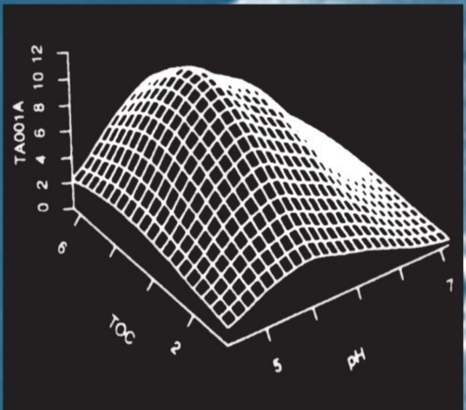
Tracking Environmental Change Using Lake Sediments

Volume 5

Data Handling and Numerical Techniques

Edited by

H. John B. Birks, André F. Lotter
Steve Juggins and John P. Smol



EXTRA MATERIALS
extras.springer.com

Springer

Tracking Environmental Change Using Lake Sediments

Developments in Paleoenvironmental Research

VOLUME 5

Aims and Scope:

Paleoenvironmental research continues to enjoy tremendous interest and progress in the scientific community. The overall aims and scope of the *Developments in Paleoenvironmental Research* book series is to capture this excitement and document these developments. Volumes related to any aspect of paleoenvironmental research, encompassing any time period, are within the scope of the series. For example, relevant topics include studies focused on terrestrial, peatland, lacustrine, riverine, estuarine, and marine systems, ice cores, cave deposits, palynology, isotopes, geochemistry, sedimentology, paleontology, etc. Methodological and taxonomic volumes relevant to paleoenvironmental research are also encouraged. The series will include edited volumes on a particular subject, geographic region, or time period, conference and workshop proceedings, as well as monographs. Prospective authors and/or editors should consult the **Series Editor John P. Smol** for more details. Any comments or suggestions for future volumes are welcomed.

For further volumes:
http://www.springer.com/series/5869

H. John B. Birks • André F. Lotter • Steve Juggins
John P. Smol
Editors

Tracking Environmental Change Using Lake Sediments

Data Handling and Numerical Techniques

Springer

Editors
H. John B. Birks
Department of Biology
University of Bergen
PO Box 7803
N-5020 Bergen
Norway

André F. Lotter
Department of Geobiology
Laboratory of Palaeobotany and Palynology
University of Utrecht
Budapestlaan 4
3584 CD Utrecht
The Netherlands

Steve Juggins
School of Geography, Politics and Sociology
Newcastle University
Daysh Building 14b
Newcastle-upon-Tyne, NE1 7 RU
United Kingdom

John P. Smol
Department of Biology
Paleoecological Environmental
Assessment and Research Laboratory
(PEARL)
Queen's University
Kingston, Ontario, K7L 3N6
Canada

Additional material to this book can be downloaded from http://extras.springer.com

ISSN 1571-5299
ISBN 978-94-007-2744-1 ISBN 978-94-007-2745-8 (eBook)
DOI 10.1007/978-94-007-2745-8
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012933106

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This book is dedicated to Cajo ter Braak,
whose work on quantifying
species-environmental relationships
underpins many of the recent advances
in quantitative palaeolimnology described
in this book

Preface

Palaeoenvironmental research has been thriving for several decades, with innovative methodologies being developed at a frenetic rate to help answer a myriad of scientific and policy-related questions. This burst in activity was the impetus for the establishment of the *Developments in Paleoenvironmental Research* (*DPER*) book series over a decade ago. The first four *DPER* volumes dealt primarily with methodologies employed by palaeolimnologists. Subsequent volumes addressed a spectrum of palaeoenvironmental applications, ranging from ice cores to dendrochronology to the study of sedimentary deposits from around the globe.

This book does not deal with the collection and synthesis of primary data, but instead it discusses the key role of data handling and numerical and statistical approaches in analysing palaeolimnological data. As summarised in our introductory chapter, palaeoenvironmental research has steadily moved from studies based on one or a few types of proxy data to large, data-rich, multi-disciplinary studies. In addition, there has been a rapid shift from simply using qualitative interpretations based on indicator species to more quantitative assessments. Although there remains an important place in palaeoenvironmental research for qualitative analyses, the reality is that many researchers now employ a wide range of numerical and statistical methodologies. It is time to review critically some of these approaches and thereby make them more accessible to the wider research community. We hope the 21 chapters making up this volume meet these goals.

Many people helped with the planning, development, and final production of this book. We would like to acknowledge the hard work and professionalism of our many reviewers, who provided constructive comments on earlier drafts of the manuscripts. We would also like to acknowledge the assistance we received from our publishers, and especially the efforts and encouragement from our main Springer colleagues—Tamara Welschot and Judith Terpos. We are grateful to Irène Hofmann for her work in the early stages of the book. We are particularly indebted to the enormous amount of work that Cathy Jenks has done in the processing and editing of the chapters, compiling and checking bibliographies and the glossary, and in the overall production of this book. Thanks are also due to our host institutions and our various funding sources, which helped facilitate our academic endeavours. We also gratefully acknowledge a variety of publishers and authors who allowed us to reproduce previously published figures. Foremost, we would like to thank the authors for their hard work and especially for their patience with the delays in completing this book. We hope that the final product was worth the wait.

Structure of the Book

This book consists of 21 chapters arranged in four parts. Part I is introductory and Chap. 1 considers the rapid development and ever-increasing application of numerical and statistical techniques in palaeolimnology. Chapter 2 provides an overview of the basic numerical and statistical approaches used in palaeolimnology in the context of data collection and assessment, data summarisation, data analysis, and data interpretation. Many of these techniques are described in more detail in chapters in Parts II and III but some important approaches such as classification, assignment and identification, and regression analysis and statistical modelling are described in Chap. 2 as they are not specifically covered elsewhere in the book. Chapter 3 describes the modern and stratigraphical data-sets that are used in some of the later chapters.

Part II considers numerical approaches that can be usefully applied to the two major types of palaeolimnological data, namely modern surface-sediment data-sets and core sediment data-sets. These approaches are exploratory data analysis and data display (Chap. 5), assessment of uncertainties associated with laboratory methods and microfossil analysis (Chap. 6), clustering and partitioning (Chap. 7), classical indirect and canonical direct ordination (Chap. 8), and a battery of techniques grouped together as statistical-learning methods in Chap. 9. These include classification and regression trees, multivariate regression trees, other types of tree-based methods, artificial neural networks and self-organising maps, Bayesian networks and genetic algorithms, principal curves, and shrinkage methods. Some other numerical techniques are not covered in these five chapters (e.g., estimating compositional turnover, richness, and species optima and tolerances, and comparing clusterings and ordinations) because the topics are not sufficiently large to warrant individual chapters. They are outlined in Chap. 4 as an introduction to Part II.

Part III contains seven chapters. They describe numerical techniques that are only applicable to the quantitative analysis of stratigraphical data-sets. Chapter 11 discusses numerical techniques for zoning or partitioning stratigraphical sequences and for detecting patterns within stratigraphical data-sets. Chapter 12 considers the essential task of establishing age-depth relationships that provide the basis for estimating rates of change and temporal patterns within and between stratigraphical sequences. Chapter 13 discusses an important but rarely used approach to core correlation by sequence-slotting. Chapter 14 discusses the quantitative reconstruction of environmental variables such as lake-water pH from, for example, fossil diatom assemblages. This general topic of environmental reconstruction has been a central focus of many palaeolimnological investigations in the last 20 years and Chap. 14 highlights the assumptions and limitations of such reconstructions, and the testing, evaluation, and validation of reconstructions. Chapter 15 considers modern analogue methods in palaeolimnology as a procedure for quantitative environmental reconstructions and for comparing fossil and modern assemblages as a tool in lake restoration and management. Chapter 16 concludes Part III by presenting new approaches to assessing temporal patterns in palaeolimnological temporal-series where the major assumptions of conventional time-series analysis are not met. Other numerical techniques such as palaeopopulation analysis, stratigraphical changes in richness, and approaches to temporal-series analysis such as LOESS smoothing and the SiZer (Significant Zero crossings of the derivative) approach and its relatives BSiZer and SiNos that are not discussed in Chaps. 11, 12, 13, 14, 15 and 16 are outlined briefly in Chap. 10, which also provides an overview and introduction to Part III.

Part IV consists of five chapters. Chapter 17 provides an introduction and overview to this Part. Three chapters (Chaps. 18, 19, 20) describe case studies where some of the numerical methods presented in Parts II and III are used to answer particular palaeolimnological research questions and to test palaeolimnological hypotheses. Chapter 18 considers limnological responses to environmental changes at inter-annual to decadal time-scales. Chapter 19 reviews the application of numerical techniques to evaluate surface-water acidification and eutrophication. Chapter 20 discusses tracking Holocene climatic change using stratigraphical palaeolimnological data and numerical techniques. The last chapter, Chap. 21, discusses eight areas of research that represent future challenges in the improved numerical analysis of palaeolimnological data.

Data-sets, figures, software, and R scripts used or mentioned in this book, links to important websites relevant to this book and its contents are available from Springer's Extras website (http://extras.springer.com).

About The Editors

H. John B. Birks is Professor in Quantitative Ecology and Palaeoecology at the Department of Biology, University of Bergen and the Bjerknes Centre for Climate Research (Norway), Emeritus Professor and Fellow at the Environmental Change Research Centre, University College London (UK), and Distinguished Visiting Fellow at the School of Geography and the Environment, University of Oxford (UK). He was Adjunct Professor in the Department of Biology at Queen's University (Kingston, Ontario, Canada) until 2011. His teaching and research include ecology, biogeography, palaeoecology, palaeolimnology, pollen analysis, numerical methods, and alpine botany.

André F. Lotter is Professor in Palaeoecology at the Institute of Environmental Biology of Utrecht University (The Netherlands), where he is head of the Laboratory of Palaeobotany and Palynology.

Steve Juggins is a Senior Lecturer and Head of Department in the School of Geography, Politics and Sociology, Newcastle University (UK), where he teaches and researches issues of aquatic pollution, diatom analysis, palaeolimnology, and quantitative palaeoecology.

John P. Smol is a Professor in the Department of Biology at Queen's University (Kingston, Ontario, Canada), where he also holds the *Canada Research Chair in Environmental Change*. He co-directs the Paleoecological Environmental Assessment and Research Lab (PEARL). John Smol was the founding editor of the international *Journal of Paleolimnology* (1987–2007) and is current editor of the journal *Environmental Reviews*, and editor of the *Developments in Paleoenvironmental Research* book series.

Contents

Contributors

N. John Anderson Department of Geography, Loughborough University, Loughborough, UK

H. John B. Birks Department of Biology and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway

Environmental Change Research Centre, University College London, London, UK

School of Geography and the Environment, University of Oxford, Oxford, UK

Maarten Blaauw School of Geography, Archaeology & Palaeoecology, Queen's University Belfast, Belfast, UK

Geoffrey S. Boulton School of GeoSciences, University of Edinburgh, Edinburgh, UK

R. Malcolm Clark School of Mathematical Sciences, Monash University, Clayton, VIC, Australia

Brian F. Cumming Paleoecological Environmental Assessment and Research Laboratory (PEARL), Department of Biology, Queen's University, Kingston, ON, Canada

Pierre Dutilleul Laboratory of Applied Statistics, Department of Plant Science, Faculty of Agricultural and Environmental Sciences, McGill University, Ste-Anne-de-Bellevue, QC, Canada

Sherylyn C. Fritz Department of Earth and Atmospheric Sciences, University of Nebraska, Lincoln, NE, USA

Roland I. Hall Department of Biology, University of Waterloo, Waterloo, ON, Canada

Einar Heegaard Department of Biology and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway

Norwegian Forest and Landscape Institute, Fana, Norway

Oliver Heiri Institute of Plant Sciences and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

Vivienne J. Jones Environmental Change Research Centre, University College London, London, UK

Steve Juggins School of Geography, Politics & Sociology, Newcastle University, Newcastle-upon-Tyne, UK

Kathleen R. Laird Paleoecological Environmental Assessment and Research Laboratory (PEARL), Department of Biology, Queen's University, Kingston, ON, Canada

Pierre Legendre Département de sciences biologiques, Université de Montréal, Montréal, QC, Canada

Melinda Lontoc-Roy Laboratory of Applied Statistics, Department of Plant Science, Faculty of Agricultural and Environmental Sciences, McGill University, Ste-Anne-de-Bellevue, QC, Canada

André F. Lotter Laboratory of Palaeobotany and Palynology, Department of Biology, University of Utrecht, Utrecht, The Netherlands

Louis J. Maher Jr. Geology and Geophysics, University of Wisconsin, Madison, WI, USA

Gavin L. Simpson Environmental Change Research Centre, University College London, London, UK

John P. Smol Paleoecological Environmental Assessment and Research Lab (PEARL), Department of Biology, Queen's University, Kingston, ON, Canada

Richard J. Telford Department of Biology and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway

Roy Thompson School of GeoSciences, University of Edinburgh, Edinburgh, UK

Dirk Verschuren Limnology Unit, Department of Biology, Ghent University, KL Ledeganckstraat 35, B-9000, Ghent, Belgium

Part I
Introduction, Numerical Overview, and Data-Sets

Chapter 1
The March Towards the Quantitative Analysis of Palaeolimnological Data

John P. Smol, H. John B. Birks, André F. Lotter, and Steve Juggins

Abstract We outline the aims of palaeolimnology and describe the major types of palaeolimnological data. The distinction between biological data derived from stratigraphical studies of cores and modern surface-sediment samples with associated environmental data is discussed. A brief history of the development of quantitative palaeolimnology is presented, starting with early applications of principal component analysis in 1975. Major developments occurred in the late 1980s, thanks to the work of Cajo ter Braak and others. The structure of the book in terms of four parts is explained. Part I is introductory and presents an overview of numerical methods and of the data-sets used. Part II presents numerical approaches

J.P. Smol (✉)
Paleoecological Environmental Assessment and Research Lab (PEARL), Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada
e-mail: smolj@queensu.ca

H.J.B. Birks
Department of Biology and Bjerknes Centre for Climate Research, University of Bergen, PO Box 7803, Bergen N-5020, Norway

Environmental Change Research Centre, University College London, Pearson Building, Gower Street, London WC1E 6BT, UK

School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK
e-mail: john.birks@bio.uib.no

A.F. Lotter
Laboratory of Palaeobotany and Palynology, Department of Biology, University of Utrecht, Budapestlaan 4, Utrecht NL-3584 CD, The Netherlands
e-mail: a.f.lotter@uu.nl

S. Juggins
School of Geography, Politics & Sociology, University of Newcastle, Newcastle-upon-Tyne NE1 7RU, UK
e-mail: stephen.juggins@ncl.ac.uk

H.J.B. Birks et al. (eds.), *Tracking Environmental Change Using Lake Sediments*, Developments in Paleoenvironmental Research 5, DOI 10.1007/978-94-007-2745-8_1,

appropriate to the analysis of modern and stratigraphical palaeolimnological data. Part III considers numerical techniques that are only applicable to stratigraphical data, and Part IV presents three case-studies and concludes with a discussion of future challenges.

Keywords Calibration • Calibration functions • Data-sets • Numerical methods • Palaeolimnology • Temporal scales • Transfer functions

Palaeolimnology

Palaeolimnology can broadly be defined as the study of the physical, chemical, and biological information stored in lake and river sediments. As such, palaeolimnology is a multi-disciplinary science with many diverse applications. The questions posed by palaeolimnologists can vary widely, ranging from applied issues, such as tracking the effects of lake acidification, eutrophication, chemical contamination, and erosion, to more fundamental scientific subjects, such as examining hypotheses regarding biogeography, evolution, natural modes of climatic change, and theoretical ecology. Common questions posed by palaeolimnologists include: Have lakes changed? If so, when and by how much? What was the cause of the change? How have species distributions and abundances changed over long time frames? Not surprisingly, given the growing interest and concern in environmental problems, and the general lack of reliable long-term monitoring data, a large portion of recent palaeolimnological research has been directed to applied issues (Smol 2008).

The overall palaeolimnological approach is relatively straightforward. The raw materials used are lake sediments[1] which, under ideal circumstances, accumulate at the bottom of a basin in an orderly and undisturbed manner. In a typical study, sediment cores can be retrieved using a variety of sampling devices, after which the sediment profiles can be sectioned into appropriate time slices, and the age-depth profile can be established using geochronological techniques (Last and Smol 2001a). Incorporated in these sediments is a diverse array of indicators and other proxy data (Last and Smol 2001b; Smol et al. 2001a, b); the palaeolimnologist's job is then to interpret these proxy data in a defendable and rigorous manner that is of interest to other scientists and the public at large.

Several recent textbooks (e.g., Cohen 2003; Smol 2008) have been published on palaeolimnology, synthesising this rapidly growing discipline. Furthermore, many palaeolimnological approaches and methods have been standardised, at least

[1] In this chapter and in this book as a whole, we will refer primarily to lake sediments; however, palaeolimnologists can also use pond, river, wetland, estuarine, and other types of sediment profiles, assuming reliable and undisturbed stratigraphic sequences can be retrieved.

at a general level (e.g., Last and Smol 2001a, b; Smol et al. 2001a, b). Over the last two decades, the amount, diversity, and quality of data generated by palaeolimnologists have been increasing steadily (Pienitz et al. 2009), with many studies producing large and complex data-sets. Parallel with these advances in data generation have been research and developments on quantifying these data in a numerically or statistically robust fashion. Not surprisingly many numerical and statistical techniques are now standard components of the palaeolimnologist's tool-kit. This book summarises some of these numerical approaches.

Types of Palaeolimnological Data

The quantity, quality, and diversity of palaeolimnological proxy data grow steadily. In the 1970s most palaeolimnological studies were largely restricted to some geochemical data, perhaps coupled with analyses of fossil diatoms and pollen grains. Today, a typical palaeolimnological study may include ten or more different types of proxy data. A scan of papers published in the international *Journal of Paleolimnology* shows a clear trend of larger and more multi-authored papers since the journal's inception in 1988.

Smol (2008) provides summaries of the commonly used palaeolimnological indicators, whilst previous volumes in this *DPER* book series (e.g., Last and Smol 2001a, b; Smol et al. 2001a, b; Battarbee et al. 2004; Francus 2004; Pienitz et al. 2004; Leng 2006) contain more detailed reviews. Sediment components are typically categorised, at least at a broad level, by their source. Allochthonous components originate from outside the lake basin, such as soil particles and pollen grains from terrestrial vegetation. Autochthonous components originate from the lake itself, including algal and aquatic invertebrates or chemical precipitates. The list of physical, chemical, and biological indicators continues to grow steadily (Pienitz et al. 2009). Amongst the biological indicators, it is true that diatom valves, chironomid head capsules, and pollen grains are most frequently used, but virtually every organism living in a lake system leaves some sort of morphological or biogeochemical (e.g., fossil pigments, lipids) indicator. These biotic indicators are either used to reconstruct past environmental conditions (e.g., lake-water pH or phosphorus, temperature, oxygen availability) in a qualitative or quantitative way or their reaction to different stressors (e.g., climate change, nutrient enrichment, heavy metal pollution) is studied using numerical methods. Similarly, a broad spectrum of inorganic and organic chemical and physical markers (e.g., metals, isotopes, persistent organic pollutants) can be used to interpret lake histories (e.g., Coolen and Gibson 2009; Francus et al. 2009; Heiri et al. 2009; Weijers et al. 2009). Although most of our examples will deal with biological indicators, numerical approaches are often equally applicable to chemical- and physical-based studies (Birks 1985, 1987; Rosén et al. 2000, 2010; Grosjean et al. 2009).

Different Temporal Scales: From Surface-Sediment Calibration Sets to Detailed Sediment-Core Studies

Palaeolimnologists use sediment cores of appropriate length (i.e., temporal range) and sectioned into time slices to answer the research questions under study (Last and Smol 2001a) and to track long-term changes within a specific lake or other aquatic system (e.g., a bay of a river). Proxy indicators (e.g., diatom valves, geochemical markers, isotopic data) contained in these sediment time-slices are isolated, identified, and counted or analysed in various ways (Last and Smol 2001b; Smol et al. 2001a, b). In studies conducted before the 1980s, most biological palaeolimnologists would have little choice but to interpret the stratigraphical changes in bio-indicators (or other types of data) qualitatively using whatever ecological data were available in the scientific literature at that time. However, the increased use of surface-sediment, modern calibration data-sets (also known as modern training-sets), beginning primarily in the late 1980s, was a major step forward in quantifying and aiding the interpretation of information preserved in sediment cores.

The concepts and assumptions underpinning surface-sediment calibration data-sets are fairly straightforward (Smol 2008), although the statistical treatment of these data is not so simple (Birks 1998, 2010; Birks et al. 2010). For example, suppose a palaeolimnologist wishes to reconstruct lake-water pH using diatoms preserved in sediment cores for a particular region. The first question one might have is how would the palaeolimnologist provide any palaeoenvironmental interpretations based on these assemblages, which may easily encompass several hundred diatom taxa, in a quantifiable and statistically defendable manner? Surface-sediment calibration sets have made this possible. A suite of calibration lakes (typically 40 or more in number) are first carefully chosen to reflect the limnological conditions that are likely to be encountered (and therefore need to be reconstructed) from the down-core sediment assemblages. For example, if this is an acidification study, and past pH may have fluctuated approximately over a pH range of about 5.5–7.0, it would be prudent to choose a calibration-set that encompasses lakes with current pH levels of about 5.0–7.5, or so. Recent limnological data for the calibration lakes are collated, which should include the limnological variables that may most likely influence assemblages (e.g., lake-water pH, nutrients, and other physical, chemical, and possibly biological factors). This represents the first data matrix: the environmental data. The second challenge is to characterise the recent biological assemblages present in the calibration lakes (in this example, diatom species composition and abundance). Because surface sediments (i.e., the top 1 or 0.5 cm) contain assemblages that lived in the lake's recent past (i.e., last few years), these data are used for the second data matrix: the species data. Numerical and statistical techniques, as described in this book, are then used to explore, define, and quantify the relationships between the two data matrices, and develop calibration or transfer functions whereby the palaeolimnologist can reconstruct past environmental conditions based on the indicators preserved in sediment profiles.

A large portion of this book addresses various numerical or statistical approaches that have been developed to deal with these complex data. Although almost all of this calibration work has focused on biological indicators, similar approaches can be used for quantifying other types of proxy data (Pienitz et al. 2009).

Palaeolimnologists, however, use a variety of temporal frameworks. Surface-sediment calibration-sets represent only one of many types of palaeolimnological applications that require numerical analyses. For example, a broad spectrum of different types of data can be generated for the down-core portion of the study. The complexity and diversity of proxy data may at times be staggering, especially in multi-proxy studies (Lotter 2003; Birks and Birks 2006). Moreover, the data can be collected and presented in different ways, including relative frequencies, concentrations, accumulation rates, and various ratios. In some cases, simple presence and absence data can be useful (e.g., Sweetman and Smol 2006). Each approach may provide some important palaeoenvironmental insights, but also requires a careful assessment and evaluation of various assumptions.

The adoption and widespread use of numerical methods in palaeolimnology has been surprisingly rapid. Early work (e.g., pre-1990s) was typically restricted to qualitative interpretations of species distributions, or the development of simple indices and ratios (for a historical review, see Battarbee et al. 1986). Some of the earliest numerical work was by Pennington and Sackin (1975), who used principal component analysis on down-core pollen and geochemical data. By the late 1970s and early 1980s, some researchers were using, for example, simple agglomerative cluster analyses to group fossil samples (e.g., Davis and Norton 1978; Jatkar et al. 1979; Norton et al. 1981; Carney 1982), or stratigraphically constrained cluster analysis to derive fossil assemblage zones (Binford 1982). Several early studies showed the value of applying numerical methods such as ordinations or clusterings to summarise patterns in modern sediment geochemistry (Dean and Gorham 1976a, b) and in modern assemblages of, for example, plant macrofossils (Birks 1973), cladocerans (Beales 1976; Hofmann 1978; Synerholm 1979), diatoms (Brugam 1980; Bruno and Lowe 1980), and ostracods (Kaesler 1966; Maddocks 1966). Similarly, palaeolimnologists began to apply numerical partitioning, clustering, or ordination techniques to sediment lithostratigraphical (Brown 1985), biological, and geochemical data (Sergeeva 1983; Peglar et al. 1984).

In his pioneering study, Binford (1982) used the results of numerical analyses of ostracod and cladoceran assemblages to reconstruct, in a semi-quantitative way, changes in water-level, salinity, and substrate over the past 12,000 years at Lake Valencia, Venezuela. Other early and creative applications of multivariate data analysis in palaeolimnology include Whiteside (1970) who used multiple discriminant analysis (= canonical variates analysis) to relate fossil assemblages of cladocerans in Danish lake sediments to modern assemblages, and Elner and Happey-Wood (1980) who used principal component and correspondence analyses to compare diatom, pollen, and geochemical stratigraphies in Holocene sedimentary records from two lakes in North Wales.

In the related field of Quaternary pollen analysis, Maher (1972a, b, 1980, 1981) emphasised the importance of considering the inherent errors in counting microfossils and showed how robust confidence intervals could be calculated for both relative percentage and concentration and accumulation-rate data (see Maher et al. 2012: Chap. 6). Maher's pioneering work on analytical and counting errors will hopefully become more used as research questions in palaeolimnology become increasingly more refined and more demanding in terms of data precision, accuracy, and uncertainties.

By the late 1980s, several multivariate techniques were being used to summarise patterns in palaeolimnological data but the techniques used were often not optimal for the research problems being addressed (e.g., unconstrained cluster analysis applied to stratigraphical time-constrained data). Several important publications in the late 1980s changed the way that many palaeolimnologists analysed their data numerically. These publications included (1) ter Braak (1986) where canonical correspondence analysis was introduced as a means of analysing species compositional and environmental data simultaneously, (2) Birks and Gordon (1985) where numerical techniques developed specifically for the analysis of Quaternary pollen-stratigraphical data were synthesised, and (3) ter Braak and Prentice (1988) where a unified theory of gradient analysis was presented with the first explicit distinction between gradient analytical techniques (e.g., regression, calibration, ordination, constrained ordination) appropriate for species data with linear or monotonic responses to the environment and for species data with unimodal responses to the environment. The scene was then set for the seminal paper by ter Braak and van Dam (1989) where two-way weighted-averaging regression and calibration were shown to be an effective and robust way of inferring lake-water pH from diatom assemblages. A year later Birks et al. (1990) demonstrated how some of these approaches could be applied in a rigorous manner to lake acidification studies, developed numerical methods for estimating sample-specific errors of prediction, and presented various numerical approaches for evaluating environmental reconstructions.

At about the same time as these publications and developments, the INQUA Commission for the Study of the Holocene, under its President-elect Brigitta Ammann, established a working group on data-handling methods in 1987. An annual or 6-monthly newsletter containing details of software, relevant literature, methodological developments, and on-going research was produced and widely distributed amongst numerical palaeoecologists and palaeolimnologists. It was originally edited by JC Ritchie (1988–1990) and then by LJ Maher (1990–1997). The newsletter was then edited by KD Bennett and the last issue was in 2003. Besides providing useful newsletters with many articles of direct interest to the rapidly evolving cohort of quantitative palaeoecologists and palaeolimnologists, Lou Maher created an invaluable file boutique of useful software for estimating confidence intervals, sequence-slotting, rarefaction analysis, etc. Although these programs were written to run under MS-DOS, they can, with care, be run under Microsoft Windows® and other operating systems. They represent a rich array of software, much of which is as relevant to the quantitative palaeolimnologist in the twenty-first century as they were to the pioneering palaeolimnologists of the mid 1980s. For details of all the newsletters, software, publications, etc., go to http://www.geology.wisc.edu/~maher/inqua.html or http://www.chrono.qub.ac.uk/inqua/index.htm. The working group on data-handling methods made a significant contribution to the spread of numerical ideas, literature, and software amongst the palaeolimnological research community in the late 1980s-late 1990s.

Opportunities and Challenges

Palaeolimnologists can justifiably be proud of their accomplishments, but there also remain many challenges and opportunities. Of course, no amount of 'statistical finesse' will ever compensate for a poor data-set or a poorly designed sampling programme. Although interpretations based on multi-proxy studies typically provide much stronger environmental and ecological interpretations, they are not without their problems (Lotter 2003; Birks and Birks 2006). Nonetheless, exciting opportunities are available as more, high quality data-sets become available. For example, as many palaeolimnological approaches are now standardised to a certain level, and because a large number of studies have been completed for some regions, it is now possible to undertake meta-analyses of large data-sets to probe various hypotheses (e.g., Smol et al. 2005; Rühland et al. 2008). Much scope remains for these types of syntheses. Moreover, in the quest for more robust calibration functions, much of the hard-earned ecological and biogeographical data contained in modern calibration-sets (e.g., Telford et al. 2006; Vyverman et al. 2007; Vanormelingen et al. 2008; Bennett et al. 2010) often remains under-utilised. It is important to keep in mind that, even in a book on numerical analyses, detailed statistical interpretations are not always needed in palaeoecological studies, and in fact elegant ecological work can often be done at a qualitative level. For example, as certain *Chaoborus* species cannot co-exist with fish, the simple presence of mandibles from one species of this taxon in a lake's profile may indicate fishless conditions (e.g., Uutala 1990).

Outline of the Book

This book focuses on numerical and statistical methods that have been widely applied in palaeolimnology (e.g., ordination methods, weighted averaging) or that have considerable potential in palaeolimnology (e.g., classification and regression trees). All the methods presented are robust and can often take account of the numerical properties of many palaeolimnological data-sets (e.g., closed percentage data, many variables, many zero values). They can all be used to answer specific research questions in palaeolimnology that can contribute to our understanding of lake history, development, and responses to a range of environmental factors. The emphasis throughout is on numerical thinking (ideas, reasons, and potentialities of numerical techniques), rather than numerical arithmetic (the actual numerical manipulations involved).

The book aims to provide an understanding of the most appropriate numerical methods for the quantitative analysis of complex multivariate palaeolimnological data. There has in the last 15–20 years been a maturation of many methods. The book provides information to what these methods can and cannot do and some guidance as to when and when not to use particular methods. It attempts to outline the major assumptions, limitations, strengths, and weaknesses of different methods.

The book is divided into four parts and contains 21 chapters. Part I is introductory and contains this chapter, an overview (Chap. 2) of the basic numerical and statistical methods (e.g., regression analysis and statistical modelling) used in palaeolimnology, and Chap. 3 that describes the data-sets used in some of the chapters. Part II includes chapters on numerical methods that can usefully be applied to the analysis of modern surface-sediment data and of core sediment data. Chapter 4 gives an introduction and overview of the methods in this part. Chapter 5 explores the essential first step of exploratory data analysis and graphical display and the important questions of identifying potential outlying data-points in large complex data-sets. Chapter 6 considers a topic that is surprisingly rarely considered in palaeolimnology, namely the assessment of uncertainties associated with laboratory methods and microfossil analysis. Chapter 7 outlines the basic techniques available for the clustering and partitioning of multivariate palaeolimnological data to detect groups or clusters and to establish the relationships of biologically defined clusters and environmental variables. Chapter 8 discusses the range of ordination techniques currently available to palaeolimnologists to detect and summarise patterns in both modern and core data. These include classical techniques such as principal component analysis and correspondence analysis and canonical or constrained techniques such as canonical correspondence analysis and redundancy analysis. The chapter also presents new and improved techniques for partitioning variation in data-sets and in detecting spatial or temporal structures at a range of scales. Chapter 9 outlines statistical-learning techniques such as the various types of classification and regression trees and artificial neural networks as useful tools in the exploration and mining of very large, heterogeneous data-sets and in developing robust, simple predictive models. It also considers powerful techniques such as principal curves and surfaces as a means of summarising patterns in complex multivariate data and shrinkage techniques such as ridge regression, the lasso, and the elastic net to help develop robust regression models based on large data-sets.

Part III is devoted to numerical techniques that are only applicable to stratigraphical data-sets and an overview of these techniques is given in Chap. 10. Chapter 11 describes basic techniques for summarising patterns in stratigraphical data by partitioning, clustering, or ordination methods, and for estimating rates of change within stratigraphical sequences. Chapter 12 discusses the critical problem of establishing age-depth relationships. It outlines procedures for calibrating radiocarbon dates first because age-depth models based on uncalibrated dates are meaningless. It reviews various age-depth modelling procedures and discusses the difficult problem of deciding which model to accept and to use. Age-depth modelling is an area where considerable advances are being made, in particular by adopting a Bayesian approach to age calibration and age-depth modelling. Chapter 13 outlines robust numerical procedures for correlating two or more cores on the basis of some measured properties (e.g., loss-on-ignition, magnetic susceptibility). These procedures can be very useful in correlating multiple undated cores with a master dated core from a lake. Chapter 14 discusses the quantitative reconstruction of past environmental variables from fossil assemblages using calibration or transfer functions. A range of such calibration-function methods is discussed, the question of which model(s) to select is explored, and the ecological problems and pitfalls of interpreting palaeoenvironmental reconstructions derived from calibration functions are considered. Chapter 15 outlines the modern analogue technique, a useful technique not only in environmental reconstruction but also in identifying potential 'reference lakes' in restoration programmes. Chapter 16 provides an introduction to robust techniques for analysing patterns in temporally ordered palaeolimnological data. The chapter cautions against the use of several conventional time-series analysis techniques (e.g., spectral analysis) that are designed for evenly spaced samples in the time-series, a property that is very rarely realised in palaeolimnology.

Part IV begins with an introduction (Chap. 17) and considers three different case studies where many of the numerical methods described in Parts II and III are used to answer particular research questions in palaeolimnology such as limnological responses to environmental changes (Chap. 18), human impacts of acidification and eutrophication (Chap. 19), and tracking climatic change using palaeolimnological techniques (Chap. 20). These three chapters highlight that the numerical methods discussed in Parts II and III are not ends in themselves but are a means to an end. In the case of palaeolimnology, the end is to improve our understanding of the timing, rates, magnitudes, and causes of limnological changes over a range of time scales. Part IV concludes with some views on where quantitative palaeolimnology has reached, what the future challenges are, and what are the limitations of our current data-sets and our numerical methods (Chap. 21).

This book makes no attempt to review the now vast literature on the application of numerical methods in palaeolimnology. Almost every paper in palaeolimnology involves at least one numerical analysis. This book similarly makes no attempt to be a complete textbook on numerical methods in ecology, environmental sciences, or palaeontology. Instead it concentrates on numerical techniques of primary interest and relevance to palaeolimnologists. For palaeolimnologists interested in textbooks on numerical ecology, environmental sciences, or palaeontology, we recommend the following:

1. Numerical ecology

Jongman RHG, ter Braak CJF, van Tongeren OFR (1987) Data analysis in community and landscape ecology. Pudoc, Wageningen, 299 pp

Legendre P, Legendre L (1998) Numerical ecology, 2nd English edn. Elsevier, Amsterdam, 853 pp

Lepš J, Šmilauer P (2003) Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge, 269 pp

McCune B, Grace, JB, Urban DL (2002) Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon, 300 pp

Zuur AF, Ieno EN, Smith GM (2007) Analysing ecological data. Springer, New York, 672 pp

2. Environmental sciences

Fielding AH (2007) Cluster and classification techniques for the biosciences. Cambridge University Press, Cambridge, 246 pp

Hanrahan G (2009) Environmental chemometrics. CRC, Boca Raton, 292 pp

Manly BFJ (2009) Statistics for environmental science and management, 2nd edn. CRC, Boca Raton, 295 pp

Qian SS (2010) Environmental and ecological statistics with R. CRC Press, Boca Raton, 421 pp

Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, 537 pp

Shaw PJA (2003) Multivariate statistics for the environmental sciences. Arnold, London, 233 pp

Sparks T (ed) (2000) Statistics in ecotoxicology. Wiley, Chichester, 320 pp

Varmuza K, Filzmoser P (2009) Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton, 321 pp

3. Palaeontology

Davis JC (2002) Statistics and data analysis in geology, 3rd edn. Wiley, New York, 646 pp

Hammer Ø, Harper DAT (2006) Paleontological data analysis. Blackwell, Oxford, 351 pp

Haslett SK (ed) (2002) Quaternary environmental micropalaeontology. Arnold, London, 340 pp

Reyment RA, Savazzi E (1999) Aspects of multivariate statistical analysis in geology. Elsevier, Amsterdam, 285 pp

Because the development of user-friendly Windows® type software for implementing specialised numerical and statistical analyses lags far behind the development of the actual numerical techniques, we give limited references in the chapters to available software for particular analyses. We believe that there will inevitably be future developments in numerical palaeolimnology but that these developments will only be available to researchers as scripts and packages in the R programming language and its vast libraries for statistical and numerical procedures. Useful introductions to R as well as to basic statistics and statistical modelling include:

Aitkin M, Francis B, Hinde J, Darnell R (2009) Statistical modelling with R. Oxford University Press, Oxford, 576 pp

Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York, 206 pp