanography, 46*(8), 1871-1880.

DeNicola, D. (1986). The representation of living diatom communities in deep-water sedimentary diatom assemblages in two Maine (U.S.A.) lakes. In J. P. Smol, R. W. Batterbee, R. B. Davis & J. Meriläinen (Eds.), *Diatoms and Lake Acidity* (pp. 73-85). Dordrecht: Dr W. Junk Publisher.

DeNicola, D. M., de Eyto, E., Wemaere, A., & Irvine, K. (2004). Using epilithic algae communities to assess trophic status in Irish lakes. *Journal of Phycology, 40* (3), 481-495.

Dillon, P. J., & Rigler, F. H. (1974). The phosphorus-Chlorophyll Relationship in Lakes. *Limnology and Oceanography, 19*, 767-773.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis: methods and application*. New York: John Wiley & Sons.

Dixit, S. S., Cumming, B. F. C., Birks, H. J. B., Smol, J. P., Kingston, J. C., Uutala, A., Charles, D. F., & Camburn, K. E. (1993). Diatom assemblages from Adirondack Lakes (New York, U.S.A.) and the development of inference models for retrospective environmental assessment. *Journal of Paleolimnology, 8*, 27-47.

Dixit, S. S., Smol, J., Kingston, J. C., & Charles, D. F. (1992). Diatoms: powerful indicators of environmental change. *Environmental Science and Technology, 26*(1), 23-33.

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*(3), 275 - 307.

Dixit, S. S., Smol, J. P., Charles, D. F., Hughes, R. M., Paulsen, S. G., & Collins, G. B. (1999). Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Canadian Journal of Fisheries and Aquatic Sciences, 56*(1), 131-152.

Dodson, S. I., & Frey, D. G. (2001). Cladocera and Other Branchiopoda. In H. J. Thorp & A. P. Covich (Eds.), *Ecology and Classification of North American Freshwater Invetrebrates* (pp. 723-786): Academic Press.

Dokulil, M. T., & Teubner, K. (2005). Do phytoplankton communities correctly track trophic changes? An assessment using directly measured and palaeolimnological data. *Freshwater Biology, 50*, 1594-1604.

Dong, X. H., Yang, X. D., Wang, R., & Pan, H. X. (2006). A diatom-total phosphorus transfer function for lakes in the middle and lower reaches of Yangtze River. *Journal of Lake Sciences, 18*(1), 1-12.

Duigan, C. (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*(3), 267-280.

Duigan, C. A., & Kovach, W. L. (1994). Relationships between littoral microcrustacea and aquatic macrophyte communities on the Isle of Skye (Scotland), with implications for the conservation of standing waters. *Aquatic Conservation: Marine and Freshwater Ecosystems, 4*(4), 307-331.

Edmondson, W. T., Anderson, G. C., & Peterson, D. R. (1956). Artificial Eutrophication of Lake Washington. *Limnology and Oceanography, 1*(1), 47-53.

Elser, J. J., Fagan, W. F., Denno, R. F., Dobberfuhl, D. R., Folarin, A., Huberty, A., Interlandi, S., Kilham, S. S., McCauley, E., Schulz, K. L., Siemann, E. H., & Sterner, R. W. (2000). Nutritional constraints in terrestrial and freshwater food webs. *Nature, 408*(6812), 578-580.

Elser, J. J., & Urabe, J. (1999). The stoichiometry of consumer-driven nutrient cycling: theory, observations, and consequences. *Ecology, 80*, 735-751.

Enache, M., & Prairie, Y. T. (2002). WA-PLS diatom-based pH, TP and DOC inference models from 42 lakes in the Abitibi clay belt area (Québec, Canada). *Journal of Paleolimnology, 27*(2), 151-171.

European Council. (2000). Directive 2000/60/EC of the European Parliment and of the Council of 23 October 2000 -Establishing a framework for community action in the field of water policy. *Official Journal of the European Communities*, 1-72.

European Union. (2000). Directive 2000/60/EC of the European Parliment and of the Council of 23 October 2000 -Establishing a framework for community action in the field of water policy. *Official Journal of the European Communities*, 1-72.

Facher, E., & Schmidt, R. (1996). A siliceous chrysophycean cyst-based pH transfer function for Central European lakes. *Journal of Paleolimnology, 16*(3), 275 - 321.

Finney, B. P., Gregory-Eaves, I., Douglas, M. S. V., & Smol, J. P. (2002). Fisheries productivity in the northeast Pacific over the past 2,200 years. *Nature, 416*, 729-733.

Flanagan, P. J., & Toner, P. F. (1975). *A Preliminary Survey of Irish Lakes*. Dublin: An Foras Forbartha (Water Resource Division).

Flower, R. J., Juggins, S., & Battarbee, R. W. (1997). Matching diatom assemblages in lake sediment cores and modern surface sediment samples: the implications for lake conservation and restoration with special reference to acidified systems. *Hydrobiologia, 344*(1-3), 27-40.

Foged, N. (1977). Freshwater diatoms in Ireland. *Bibliotheca Phycologica, 34*, 221.

Fox, J. (2002). *An R and S-Plus Companion to Applied Regression*. Thousand Oaks, California: Sage Publications.

Free, G., Little, R., Donnelly, K., Tierney, D., Caroni, R., Bowman, J., & McGarrigle, M. (2005). *Summary notes of Irish lake typology to be applied in Ireland's River Basin Districts*. Wexford: Environmental Protection Agency.

Free, G. N. (2002). *The relationship between catchment characteristics and lake chemistry in the Republic of Ireland.* Unpublished Ph.D. thesis, Trinity College, University of Dublin, Dublin.

Frey, D. G. (1958). The later-glacial cladoceran fauna of a small lake. *Archiv Fur Hydrobiologie, 54*, 209-275.

Frey, D. G. (1959). The taxonomic and phylogenetic significance of the head pores of the Chydoridae (Cladocera). *Internationale Revue der Gesamten Hydrobiologie, 44*, 27-50.

Frey, D. G. (1960). 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*(6), 1133-1156.

Frey, D. G. (1962b). Supplement to the taxonomic and 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*(2), 159-173.

Frey, D. G. (1965). A new genus of Chydoridae (Cladocera). *Internationale Revue der Gesamten Hydrobiologie, 50*, 153-168.

Frey, D. G. (1976). Interpretation of Quaternary paleoecology from Cladocera and midges, and prognosis regarding usability of other organisms. *Canadian Journal of Zoology, 54*, 2208-2226.

Frey, D. G. (1986). Cladocera analysis. In B. E. Berglund (Ed.), *Handbook of Holocene Palaeoecology and Palaeohydrology* (pp. 667-692). New York: John Wiley & Sons.

Frey, D. G. (1988a). Littoral and offshore communities of diatoms, cladocerans and dipterous larvae, and their interpretation in paleolimnology. *Journal of Paleolimnology, 1*, 179-191.

Frey, D. G. (1988b). What is paleolimnology? *Journal of Paleolimnology, 1*, 5-8.

Freyer, 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 B., 254*, 221-385.

Fritz, S. C. (1990). Twentieth-Century Salinity and Water-Level Fluctuations in Devils Lake, North Dakota: Test of a Diatom-Based Transfer Function. *Limnology and Oceanography, 35*(8), 1771-1781.

Fritz, S. C., Kingston, J. C., & Engstrom, D. R. (1993). Quantitative trophic reconstruction from sedimentary diatom assemblages: A cautionary tale. *Freshwater biology, 30*(11), 1-23.

Gabriel, K. R. (1971). The biplot graphic-display of matrices with application to principal component analysis. *Biometrika, 58*, 453-467.

Gasse, F., Juggins, S., & Khelifa, L. B. (1995). Diatom-based transfer functions for inferring past hydrochemical characteristics of African lakes. *Palaeogeography, Palaeoclimatology, Palaeoecology, 117*(1-2), 31-54.

Gibson, C. E. (1991). Contribufions to the regional limnology of Northem Ireland:(4) the lakes of Co. Tyrone. *Irish naturalists' journal, 23*, 429.

Gibson, C. E., Anderson, N. J., & Haworth, E. Y. (2003). Aulacoseira subarctica: taxonomy, physiology, ecology and palaeoecology. *European Journal of Phycology, 38*(2), 83 -101.

Gibson, C. E., Foy, R. H., & Bailey-Watts, A. E. (1996). An analysis of the total phosphorus cycle in some temperate lakes: the response to enrichment. *Freshwater Biology, 35*, 525-532.

Gibson, C. E., Wu, Y., & Pinkerton, D. (1995). Substance budgets of an upland catchment: the significance of atmospheric phosphorus inputs. *Freshwater Biology, 33*, 385-392.

Godínez-Domínguez, E., & Freire, J. (2003). Information-theoretic approach for selection of spatial and temporal models of community organization. *Marine Ecology Progress series, 253*, 17-24.

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.

Green, J. (1995). Altitudinal distribution of tropical planktonic Cladocera. *Hydrobiologia, 307*, 75-84.

Gregory-Eaves, I., Smol, J. P., Finney, B. F., & Edwards, M. E. (1999). Diatom-based Transfer Functions for Inferring Past Climatic and Environmental Changes in Alaska, U.S.A. *Arctic, Antarctarctic and Alpine Research, 31*(4), 353-365.

Guy, M., Taylor, W. D., & Carter, J. C. H. (1994a). Decline in total phosphorus in the surface waters of lakes during summer stratification, and its relationship to size distribution of particles and sedimentation. *Canadian Journal of Fisheries and Aquatic Sciences, 51*(6), 1330-1337.

Guy, M., Taylor, W. D., & Carter, J. C. H. (1994b). Decline in total phosphorus in the surface waters of lakes during summer stratification, and its relationship to size distribution of particles and sedimentation. *Canadian Journal of Fisheries and Aquatic Sciences, 51*(6), 1330-1337.

Haberyan, K. A. (1990). The misrepresentation of the planktonic diatom assemblage in traps and sediments: Southern Lake Malawi, Africa. *Journal of Paleolimnology, 3*(1), 35-44.

Håkansson, H. (2002). A compilation and evaluation of species in the general Stephanodiscus, Cyclostephanos and Cyclotella with a new genus in the family Stephanodiscaceae. *Diatom Research, 17*, 1-139.

Hall, R. I., & Smol, J. (1999). Diatoms as indicators of lake eutrophication. In E. F. Stoermer & J. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences* (pp. 128-168). Cambridge: Cambridge University Press.

Hall, R. I., & Smol, J. P. (1992). A Weighted-Averaging Regression and Calibration Model for Inferring Total Phosphorus Concentration from Diatoms in British Columbia (Canada) Lakes. *Freshwater Biology, 27*, 417-434.

Hall, R. I., & Smol, J. P. (1996). Paleolimnological assessment of long-term water-quality changes in south-central Ontario lakes affected by cottage development and acidification. *Canadian Journal of Fisheries and Aquatic Sciences, 53*(1), 1-17.

Hann, B. J. (1989). Cladocera. Methods in Quaternary Ecology. *Geoscience Canada, 16*, 17-26.

Hann, B. J., Leavitt, P. R., & Chang, P. S. S. (1994). Cladocera community response to experimental eutrophication in Lake 227 as recorded in laminated sediments. *Canadian Journal of Fisheries and Aquatic Sciences, 51*(10), 2312-2321.

Harmsworth, R. V. (1968). The developmental history of Blelham Tarn(England) as shown by animal microfossils, with special ref-erence to the Cladocera. *Ecological Monographs, 38*, 223-241.

Hasler, A. D. (1947). Eutrophication of Lakes by Domestic Drainage. *Ecology, 28*(4), 383-395.

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology, 54*(2), 427-432.

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

Hofmann, W. (1987a). Cladocera In Space And Time - Analysis Of Lake-Sediments. *Hydrobiologia, 145*, 315-321.

Hofmann, W. (1987b). Cladocera in space and time: Analysis of lake sediments. *Hydrobiologia, 145*, 315-321.

Hofmann, W. (1996). Empirical relationships between cladoceran fauna and trophic state in thirteen northern German lakes: analysis of surficial sediments. *Hydrobiologia, 318*(3), 195-201.

Hustedt, F. (1927-1966). *Kryptogamen-Flora von Deutschland, Österreich und der Schweiz Band VII Die Kieselalgen*. Leipzig: Akademische Verlagsgesellschaft.

Hustedt, F. (1930). *Bacillariophyta (Diatomeae)* (Vol. 10). Jena: Fischer.

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299--314.

Irvine, K., Allott, N., de Eyto, E., Free, G., White, J., Caroni, R., Kennelly, C., Keaney, J., Lennon, C., Kemp, A., Barry, E., Day, S., Mills, P., O' Riain, G., Quirke, B., Twomey, H., & Sweeney, P. (2001). *The ecological assessment of Irish lakes: The development of a new methodology suited to the needs of the EU Directive for Surface Waters*. Wexford: Environmental Protection Agency.

Irvine, K., Boelens, R., Fitzsimmons, J., Kemp, A., & Johnston, P. (2002). *Review of monitoring and Research to meet the needs of the EU Water Framework Directive*. Wexford: Environmental Protection Agency, Ireland.

Jackson, D. A. (1993). Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology, 74*(8), 2204-2214.

Jacques, J. M. S., Douglas, D. J., Price, N., Drakulic, N., & Gubala, C. P. (2005). The effect of fish intruductions on the diatom and cladoceran communities of Lake Penongo, Ontario, Canada. *Hydrobiologia, 549*, 99-113.

Jennings, E., Mills, P., Jordan, P., Jensen, J. P., Søndergaard, M., Barr, A., Glasgow, G., & Irvine, K. (2003). *Eutrophication from agricultural sources. Seasonal patterns & Effects of phosphorus*. Wexford: EPA Ireland.

Jeppesen, E., Jensen, J. P., Amsinck, S., Landkildehus, F., Lauridsen, T. L., & Mitchell, S. F. (2002). Reconstructing the historical changes in Daphnia mean and planktivorous fish abundance in lakes from the size of *Daphnia* ephippia in the sediment. *Journal of Paleolimnology, 27*(1), 133-143.

Jeppesen, E., Jensen, J. P., Lauridsen, T. L., Amsinck, S. L., Christoffersen, K., Søndergaard, M., & Mitchell, S. F. (2003). Sub-fossils of cladocerans in the surface sediment of 135 lakes as proxies for community structure of zooplankton, fish abundance and lake temperature. *Hydrobiologia, 491*(1 - 3), 321 - 330.

Jeppesen, E., Jensen, J. P., SØndergaard, M., Lauridsen, T., & Landkildehus, F. (2000). Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. *Freshwater Biology, 45*, 201-218.

Jeppesen, E., Jensen, J. P., Søndergaard, M., Lauridsen, T., Pedersen, L. J., & Jensen, L. (1997). Top-down control in freshwater lakes: the role of nutrient state, submerged macrophytes and water depth. *Hydrobiologia, 342-343*, 151 - 164.

Jeppesen, E., Leavitt, P. R., de Meester, L., & Jensen, J. P. (2001). Functional ecology and palaeolimnology: using cladoceran remains to reconstruct anthropogenic impact. *Trends in Ecology and Evolution, 16*, 191-198.

Jeppesen, E., Madsen, E. A., Jensen, J. P., & Anderson, N. J. (1996). Reconstructing the past density of planktivorous fish and trophic structure from sediment zooplankton fossils - a surface sediment calibration data set from shallow lakes. *Freshwater Biology, 36*, 115-127.

Jordan, C. (1997). Mapping of rainfall chemistry in Ireland 1972-94. *Biology and Environment: Proceedings of the Royal Irish Academy, 97B*(1), 55-73.

Juggins, S. (2003). *C2 User Guide-Software for ecological and palaeoecological data analysis and visualisation.* Newcastle upon Tyne, UK: University of Newcastle.

Kattel, G. R., Mackay, A., & Battarbee, R. W. (2006). Variation in patterns of composition and abundance of cladocerans in littoral and open water habitats of Loch Coire Fionnaraich, Scotland. *International Journal of Lakes and Rivers, 1*(1).

Kauppila, T., Moisio, T., & Salonen, V. P. (2002). A diatom-based inference model for autumn epilimnetic total phosphorus concentration and its application to a presently eutrophic boreal lake. *Journal of Paleolimnology, 27*, 261-273.

King, J. J., & Champ, W. S. T. (2000). Baseline water quality investigations on Lough Carra, Western Ireland, with reference to water chemistry, phytoplankton and aquatic plants. *Biology and Environment: Proceedings of the Royal Irish Academy, 100 (B)*, 13-25.

King, L., Barker, P., & Jones, R. I. (2000). Epilithic algal communities and their relationship to environmental variables in Lakes of the English Lake District. *Freshwater Biology, 45*, 425-442.

Korhola, A. (1999). Distribution patterns of Cladocera in subarctic Fennoscandian lakes and their potential in environmental reconstruction. *Ecography, 22*(4), 357-373.

Korhola, A., Olander, H., & Blom, T. (2000). Cladoceran and chironomid assemblages as qualitative indicators of water depth in subarctic Fennoscandian lakes. *Journal of Paleolimnology, 24*(1), 43 - 54.

Korhola, A., & Rautio, M. (2001). Cladocera and other branchiopod crustaceans. In J. P. Smol, H. J. B. Birks & W. M. Last (Eds.), *Tracking Environmental Change Using Lake Sediments* (Vol. 4, pp. 5-41). Dordrecht: Kluwer Academic.

Koster, D., Racca, J. M. J., & Pienitz, R. (2004). Diatom-based inference models and reconstructions revisited: methods and transformations. *Journal of Paleolimnology, 32*(3), 233-246.

Krammer, K., & Lange-Bertalot, H. (1986). *Bacillariophyceae. 1. Teil: Naviculaceae*. Heidelberg & Berlin: Spektrum Akademischer Verlag.

Krammer, K., & Lange-Bertalot, H. (1988). *Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae*. Heidelberg & Berlin: Spektrum Akademischer Verlag.

Krammer, K., & Lange-Bertalot, H. (1991a). *Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae, Eunotiaceae*. Heidelberg & Berlin: Spektrum Akademischer Verlag.

Krammer, K., & Lange-Bertalot, H. (1991b). *Bacillariophyceae. 4. Teil: Achnanthaceae, Literaturverzeichnis*. Heidelberg & Berlin: Spektrum Akademischer Verlag.

Krammer, K., & Lange-Bertalot, H. (2000). *Bacillariophyceae. 5. Teil: English and French translation of the keys.* Heidelberg & Berlin: Spektrum Akademischer Verlag.

Krause-Dellin, D., & Steinberg, C. (1986). Cladoceran remains as indicators of lake acidification. *Hydrobiologia, 143*(1), 129.

Lampert, W., & Sommer, U. (1997). *Limnoecology: the Ecology of Lakes and Streams* (J. F. Haney, Trans.). New York: Oxford University Press.

Lange-Bertalot, H., & Metzeltin, D. (1996). *Indicators of oligotrophy: 800 taxa representative of three ecologically distinct lake types: carbonate buffered-oligodystrophic-weakly buffered soft water* (Vol. 2). Königstein: Koeltz Scientific Books.

Leavitt, P. R., Carpenter, S. R., & Kitchell, J. F. (1989). Whole lake experiments: the annual record of fossil pigments and zooplankton. *Limnology and Oceanography, 34*, 700-717.

Leavitt, P. R., Sanford, P. R., Carpenter, S., & Kitchell, J. F. (1994). An annual fossil record of production, planktivory and piscivory during whole-lake manipulations. *Journal of Paleolimnology, 11*, 133-149.

Legendre, P., & Legendre, L. (1998). *Numerical Ecology* (2nd English Edition ed.). Amsterdam: Elsevier.

Leira, M., Jordan, P., Taylor, D., Dalton, C., Bennion, H., Rose, N., & Irvine, K. (2006). Assessing the ecological status of candidate reference lakes in Ireland using palaeolimnology. *Journal of Applied Ecology, 43*(4), 816-827.

Lepš, J., & Šmilauer, P. (2003). *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge, UK: Cambridge University Press.

Lieder, U. (1983). Revision of the genus Bosmina Baird, 1845 (Crustacea, Cladocera). *Internationale Revue der Gesamten Hydrobiologie, 68*, 121-139.

Lotter, A. F. (2005). Multi-proxy climatic reconstructions. In A. MacKay, R. Battarbee, J. Birks & F. Oldfield (Eds.), *Global Change in the Holocene* (pp. 373-383): Hodder Arnold.

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.

Lotter, A. F., Birks, H. J. B., Hofmann, W., & Marchetto, A. (1998). Modern diatom, cladocera, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps. II. Nutrients. *Journal of Paleolimnology, 19*, 443-463.

Lotter, A. F., Walker, I. R., Brooks, S. J., & Hofmann, W. (1999). An intercontinental comparison of chironomid palaeotemperature inference models: Europe vs North America. *Quaternary Science Reviews, 18*(6), 717-735.

Lowe, J. J., & Walker, M. J. C. (1997). *Reconstructing Quaternary Environments* (Second ed.). Harlow, England: Longman.

Lucey, J., Bowman, J. J., Clabby, K. J., Cunningham, P., Lehane, M., MacCárthaigh, M., McGarrigle, M. L., & Toner, P. F. (1999). *Water Quality in Ireland 1995-1997*. Wexford: Environmental Protection Agency.

Lund, J. W. G. (1954). The Seasonal Cycle of the Plankton Diatom, Melosira Italica (EHR.) Kutz. Subsp. Subarctica O. Mull. *The Journal of Ecology, 42*(1), 151-179.

Lund, J. W. G. (1979). Changes in the Phytoplankton of an English Lake, 1945-1977. *Hydrobiological Journal, 14*, 6-21.

Lynch, M. (1980). The evolution of cladoceran life histories. *The Quarterly Review of Biology, 55*, 23-42.

Mackay, A. W., Jones, V. J., & Battarbee, R. W. (2003). Approaches to Holocene climate reconstruction using diatoms. In A. W. Mackay, R. W. Battarbee, H. J. B. Birks & F. Oldfield (Eds.), *Global Change in the Holocene* (pp. 294-309). London: Hodder Arnold.

Margaritora, F. G. (1985). *Cladocera*. Bologna, Italia: Edizioni Calderini.

Mason, C. F. (2002). *Biology of Freshwater Pollution*. Essex: Pearson Education.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (Second ed.). London, New York: Chapman and Hall.

McGarrigle, M. L. (2001). *Eutrophication of Irish Waters -from Science to Legisilation and Management*. Dublin: Royal Dublin Society.

McGarrigle, M. L., Bowman, J. J., Clabby, K. J., Lucey, J., Cunningham, P., MacCárthaigh, M., Keegan, M., Cantrell, B., Lehane, M., Clenaghan, C., & Toner, P. F. (2002). Water Quality in Ireland: 1998-2000.

Melo, R. D., & Hebert, P. D. N. (1994). A taxonomic reevaluation of North American Bosminidae. *Canadian Journal of Zoology, 72*, 1808-1825.

Mezquita, F., Roca, J. R., Reed, J. M., & Wansard, G. (2005). Quantifying species-environment relationships in non-marine Ostracoda for ecological and palaeoecological studies: Examples using Iberian data. *Palaeogeography, Palaeoclimatology, Palaeoecology, 225*(1-4), 93-117.

Miettinen, J. O. (2003). A diatom-total phosphorus transfer function for freshwater lakes in southeastern Finland, including cross-validation with independent test lakes. *Boreal Environment Research, 8*, 215-228.

Morgan, M. A. (1977). Enrichment of water with nutrients derived from agriculture. In W. K. Downey & G. Nı´ Uid (Eds.), *Pollution prevention by eutrophication control* (pp. 111-118). Dublin: Stationery Office.

Moss, B. (1998). *Ecology of Fresh Waters-Man and Medium, Past to Future* (Third ed.). Oxford: Blackwell Science.

Murray, D. A., & Douglas, D. J. (1977). *Eutrophication: past, present- and future?* Paper presented at the the Seminar of Lake Pollution Prevention by Eutrophication Control, Killarney, Ireland.

Nilssen, J. P., & Sandoy, 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 B., 327*, 299-309.

Nilsson, M. B., Dabakk, E., Korsman, T., & Renberg, I. (1996). Quantifying Relationships between Near-Infrared Reflectance Spectra of Lake Sediments and Water Chemistry. *Environment Science and Technology, 30*(8), 2586 - 2590.

Noy-Meir, I., Walker, D., & Williams, W. T. (1975). Data Transformations in Ecological Ordination: II. On the Meaning of Data Standardization. *The Journal of Ecology, 63*, 779-800.

Odland, A., Birks, H. J. B., & Line, J. M. (1995). Ecological optima and tolerances of Thelypteris limbosperma, Athyrium distentifolium, and Matteuccia struthiopteris along environmental gradients in Western Norway. *Vegetatio, 120*, 115-129.

OECD. (1982). *Eutrophication of waters. Monitoring, assessment and control*. Paris: OECD.

Oksanen, J. (2004). Multivariate Analysis in Ecology-Lecture Notes: Department of Biology, University of Oulu, Finland.

Oksanen, J. (2005a). Frequently Asked Questions (Vegan package of R program). Retrieved February 4, 2005

Oksanen, J. (2005b). Multivariate Analysis of Ecological Communities in R: vegan tutorial. Retrieved September 26, 2005

Osborne, P. L., & Moss, B. (1977). Paleolimnology and trends in the phosphorus and iron budgets of an old man-made lake, Barton Broad, Norfolk. *Freshwater Biology, 7*, 213-233.

Owen, R. B., & Crossley, R. (1992). Spatial and temporal distribution of diatoms in sediments of Lake Malawi, Central Africa, and ecological implications. *Journal of Paleolimnology, 7*(1), 55 - 71.

Pennak, R. W. (1989). Fresh-Water Invertebrates of the United States: Protozoa to Mollusca. In (pp. 369-409). Toronto, Canada: John Wiley and Sons.

Philibert, A., & Prairie, Y. T. (2002). Diatom-based transfer functions in western Quebec lakes (Abitibi and Haute Mauricie): the possible role of epilimnetic CO2 concentration in influencing diatom assemblages. *Journal of Paleolimnology, 27*, 465-480.

Pollard, P., & Huxham, M. (1998). The European Water Framework Directive: a new era in the management of aquatic ecosystem health? *Aquatic Conservation: Marine and Freshwater Ecosystems, 8*(6), 773-792.

Prygiel, J., & Coste, M. (2000). *Guide méthodologique pour la mise en œuvre de l'Indice Biologique Diatomées*. Bordeaux: Agences de l'Eau - Cemagref.

Racca, J. M. J., Gregory-Eaves, I., Pienitz, R., & Prairie, Y. T. (2004). Tailoring palaeolimnological diatom-based transfer functions. *Canadian Journal of Fisheries and Aquatic Sciences, 61*(12), 2440-2454.

Ramstack, J. M., Fritz, S. C., Engstrom, D. R., & Heiskary, S. A. (2003). The Application of a Diatom-based Transfer Function to Evaluate Regional Water-Quality Trends in Minnessota Since 1970. *Journal of Paleolimnology, 29*(1), 79 - 94.

Rautio, M., Sorvari, S., & Korhola, A. (2000). Diatom and crustacean zooplankton communities, their seasonal variability and representation in the sediments of subarctic Lake Saanajarvi. *Journal of Limnology, 59*(Supp. 1), 81-96.

Reavie, E. D., Hall, R. I., & Smol, J. P. (1995). An expanded weighted-averaging model for inferring past total phosphorus concentrations from diatom assemblages in eutrophic British Columbia (Canada) lakes. *Journal of Paleolimnology*(14), 49 - 67.

Reavie, E. D., & Smol, J. P. (2001). Diatom-environmental relationships in 64 alkaline southeastern Ontario (Canada) lakes: a diatom-based model for water quality reconstructions. *Journal of Paleolimnology, 25*(1), 25-42.

Reavie, E. D., Smol, J. P., & Dillon, P. J. (2002). Inferring long-term nutrient changes in southeastern Ontario lakes: comparing paleolimnological and mass-balance models. *Hydrobiologia, 481*(1 - 3), 61 - 74.

Renberg, I. (1991). The HON-Kajak sediment corer. *Journal of Paleolimnology, 6*(2), 167 - 170.

Reynolds, C. S. (1984). *The Ecology of Freshwater Phytoplankton*. Cambridge; New York: Cambridge University Press.

Reynolds, J. D. (1998). *Ireland's Freshwaters*. Dublin: The Marine Institute.

Round, F. E. (1981). *The Ecology of Algae*. Cambridge, U.K.: Cambridge University Press.

Round, F. E., Crawford, R. M., & Mann, D. G. (1990). *Diatoms: Biology and Morphology of the Genera*: Cambridge University Press.

Sayer, C. D. (2001). Problems with the application of diatom-total phosphorus transfer functions: Examples from a shallow English Lake. *Freshwater Biology, 46*, 743-757.

Scheffer, M. (1998). *Ecology of Shallow Lakes*. Dordrecht: Kluwer Academic.

Scheffer, M., Hosper, S. H., Meijer, M. L., Moss, B., & Jeppesen, E. (1993). Alternative equilibria in shallow lakes. *Trends in Ecology and Evolution, 8*(8), 275-279.

Schulz, K. L., & Sterner, R. W. (1999). Phytoplankton Phosphorus Limitation and Food Quality for Bosmina. *Limnology and Oceanography, 44*(6), 1549-1556.

Scourfield, D. J., & Harding, J. P. (1966). *A Key to the British Freshwater Cladocera*: Freshwater Biological Association.

Shaw, P. J. A. (2003). *Multivariate Statistics for the Environmental Sciences*. London: Hodder Arnold.

Shumate, B. C. A., Schelske, C. L. A., Crisman, T. L. A., & Kenney, W. F. A. (2002). Response of the cladoceran community to trophic state change in Lake Apopka, Florida. *Journal of Paleolimnology, 27*(1), 71 - 77.

Simpson, G. L. (2005a). *Defining restoration targets for acidified upland lakes using diatom and cladoceran sub-fossil remains and the modern analogue technique.* Unpublished Ph.D. thesis, University College London, London.

Simpson, G. L. (2005b). Practicals for the course of Numerical Analysis of Biological and Environmental Data: ECRC, University College London.

Smilauer, P. B., H.J.B. (1995). The use of generalised additive models in the description of diatom-environment response surfaces. In (Vol. 7, pp. 42-47): GEUS.

Smirnov, N. N. (1974). *Chydoridae* (Vol. 1). Jerusalem: Keter Press.

Smirnov, N. N. (1996). *Cladocera: the Chydorinae and Sayciinae (Chydoridae) of the World* (Vol. 11). Leiden: Backhuys Publishers.

Smol, J. P. (2002). *Pollution of Lakes and Rivers: A paleoenvironmental Perspective*. London: Arnold Publishers.

Søndergaard, M., Jensen, J. P., & Jeppesen, E. (1999). Internal phosphorus loading in shallow Danish lakes. *Hydrobiologia, 408-409*(145 - 152).

Søndergaard, M., Jeppesen, E., Jensen, J. P., & Amsinck, S. L. (2005). Water Framework Directive: ecological classification of Danish lakes. *Journal of Applied Ecology, 42*, 616-629.

Stansfield, J. H., Perrow, M. R., Tench, L. D., Jowitt, A. J. D., & Taylor, A. A. L. (1997). Submerged macrophytes as refuges for grazing Cladocera against fish [-3pt]predation: observations on seasonal changes in relation to macrophyte cover and predation pressure. *Hydrobiologia, 342-343*, 229-240.

Steinberg, C., Hartmann, H., Arzet, K., & Krause-Dellin, D. (1988). Paleoindication of acidification in Kleiner Arbersee (Federal Republic of Germany, Bavarian Forest) by chydorids, chrysophytes, and diatoms. *Journal of Paleolimnology, 1*(2), 149 - 157.

Sterner, R. W., & Hessen, D. O. (1994). Algal Nutrient Limitation and the Nutrition of Aquatic Herbivores. *Annual Review of Ecology and Systematics, 25*, 1-29.

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., Hawworth, 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*. London: ENSIS Publishing.

Stevenson, R. J., & Pan, Y. (1999). Assessing environmental conditions in rivers and streams with diatoms. In E. F. Stoermer & J. P. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences* (pp. 11-40). Cambridge, UK: Cambridge University Press.

Stoermer, E. F. (1984). Qualitative characteristics of phytoplankton assemblages. In L. E. Shubert (Ed.), *Algae as Ecological Indicators* (pp. 49-67): Academic Press.

Stoermer, E. F., & Smol, J. (1999). *The Diatom: Applications for the Environmental and Earth Sciences*. Cambridge: Cambridge University Press.

Sweetman, J., & Smol, J. (2006). Patterns in the Distribution of Cladocerans (Crustacea: Branchiopoda) in Lakes Across a Northâ€“south Transect in Alaska, USA. *Hydrobiologia, 553*(1), 277.

Sylvestre, F., Servant-Vildary, S., & Roux, J. C. (2001). Diatom based ionic concentration and salinity models from the South Bolivian Altiplano. *Journal of Paleolimnology, 25*, 279-295.

Talling, J. F., & Heaney, S. I. (1988). Long-term changes in some English (Cumbrian) lakes subjected to increased nutrient inputs. In F. E. Round (Ed.), *Algae and the Aquatic Environment* (pp. 1-29). Bristol: Biopress.

Taylor, D., Dalton, C., Leira, M., Jordan, P., Irvine, K., Bennion, H., Magee, E., & León-Vintro., L. (2006). *Identification of reference-status for Irish lake typologies using palaeolimnological methods and techniques-IN-SIGHT Final report*. Dublin: School of Natural Sciences, Trinity College, University of Dublin.

Telford, R. J., & Birks, H. J. B. (2005). The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quaternary Science Reviews, 24*(20-21), 2173-2179.

ter Braak, C. J. F. (1987a). calibration. In R. H. G. Jongman, C. J. F. ter Braak & O. F. R. van Tongeren (Eds.), *Data Analysis in Community and landscape Ecology* (pp. 78-90). Wageningen: Pudoc.

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

ter Braak, C. J. F. (1988). *CANOCO - a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis. Version 2.1*. Wageningen: Agricultural Mathematics Group, Ministry of Agriculture and Fisheries.

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*(1), 485 - 502.

ter Braak, C. J. F., Juggins, S., H.J.B., B., & van der Voet, H. (1993). Weighted-averaging-partial least square regression (WA-PLS): Definition and comparison with other methods for species-environmental calibration. In G. P. Patil & C. R. Rao (Eds.), *Multivariate Environmental Statistics* (pp. 525-560). Amsterdam: Elsevier Science Publishers.

ter Braak, C. J. F., & Looman, C. W. N. (1986). Weighted averaging, logistic regression and the Gaussian response model. *Plant Ecology, 65*(1), 3-11.

ter Braak, C. J. F., & Looman, C. W. N. (1987). Regression. In R. H. G. Jongman, C. J. F. ter Braak & O. F. R. van Tongeren (Eds.), *Data Analysis in Community and landscape Ecology* (pp. 29-77). Wageningen: Pudoc.

ter Braak, C. J. F., & Prentice, I. C. (1988). A theory of gradient analysis. *Advances in Ecological Research, 18*, 271-313.

ter Braak, C. J. F., & Šmilauer, P. (2002). *CANOCO reference manual and CanoDraw for Windows User's guide: Software for Canonical Community Ordination (version 4.5)*. Ithaca, NY, USA: Microcomputer Power.

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

Thoms, M. C., Ogden, R. W., & Reid, M. A. (1999). Establishing the condition of lowland floodplain rivers: a palaeo-ecological approach. *Freshwater Biology, 41*(2), 407-423.

Tibby, J. (2004). Development of a diatom-based model for inferring total phosphorus in south-eastern Australian water storages. *Journal of Paleolimnology, 31*(1), 23-36.

Tilman, D., Kilham, S. S., & Kilham, P. (1982). Phytoplankton Community Ecology: The Role of Limiting Nutrients. *Annual Review of Ecology and Systematics, 13*, 349-372.

Toner, P., Bowman, J., Clabby, K., Lucey, J., McGarrigle, M., Concannon, C., Clenaghan, C., Cunningham, P., Delaney, J., O'Boyle, S., MacCarthaigh, M., Craig, M., & Quinn, R. (2005). *Water Quality in Ireland 2001-2003*. Wexford: Environmental Protection Agency of Ireland.

Toner, P. F. (1977). Investigation of eutrophication in Irish lakes. In W. K. Downey & G. Nı´ Uid (Eds.), *Pollution prevention by eutrophication control* (pp. 15-49). Dublin: Stationery Office.

Toner, P. F., Clabby, K. J., Bowman, J. J., & McGarrigle, M. L. (1986). *Water Quality in Ireland- The Current Position. Part One: General Assessment*. Dublin: An Foras Forbartha.

Trifonova, I. S. (1987). Oligotrophic-eutrophic succession of lake phytoplankton. In R. F. (Ed.), *Algae and the Aquatic Environment* (pp. 106-124). Bristol: Biopress.

Uimonensimola, P., & Tolonen, K. (1987). Effects Of Recent Acidification On Cladocera In Small Clear-Water Lakes Studied By Means Of Sedimentary Remains. *Hydrobiologia, 145*, 343-351.

Urabe, J., Clasen, J., & Sterner, R. W. (1997). Phosphorus Limitation of Daphnia Growth: Is it Real? *Limnology and Oceanography, 42*(6), 1436-1443.

Vadeboncouer, Y., Jeppesen, E., Vander Zanden, M. J., Schierup, H. H., Christoffersen, K., & Lodge, D. M. (2003). From Greenland to green lakes: cultural eutrophication and the loss of benthic pathways in lakes. *Limnology and Oceanography, 48*(4), 1408-1418.

Van Dam, H., Mertens, A., & Sinkeldam, J. (1994). A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology, 28*(1), 117-133.

van Tongeren, O. F. R. (1995). Cluster analysis. In R. H. G. Jongman, C. J. F. ter Braak & O. F. R. van Tongeren (Eds.), *Data Analysis in Community and Landscape Ecology* (pp. 174-212). Cambridge, U.K.: Cambridge University Press.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer.

Venables, W. N., Smith, D. M., & Team, t. R. D. C. (2005). An introduction to R: notes on R: A programming Environment for Data Analysis and Graphics (version 2.2.0) (2005-10-06).

Wemaëre, A. (2005). *Influence of catchment characteristics on the relationship between land use and lake water quality in County Clare.* Unpublished Ph.D. thesis, Trinity College, University of Dublin, Dublin.

Werner, P., & Smol, J. P. (2005). Diatom-environmental relationships and nutrient transfer functions from contrasting shallow and deep limestone lakes in Ontario, Canada. *Hydrobiologia, 533* (1), 145-173.

Wetzel, R. (2001). *Limnology-Lake and River Ecosystems*. San Diego, CA, USA: Academic Press.

Whiteside, M. C. (1970). Danish chydorid Cladocera: modern ecology and core studies. *Ecological Monographs, 40*, 79-118.

Wood, S. (2006). The mgcv Package (Version 1.3-13).

Wunsam, S., & Schmidt, R. (1995). A diatom-phosphorus transfer function for alpine and pre-alpine lakes. *Memorie dell'Istituto Italiano di Idrobiologia, 53*, 85-99.

Appendix A: Summary of Study Sites

Basic geographical information with site codes for 75 study lakes used in this thesis. NA means no related information are collected.

Site Code	Lake Name	County	LakeCodeV6	Grid Reference
ANN	Annaghmore	Roscommon	26-0155a-1060-000	M 900 837
ANS	Anascaul	Kerry	22-00200-0010-000	Q 585 052
ARD	Arderry	Galway	31-000r4-1120-000	L 995 457
ATE	Atedaun	Clare	27-00158-0980-000	R 295 885
BAA	Ballyallia	Clare	28-00154-0110-000	R 342 809
BAB	Ballybeg	Clare	27-00158-0030-000	R 330 739
BAC	Ballycar	Clare	27-0155c-0110-000	R 414 690
BAF	Barfinnihy	Kerry	21-00216-0020-000	V 850 768
BAL	Ballynakill (Gorumna)	Galway	31-000r4-2040-000	L 856 225
BAN	Bane	Westmeath	07-00159-0930-000	N 550 712
BAR	Barra	Donegal	38-00048-0490-000	B 935 120
BAT	Ballyteige	Clare	27-00158-0900-000	R 348 888
BEA	Beaghcauneen	Galway	32-t4_32-0270-000	L 680 472
BRL	Bray Lower	Wicklow	10-00169-0040-000	O 137 161
BUN	Bunny	Clare	27-00158-1760-000	R 375 967
CAA	Carra	Mayo	30-00143-1690-000	M 180 710
CAR	Caragh	Kerry	22-00208-0020-000	V 725 905
CAS	Castle	Clare	27-00158-1870-000	R 486 690
CAU	Caum	Clare	23-00199-0030-000	R 182 810
CLO	Cloonaghlin	Kerry	21-00213-0090-000	V 610 709
CRA	Crans	Tyrone	NA	H 711 568
CUL	Cullaun	Clare	27-00158-1190-000	R 315 905
CUY	Cullaunyheeda	Clare	NA	R 464 747
DAN	Dan	Wicklow	10-00171-0070-000	O 150 040
DOC	Doo (CE)	Clare	28-00152-0050-000	R 120 721
DOO	Doo (DL)	Donegal	39-00008-0010-000	C 359 394
DRO	Dromore	Clare	27-00158-0560-000	R 346 859
DRU	Drumanure	Clare	NA	R 215842
DUN	Dunglow	Donegal	38-00016-0070-000	B 782 117
EAS	Easky	Sligo	35-00114-0150-000	G 442 225
EFF	Effernan	Clare	27-0155e-0080-000	R 222 558
FAD	Fad Inishowen East	Donegal	40-0000c-0020-000	C 539 439
FEE	Fee	Galway	32-00132-0020-000	L 790 613
FEG	Feeagh	Mayo	32-00107-0070-000	F 965 000
GAR	Garvillaun	Clare	28-00149-0240-000	R 248 829
GOR	Gortaganniv	Clare	27-00158-0420-000	R 251 759
HOH	Moher	Mayo	32-00126-0050-000	L 977 766
INC	Inchichronan	Clare	27-00158-0610-000	R 391853
INQ	Inchiquin	Clare	21-00220-0050-000	R 268 897
KEE	Keel (Rosses)	Donegal	38-00022-0080-000	B 847 162
KIL	Kiltooris	Donegal	38-000j6-0050-000	G 676 972

Site Code	Lake Name	County	LakeCodeV6	Grid Reference
KIN	Kindrum	Donegal	38-u6_38-0110-000	C 185 430
KYL	Kylemore	Galway	32-00133-0050-000	L 770 552
LEN	Lene	Westmeath	07-00159-1150-000	N 510 685
LIC	Lickeen	Clare	28-00149-0080-000	R 176 909
LIS	Lisnahan	Clare	NA	Q 900 617
MAU	Maumwee	Galway	30-00143-1460-000	L 977 484
MCN	McNean	Leitrim	36-00123-1070-000	H 040 400
MOA	Moanmore	Clare	NA	Q 979 611
MOO	Mooghna	Clare	28-00149-0040-000	R 137 842
MOR	Morgans	Clare	NA	R 255 835
MUC	Muckanagh	Clare	27-00158-1470-000	R 370 925
MUL	Mullagh	Cavan	07-00159-0220-000	N 677 855
MUN	Muckno	Monaghan	06-00094-0280-000	H 856 175
NAB	Nambrackkeagh	Galway	32-00132-0040-000	L 821 603
NAH	Nahasleam	Galway	31-000r4-0720-000	L 971 244
NAM	Naminn	Donegal	40-00004-0020-000	C 396 419
NAN	Naminna	Clare	28-00152-0060-000	R 176 710
OFL	O'Flynn	Roscommon	26-00156-0570-000	M 585 795
OOR	Oorid	Galway	31-00136-0670-000	L 930 460
OUG	Oughter	Cavan	36-00123-3230-000	H 342 075
OWE	Owel	Westmeath	26-00157-0260-000	N 400 581
POL	Pollaphuca	Wicklow	09-00168-0230-000	N 985 086
RAM	Ramor	Cavan	07-00159-0600-000	N 603 868
REA	Rea	Galway	29-00145-0180-000	M 615 155
ROS	Rosconnell	Clare	NA	R 222 793
RUS	Rushaun	Clare	NA	R 253 791
SHI	Shindilla	Galway	31-000r4-0950-000	L 960 460
SIL	Sillan	Monaghan	36-00123-5720-000	H 700 070
TAL	Talt	Sligo	34-00110-0630-000	G 398 150
TAY	Tay	Wicklow	10-00171-0090-000	O 160 750
TUL	Tullabrack	Clare	28-00154-0030-000	R 018 597
UPE	Upper	Kerry	22-00207-0260-000	V 900 817
VEA	Veagh	Donegal	38-00027-0210-000	C 022 215

Appendix B: Summary of Physical Data

Data for six physical variables of 75 study lakes and NA means no related information are collected.

Site Code	Lake Name	Altitude (m)	Catchment Area (ha)	Lake Area (ha)	Catchment Area: Lake Area	Max Depth (m)	Mean Depth (m)
ANN	Annaghmore	46	395.8	53.1	7.5	16.0	5.8
ANS	Anascaul	NA	NA	< 50	NA	24.0	15.6
ARD	Arderry	37	1442.1	81.1	17.8	11.7	5.7
ATE	Atedaun	22	28250.0	38.0	743.6	13.0	1.4
BAA	Ballyallia	10	2473.0	32.6	75.9	16.5	6.3
BAB	Ballybeg	10	414.0	19.7	21.0	5.7	2.7
BAC	Ballycar	30	367.0	3.0	122.3	NA	2.3
BAF	Barfinnihy	249	79.3	13.6	5.8	16.7	9.8
BAL	Ballynakill	13	140.9	23.9	5.9	19.0	6.7
BAN	Bane	112	470.1	75.4	6.2	16.5	10.9
BAR	Barra	90	1968.1	62.6	31.4	12.0	4.4
BAT	Ballyteige	20	29042.0	14.2	2045.2	7.9	4.1
BEA	Beaghcauneen	NA	NA	26.0	NA	16.0	8.0
BRL	Bray Lower	378	142.8	24.8	5.8	45.7	19.8
BUN	Bunny	20	7750.0	102.6	75.5	11.6	1.4
CAA	Carra	21	10400.0	1438.0	7.2	18.0	1.8
CAR	Caragh	15	16107.8	499.4	32.3	40.0	11.6
CAS	Castle	20	13235.0	23.1	572.9	NA	3.4
CAU	Caum	51	519.0	6.8	76.3	8.2	3.1
CLO	Cloonaghlin	109	1023.5	127.7	8.0	27.0	12.0
CRA	Crans	95	59.5	8.5	7.0	12.0	6.7
CUL	Cullaun	25	8132.0	49.7	163.6	25.0	13.4
CUY	Cullaunyheeda	30	2943.0	152.8	19.3	25.0	7.8
DAN	Dan	200	6313.3	102.9	61.4	33.5	13.5
DOC	Doo (CE)	91	2275.0	130.2	17.5	14.9	3.3
DOO	Doo (DL)	283	427.9	9.0	47.4	16.5	5.3
DRO	Dromore	20	167.0	49.1	3.4	19.0	6.0
DRU	Druminure	100	68.0	2.8	24.3	4.0	2.3
DUN	Dunglow	13	3767.2	61.2	61.5	8.0	5.5
EAS	Easky	180	1160.6	119.2	9.7	11.0	2.4
EFF	Effernan	60	276.0	10.3	26.8	8.3	4.5
EGI	Egish	162	784.3	121.7	6.4	12.0	5.0
FAD	Fad Inishowen	233	60.7	12.3	5.0	13.6	5.6
FEE	Fee	47	1575.9	173.7	9.1	32.0	15.5
FEG	Feeagh	11	10033.3	394.8	25.4	45.3	14.5
GAR	Garvillaun	100	70.0	2.3	30.4	4.2	3.0
GOR	Gortaganniv	80	203.0	3.2	63.4	4.5	3.0
HOH	Moher	75	934.4	40.4	23.1	13.4	2.9
INC	Inchichronan	30	3334.0	116.7	28.6	18.1	6.5
INQ	Inchiquin	35	14714.0	106.9	137.6	31.0	12.2
KEE	Keel	136	396.4	11.4	34.7	10.0	5.2
KIL	Kiltooris	7	554.5	43.5	12.8	13.5	<4
KIN	Kindrum	8	366.8	60.8	6.0	12.5	6.6
KYL	Kylemore	35	2090.5	132.2	15.8	30.0	11.7
LEN	Lene	93	1169.0	416.2	2.8	19.7	8.5
LIC	Lickeen	71	867.0	84.4	10.3	23.6	3.9
LIS	Lisnahan	54	77.0	5.9	13.1	2.3	1.4
MAU	Maumwee	46	425.2	27.6	15.4	8.3	2.0

Site Code	Lake Name	Altitude (m)	Catchment Area (ha)	Lake Area (ha)	Catchment Area: Lake Area	Max Depth (m)	Mean Depth (m)
MCN	McNean	50	12036.9	977.8	12.3	14.5	6.7
MOA	Moanmore	10	335.0	12.1	27.7	1.1	0.9
MOO	Mooghna	91	127.0	3.3	38.5	9.7	4.8
MOR	Morgans	100	30.0	1.2	25.0	4.8	3.5
MUC	Muckanagh	20	3935.0	96.1	40.9	15.0	2.9
MUL	Mullagh	120	114.2	35.1	3.3	8.1	2.3
MUN	Muckno	90	16072.4	364.4	44.1	27.0	5.9
NAB	Nambrackkeagh	65	55.6	6.7	8.3	8.5	4.1
NAH	Nahasleam	33	2278.0	28.1	81.1	5.6	2.1
NAM	Naminn	150	110.2	15.0	7.3	7.8	3.9
NAN	Naminna	170	207.0	20.0	10.4	10.9	4.4
OFL	O'Flynn	77	1834.7	137.5	13.3	14.5	2.8
OOR	Oorid	45	739.5	60.5	12.2	14.0	5.5
OUG	Oughter	48.9	147874.0	1105.5	133.8	14.0	2.2
OWE	Owel	97.8	4694.3	1029.4	4.6	22.8	7.2
POL	Pollaphuca	180	30265.0	1973.9	15.3	NA	6.8
RAM	Ramor	83.4	25150.2	741.2	33.9	4.0	3.0
REA	Rea	81	1353.0	301.1	4.5	20.9	14.5
ROS	Rosconnell	50	129.0	9.0	14.3	9.2	5.9
RUS	Rushaun	71	123.0	3.4	36.2	7.2	4.2
SHI	Shindilla	38	965.7	70.2	13.8	22.0	8.1
SIL	Sillan	94	NA	140.0	NA	12.0	6.0
TAL	Talt	130	482.8	97.3	5.0	23.0	8.9
TAY	Tay	250	2002.9	50.0	40.1	32.8	17.1
TUL	Tullabrack	40	44.0	2.5	17.9	1.1	0.7
UPE	Upper	18	11307.6	169.9	66.6	36.1	14.5
VEA	Veagh	40	3687.7	260.9	14.1	28.0	>4

Appendix C: Summary of Land Cover Data

Land cover data provided by the Irish EPA were collected from the CORINE 2000 dataset for all 75 lakes except for four lakes (marked with *) from the CORINE 1990 dataset.

Site Code	Lake Name	Agriculture (%)	Forest (%)	Pasture (%)	Peat (%)	Urban (%)
ANN	Annaghmore	0.0	10.7	78.7	0.0	0.0
ANS	Anascaul	0.0	0.0	1.6	62.6	0.0
ARD	Arderry	0.0	12.0	0.0	78.2	0.0
ATE	Atedaun	11.3	0.8	38.4	3.1	0.3
BAA	Ballyallia	11.0	4.6	45.4	5.6	0.4
BAB	Ballybeg	4.0	22.6	53.7	0.0	10.9
BAC*	Ballycar	0.0	0.0	69.0	0.0	0.0
BAF	Barfinnihy	0.0	0.0	0.0	64.5	0.0
BAL	Ballynakill	84.7	0.0	9.5	5.8	0.0
BAN	Bane	3.6	0.0	96.0	0.0	0.0
BAR	Barra	0.0	0.0	0.0	72.3	0.0
BAT	Ballyteige	10.5	0.8	39.5	3.7	0.2
BEA	Beaghcauneen	0.0	0.0	0.0	100.0	0.0
BRL	Bray Lower	0.0	0.0	0.0	99.5	0.0
BUN	Bunny	4.4	0.0	32.1	1.6	0.0
CAA	Carra	24.5	4.2	53.0	10.2	0.0
CAR*	Caragh	6.2	4.3	5.2	64.4	0.0
CAS*	Castle	0.0	16.0	65.0	14.0	0.0
CAU	Caum	0.0	23.1	0.0	75.4	0.0
CLO	Cloonaghlin	0.0	0.0	0.0	70.0	0.0
CRA	Crans	0.0	15.0	85.0	0.0	0.0
CUL	Cullaun	6.5	0.0	42.8	0.5	0.0
CUY	Cullaunyheeda	8.1	5.7	68.4	7.7	6.4
DAN	Dan	0.2	13.1	1.4	80.3	0.0
DOC	Doo (CE)	6.1	5.5	24.7	63.8	0.0
DOO	Doo (DL)	0.0	0.0	0.0	100.0	0.0
DRO	Dromore	12.9	9.1	54.5	4.0	1.5
DRU	Druminure	0.0	0.0	20.4	0.0	0.0
DUN	Dunglow	5.7	0.0	0.0	86.3	0.0
EAS	Easky	0.0	0.0	0.0	98.0	0.0
EFF	Effernan	52.1	15.6	19.7	12.6	0.0
EGI	Egish	10.9	0.0	86.4	0.0	2.7
FAD	Fad Inishowen	0.0	0.0	0.0	100.0	0.0
FEE	Fee	0.0	7.6	0.0	72.9	0.0
FEG	Feeagh	2.9	22.8	0.0	62.0	0.0
GAR	Garvillaun	0.0	0.0	24.5	0.0	0.0
GOR	Gortaganniv	14.8	0.0	85.2	0.0	0.0
HOH	Moher	41.1	9.9	0.0	49.1	0.0
INC	Inchichronan	18.5	9.9	32.5	21.0	0.0
INQ	Inchiquin	8.0	6.0	32.0	2.0	0.0
KEE	Keel	0.0	0.0	0.0	100.0	0.0
KIL	Kiltooris	0.0	0.0	59.5	33.5	0.0
KIN	Kindrum	29.0	0.0	20.1	50.9	0.0
KYL	Kylemore	0.0	8.0	0.0	66.7	0.0
LEN	Lene	0.4	0.0	87.2	0.0	3.2
LIC	Lickeen	19.0	0.0	50.5	30.3	0.0
LIS	Lisnahan	36.9	0.0	62.9	0.0	0.0

Site Code	Lake Name	Agriculture (%)	Forest (%)	Pasture (%)	Peat (%)	Urban (%)
MAU	Maumwee	0.0	0.0	0.0	99.7	0.0
MCN	McNean	18.7	14.9	16.6	34.1	0.4
MOA	Moanmore	0.0	0.0	6.6	93.4	0.0
MOO	Mooghna	0.0	0.0	100.0	0.0	0.0
MOR	Morgans	43.2	0.0	56.8	0.0	0.0
MUC	Muckanagh	12.3	0.0	51.7	11.0	0.0
MUL	Mullagh	4.3	0.0	95.7	0.0	0.0
MUN	Muckno	5.3	0.3	89.1	1.5	2.3
NAB	Nambrackkeagh	0.0	31.7	0.0	57.8	0.0
NAH	Nahasleam	0.0	9.5	0.0	78.1	0.0
NAM	Naminn	0.0	0.0	0.0	100.0	0.0
NAN	Naminna	0.0	43.9	0.0	24.6	0.0
OFL	O'Flynn	7.6	2.9	41.0	45.1	1.8
OOR	Oorid	5.8	2.9	0.0	91.4	0.0
OUG	Oughter	3.2	1.1	87.3	3.5	0.8
OWE	Owel	14.0	1.1	80.6	0.0	0.0
POL	Pollaphuca	4.8	13.0	25.6	46.8	2.0
RAM	Ramor	12.1	2.0	81.7	0.6	0.8
REA	Rea	6.0	0.0	86.5	0.0	7.5
ROS*	Rosconnell	0.0	10.0	72.0	18.0	0.0
RUS	Rushaun	13.4	0.0	69.8	0.0	0.0
SHI	Shindilla	0.0	5.7	0.0	94.3	0.0
SIL	Sillan	18.0	0.0	79.8	0.0	1.3
TAL	Talt	0.0	1.6	26.9	70.9	0.0
TAY	Tay	0.0	6.6	0.0	88.0	0.0
TUL	Tullabrack	0.0	0.0	100.0	0.0	0.0
UPE	Upper	2.4	7.2	0.5	73.2	0.0
VEA	Veagh	0.0	39.6	0.0	58.0	0.0

Appendix D: Summary of Hydrochemical Data

Seventy-five lakes were ordered in the alphabetic sequence of site codes and lake names for site codes are listed in Appendix A. Unpublished data of candidate reference lakes were provided by the Irish EPA in 2003 and were cited as EPA CRL 2003. NA means no information is available when this thesis is being written.

Site Code	Year Sampled	Month Sampled	No. of Samples	Alkalinity (mg $L^{-1}CaCO_3$)	Chlorophyll a ($\mu g\ l^{-1}$)	Colour (mg l^{-1} PtCo/Hazen)	Conductivity ($\mu S\ cm^{-1}$)	pH	TP ($\mu g\ l^{-1}$)	Data Source
ANN	NA	NA	NA	159.4	0.4	19.0	351.0	8.5	6.5	EPA CRL 2003
ANS	1996/7	7	1	5.3	3.1	30.0	73.0	6.7	4.0	Irvine et al., 2001
ARD	NA	NA	NA	6.1	1.8	78.0	84.0	6.3	6.0	EPA CRL 2003
ATE	2000	4,6,7 9	4	135.4	5.8	29.8	334.5	8.0	36.7	Wemaere, 2005
BAA	2000	4,6,7 9	4	157.7	8.1	21.3	381.8	8.2	21.0	Wemaere, 2005
BAB	2001	1,4,5,6,7,8,9,10	8	128.0	29.3	26.7	299.4	7.9	84.3	Wemaere, 2005
BAC	2000	4,6,7 9	4	186.0	9.4	30.8	436.8	8.1	35.3	Wemaere, 2005
BAF	NA	NA	NA	4.2	3.2	3.0	56.0	6.8	4.1	EPA CRL 2003
BAL	NA	NA	NA	20.0	3.7	20.0	244.0	7.1	5.0	EPA CRL 2003
BAN	NA	NA	NA	132.5	1.4	1.0	297.0	8.4	4.6	EPA CRL 2003
BAR	NA	NA	NA	3.8	0.9	29.0	54.0	6.3	5.0	EPA CRL 2003
BAT	2000/1	9; 8	2	116.3	9.1	36.0	324.5	7.9	39.5	Wemaere, 2005
BEA	2003	1,2,3,4,6,9,11,12	9	15.0	5.3	NA	NA	6.3	19.6	2003
BRL	1996/7	4,6,7,9;1,4,6,7,9,12	10	-1.0	17.1	48.0	45.0	5.1	12.0	Irvine et al., 2001
BUN	NA	NA	NA	156.2	1.4	9.0	361.0	8.5	5.4	EPA CRL 2003
CAA	NA	NA	NA	138.0	2.6	12.8	320.1	8.4	11.6	Taylor et al., 2006
CAR	NA	NA	NA	3.3	2.4	19.0	74.0	6.7	5.5	EPA CRL 2003
CAS	2001	4,5,6,7,8,9,10	7	121.1	15.9	39.4	267.9	8.0	27.0	Wemaere, 2005
CAU	2000/1	8; 8	2	17.9	15.2	186.0	106.0	6.7	54.7	Wemaere, 2005
CLO	NA	NA	NA	2.0	3.6	15.0	62.0	6.8	5.3	EPA CRL 2003
CRA	NA	NA	NA	78.0	48.0	NA	316.0	8.5	89.0	Taylor et al., 2006
CUL	NA	NA	NA	172.0	0.8	16.0	393.0	8.4	5.6	EPA CRL 2003
CUY	2001	1,5,6,7,8,9,10	7	173.3	3.1	36.3	375.0	8.3	25.6	Wemaere, 2005
DAN	NA	NA	NA	-0.1	0.8	108.0	42.0	5.1	6.3	EPA CRL 2003
DOC	1996/7	3,4,5,6,7,8,9,10;1,3, 4,5,6,7,8,9,10,12	18	5.0	6.6	80.0	101.0	6.8	16.0	Irvine et al., 2001
DOO	NA	NA	NA	2.0	2.3	85.0	78.1	5.9	12.0	EPA CRL 2003
DRO	2001	1,5,6,7,8,9,10	7	163.3	9.2	22.9	343.9	8.1	20.8	Wemaere, 2005
DRU	2000/1	8; 8	2	20.1	21.4	96.5	119.0	7.4	51.8	Wemaere, 2005
DUN	NA	NA	NA	59.6	1.8	33.0	100.0	5.7	6.0	EPA CRL 2003
EAS	NA	NA	NA	4.0	2.9	32.0	48.0	6.5	7.0	EPA CRL 2003
EFF	2000/1	8; 8	2	22.4	10.0	69.0	164.5	7.4	30.1	Wemaere, 2005
EGI	1996/7	6,7,9;1,4,6,7,9,12	9	69.0	35.0	25.0	229.0	8.1	344.0	Irvine et al., 2001
FAD	NA	NA	NA	5.0	1.4	53.0	80.9	6.4	7.0	EPA CRL 2003
FEE	NA	NA	NA	3.1	0.9	27.0	62.0	6.6	9.0	EPA CRL 2003
FEG	NA	NA	NA	9.6	1.3	54.0	86.0	7.4	8.0	EPA CRL 2003
GAR	2000/1	8; 8	2	67.4	13.3	50.0	197.0	7.6	76.4	Wemaere, 2005
GOR	2000/1	8; 8	2	88.1	20.4	50.0	229.0	7.9	55.5	Wemaere, 2005
HOH	1996/7	3,4,5,6,7,8,9,10;1,3, 4,5,6,7,8,9,10,12	18	17.8	5.0	33.0	125.0	7.2	10.0	Irvine et al., 2001
INC	2001	4, 6, 8, 9	4	131.7	8.4	36.3	294.0	8.0	21.9	Wemaere, 2005
INQ	2001	1,4,5,6,7,8,9,10	8	161.8	5.1	23.0	334.0	8.2	19.3	Wemaere, 2005
KEE	NA	NA	NA	2.4	2.1	47.0	135.0	5.3	8.0	EPA CRL 2003
KIL	NA	NA	NA	27.4	1.4	33.0	205.0	7.2	14.0	EPA CRL 2003

Site Code	Year Sampled	Month Sampled	No. of Samples	Alkalinity (mg $L^{-1}CaCO_3$)	Chlorophyll a ($\mu g\ l^{-1}$)	Colour (mg l^{-1} PtCo/Hazen)	Conductivity ($\mu S\ cm^{-1}$)	pH	TP ($\mu g\ l^{-1}$)	Data Source
KIN	NA	NA	NA	69.5	5.5	26.0	318.0	8.3	11.0	EPA CRL 2003
KYL	NA	NA	NA	7.0	0.6	21.0	72.0	6.6	6.0	EPA CRL 2003
LEN	NA	NA	NA	104.9	3.4	4.0	250.0	8.5	6.1	EPA CRL 2003
LIC	1996/7	3,4,5,6,7,8,9,10;1,3,4,5,6,7,8,9,10,12	18	21.0	13.1	57.0	157.0	7.5	16.0	Irvine et al., 2001
LIS	2001	6,8,9,10	4	47.5	14.7	36.0	304.0	7.8	34.7	Wemaere, 2005
MAU	NA	NA	NA	6.3	1.3	18.0	72.0	6.1	5.0	EPA CRL 2003
MCN	NA	NA	NA	23.6	6.9	80.0	116.0	7.6	17.0	EPA CRL 2003
MOA	2001	6,8,9,10	4	15.5	13.5	208.5	156.8	7.1	44.8	Wemaere, 2005
MOO	2000/1	8; 8	2	46.8	13.1	87.0	176.5	7.4	48.7	Wemaere, 2005
MOR	2000/1	8; 8	2	43.0	16.7	115.0	171.5	7.5	142.3	Wemaere, 2005
MUC	NA	NA	NA	208.6	0.8	26.0	462.0	8.5	4.8	EPA CRL 2003
MUL	1996/7	4,6,7,9;1,4,6,7,9,12	10	52.0	32.5	22.0	171.0	7.8	55.0	Irvine et al., 2001
MUN	1996/7	4,6,7,9;1,4,6,7,9,12	10	45.0	12.7	33.0	213.0	7.8	33.0	Irvine et al., 2001
NAB	NA	NA	NA	2.3	0.5	43.0	101.0	6.0	10.0	EPA CRL 2003
NAH	NA	NA	NA	9.6	1.3	37.0	100.8	6.5	7.0	EPA CRL 2003
NAM	NA	NA	NA	7.0	0.6	40.0	112.0	6.6	10.0	EPA CRL 2003
NAN	NA	NA	NA	0.7	3.8	50.0	77.0	6.0	7.6	EPA CRL 2003
OFL	NA	NA	NA	138.9	0.8	63.0	333.0	8.5	10.1	EPA CRL 2003
OOR	NA	NA	NA	8.1	1.4	23.0	65.0	6.4	7.0	EPA CRL 2003
OUG	1996/7	3,4,6,7,9;1,4,6,7,9,12	11	70.0	20.3	49.0	233.0	7.9	72.0	Irvine et al., 2001
OWE	1996/7	3,4,5,6,7,8,9,10;1,3,4,5,6,7,8,9,10,12	18	96.0	6.3	6.0	254.0	8.3	10.0	Irvine et al., 2001
POL	1996/7	4,6,7,9;1,4,6,7,9,12	10	21.0	5.2	73.0	86.0	7.5	8.0	Irvine et al., 2001
RAM	1996/7	3,4,5,6,7,8,9,10;1,3,4,5,6,7,8,9,10,12	18	50.0	58.1	49.0	194.0	8.0	88.0	Irvine et al., 2001
REA	NA	NA	NA	128.5	2.4	3.0	308.0	8.5	6.1	EPA CRL 2003
ROS	2001	8	1	25.8	62.7	111.0	107.0	7.9	75.2	Wemaere, 2005
RUS	2000	8	1	69.7	18.5	76.0	206.0	7.9	31.3	Wemaere, 2005
SHI	NA	NA	NA	6.2	1.5	21.0	73.0	6.5	4.0	EPA CRL 2003
SIL	NA	NA	NA	140.0	4.5	36.0	354.0	8.3	141.0	Taylor et al., 2006
TAL	NA	NA	NA	85.1	1.9	15.0	190.0	8.0	8.0	EPA CRL 2003
TAY	NA	NA	NA	-0.3	0.6	134.0	40.0	5.1	8.1	EPA CRL 2003
TUL	2000	8	1	13.8	6.1	102.0	165.0	7.1	99.5	Wemaere, 2005
UPE	NA	NA	NA	2.8	1.8	22.0	58.0	6.4	4.7	EPA CRL 2003
VEA	NA	NA	NA	2.2	1.7	33.0	33.0	6.3	0.0	EPA CRL 2003

Appendix E: Diatom List

The common 233 diatom taxa with maximum abundance at least 1% and occurrences of at least 3 sites were listed with authorities for the 72-lake diatom training set.

Taxon Code	Taxon Name and Authority
AAEQ	Amphora aequalis Krammer
AAMB	Aulacoseira ambigua (Grun.) Simonsen
AATG	Achnanthidium alteragracillima (Lange-Bertalot)Round & Bukhtiyarova
ACNP	Achnanthidium pusillum (Grun.in Cl. & Grun) Czarnecki
ACON	Achnanthes conspicua A.Mayer
ACOP	Amphora copulata (Kutz) Schoeman & Archibald
ACUR	Achnanthes curtissima Carter
ADCA	Achnanthidium caledonicum(Lange-Bertalot)Lange-Bertalot
ADMF	Achnanthidium minutissima (Kütz.)Czarn.var. affinis(Grun.) Bukht.
ADMI	Achnanthidium minutissimum (Kütz.) Czarnecki
ADMS	Adlafia minuscula (Grunow) Lange-Bertalot
ADSA	Achnanthidium saprophila (Kobayasi et Mayama) Round & Bukhtiyarova
ADSU	Achnanthidium subatomus (Hustedt) Lange-Bertalot
AFOR	Asterionella formosa Hassall
AINA	Amphora inariensis Krammer
AIPX	Achnanthes impexa Lange-Bertalot
ALIO	Achnanthes linearioides Lange-Bertalot
ALIR	Aulacoseira lirata (Ehr.) Ross in Hartley
AMJA	Achnanthes minutissima Kutzing var.jackii(Rabenhorst) Lange-Bertalot
APED	Amphora pediculus (Kutzing) Grunow
APET	Achnanthes petersenii Hustedt KLB91p67f37/24-40
ARAL	Asterionella ralfsii W.Smith var. ralfsii
ATHU	Amphora thumensis (Mayer)A.Cleve-Euler
AUAL	Aulacoseira alpigena(Grunow) Krammer
AUDI	Aulacoseira distans (Ehr.)Simonsen
AUGA	Aulacoseira granulata (Ehr.) Simonsen var.angustissima (O.M.)Simonsen
AUGR	Aulacoseira granulata (Ehr.) Simonsen
AUIS	Aulacoseira islandica(O.Muller)Simonsen
AUIT	Aulacoseira italica (Ehr.)Simonsen
AUSU	Aulacoseira subarctica (O.Muller) Haworth
BBRE	Brachysira brebissonii Ross in Hartley ssp. brebissonii
BEXI	Brachysira exilis Round & Mann
BGAR	Brachysira garrensis (Lange-Bertalot & Krammer) Lange-Bertalot
BPRO	Brachysira procera Lange-Bertalot & Moser
BVIT	Brachysira vitrea (Grunow) Ross in Hartley
CAFF	Cymbella affinis Kutzing var.affinis
CAGR	Cyclotella atomus var. gracilis Genkal & Kiss
CCMS	Cyclotella comensis Grunow in Van Heurck
CCOC	Cavinula cocconeiformis (Gregory ex Greville) Mann & Stickle
CCYM	Cymbella cymbiformis Agardh
CDEL	Cymbella delicatula Kutzing
CDTG	Cyclotella distinguenda var.distinguenda Hustedt
CDUB	Cyclostephanos dubius (Fricke) Round
CGOR	Cyclotella gordonensis Kling & Hakansson
CHAL	Craticula halophila (Grunow ex Van Heurck) Mann
CHEL	Cymbella helvetica Kutzing
CHEV	Chamaepinnularia evanida (Hustedt) Lange-Bertalot
CHME	Chamaepinnularia mediocris (Krasske) Lange-Bertalot
CHSP	Chamaepinnularia sp.
CINV	Cyclostephanos invisitatus(Hohn & Hellerman)Theriot Stoermer & Hakansson
CJAR	Cavinula jaernefeltii (Hustedt) Mann & Stickle
CKRM	Cyclotella krammeri HÅkansson
CLAE	Cymbella laevis Naegeli in Kutzing var.laevis
CMEN	Cyclotella meneghiniana Kutzing
CNDI	Cocconeis neodiminuta Krammer
CNTH	Cocconeis neothumensis Krammer
COCE	Cyclotella ocellata Pantocsek
CPED	Cocconeis pediculus Ehrenberg
CPLA	Cocconeis placentula Ehrenberg var. placentula
CPOL	Cyclotella polymorpha Meyer & Hakansson
CPST	Cyclotella pseudostelligera Hustedt

Taxon Code	Taxon Name and Authority
CSAE	Cymbella subaequalis Grunow
CSCM	Cyclotella schumanni (Grunow) Håkansson
CSTR	Cyclotella striata(Kutzing)Grunow 1880 in Cleve & Grunow
CVMO	Cavinula mollicula (Hust.) Lange-Bertalot
CVSO	Cavinula scutelloides (W.Smith) Lange-Bertalot
CYDE	Cyclotella delicatula Hustedt
CYMS	Cymbella species
DELL	Diploneis elliptica (Kutzing) Cleve
DITE	Diatoma tenuis Agardh
DOBL	Diploneis oblongella (Naegeli) Cleve-Euler
DOVA	Diploneis ovalis (Hilse) Cleve
DPET	Diploneis peterseni Hustedt
DTEN	Denticula tenuis Kutzing
EARC	Eunotia arcus Ehrenberg var. arcus
EBIL	Eunotia bilunaris (Ehr.) Mills var. bilunaris
ECES	Encyonopsis cesatii (Rabenhorst) Krammer
ECPM	Encyonopsis minuta Krammer & Reichardt
EELE	Eunotia elegans Oestrup
EEXI	Eunotia exigua (Brebisson ex Kützing) Rabenhorst
EFAB	Eunotia faba Grunow
EGAE	Encyonema gaeumanii (Meister) Krammer
EGOE	Epithemia goeppertiana Hilse
EHEB	Encyonema hebridicum Grunow ex Cleve
EIMP	Eunotia implicata Nörpel, Lange-Bertalot & Alles
EINC	Eunotia incisa Gregory var.incisa
EMBI	Eunotia monodon Ehrenberg var.bidens (Gregory) Hustedt
EMIN	Eunotia minor (Kutzing) Grunow in Van Heurck
ENCM	Encyonopsis microcephala (Grunow) Krammer
ENNG	Encyonema neogracile Krammer
ENPE	Encyonema perpusillum (A. Cleve) D.G. Mann
EOMI	Eolimna minima(Grunow) Lange-Bertalot
EPEC	Eunotia pectinalis (Dyllwyn) Rabenhorst var.pectinalis
EPUN	Eunotia pectinalis(Kutz.)Rabenhorst var.undulata (Ralfs) Rabenhorst
ERHO	Eunotia rhomboidea Hustedt
ERHY	Eunotia rhynchocephala Hustedt var. rhynchocephala
ESLE	Encyonema silesiacum (Bleisch in Rabh.) D.G. Mann
ESMI	Epithemia smithii Carruthers 1864
ESPP	Encyonema subperpusillum Krammer
ESUB	Eunotia subarcuatoides Alles Nörpel & Lange-Bertalot
EUAL	Eucocconeis alpestris(Brun) Lange-Bertalot
EUPA	Eunotia paludosa Grunow in Van Heurck var. paludosa
FCAP	Fragilaria capucina Desmazieres var.capucina
FCME	Fragilaria capucina Desmazieres var.mesolepta (Rabenhorst) Rabenhorst
FCRO	Fragilaria crotonensis Kitton
FCRP	Fragilaria capucina Desm. var. rumpens (Kütz.) Lange-Bert. ex Bukht.
FCVA	Fragilaria capucina Desmazieres var.vaucheriae(Kutzing)Lange-Bertalot
FDEL	Fragilaria delicatissima (W.Smith) Lange-Bertalot
FERI	Frustulia erifuga Lange-Bertalot & Krammer
FGRA	Fragilaria gracilis Østrup
FLAP	Fragilaria lapponica Grunow in van Heurck (Staurosirella)
FLEN	Fallacia lenzi(Hustedt) Van de Vijver & al. nov. comb.
FNAN	Fragilaria nanana Lange-Bertalot
FOLD	Fragilaria oldenburgiana Hustedt
FPCO	Fragilaria pseudoconstruens Marciniak
FPLA	Fragilaria pinnata Ehrenberg var.lancettula (Schumann) Hustedt
FRHO	Frustulia rhomboides(Ehr.)De Toni
FROB	Fragilaria robusta (Fusey) Manguin
FSAX	Frustulia saxonica Rabenhorst
FSPP	Fragilaria sp.
FTEN	Fragilaria tenera (W.Smith) Lange-Bertalot
FUAC	Fragilaria ulna(Nitzsch.)Lange-Bertalot var.acus(Kutz.)Lange-Bertalot
FUAN	Fragilaria ulna Sippen angustissima(Grun.)Lange-Bertalot
FVIR	Fragilaria virescens Ralfs
GACU	Gomphonema acuminatum Ehrenberg
GANT	Gomphonema angustum Agardh
GBAV	Gomphonema bavaricum Reichardt & Lange-Bertalot
GEXL	Gomphonema exilissimum(Grun.) Lange-Bertalot & Reichardt
GGRA	Gomphonema gracile Ehrenberg
GHEB	Gomphonema hebridense Gregory

Taxon Code	Taxon Name and Authority
GLAT	Gomphonema lateripunctatum Reichardt & Lange-Bertalot
GMCU	Gomphonema minutum f.curtum (Hustedt) Lange-Bertalot & Reichardt
GOLI	Gomphonema olivaceum (Hornemann) Brébisson var. olivaceum
GOMS	Gomphonema species
GPAR	Gomphonema parvulum (Kützing) Kützing var. parvulum f. parvulum
GPRC	Gomphonema procerum Reichardt & Lange-Bertalot
GPUM	Gomphonema pumilum (Grunow) Reichardt & Lange-Bertalot
GPVL	Gomphonema parvulius Lange-Bertalot & Reichardt
GSUB	Gomphonema subtile Ehr.
KCLE	Karayevia clevei(Grun. in Cl. & Grun.) Round & Bukhtiyarova
KLAT	Karayevia laterostrata(Hust.) Kingston
KOSU	Kobayasiella subtilissima (Cleve) Lange-Bertalot
KSUC	Kolbesia suchlandtii (Hustedt) Kingston
MAAT	Mayamaea atomus (Kutzing) Lange-Bertalot
MAGR	Mayamaea agrestis(Hustedt) Lange-Bertalot
MCCO	Meridion circulare (Greville) Agardh var.constrictum (Ralfs) Van Heurck
MCIR	Meridion circulare (Greville) C.A.Agardh var. circulare
MELL	Mastogloia elliptica (C.A. Agardh) Cleve
MLAC	Mastogloia lacustris (Grunow) van Heurck
MSMI	Mastogloia smithii Thwaites
NARV	Navicula arvensis Hustedt
NBCL	Nitzschia bacillum Hustedt
NCAR	Navicula cari Ehrenberg
NCRY	Navicula cryptocephala Kutzing
NCTE	Navicula cryptotenella Lange-Bertalot
NCTO	Navicula cryptotenelloides Lange-Bertalot
NDEN	Nitzschia denticula Grunow
NDPA	Navicula parabryophila Lange-Bertalot
NELO	Naviculadicta elorantana Lange-Bertalot
NESP	Neidium species in Metzeltin & Lange Bertalot
NGRE	Navicula gregaria Donkin
NHMD	Navicula heimansioides Lange-Bertalot
NIAR	Nitzschia archibaldii Lange-Bertalot
NIFR	Nitzschia frustulum(Kutzing)Grunow var.frustulum
NIGR	Nitzschia gracilis Hantzsch
NILA	Nitzschia lacuum Lange-Bertalot
NIPM	Nitzschia perminuta(Grunow) M.Peragallo
NIVA	Nitzschia valdestriata Aleem & Hustedt
NLCN	Navicula lucinensis Hustedt
NLST	Navicula leptostriata Jorgensen
NPAE	Nitzschia paleacea (Grunow) Grunow in van Heurck
NPAL	Nitzschia palea (Kutzing) W.Smith
NPHP	Navicula phylleptosoma Lange-Bertalot
NPSL	Navicula pseudolanceolata Lange-Bertalot
NRAD	Navicula radiosa Kützing
NRCS	Navicula recens (Lange-Bertalot) Lange-Bertalot
NREC	Nitzschia recta Hantzsch in Rabenhorst
NRFA	Navicula radiosafallax Lange-Bertalot
NRHY	Navicula rhynchocephala Kutzing
NSBR	Navicula subrotundata Hustedt
NSMU	Navicula submuralis Hustedt
NSPP	Navicula seippiana Lange-Bertalot & Steindorf
NSUA	Nitzschia subacicularis Hustedt in A.Schmidt et al.
NTPT	Navicula tripunctata (O.F.Müller) Bory
NVDS	Navicula(dicta) seminulum (Grunow) Lange Bertalot
NVEN	Navicula veneta Kutzing
NVIO	Navicula vitiosa Schimanski
PALT	Psammothidium altaicum Bukhtiyarova
PBBI	Pseudostaurosira brevistriata Grun.v.binodis(Pant.)Andresen Stoermer&Kreis
PCHL	Psammothidium chlidanos (Hohn & Hellerman) Lange-Bertalot
PCLT	Placoneis clementis (Grun.) Cox
PFIB	Peronia fibula (Breb.ex Kutz.)Ross
PHEL	Psammothidium helveticum (Hustedt) Bukhtiyarova et Round
PIRR	Pinnularia irrorata (Grunow) Hustedt
PLFR	Planothidium frequentissimum(Lange-Bertalot)Lange-Bertalot
PLVD	Psammothidium levanderi (Hustedt)Czarnecki in Czarn. et Edlund
PMAJ	Pinnularia maior (Kutzing) Rabenhorst
POBG	Psammothidium oblongellum(Oestrup) Van de Vijver
PPSW	Psammothidium pseudoswazi (Carter) Bukht. et Round

Taxon Code	Taxon Name and Authority
PRAD	Puncticulata radiosa (Lemmermann) Håkansson
PSAC	Psammothidium sacculum (Carter) Bukhtiyarova et Round
PSAT	Psammothidium subatomoides (Hustedt) Bukht.et Round
PSBR	Pseudostaurosira brevistriata (Grun.in Van Heurck) Williams & Round
PSCA	Pinnularia subcapitata Gregory var. subcapitata
PSIL	Pinnularia silvatica Petersen
PTDE	Planothidium delicatulum(Kutz.) Round & Bukhtiyarova
PTHA	Planothidium hauckianum (Grun.) Round & Bukhtiyarova
PTLA	Planothidium lanceolatum(Brebisson ex Kützing) Lange-Bertalot
PTOE	Planothidium oestrupii(Cleve-Euler)Round & Bukhtiyarova
PUCO	Puncticulata comta(Ehr.)Håkansson
RABB	Rhoicosphenia abbreviata (C.Agardh) Lange-Bertalot
SAGA	Stephanodiscus agassizensis Hakansson & Kling
SALP	Stephanodiscus alpinus Hustedt in Huber-Pestalozzi
SCON	Staurosira construens Ehrenberg
SCVE	Staurosira construens Ehr. var venter (Ehr.) Hamilton
SDPA	Synedrella parasitica (W.Sm.) Round & Maidana
SDSU	Synedrella subconstricta (Grunow in Van Heurck) Round & Maidana
SELI	Staurosira elliptica (Schumann) Williams & Round
SEXG	Stauroforma exiguiformis Flower Jones et Round
SHAN	Stephanodiscus hantzschii Grunow in Cl. & Grun. 1880
SHTE	Stephanodiscus hantzschii fo.tenuis(Hustedt)Hakansson et Stoermer
SLEP	Staurosirella leptostauron (Ehr.) Williams & Round
SMAT	Staurosira martyi (Heribaud) Lange-Bertalot
SMED	Stephanodiscus medius Håkansson
SNEO	Stephanodiscus neoastraea Hakansson et Hickel
SPAV	Stephanodiscus parvus Stoermer et Hakansson
SPIN	Staurosirella pinnata (Ehr.) Williams & Round
SPUP	Sellaphora pupula (Kutzing) Mereschkowksy
STMI	Stephanodiscus minutulus (Kutzing) Cleve & Moller
STPI	Staurosirella pinnata var. intercedens (Grunow in V.Heurck) Hamilton
TFLO	Tabellaria flocculosa(Roth)Kutzing
TQUA	Tabellaria quadriseptata Knudson
UULN	Ulnaria ulna (Nitzsch.) Compère

Appendix F: Cladocera List

Species names with authorities are listed in the order of the taxon codes used in the 33-lake Cladocera training set.

Taxon Code	Taxon Name	Authority
ACRHAR	*Acroperus harpae*	(Baird, 1834)
ALOAFF	*Alona affinis*	(Leydig, 1860)
ALOCOS	*Alona costata*	G.O. Sars, 1862
ALOELO	*Alonopsis elongata*	(G.O. Sars, 1862)
ALOEXC	*Alonella excisa*	(Fischer, 1854)
ALOEXI	*Alonella exigua*	(Lilljeborg, 1853)
ALOGR	*Alona guttata/rectangula*	G.O. Sars, 1862/ G.O. Sars, 1862
ALOINT	*Alona intermedia*	G.O. Sars, 1862
ALONAN	*Alonella nana*	(Baird, 1843)
ALOQUA	*Alona quadrangularis*	(O.F. Müller, 1776)
ALORUS	*Alona rustica*	Scott, 1895
ANCEMA	*Anchistropus emarginatus*	G.O. Sars, 1862
BOSCOR	*Bosmina coregoni*	(Baird, 1857)
BOSLOR	*Bosmina longirostris*	(O.F. Müller, 1776)
BOSLOS	*Bosmina longispina*	(Leydig, 1860)
CAMREC	*Camptocercus rectirostris*	Schödler, 1862
CHYPIG	*Chydorus piger*	G.O. Sars, 1862
CHYSPH	*Chydorus Sphaericus*	(O.F. Müller, 1776)
DAPLOG	*Daphnia longispina group*	
DAPPUG	*Daphnia pulex group*	
EURLAM	*Eurycercus lamellatus*	(O.F. Müller, 1776)
GRATES	*Graptoleberis testudinaria*	(S. Fischer, 1848)
HOLGIB	*Holopedium gibberum*	Zaddach, 1855
ILYSIL	*Ilyocryptus silvaeducensis*	Romijn, 1919
KURLAT	*Kurzia latissima*	(Kurz, 1875)
LATSET	*Latona setifera*	(O.F. Müller , 1785)
LEPKIN	*Leptodora kindti*	Focke, 1844
LEYACA	*Leydigia acanthocercoides*	(S. Fischer, 1854)
LEYLEY	*Leydigia leydigii*	(Schödler, 1862)
MONDIS	*Monospilus dispar*	G.O. Sars, 1862
OXYTEN	*Oxyurella tenuicaudis*	(G.O. Sars, 1862)
PHRROS	*Phrixura rostrata*	(Koch, 1844)
PLEADU	*Pleuroxus aduncus*	(Jurine, 1820)
PLEDEN	*Pleuroxus denticulatus*	Birge, 1879
PLELAE	*Pleuroxus laevis*	G.O. Sars, 1862
PLESP	*Pleuroxus sp.*	
PLETRI	*Pleuroxus trigonellus*	(O.F. Müller, 1785)
PLETRU	*Pleuroxus truncatus*	(O.F. Müller, 1785)
PLEUNC	*Pleuroxus uncinatus*	Baird, 1850
POLPED	*Polyphemus pediculus*	(Linnaeus, 1758)
RHYFAL	*Rhynchotalona falcata*	(G.O. Sars, 1862)
SCAMUC	*Scapholeberis mucronata*	(O.F. Müller, 1776)
SIDCRY	*Sida crystallina*	(O.F. Müller, 1776)
TREAMB	*Tretocephala ambigua*	(Lilljeborg, 1900)

Appendix G: TP Optima of Diatom Taxa

Weighted averaging TP optima and tolerance of the 233 diatom taxa (≥ 1% at 3 sites) in the 72-lake training set are listed in the order of species name (number of occurrence (count), maximum abundance and Hill's effective number of occurrence (N2) of each taxon are also shown; both TP optima and tolerance are back-transformed to μg l^{-1} units).

Taxon	Count	Max (%)	N2	TP (μg l^{-1})	Tolerance (μg l^{-1})
Achnanthes altaica	13	4.62	6.39	7.92	1.35
Achnanthes biasolettiana var. *atomus*	5	1.29	4.12	8.72	1.48
Achnanthes chlidanos	3	1.10	2.23	7.06	1.28
Achnanthes clevei	11	5.21	4.76	21.60	1.88
Achnanthes conspicua	9	2.95	5.11	15.17	2.28
Achnanthes delicatula	8	1.53	4.88	14.21	2.29
Achnanthes delicatula var. *hauckiana*	5	1.57	4.44	7.89	1.67
Achnanthes flexella var. *alpestris*	4	1.59	2.16	7.13	1.32
Achnanthes helvetica	13	34.30	2.16	12.32	2.01
Achnanthes lanceolata	21	2.90	12.48	29.94	2.37
Achnanthes lanceolata var. *frequentissima*	4	1.49	3.08	10.64	1.59
Achnanthes laterostrata	12	12.50	2.24	8.96	1.45
Achnanthes levanderii	11	2.30	5.97	9.52	1.83
Achnanthes linearis	3	2.99	1.61	12.12	1.61
Achnanthes minutissima var. *affinis*	17	4.42	8.49	7.16	1.49
Achnanthes minutissima var. *gracillima*	5	3.28	3.94	6.56	1.38
Achnanthes minutissima var. *jackii*	12	2.39	9.07	6.76	1.40
Achnanthes minutissima var. *minutissima*	68	30.15	37.11	14.62	2.42
Achnanthes minutissima var. *saprophila*	12	2.69	7.47	7.92	1.70
Achnanthes oblongella	16	4.36	8.98	11.06	2.77
Achnanthes oestrupii	4	3.62	1.67	14.28	1.59
Achnanthes petersenii	8	4.50	3.56	7.72	1.63
Achnanthes pseudoswazii	9	3.79	3.88	7.39	1.36
Achnanthes pusilla	27	5.80	9.48	15.83	2.77
Achnanthes scotica	20	6.18	9.65	6.59	1.56
Achnanthes sp cf curtissima	3	1.23	2.67	5.75	1.35
Achnanthes sp cf saccula	10	5.64	3.42	7.16	1.94
Achnanthes subatomoides	19	6.78	7.80	12.88	2.21
Achnanthes suchlandtii	10	2.60	5.01	18.02	2.73
Amphora aequalis	3	1.49	2.65	7.93	1.31
Amphora inaeriensis	17	6.44	9.29	8.95	1.98
Amphora lybica	24	1.19	16.85	12.91	2.47
Amphora pediculus	30	23.99	13.06	10.91	1.89
Amphora thumensis	8	1.87	5.39	7.96	1.41
Anomoeoneis garrensis	18	4.94	10.24	6.53	1.33
Anomoeoneis neoexilis	26	6.16	16.76	6.74	1.30
Anomoeoneis procera	13	2.09	9.70	6.01	1.34
Anomoeoneis vitrea	10	15.30	2.24	6.64	2.23
Asterionella formosa	40	37.70	16.97	36.02	2.61
Asterionella ralfsii	11	64.72	3.07	7.74	1.24
Aulacoseira alpigena	9	4.90	4.23	13.35	2.62
Aulacoseira ambigua	20	21.98	7.99	22.66	3.19
Aulacoseira distans	6	1.10	4.96	32.11	3.12
Aulacoseira granulata	11	24.60	5.86	35.07	1.99
Aulacoseira granulata var. *angustissima*	3	40.83	1.98	77.91	1.86
Aulacoseira islandica var. *islandica*	12	18.40	5.37	33.11	2.16
Aulacoseira italica	2	0.90	1.60	36.49	1.32

Taxon	Count	Max (%)	N2	TP (µg l^{-1})	Tolerance (µg l^{-1})
Aulacoseira lirata	3	3.39	1.37	7.43	1.35
Aulacoseira subarctica	30	67.00	14.99	29.10	2.29
Brachysira brobissonii	21	3.45	10.36	7.19	1.51
Cocconeis neodiminuta	3	1.58	2.23	13.31	1.38
Cocconeis neothumensis	8	10.42	3.21	10.12	1.61
Cocconeis pediculus	4	1.99	2.27	22.45	1.53
Cocconeis placentula	43	26.00	11.61	27.30	2.11
Cyclostephanos dubius	18	19.91	5.95	29.31	1.98
Cyclostephanos invisitatus	11	9.30	4.59	72.18	1.74
Cyclotella aff schumannii	2	1.08	1.99	7.00	1.91
Cyclotella atomus var. *gracilis*	3	6.67	1.86	11.48	3.87
Cyclotella comensis	27	35.74	13.60	7.91	1.45
Cyclotella comta	12	16.80	3.04	5.63	2.23
Cyclotella delicatula	5	2.19	2.81	5.03	1.25
Cyclotella distinguenda	9	11.01	2.02	7.31	2.24
Cyclotella gordonensis	6	28.38	2.13	6.24	1.36
Cyclotella kuetzingiana	19	24.04	8.27	6.27	1.44
Cyclotella kuetzingiana cf striata	3	7.26	2.12	7.07	1.18
Cyclotella kuetzingiana cf polymorpha	4	2.44	2.33	6.73	1.83
Cyclotella meneghiniana	18	10.70	4.95	35.22	1.86
Cyclotella ocellata	6	7.01	2.05	8.46	1.77
Cyclotella pseudostelligera	30	54.59	4.76	12.65	3.15
Cyclotella radiosa	43	11.33	22.41	10.86	2.15
Cymbella affinis	13	1.85	9.99	9.35	2.43
Cymbella cesattii	14	2.23	10.32	7.38	1.72
Cymbella cymbiformis	7	1.77	5.45	7.86	1.47
Cymbella delicatula	7	2.52	4.55	6.57	1.29
Cymbella gaeumannii	6	4.50	3.56	6.09	1.34
Cymbella gracilis	32	9.57	14.86	7.89	1.77
Cymbella hebridica	6	1.27	4.67	6.89	1.21
Cymbella helvetica	6	8.59	1.84	6.92	1.41
Cymbella laevis var. *capitata*	6	5.66	4.68	6.15	1.32
Cymbella microcephala	27	7.32	16.80	7.44	1.57
Cymbella minuta	32	2.36	20.18	14.67	2.84
Cymbella perpusilla	13	4.03	7.66	7.48	1.31
Cymbella silesiaca	25	3.60	15.66	13.12	2.59
Cymbella spp	9	1.84	6.39	7.13	1.47
Cymbella subaequalis	5	1.23	3.12	6.42	1.43
Denticula tenuis	12	6.15	6.46	7.86	1.40
Diatoma tenuis	13	4.00	5.86	43.72	2.35
Diploneis elliptica	9	4.96	3.42	6.19	1.50
Diploneis oblongella	15	4.26	5.35	12.09	3.03
Diploneis ovalis	7	1.19	4.91	9.05	1.49
Diploneis petersenii	7	3.07	2.34	5.31	1.40
Epithemia muelleri	5	1.89	2.71	5.41	1.10
Epithemia smithii	3	1.49	2.13	10.03	1.44
Eunotia arcus	10	3.28	5.27	7.28	1.58
Eunotia bilunaris	22	2.13	13.24	13.94	2.75
Eunotia elegans	3	1.23	1.86	6.11	1.52
Eunotia exigua	15	3.20	8.63	8.38	1.72
Eunotia faba	11	4.43	5.16	10.36	2.99
Eunotia implicata	29	2.69	18.56	8.32	1.91
Eunotia incisa	35	11.53	14.27	8.24	1.79
Eunotia minor	10	1.70	6.89	9.44	1.69
Eunotia monodon var. *bidens*	3	2.87	1.38	6.93	1.29

Taxon	Count	Max (%)	N2	TP (µg l^{-1})	Tolerance (µg l^{-1})
Eunotia paludosa	7	1.10	5.33	9.78	1.51
Eunotia pectinalis	3	1.80	2.57	7.24	2.00
Eunotia pectinalis var. *undulata*	16	7.77	8.13	7.33	1.60
Eunotia rhomboidea	16	3.55	10.48	7.99	1.32
Eunotia rhynchocephala var. *rhynchocephala*	5	1.30	3.37	6.50	1.53
Eunotia sp cf subarcuatoides	6	2.46	2.54	5.64	1.27
Eunotia sp.	12	4.20	4.69	12.92	2.64
Fragilaria brevistriata	34	13.73	12.05	8.90	2.05
Fragilaria brevistriata var. *binodis*	6	6.23	1.68	9.62	2.78
Fragilaria capucina	21	8.85	9.78	21.25	2.64
Fragilaria capucina var. *gracilis*	31	9.80	14.95	25.62	2.48
Fragilaria capucina var. *mesolepta*	7	3.80	4.46	53.29	2.22
Fragilaria capucina var. *rumpens*	10	6.20	4.90	25.27	2.87
Fragilaria construens f. *construens*	15	2.90	8.61	29.28	2.11
Fragilaria construens f. *venter*	40	60.90	8.01	24.91	2.70
Fragilaria crotonensis	6	15.85	2.11	9.30	3.47
Fragilaria delicatissima	4	6.80	1.24	10.73	3.08
Fragilaria elliptica	13	6.78	4.66	16.62	4.32
Fragilaria exigua	41	25.80	17.13	9.71	2.11
Fragilaria lapponica	5	3.94	2.91	6.24	1.21
Fragilaria leptostauron var. *leptostauron*	9	1.23	7.01	9.06	1.94
Fragilaria leptostauron var. *martyi*	4	8.36	2.16	6.76	2.13
Fragilaria nanana	15	9.60	4.14	26.07	3.64
Fragilaria oldenburgiana	7	2.02	4.37	14.52	2.34
Fragilaria parasitica	13	1.42	8.24	13.25	2.96
Fragilaria parasitica f. *subconstricta*	3	6.00	1.20	97.80	1.14
Fragilaria pinnata	34	14.20	13.31	23.11	2.40
Fragilaria pinnata var. *intercedens*	3	1.23	2.64	6.72	1.66
Fragilaria pinnata var. *lancetttula*	3	1.79	1.99	5.12	1.13
Fragilaria robusta	5	1.80	3.04	40.41	3.90
Fragilaria sp.	10	2.30	6.69	16.24	2.74
Fragilaria tenera	6	3.70	1.96	22.50	2.18
Fragilaria ulna var. *acus*	3	2.10	1.95	73.83	1.75
Fragilaria ulna var. *angustissima*	5	22.70	2.31	64.26	2.01
Fragilaria vaucheriae	28	4.40	16.68	26.62	2.56
Fragilaria virescens	7	23.06	1.68	6.87	1.41
Frustulia erifuga	5	1.62	3.72	7.61	1.13
Frustulia rhomboides	28	8.30	12.01	7.19	1.48
Frustulia saxonica	20	6.69	11.97	7.31	1.32
Gomphonema acuminatum	17	1.59	9.96	16.73	2.90
Gomphonema angustum	5	2.11	2.89	8.92	2.22
Gomphonema bavaricum	3	1.82	2.31	10.17	1.53
Gomphonema gracile	14	1.18	10.30	9.38	2.26
Gomphonema hebridense	5	1.19	3.06	5.91	1.33
Gomphonema lateripunctatum	16	10.61	9.35	7.21	1.34
Gomphonema minutum	9	3.70	7.61	32.81	2.00
Gomphonema olivaceum var. *olivaceum*	11	1.60	7.41	31.31	1.81
Gomphonema parvulum var. *exilissimum*	16	7.09	7.24	7.46	1.36
Gomphonema parvulum var. *parvulius*	9	1.45	6.93	7.79	1.35
Gomphonema parvulum var. *parvulum*	32	5.10	18.73	34.90	2.61
Gomphonema procerum	5	3.80	2.59	9.49	1.16
Gomphonema pumilum	21	10.70	8.00	21.58	2.52
Gomphonema spp	3	1.80	2.01	5.96	2.66
Gomphonema subtile	3	1.26	2.32	8.49	1.97
Mastogloia elliptica	3	9.82	1.19	6.74	1.35

Taxon	Count	Max (%)	N2	TP (µg l^{-1})	Tolerance (µg l^{-1})
Mastogloia lacustris	8	13.33	5.05	5.82	1.13
Mastogloia smithii	5	4.36	3.14	7.84	1.55
Meridion circulare	5	8.30	1.28	34.82	2.67
Meridion circulare var. *constrictum*	3	1.18	1.94	8.60	1.37
Navicula agretis	5	1.30	3.68	23.00	2.09
Navicula arvensis var arvensis	4	1.35	3.10	9.68	1.86
Navicula atomus	10	3.00	5.31	14.13	2.47
Navicula cari	6	4.60	1.74	26.01	1.93
Navicula cf Chamaepinnularia	6	1.18	4.65	10.91	3.18
Navicula clementis	6	1.00	4.63	36.74	1.39
Navicula cocconeiformis	14	2.16	6.91	8.65	1.53
Navicula cryptocephala	16	4.90	7.14	35.45	1.83
Navicula cryptotenella	19	3.78	7.54	24.92	3.17
Navicula cryptotenelloides	7	1.79	4.54	10.90	1.72
Navicula elorantana	3	2.16	2.28	5.01	1.10
Navicula evanida	3	1.20	1.68	90.08	1.39
Navicula gregaria	7	4.50	1.82	32.77	2.52
Navicula halophila	7	1.10	3.98	24.93	2.23
Navicula heimansioides	12	1.85	9.06	7.98	1.53
Navicula impexa	10	4.80	4.28	24.51	2.53
Navicula jaernefeldtii	3	1.19	2.47	11.81	1.22
Navicula lenzii	3	1.18	2.58	6.50	1.55
Navicula leptostriata	5	6.60	2.07	7.22	2.10
Navicula lucinensis	3	3.07	2.12	5.09	1.17
Navicula mediocris	13	3.20	6.42	7.16	1.40
Navicula minima	21	3.20	12.53	13.52	2.62
Navicula minuscula	8	3.20	3.44	41.37	1.72
Navicula mollicula	8	1.59	5.43	7.61	1.43
Navicula parabryophila	3	1.85	2.45	7.73	1.30
Navicula phylleptosoma	5	1.60	2.12	39.12	1.58
Navicula pseudoconstruens	15	5.06	8.24	10.47	2.21
Navicula pseudolanceolata	5	1.18	4.16	9.45	2.81
Navicula pupula	25	3.00	11.13	29.82	3.34
Navicula radiosa	32	9.20	8.31	23.43	3.42
Navicula radiosafallax	6	1.00	4.22	44.12	2.07
Navicula rhyncocephala	15	4.60	4.23	42.71	1.83
Navicula scutelloides	7	2.38	3.46	14.87	2.16
Navicula seminulum	7	1.70	5.42	17.56	2.92
Navicula sp cf recens	6	1.20	4.42	14.79	2.62
Navicula spp	16	2.80	6.97	15.03	3.14
Navicula submuralis	12	2.47	7.79	13.46	2.83
Navicula subrotundata	4	1.65	2.86	7.21	2.78
Navicula subtilissima	9	2.60	5.57	6.64	1.53
Navicula tripunctata	9	1.10	6.29	34.56	1.59
Navicula veneta	13	2.00	8.18	42.97	1.73
Navicula vitiosa	10	2.30	6.33	11.12	2.35
Neidium sp.	4	1.70	1.75	11.25	6.07
Nitzschia archibaldii	3	1.25	2.42	17.44	4.06
Nitzschia bacillum	6	1.80	3.94	10.64	2.23
Nitzschia denticula	8	3.77	5.88	7.06	1.38
Nitzschia frustulum var frustulum	7	1.00	5.17	36.07	2.24
Nitzschia gracilis	4	4.00	1.78	31.60	2.50
Nitzschia lacuum	20	1.42	15.42	11.10	2.07
Nitzschia palea	26	2.40	18.58	33.86	2.15
Nitzschia paleacea	17	2.50	10.41	48.04	1.59

Taxon	Count	Max (%)	N2	TP (µg l^{-1})	Tolerance (µg l^{-1})
Nitzschia perminuta	17	5.40	7.00	13.36	2.48
Nitzschia recta	11	1.50	8.08	17.43	2.39
Nitzschia subacicularis	3	1.40	1.85	60.77	3.09
Nitzschia valdestriata	3	1.19	1.89	16.37	3.71
Peronia fibula	11	6.40	4.00	5.70	1.25
Pinnularia irrorata	13	4.34	7.07	8.28	1.33
Pinnularia maior	3	1.27	1.87	7.54	1.48
Pinnularia silvatica	3	1.18	2.02	6.98	1.34
Pinnularia subcapitata	9	11.20	3.15	9.26	1.67
Rhoicosphenia abbreviata	10	1.99	5.99	34.57	2.62
Stephanodiscus agassizensis	2	1.30	1.76	37.63	2.02
Stephanodiscus alpinus	13	1.77	9.99	16.37	2.35
Stephanodiscus hantzschii	18	7.06	7.25	43.06	2.24
Stephanodiscus medius	7	3.10	4.76	25.57	2.27
Stephanodiscus minutulus	13	20.80	3.10	58.32	1.77
Stephanodiscus neoastreae	13	13.12	4.84	19.73	2.83
Stephanodiscus parvus	34	33.27	11.95	46.68	2.83
Stephanodiscus tenuis	3	13.98	1.26	83.56	2.23
Synedra ulna	22	3.10	11.57	23.10	2.64
Tabellaria flocculosa	47	25.40	20.90	13.34	2.19
Tabellaria quadriseptata	5	11.27	1.58	10.39	1.53