From:

Maddy, D. & Brew, J.S. (1995)
Statistical Modelling of Quaternary Science Data.
Technical Guide 5, Quaternary Research Association,
Cambridge. 271pp.

Chapter 6

QUANTITATIVE PALAEOENVIRONMENTAL RECONSTRUCTIONS

H. John B. Birks

INTRODUCTION

The aim of many Quaternary palaeoecological studies is to reconstruct features of the past environment from fossil assemblages preserved in lake, bog, or ocean sediments. Although the fossil assemblages are usually studied quantitatively with individual fossil pollen, diatoms, ostracods, foraminifera, etc. being routinely counted, many environmental reconstructions are qualitative and are presented in terms such as "cool", "temperate", "moist", "dry", etc. In 1971, Imbrie and Kipp revolutionised Quaternary palaeoecology by presenting, for the first time, a procedure for the quantitative reconstruction of past environmental variables from fossil assemblages involving so-called transfer or calibration functions. Since this pioneering work on marine foraminifera in relation to ocean surface temperatures and salinity, the general approach of quantitative palaeoenvironmental reconstruction has been adopted in many areas of palaeoecology with different groups of organisms, including pollen, diatoms, chrysophytes, coleoptera, chironomids, ostracods, mosses, radiolaria, and foraminifera.

The primary aim throughout is to express the value of an environmental variable (e.g. lake-water pH) as a function of biological data (e.g. diatom assemblages), so-called environmental "proxy data". This function is termed the transfer function or biotic index (ter Braak, 1987a), and its construction is called calibration by statisticians (ter Braak, 1987b; ter Braak and Prentice, 1988). Calibration is the opposite of regression where the aim is to model the response of a biological taxon as a function of one or more environmental variables. Calibration and regression differ because the causal and statistical relationships between taxa and their environment are not symmetrical (ter Braak, 1987a). The environmental "proxy data" (= biological assemblages) are usually quantitative counts, commonly expressed as percentages or proportions. They are, more rarely, ordinal estimates (e.g. a 1-5 abundance scale) or presence/absence data. All fossil data are usually from different stratigraphical intervals and hence different times; occasionally the fossil data may be contemporaneous (or nearly so) but from many sites (e.g. Bartlein and Webb, 1985).

The basic idea of quantitative environmental reconstruction is shown in Figure 6.1. There is one or more environmental variable X_0 that we wish to reconstruct from fossil biological data Y_0 consisting of *m* taxa in *t* samples. To estimate values of X_0, we have to model the responses of the same *m* taxa today in relation to the environmental variable(s) of interest. This involves a modern "training set" of *m* taxa at *n* sites (Y) studied as assemblages preserved in surface sediments such as surficial lake muds, ocean sediments, peats, or soil and with an associated set of modern environmental variables (X) for the same *n* sites. The modern relationships between Y and X are modelled statistically and the resulting function is then used as a transfer function to transform the fossil data Y_0 into quantitative estimates of the past environmental variable(s) X_0.

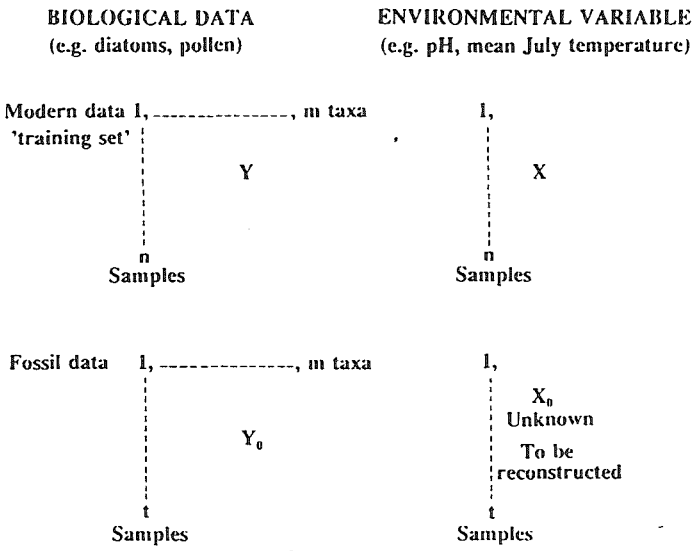

Figure 6.1 *Principles of quantitative palaeoenvironmental reconstruction showing X_0, the unknown environmental variable to be reconstructed from fossil assemblage Y_0, and the role of a modern training set consisting of modern biological X and environmental data Y.*

Since the initial study of Imbrie and Kipp (1971), several numerical techniques have been developed to reconstruct palaeoenvironmental variables quantitatively. Some of these have a stronger theoretical basis, either statistically, ecologically, or both, than others and some are therefore more appropriate for a particular research problem than others. The fundamental distinction between the existing methods concerns the underlying taxon-environment response model that is assumed by the different methods.

There are methods that assume a linear response model of taxa to their environment, so-called linear-based methods, and there are methods that assume a unimodal (Gaussian-like) response model of taxa to their environment, so-called unimodal-based methods (ter Braak and Prentice, 1988). Within the linear-based and the unimodal-based methods, different methods result from how the parameters of the models are estimated (maximum likelihood, least squares, weighted averaging, partial least squares, etc.).

The structure of the chapter is as follows. After outlining the general theory and assumptions of quantitative environmental reconstructions and the numerical properties of biological and environmental data, I discuss the basic biological and statistical requirements of quantitative reconstruction procedures. I then consider how the predictive performance of modern training sets can be assessed statistically. The fundamental distinction between linear-based and non-linear unimodal-based methods is outlined and methods are discussed to evaluate which method to use with a given data-set. In the section on linear-based methods, I review classical linear regression, inverse linear regression, restricted inverse linear regression, principal components regression, partial least squares regression, and a group of linear techniques for reconstructing more than one environmental variable simultaneously. In the section on non-linear unimodal-based methods, I discuss maximum likelihood regression and calibration (Gaussian logit and multinomial logit models), weighted averaging regression and calibration, correspondence analysis regression, weighted averaging partial least squares regression, and canonical correspondence analysis and multivariate weighted averaging partial least squares regression. To conclude the discussion of existing methods that do not fall into either the linear- or unimodal-based distinctions, I review modern analogue techniques and response surfaces. I then consider how palaeoenvironmental reconstructions can be evaluated numerically and ecologically, and present, for illustrative purposes, an example of numerical reconstructions of three environmental variables using the original data of Imbrie and Kipp (1971) and several reconstruction techniques. As all reconstruction methods are totally computer-dependent, I give details of available PC software. I conclude the chapter with a discussion of possible future research needs and directions.

Throughout this chapter I have tried to keep the mathematical detail to a minimum and to concentrate on the ecological, palaeoecological, and general statistical principles. Inevitably the chapter involves some statistical concepts and terminology. Any reader unfamiliar with these is recommended to consult the excellent text-books on ecological data analysis by Jongman *et al.* (1987) and on calibration by Martens and Næs (1989) and the review by ter Braak and Prentice (1988). In preparing this chapter, I have drawn extensively on the many contributions to the theory and methodology of environmental reconstruction by ter Braak and his associates, in particular ter Braak (1987a, 1987b,

1987c, 1995), ter Braak and Looman (1986, 1987), ter Braak and Barendregt (1986), ter Braak and van Dam (1989), Birks *et al.* (1990a), ter Braak and Juggins (1993), and ter Braak *et al.* (1993).

GENERAL THEORY, ASSUMPTIONS, AND DATA PROPERTIES

Notation

The following notation is used throughout: **Y** is the $n \times m$ matrix of abundances of taxa ("biological responses") estimated within defined spatial and temporal domains with y_{ik} representing the abundance of taxon k in sample i ($y_{ik} \geq 0$) (i = 1, ..., n sites and k = 1, ..., m taxa) ; **X** is the $n \times p$ matrix of p physico-chemical variables of the environment ("environmental predictors") measured over the same space-time domains as **Y** and assumed to be related functionally to **Y**; x is the environmental variable of interest to be reconstructed ("calibrated") on the basis of a modern training data set consisting of the matrix **Y** and the vector of the environmental variable $\mathbf{x} = (x_i)$, with x_i the value of the environmental variable (e.g. lake-water pH) in site i; **B** is the matrix of additional physical, chemical, and biological variables that together with **X** completely determine **Y** (e.g. variables difficult or impossible to measure such as nutrient availability, predation, disease, other biotic factors, etc.); $\hat{x}_i$ is the estimated or inferred value of x in site i; the subscript index 0 indicates a fossil assemblage (y_0) for which the environmental variable is to be inferred (x_0).

The abundances of the taxa are, almost invariably in palaeoecology, expressed as percentages or proportions with respect to all fossils counted in site i. Thus matrix **Y** contains compositional data that have a constant sum constraint, namely $\sum_{k=1}^{m} y_{ik} = 1$

A "+" replacing a subscript denotes summation over that subscript, e.g. $y_{i+} = \sum_{k=1}^{m} y_{ik}$

General theory

The general theory of quantitative palaeoenvironmental reconstructions from palaeoecological data is presented by Imbrie and Kipp (1971) and Imbrie and Webb (1981), on which this section is based.

If **Y** is totally explicable as responses to the variables represented by **X** and **B**, namely

$$\mathbf{Y} = \mathbf{XB}$$

we have a deterministic model, as no allowance is made for random stochastic or historical factors.

If **B** is assumed to be constant or to be 0, we can model **Y** in terms of **X** and $\mathbf{R}_c$, a set of ecological response functions, as

$$\mathbf{Y} = \mathbf{R}_c\,(\mathbf{X})$$

If **B** is not constant, the model must include the total response function $\mathbf{R}_t$, as

$$\mathbf{Y} = \mathbf{R}_t\,(\mathbf{XB})$$

The task in quantitative palaeoecology is to derive a set of ecological response functions $\mathbf{R}_c$ so that **X** can be estimated from **Y** by

$$\mathbf{X} = \mathbf{R}_c^{-1}\,(\mathbf{Y})$$

or, more specifically, to estimate $\hat{\mathbf{X}}_0$ given a fossil assemblage $\mathbf{Y}_0$ using modern response functions $\mathbf{R}_c$ that are assumed to be invariant in space and time, namely

$$\hat{\mathbf{X}}_0 = \mathbf{R}_c^{-1}\,(\mathbf{Y}_0)$$

If the ecological responses of the taxa in question and the processes that determine their distribution and abundance were understood, $\mathbf{R}_c$ could be derived deductively from modern ecological studies and an explanatory model constructed. Given our poor current ecological knowledge, there are inadequate data for the construction of an explanatory model. Instead, we have to use direct empirical models based on the observed patterns and covariances of the m taxa in n modern surface-samples (**Y**) in relation to **X**, to derive $\hat{\mathbf{U}}$, our empirical calibration functions

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

and then infer the past environment, $\mathbf{X}_0$, from

$$\hat{\mathbf{X}}_0 = \hat{\mathbf{U}}^{-1}\,(\mathbf{Y}_0)$$

Although **U** and $\mathbf{X}_0$ can be estimated or inferred in many different ways, in practice nearly all quantitative palaeoenvironmental reconstructions involve two stages. First, the responses of modern taxa to contemporary environment are modelled. This is a regression problem (ter Braak and Prentice, 1988) and involves a modern training set of taxon assemblages ("response" variables) from surface samples with associated

environmental data ("predictor" variables) (Figure 6.1). Second, the modelled responses are used to infer the past environment variables from the composition of fossil assemblages. This is a calibration problem (ter Braak and Prentice, 1988). These two steps can be summarised as follows:

(1) Regression step in which we estimate U, the modern ecological response functions or so-called calibration functions or regression coefficients, from the modern training set of Y and X

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

or, as in the "inverse regression" approach (ter Braak, 1987a),

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

(2) Calibration step in which we infer $\hat{X}_0$, the past environmental variable, from palaeoecological data Y_0 and our estimates of U derived in the regression step

$$\hat{X}_0 = \hat{U}^{-1}(Y_0)$$

or, as in the "inverse" regression approach (ter Braak, 1987a),

$$\hat{X}_0 = \hat{U}(Y_0)$$

Classical and inverse approaches

There are, as mentioned above, two conceptually different approaches to quantitative calibration and palaeoenvironmental reconstructions. There are discussed in detail by ter Braak (1995). This section is based on his discussion. The two approaches are the classical and the inverse approaches (Osborne, 1991). Martens and Næs (1989) also call the inverse approach a forward or direct approach and the classical approach an indirect or reverse approach.

In the classical approach, the empirical calibration functions U are estimated from the training set by regressing Y on X (Sundberg, 1985). This can, if required, be a linear or non-linear regression and a univariate or a multivariate regression. The estimated U is then "inverted" to estimate the unknown environmental variable x_0 from a fossil sample y_0. The steps are

(1) Regression $Y = \hat{U}(X)$

(2) Calibration $\hat{X}_0 = \hat{U}^{-1}(Y_0)$

Note that with the exception of simple linear calibration the inverse of U does not exist, and what is attempted instead is to find values of x_0 so that the two sides of the calibration equation are most similar in some statistical sense (ter Braak, 1995). In practice, values of x_0 are sought that have the highest probability of producing the observed fossil assemblage y_0 if the estimated value of x_0 was the true value.

In the inverse approach, the empirical calibration functions U are estimated directly from the training set by regressing X on Y and the unknown environmental variable x_0 is then estimated directly from the modern regression equation.

(1) Regression $X = \hat{U}(Y)$

(2) Calibration $\hat{X}_0 = \hat{U}(Y_0)$

There is a substantial literature on the relative statistical advantages and disadvantages of these two approaches (e.g. Krutchkoff, 1967; Lwin and Maritz, 1982; Martinelle, 1970; Sundberg, 1985; Martens and Næs, 1989; Osborne, 1991). Ter Braak (1995) summarises this on-going debate among statisticians by suggesting that the inverse approach performs slightly better when the fossil samples are from the central part of the distribution of the modern training set whereas the classical approach may perform better at the extremes and with slight extrapolation (Sundberg, 1985; Birks *et al.*, 1990a; Gasse *et al.*, 1995).

From a practical viewpoint, the inverse approach considers each environmental variable individually, because a multivariate regression reduces to a series of multiple regressions of individual dependent variables (ter Braak, 1995). Thus if interest is on one environmental variable only, the others need not be considered (Brown, 1982; Lorber *et al.*, 1987). This naturally makes the construction of modern training sets easier. As ter Braak (1995) notes, this is a surprising property of inverse regression, particularly as other environmental variables may be ecologically important in influencing the composition and abundance of the biological assemblages.

Of the numerical techniques widely used in quantitative palaeoenvironmental reconstructions and discussed in this chapter, the following fall within the general classical approach (ter Braak, 1995): Gaussian logit regression, multinomial logit regression, and response surfaces. In contrast, restricted inverse linear regression, principal components regression, partial least squares regression, weighted averaging, correspondence analysis regression, weighted averaging partial least squares regression, and modern analogue techniques all lie within the inverse approach (ter Braak, 1995).

Assumptions

There are at least five major assumptions in quantitative palaeoenvironmental reconstructions (Imbrie and Kipp, 1971; Imbrie and Webb, 1981; Birks *et al.* 1990a). See also Webb and Clark (1977) and Sachs *et al.* (1977) for a discussion of the major assumptions.

(1) The taxa in the modern training set (Y) are systematically related to the environment (X) in which they live (see Le ,1992 for a critical discussion of this assumption).

(2) The environmental variable(s) to be reconstructed (x) is, or is linearly related to, an ecologically important determinant in the ecological system of interest (see Korsman and Birks, 1995 for a discussion of the assumption).

(3) The taxa in the training set (Y) are the same biological entities as in the fossil data (Y_0) and their ecological responses ($\hat{U}$) to the environment variable(s) of interest have not changed significantly over the time span represented by the fossil assemblage. Contemporary patterns of taxon abundance in relation to X can thus be used to reconstruct changes in X through time.

(4) The mathematical methods in regression and calibration adequately model the biological responses to the environmental variable(s) of interest and yield calibration functions with sufficient predictive powers to allow useful, accurate, and unbiased reconstructions of X.

(5) Other environmental variables than the one of interest have negligible influence, or their joint distribution with the environmental variable in the fossil set is the same as in the training set (see Le and Shackleton, 1994 for simulation studies that assess this assumption).

Data properties

Modern biological data in training sets and fossil biological data usually contain many taxa (e.g. 50-300 taxa). The data are either binary (presence/absence of a taxon at a site) or quantitative and usually contain many zero values for sites where a taxon is absent. Quantitative data are commonly expressed as percentages or proportions of the total count in a sample and they are thus closed, multivariate compositional data with a constant-sum constraint. Quantitative data are often highly variable and invariably show a skewed distribution. The biological data are thus complex and invariably show noise, some redundancy, and internal correlations, and often contain outliers. Taxa generally have a non-linear relationship with quantitative environmental variables. Taxon

abundance or probability of occurrence is often a unimodal function of the environmental variables (ter Braak, 1987b).

Modern environmental data usually contain a smaller number of variables (*ca.* 10-30) than the corresponding biological data. Environmental data can be binary (presence/absence of a variable at a site), ordinal or ranked (e.g. dry, medium, wet, very wet), or quantitative. If ordinal or quantitative, such data rarely, if ever, contain any zero values. Quantitative environmental variables commonly follow a log-normal distribution (see Ott, 1990), namely logarithms of the variable show a normal distribution (Jager and Looman, 1987). Environmental variables commonly show linear relationships and high correlations between variables (e.g. lake-water pH, Ca, alkalinity). There is thus often considerable redundancy in environmental data.

BASIC REQUIREMENTS

There are at least seven major requirements in any quantitative palaeoenvironmental reconstruction (Birks, 1994). These are as follows.

(1) A biological system is required that produces abundant identifiable fossils and that is responsive and sensitive to the environmental variable(s) of interest today at the spatial and temporal scales of study.

(2) A large high-quality training set of modern surface-samples and associated environmental data is available. This should be representative of the likely range of environmental variables in the past, be of consistent taxonomy and nomenclature, be of comparable quality (count size, sampling methodology, preparation procedures, counting techniques, etc.), and be from the same type of sedimentary environment (e.g. lakes) and hence have comparable taphonomies.

(3) The fossil data-sets used for reconstruction purposes should be of comparable taxonomy and nomenclature, quality, and sedimentary environment as the modern training set.

(4) Good independent chronological control is required for the fossil data-sets to permit correlations and comparisons and, if required, to allow an assessment of rates of biotic or environmental change.

(5) Robust statistical methods for regression and calibration are required that can adequately model the complex, non-linear, often unimodal relationships between modern taxa and their environment.

(6) Reliable and realistic statistical estimation of standard errors of prediction for the modern training set as a whole and for each reconstructed value is required. As the reliability of the reconstructed environmental values may vary from one fossil sample to another, depending on, for example, composition, preservation, etc. sample-specific standard errors of prediction are needed.

(7) Critical ecological and statistical evaluations of all reconstructions are needed, as any statistical regression and calibration procedure is designed to produce a result. It is essential, however, to evaluate if the result is ecologically and statistically reliable.

This chapter concentrates on requirements (5), (6), and (7), namely the statistical aspects of palaeoenvironmental reconstruction. Before turning to specific statistical methods and the evaluation of reconstructions, it is appropriate to discuss at this stage the general principles of estimating standard errors of prediction and of assessing the performance of training sets.

ASSESSING THE PERFORMANCE OF MODERN TRAINING SETS

The standard error or root mean square error (RMSE) of $(x_i - \hat{x}_i) = \left[1/n\,(x_i - \hat{x}_i)^2\right]^{1/2}$ is commonly calculated for the training set and quoted as a measure of the predictive abilities of the training set. It assesses the prediction errors. Wallach and Goffinet (1989) discuss the values of RMSE as a means of evaluating how well a model can be expected to function as a predictive tool (see also Mayer and Butler, 1993 and Power, 1993 for a discussion of predictive assessment). The correlation (r) and/or the coefficient of determination (r^2) between x_i and $\hat{x}_i$ are also often calculated. These are measures of the strength of the relationship between observed and inferred values and permit comparisons between transfer functions for different environmental variables (Gasse *et al.*, 1995). As RMSE is invariably under-estimated and r and r^2 are over-estimated when based solely on the training set (Oksanen *et al.*, 1988; ter Braak and van Dam, 1989; Birks *et al.*, 1990a; ter Braak and Juggins, 1993), some form of split-sampling or cross-validation (Stone, 1974; Snee, 1977; Picard and Cook, 1984; Osten, 1988; Martens and Næs, 1989) is required to derive a more reliable and realistic estimate of prediction error and hence to evaluate the predictive abilities of a training set than is given by RMSE, r, or r^2 when x_i and $\hat{x}_i$ are included in the comparisons.

Split-sampling involves randomly splitting the modern data into a training set and a test set and using the training set to predict x_i for all samples in the test set. Comparison of observed x_i and predicted x_i for the samples in the test set provides a realistic assessment of the root mean square error of prediction (RMSEP), r, and r^2 (e.g. Bartlein *et al.*, 1984;

Andrews *et al.*, 1980; D'Antoni and Spanner 1993; Kipp, 1986; Norton *et al.*, 1986; Webb and Bryson, 1972; Birks *et al.*, 1990a; D'Antoni, 1993). This RMSEP is, as ter Braak and van Dam (1989, p. 210) emphasise, the "appropriate benchmark to compare methods" as all errors are considered. In many studies, however, large test sets are not available. Instead the predictive errors for a test set can be simulated by statistical cross-validation.

The simplest cross-validation approach is jack-knifing or "leave-one-out" (ter Braak and Juggins, 1993) where the reconstruction procedure is applied n times using a training set of size (n - 1). In each of the n predictions, one sample is left out in turn and the calibration function based on the (n - 1) sites in the training set is applied to the one sample in the test set and omitted from the training set, giving for that sample a predicted ($\hat{x}_i$) and, by subtracting this from the observed value (x_i), a prediction error for the sample. Thus in each prediction, the individual samples act in turn as a test set, each time of size 1. The prediction errors are accumulated to form a "leave-one-out" RMSEP or $\text{RMSEP}_{(\text{jack})}$ which can, in some instances (cf. ter Braak, 1995), be a consistent estimate of the true RMSEP (ter Braak and Juggins, 1993). In addition, r and r^2 can be calculated for each x_i when sample i is included in the test set but excluded from the training set.

It is important to distinguish between RMSE, r, and r^2 based on the comparison of x_i and $\hat{x}_i$ when sample i is part of the training set (apparent RMSE, apparent r, apparent r^2) and RMSEP, r, and r^2 based on x_i and $\hat{x}_i$ when sample i is excluded from the training set and forms a test-set. These statistics are called $\text{RMSEP}_{(\text{jack})}$, $r_{(\text{jack})}$, and $r^2_{(\text{jack})}$.

In evaluating the performance of a training set it is important to consider not only the random differences in the predictions but also systematic differences in the predictions (Oksanen *et al.*, 1988; ter Braak and Juggins, 1993; ter Braak, 1995). Random differences or "error" (*sensu* Altman and Bland, 1983) are quantified by the standard error of (x_i - $\hat{x}_i$), namely RMSE. Systematic differences or relative or mean bias in the predictions can be estimated by the mean of (x_i - $\hat{x}_i$). In addition, maximum bias can be measured (ter Braak and Juggins, 1993). For this, the sampling interval of x_i is subdivided into equal intervals (conventionally 10), the mean bias per interval is calculated, and the largest absolute value of mean bias for an interval is used as a measure of maximum bias. Power, 1993 discusses other measures of bias. Just as RMSE can be based on the training set alone (apparent RMSE) or on a jack-knifed cross-validation with a test set ($\text{RMSE}_{(\text{jack})}$), mean and maximum bias can be based on either the training set alone or on a test set in a jack-knifed cross validation.

RMSEP is a measure of the overall predictive abilities of training set. It does not provide sample-specific errors of prediction for each individual fossil sample, as clearly the observed x_i is not known for the fossil samples. One approach has been to calculate approximate prediction confidence regions that are largely based on the apparent mean square error of prediction of the training set for linear-regression-based calibrations (e.g. Howe and Webb, 1983; Bartlein and Whitlock, 1993) or other confidence regions. These are almost certainly "an underestimate of the true uncertainty" (Bartlein and Whitlock, 1993, p. 278).

As discussed below, bootstrapping, a computer-intensive resampling procedure (Efron, 1982; Efron and Gong, 1983; Diaconis and Efron, 1983; Wallach and Goffinet, 1989; Efron and Tibshirani, 1993) can be used to derive RMSEP for individual fossil samples, as well as for all individual modern training samples. It also provides another estimate of the RMSEP of the training set (Birks *et al.*, 1990a). RMSEP for an environmental variable in the past can vary from sample to sample depending on the composition of the fossil assemblage and thus the presence or absence of taxa with a particularly strong "signal" for the environmental variable of interest. Bootstrapping for the estimation of sample-specific errors has been implemented for weighted averaging regression and calibration (Birks *et al.*, 1990a; Line *et al.*, 1994).

STATISTICAL METHODS

Introduction

It is almost a general law of nature that the relationships between taxa and quantitative environmental variables are non-linear and the abundance of taxa is often a unimodal function of the environmental variables. Each taxon grows best at a particular optimal value of an environmental variable and cannot survive where the value of the environmental variable is either too low or too high (ter Braak, 1987b). Thus all taxa tend to occur over a characteristic but limited environmental range and within this range to be most abundant at or near their environmental optimum. Gause (1930, p. 307) proposed that "the law of Gauss is the basis of ecological curves" and Gauch and Whittaker (1972) proposed the Gaussian curve as a simple model for unimodal relationships in ecology. The equation for such a curve (Figure 6.2a) is (ter Braak, 1987b)

$$Ey_{ik} = c_k e^{-\frac{1}{2}(x_i - u_k)^2 / t_k^2}$$

where c_k is the maximum of the response curve for taxon k,

u_k is the optimum of taxon k, i.e. the value of x at which the maximum (mode) of the response curve occurs,

t_k is the tolerance of taxon k, i.e. a measure of the breadth of the response curve or ecological amplitude,

Ey_{ik} is the expected abundance of taxon k in sample i,

(e^a = exp (a) = exponential function or antilog of a, i.e. $e^{\log(a)} = a$).

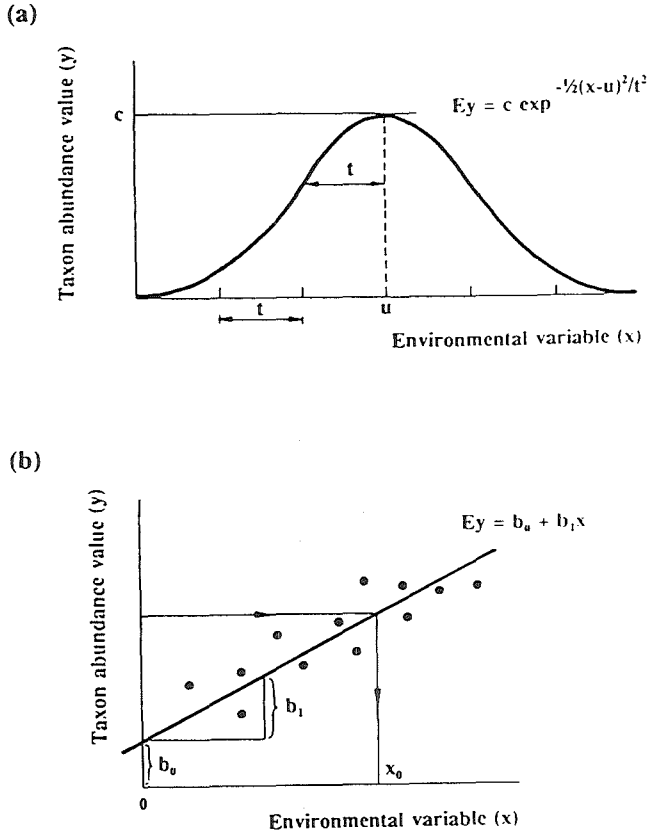

Figure 6.2 *Taxon-environment response models. (a) Gaussian unimodal relationship between the abundance (y) of a taxon and an environmental variable (x). The three important ecological parameters of the model are shown - u = optimum, t = tolerance, c = maximum. The equation for the expected value of the taxon's abundance y is given for the Gaussian response model (modified from ter Braak, 1987c). (b) Linear relationship between the abundance (y) of a taxon and an environmental variable (x). The equation for the expected value of the taxon's abundance y in relation to x is given for the linear response model. (b_0 = intercept, b_1 = slope or regression coefficient) (modified from ter Braak, 1987c).*

A unimodal curve will appear monotonic and approximately linear, however, if only a limited range of the environmental variable x is sampled. In such cases the estimates of

the parameters in the above equation will be inadequately determined and a statistical model that assumes a linear or monotonic response (e.g. a linear model that fits a straight line) would clearly be more appropriate (Figure 6.2b). Standard statistical methods that assume linear or monotonic relationships between taxa and their environment are well developed (ter Braak and Prentice, 1988) but they are often of limited value because of the generally non-linear non-monotonic responses of taxa to environmental variables. ter Braak and Prentice (1988) make the fundamental distinction between statistical methods for the analysis of multivariate biological and environmental data according to the underlying implicit response model, namely linear and unimodal models (see also ter Braak, 1987c).

The first step in any quantitative environmental reconstruction is thus to discover whether linear or unimodal statistical methods are appropriate for the available training set in relation to the environmental variable(s) of interest. Hill and Gauch (1980) defined the length of the gradients of variation in biological data to be the range of the sample scores in detrended correspondence analysis (DCA). This length is expressed in so-called standard deviation units (SD) of biological turnover. The tolerance of a taxon's curve along the scaled DCA axis is close to 1, and each taxon therefore appears, rises to its mode, falls, and then disappears over a distance of about 4 SD units. Samples that differ by 4 SD units can thus be expected to have no taxa in common. A complete turnover in taxonomic composition of samples occurs in about 4 SD units. A change of 50% in sample composition (called a half-change) occurs in about 1 SD units or slightly more. Økland (1986) and Peet *et al.* (1988) discuss further the value of scaling DCA axes in terms of compositional turnover.

For reconstruction purposes, it is essential to estimate the gradient length for the environmental variable(s) of interest. Detrended canonical correspondence analysis (DCCA) (ter Braak, 1986, 1987c) with x as the only environmental or external predictor variable and detrending by segments and non-linear rescaling provides an estimate (as the length of DCCA axis 1) of the gradient length in relation to x in SD units. The length may be different for different environmental variables and the same biological data (e.g. pH 2.62 SD units, alkalinity 2.76 SD units, lake colour 1.52 SD units in a modern diatom-lake chemistry data set; Korsman and Birks, 1995).

If the gradient length is short (2 SD units or less), taxa are generally behaving monotonically along the gradient and linear regression and calibration methods are thus appropriate. If the gradient length is longer (2 SD units or more), several taxa will have their optima located within the gradient and unimodal-based methods of regression and calibration are appropriate.

It should be emphasised that although ecological response curves are commonly thought to be more complex than is implied by the Gaussian-type unimodal models (e.g. Austin, 1980, 1987, 1992; Austin *et al.*, 1994; Austin and Gaywood, 1994), "these models are nevertheless useful in developing statistical descriptive techniques for data showing mostly unimodal responses, just as linear models are useful in statistical analysis of data that are only approximately linear" (ter Braak and Prentice, 1988, p. 284). The Gaussian unimodal model is a convenient and robust approximation for the analysis of biological data that span gradients in excess of 2 SD units.

LINEAR-BASED TECHNIQUES

Classical linear regression

Consider the simple case of inferring an unknown value of a quantitative environmental variable (x_0) from the quantitative response of a single taxon (y) (ter Braak, 1987a). We first model the modern relationship between the abundance of the taxon and the environmental variable as an ecological response curve consisting of a systematic component and a random (error) component. Such a curve is fitted to an appropriate and representative training set by regression analysis. The response curve and its assumed error structure form a statistical model of the taxon's abundance in relation to the environmental variable (ter Braak, 1987a). Let the response function be $Ey = f(x)$ (where Ey is the expected value of the random variable y) and let the error component be normally distributed. For a given range of values for the environmental variable, the probability of observing a particular abundance of the taxon is now known. What we want to do in environmental reconstructions is to infer or predict the value of the environmental variable (x_0) from the taxon's abundance in a particular sample (e.g. a fossil sample) when the value of x is not known. This model of the response curve and its error structure can be used to calculate the probability that a particular value of the environmental variable would occur with a particular taxon abundance over the likely range of environmental values (ter Braak, 1987a).

The principle of maximum likelihood provides the estimate of x_0 that has the highest probability of producing that particular observed taxon abundance if the estimated value of x was the true value (ter Braak, 1987a). This maximum likelihood estimate is obtained by solving the equation $y = f(x_0)$ for x_0. In a graph of the response curve (Figure 6.2b), this involves drawing a horizontal line at the appropriate value of y for fossil sample 0 and reading off the value of x_0 where the line cuts the response curve (ter Braak, 1987a). For a linear response curve (Figure 6.2b), this gives the estimate

$$\hat{x}_0 = (y - b_0)/b_1$$

where b_0 is the intercept and b_1 is the slope parameter of the response curve. If the response curve is unimodal (Figure 6.2a), this horizontal line will intersect the response curve twice, resulting in two estimates for x_0. However, if more than one taxon is used for calibration, this problem may disappear (Imbrie and Kipp, 1971; Brown, 1982; ter Braak, 1987a).

If each of the m taxa shows a linear relationship with the environmental variable and we wish to reconstruct x_0 from the abundance values of the m taxa in a fossil sample, reading off the m possible estimates of x_0 gives different inferred values of x_0 from the model $x_0 = (y - b_0)/b_1$ (ter Braak, 1987a). By minimising the sum-of-squares differences between the observed and expected responses, the combined estimate for $\hat{x}_i$ is

$$\hat{x}_0 = \frac{\sum_{k=1}^{m} (y_k - b_{0k}) b_{1k}}{\sum_{k=1}^{m} b_{1k}^2}$$

This is the maximum likelihood (and least-squares) estimate of $\hat{x}_i$ if the errors follow a normal distribution and the m taxa are independent and have equal variances (ter Braak, 1987a; ter Braak and Prentice, 1988).

These conditions are unlikely to be met with real biological and environmental data, because other environmental variables may cause correlations between taxon abundances and because the errors are often different for different taxa (ter Braak and Prentice, 1988). Residual correlations and variances can be estimated from the residuals of the regressions used for estimating the b_0 and b_1 parameters (ter Braak and Prentice, 1988). Deriving the likelihood maximum with respect to x_0 results in a general weighted least-squares problem (Brown, 1979; Brown, 1982; ter Braak and Prentice, 1988). This general approach can be extended to infer more than one environmental variable simultaneously from taxon abundances (Brown, 1982). Calibration of a single environmental variable can also be attempted using polynomial response functions rather than linear response functions (e.g. Scheffé, 1973; Schwartz, 1977; Brown, 1982).

This approach to reconstructing x_0 from biological data using so-called classical regression and the maximum likelihood principle does not appear to have been widely used in palaeoecology, except for inferring sea-surface temperatures from single variables such as foraminiferal test sizes (e.g. Malmgren and Kennett, 1978a, 1978b; Malmgren and Healy-Williams, 1978). In contrast, the alternative approach of inverse

regression has been widely used, particularly for reconstructing past climate from pollen-analytical data.

Inverse linear regression

In inverse linear regression, the training set is not used to construct response curves by regressing the abundances of the taxa on the environmental variable. Instead, as is implied by the name inverse regression, the environmental variable of interest is used as the "response variable" and the taxon abundances are used as "predictor", "explanatory variables" (ter Braak, 1987a).

The resulting regression equation is directly the transfer function used for calibration or reconstructing x_0.

$$x = b_0 + b_1 y \qquad \text{for the modern training set}$$

$$\hat{x}_0 = b_0 + b_1 y_0 \; \hat{x}_0 \qquad \text{for fossil sample 0}$$

If the distribution of the environmental variable being reconstructed equals its distribution in the modern training set, inverse regression is a statistically efficient procedure (Brown, 1979; ter Braak 1987a). There is, however, a substantial literature on the relative *pros* and *cons* of inverse and classical regression in linear calibration, as reviewed by Osborne (1991) and ter Braak (1995) and discussed earlier in this chapter.

The approach of inverse linear regression can be readily extended to reconstruct x_0 using the abundances of more than one taxon. Each taxon becomes an explanatory variable, and the inverse regression is simply a multiple least-squares regression of the environmental variable on the abundances of the m taxa in the modern training set, i.e.

$$x = b_0 + b_1 y_1 + b_2 y_2 + b_3 y_3 + \ldots\ldots + b_m y_m; \qquad \text{for the modern training set.}$$

Reconstructions for x_0 are obtained directly from the regression parameters (b_0, b_1 ..... b_m) applied to the abundances of the m taxa in the fossil sample 0. Thus the reconstruction of x_0 is based on the linear combination of the taxon abundances in fossil sample 0 and the regression coefficients estimated by multiple regression of x and y in the training set, i.e.

$$\hat{x} = b_0 + b_1 y_1 + b_2 y_2 + b_3 y_3 + \ldots\ldots + b_m y_m \qquad \text{for fossil sample 0}$$

The approach is most efficient, in a statistical sense, if the relationship between each taxon and the environmental variable is linear with a normal error distribution, if the environmental variable also follows a normal distribution, and if the prior distribution of

the environmental variable to be reconstructed equals its distribution in the training set (Brown, 1982; ter Braak 1987a).

Inverse linear regression is generally not useable in this form because the resulting estimates of the regression coefficients (b) are usually poorly and unreliably estimated. This arises because (1) the abundances of ecologically similar taxa commonly show high correlations or multicollinearity (Montgomery and Peck, 1982), resulting in the regression parameter estimates having highly inflated variances and being unstable and far from their target values, (2) there are often many (>100) taxa in the training and fossil sets, and (3) there are invariably many zero values in these data. Under these circumstances, particularly having unstable regression coefficients, predictions using a regression model are often poor and sometimes not possible. Moreover, inverse regression assumes a basically linear response model for taxa and their environment, an assumption that is contradicted by the observation that taxa have ecological optima and generally show unimodal responses to their environment.

Although used in its basic form to reconstruct past climates from pollen-analytical data for relatively limited geographical areas by, for example, Webb and Clark (1977), Andrews *et al.* (1980, 1981), Heusser and Streeter (1980), Bernabo (1981), Andrews and Nichols (1981), Swain *et al.* (1983), Kay and Andrews (1983), Norton *et al.* (1986), and D'Antoni (1993), to calibrate modern pollen and remote-sensing data by D'Antoni and Spanner (1993), and to relate modern biological and limnological data by Whitmore (1989, 1991), Brenner *et al.* (1993), Binford *et al.* (1987), and ter Braak and van Dam (1989), various approaches have been developed to make inverse linear regression more robust and reliable statistically by overcoming some of the problems inherent in the method. These approaches include restricted inverse linear regression, principal components regression, and partial least squares regression.

A variant of inverse linear regression commonly used in palaeolimnology that avoids the problems of multicollinearity inherent when using large numbers of biological taxa as predictor variables is to group the taxa *a priori* into ecological categories, such as pH ecological groups defined from the literature or from an initial cluster analysis of the taxa (e.g. Charles, 1985; Davis and Anderson, 1985; Flower, 1986, Dixit, 1986, Charles and Smol, 1988; ter Braak and van Dam, 1989).

Restricted inverse linear regression

Restricted inverse linear regression (= segmented inverse linear regression - ter Braak, 1995) has been developed for the reconstruction of past climatic variables from pollen-analytical data over large geographical areas. Pollen types today, when expressed as percentages of some pollen sum, commonly show strongly non-linear responses to

climatic variables such as mean January and mean July temperatures and annual precipitation over large geographical areas (e.g. subcontinents) and hence over long environmental gradients (e.g. Bartlein *et al.*, 1986; Webb *et al.*, 1987; Prentice *et al.*, 1991; Anderson *et al.*, 1991; R. Webb *et al.*, 1993). In restricted inverse linear regression, an attempt is made to convert these responses into approximately linear or at least monotonic responses by (a) restricting the geographical area and (b) transforming the pollen percentage data using power transformations of -2, -1, -0.5, -0.25, log y, 0.25, 0.5, 1, and 2. The effect of restricting the geographical area so that the pollen responses are monotonic is to reduce the lengths of the underlying gradients from where taxon-environment responses are unimodal (> 2 SD units) to shorter gradients (< 2 SD units) where the responses are monotonic and approximately linear. For example, Bartlein and Webb (1985) (see also Bartlein *et al.*, 1984; Bartlein and Whitlock, 1993) subdivided eastern North America into subregions for inverse regression purposes and then derived separate inverse regression equations for particular environmental variables in each subregion. The regression equation used for reconstruction of a particular fossil sample is selected by analogue matching using a dissimilarity measure between the fossil assemblage and the modern assemblages in the total training set to identify which modern samples are, and hence which subregion is, most similar to the fossil sample of interest. Uses of separate inverse regression equations from Bartlein and Webb (1985) and/or Bartlein and Whitlock (1993) for palaeoclimatic reconstructions at individual sites include Shane and Anderson (1993), McAndrews and Campbell (1993), and Keen and Shane (1990).

Considerable care is taken in these restricted inverse linear regressions in selecting the subregions, in the choice of pollen taxa and pollen sums to be used, in detecting and screening for outliers or "rogue" samples, in finding appropriate power transformations for different pollen taxa, in selecting the most useful set of explanatory pollen variables using the "best possible subsets" approach (Weisberg, 1985), in assessing the extent of extrapolation in reconstruction (Weisberg, 1985), and thus in explicitly examining the statistical assumptions inherent in inverse regression. Every attempt is made to minimise violations of these assumptions (Howe and Webb, 1977, 1983; Bartlein and Webb, 1985; Bartlein and Whitlock, 1993). Prediction errors are estimated either by split-sampling cross-validation (Bartlein *et al.*, 1984) or, less realistically, from the R^2 values for the individual inverse regression equations for the training subsets and from estimates of 95% prediction confidence intervals for individual inferred values based on standard equations that assume normal distributions and primarily involve the mean squared error of prediction of the modern training set (e.g. Howe and Webb, 1983; Bartlein and Whitlock, 1993).

When applied to modern eastern North America pollen data, restricted inverse linear regression resulted in 13 subregions for reconstructing mean July temperatures using a subset of 3-7 (median = 5) pollen taxa based on a pollen sum of 6-15 (median = 12) taxa (Bartlein and Webb, 1985). In Europe, Huntley and Prentice (1988) used 4 subregions and regression equations based on a subset of 1-7 pollen types (median = 5) and pollen sums including 6-15 (median = 8) taxa.

Problems in the approach include how to select the subregions, how to choose which pollen taxa to include in the pollen sum, how to select appropriate transformations of the pollen percentages, how to decide which taxa to include in the inverse regression equation, how to estimate reliable prediction errors for individual samples, how to select which subregion and hence which regression equation to use for reconstructing past environmental variables from fossil data, and how to avoid the problems of multicollinearity and hence the use of a very small and potentially ecologically unrepresentative set of taxa in the regression equations. Grimm (1988, pp. 69-70) suggests that the restricted inverse linear regression approach "probably satisfies many of the criticisms of past attempts at quantitative paleoclimatic reconstruction, but it is cumbersome", whereas Huntley (1993, p. 219) proposes that "conventional regression analysis (Huntley & Prentice, 1988) should be avoided." Restricted inverse linear regression, when used critically and with checks to safeguard against outliers, unjustified extrapolation, and non-linearities using regression diagnostics (e.g. Howe and Webb, 1983, Bartlein and Webb, 1985, Cook and Weisberg, 1982; Hocking, 1983; Weisburg, 1985; Montgomery and Peck, 1982) can provide palaeoenvironmental reconstructions that are generally consistent with reconstructions using other methods (e.g. Bartlein and Whitlock, 1993). It is, however, "cumbersome" (*sensu* Grimm, 1988) and liable to be prone to user-error. The approach has the "intrinsic merit of being statistically optimal when the assumptions underlying the approach are not violated" (Bartlein and Whitlock, 1993, p. 280). However, as Howe and Webb (1983) and Birks and Gordon (1985) discuss, the underlying assumptions cannot always be completely justified when using pollen percentage data because the pollen percentage values, used as independent explanatory variables in inverse regression, are not known without error, and the error component in the environmental (e.g. climatic) data, used as the dependent response variable, is unlikely to be statistically independent of the error term in all the other observations because of inherent spatial autocorrelation in a geographical array of climatic observations. As taxon abundances invariably show multicollinearity and their regression coefficients are unstable, and as palaeoecological data usually contain many taxa, other approaches using linear-based methods attempt to include more taxa by representing them as a few, major uncorrelated assemblages, such as principal component axes, in principal components regression.

Principal components regression

A reduction in the number of taxa used as predictors in inverse regression and a prevention of multicollinearity among the predictor variables can be achieved by carrying out a principal components analysis (PCA) of the modern biological data and then using a limited number, r say, orthogonal principal component axes ($v_1, v_2, \ldots, v_r$) as the variables to be used in the inverse regression,

$$x = b_0 + b_1 v_1 + b_2 v_2 + \ldots\ldots + b_r v_r \qquad \text{for the modern training set.}$$

Reconstructions for x_0 are obtained by first transforming the fossil biological data into principal components. This is done by multiplying the fossil data (after appropriate centrings, standardisations, and/or transformations) by the relevant eigenvectors or component loadings for the r axes for the taxa in the modern training set. Given the fossil data transformed to principal components, x_0 is reconstructed by

$$\hat{x}_0 = b_0 + b_1 v_1 + b_2 v_2 + \ldots\ldots + b_r v_r \qquad \text{for fossil sample 0.}$$

This approach is known as principal components regression (PCR) (Jolliffe, 1986; Næs and Martens, 1988; Morzuch and Ruark, 1991; Jackson, 1991).

PCA redefines the original variables in the biological data as linear combinations (PCA axes), concentrates the major patterns of variation within the biological data into the first few components, and relegates the less coherent patterns and "noise" in the data to the later components (Gauch, 1982; ter Braak, 1987d; Kovach, this volume). The principal components are mutually orthogonal which means that each principal component used as a predictor variable in the inverse regression is, by definition, uncorrelated to the other predictor variables. Multicollinearity thus does not arise in PCR.

PCR has been used to reconstruct past climate from pollen-analytical data by Cole (1969), Webb and Clark (1977), Kay (1979), Kay and Andrews (1983), and Norton *et al.* (1986).

While PCA may provide an informative low-dimensional summary of modern biological data when the data have compositional gradients of 2 or less SD units (ter Braak and Prentice, 1988), it is not guaranteed that the major directions of variation in the biological data will necessarily always represent the most important factors for predicting particular environmental variables (Birks and Gordon, 1985). In an attempt to enhance the interpretability of the principal components, the first r components can be further rotated according to the "varimax" criterion (Davis, 1986). The aim is to re-orientate the axes so that each has high loadings for a few dominant taxa. Varimax rotation leaves the r-dimensional configuration of samples unaltered but it can make the individual axes easier to characterise in terms of a few dominant taxa (Prentice, 1980).

A variant of PCR that has been widely used in palaeoenvironmental reconstructions is the method of Imbrie and Kipp (1971). It involves a PCA of a *non*-centred cross-products matrix between variables after each object in the biological data-set is normalised to unit sum-of-squares and a subsequent varimax rotation around the zero point to derive "assemblages" for subsequent inverse regression. The normalisation transforms each sample into a vector of unit length, thereby changing the implicit Euclidean distance measure in conventional PCA into a cosine-theta distance of $(2 - 2\cos\theta)^{1/2}$ where $\cos\theta$ is the cosine-theta coefficient of proportional similarity (Klovan and Imbrie, 1971) and θ is the angle between the vectors representing samples (Davis, 1986). This distance measure gives maximum weight to large-valued taxa and little or no weight to small-valued taxa. This so-called "Q-mode factor analysis" of Klovan and Imbrie (1971) produces component axes dominated by the large-valued taxa. The Imbrie and Kipp (1971) approach of Q-mode factor analysis and inverse regression is widely used in marine palaeoecology (e.g. Kipp, 1976; Imbrie *et al.*, 1982; Hutson, 1977, 1978; Morley, 1979, 1989a; Molfino *et al.*, 1982; Karpuz and Schrader, 1990; Le, 1992; Schrader *et al.*, 1993) but it has only occasionally been used to reconstruct past climate from terrestrial pollen data (e.g. Webb and Clark, 1977; Andrews *et al.*, 1980, 1981; Andrews and Nichols, 1981; Heusser and Heusser, 1980, 1981; Heusser *et al.*, 1980; Mathewes and Heusser, 1981).

Besides a linear inverse regression, Imbrie and Kipp (1971) also used a curvilinear regression with cross-product and square terms of the varimax component axes being used as predictor variables in addition to the components themselves. If $r = 3$ components, the inverse regression is now

$$x = \hat{b}_0 + b_1 v_1 + b_2 v_2 + b_3 v_3 + b_4 v_1 v_2 + b_5 v_1 v_3 + b_6 v_2 v_3 + b_7 v_1^2 + b_8 v_2^2 + b_9 v_3^2$$

Thus for r components, the resulting equation contains $(r^2 + 3r)/2$ predictor variables. When many components are used, the equation becomes large and some selection of predictor variables (e.g. by forward selection) in the regression step is required.

There are four major problems in the use of PCR in palaeoenvironmental reconstructions. First, there is the question of how many principal components axes should be included in the regression (Hawkins, 1973; Mansfield *et al.*, 1977; Jolliffe, 1982; 1986; Næs and Martens, 1988; Jackson, 1991). Wartenberg (1985a, 1985b, 1985c) presents various methods for assessing the number of components to retain in an analysis of the original Imbrie and Kipp (1971) modern foraminiferal data involving spatial autocorrelation, canonical trend surface analysis, and multiple spatial correlation. Næs and Martens (1988) discuss various cross-validation approaches for PCR. Second, PCA results are inevitably influenced by the choice of data transformation and standardisation

and of axis scaling (e.g. Prentice, 1980; ter Braak, 1987d). Third, PCR is a linear-based method, as the latent variables represented as PCA axes are constructed under the assumption of an underlying linear response model between taxa and their environment (ter Braak, 1987d). It is thus only appropriate for relatively short gradients (< 2 SD units). Fourth, in PCR the components are chosen irrespective of their predictive value for the environmental variable x to represent the maximum variance within the biological data, whereas the components should be selected to maximise the covariance with the response variable x and the number of components should be decided by a cross-validation procedure (Stone and Brooks, 1990). This leads logically to partial least squares (PLS) regression, a technique developed primarily in chemometrics for calibration purposes (e.g. Næs and Martens, 1985; Næs *et al.*, 1986).

Partial least squares regression

Within chemometrics, in particular the extraction of chemical information from near infra-red spectroscopy, regression and calibration methods started, as in palaeoecology (Cole, 1969; Imbrie and Kipp, 1971) with principal components regression (PCR). However, subsequent developments followed different directions (Martens and Næs, 1989), with the development of partial least squares (PLS) regression (Wold *et al.*, 1984) in chemometrics and of restricted inverse linear regression in palaeoecology (see above).

In PCR the components are selected to capture the maximum variance within the predictor variables irrespective of their predictive value for the environmental "response" variable. In PLS regression the components are chosen to maximise the covariance with the response variable (Stone and Brooks, 1990). As a result, PLS usually requires fewer components and gives a lower prediction error than PCR. Both PCR and PLS are biased inverse regression methods that guard against multicollinearity among predictor variables by selecting a limited number of uncorrelated orthogonal components. The number of PLS components is estimated by cross-validation on the basis of the empirical predictive power, i.e. the prediction error sum-of-squares (PRESS) criterion (Osten, 1988). This is necessary to avoid developing a regression model with so many parameters that, although fitting the data perfectly, may have little predictive power (Martens and Næs, 1989). Ter Braak and Juggins (1993) show with a simulated training set and an independent evaluation test set how the optimal number of PLS components should not be based on model fit and "apparent" errors but on the prediction error derived by cross-validation ($RMSEP_{(jack)}$).

PLS and PCR are the two most used methods for biased regression, whereas ordinary least-squares regression is the most used unbiased regression technique (Martens and Næs, 1989). PCR and PLS are termed biased methods because some of the data are

discarded and so a biased least-square estimate of the model parameters is calculated (Kowalski and Seasholtz, 1991). For a single response variable, the three techniques fall within Stone and Brooks's (1990) more general continuum regression. Depending on the value of a parameter (α), continuum regression ranges from one extreme ($\alpha = 0$; ordinary least squares) to the other extreme ($\alpha = 1$; PCR), via the intermediate PLS ($\alpha = 0.5$). Each seeks linear combinations of the predictor variables to maximise a particular mathematical criterion with the restriction that the linear combinations or components must be orthogonal. In least-squares regression, the criterion being maximised for the sole linear combination is its correlation with the response variable. In PCR the components do not depend on the response variable and the maximisation criterion is the variance of the components of the predictor variables. In PLS the criterion for each component is the covariance of the response variable with the PLS component. PLS is thus, in some ways, a compromise between inverse regression and PCR. PLS probably derives its generally good empirical performance (e.g. Thomas and Haaland, 1990; Sjöström *et al.*, 1983; ter Braak *et al.*, 1993) by combining the desirable properties of inverse regression (high correlation) and PCR (stable predictors of high variance) into one technique (de Jong, 1993). A PLS model always gives a better fit, as assessed by the coefficient of determination (r^2), than a PCR model with the same number of components (de Jong, 1993).

It is surprising, therefore, that PLS has hardly been used for palaeoenvironmental reconstructions, with the exception of the methodological studies of ter Braak and Juggins (1993) and ter Braak *et al.* (1993), and by Korsman *et al.* (1992) for calibrating near infra-red spectra of lake sediments to pH, and by Korsman (1994) and Korsman and Birks (1995) for calibrating and diatom assemblages to pH and other water chemistry variables. PLS has also been used to model methane and carbon dioxide concentrations in mires in relation to peat composition (Nilsson and Bohlin, 1993).

Good introductions to PLS include Beebe and Kowalski (1987), de Jong (1991), Martens and Næs (1984, 1989), Næs and Martens (1984), Geladi and Kowalski (1986), and Haaland and Thomas (1988a). Reviews of its algebraic and conceptual relationships to other statistical techniques are given by Lorber *et al.* (1987), Frank and Friedman (1983), Stone and Brooks (1990), Geladi (1988), Osborne (1991), Helland (1988, 1990), Kowalski and Seasholtz (1991), and de Jong and Kiers (1992).

Ter Braak and Juggins (1993) compare the basic algorithms of PCA regression and PLS regression and show that there are subtle but important differences between the two methods. In PLS the values of the environmental variable are used at the outset whereas in PCA these values are not used initially. The environmental variable is regressed on the current PLS component of the predictor variables and the fitted values are used as the

current estimates of x, whereas in PCR the fitting is done after the PCA. The residuals of this regression are then used in PLS as new initial sample scores for the extraction of the next component. Each component is made uncorrelated with previous components by orthogonalisation. The number of components retained is that number that minimises PRESS in cross-validation by jack-knifing. If as many PLS components are extracted as there are taxa, PLS reduces to a multiple regression of x on all the predictor taxon variables.

In addition to univariate PLS (PLS-1) discussed above involving one environmental variable, there is multivariate PLS (PLS-2) where two or more environmental variables are considered simultaneously. Linear-based methods that consider more than one environmental variable in relation to many multivariate predictor variables include PLS-2, canonical correlation analysis, and redundancy analysis.

Canonical correlation analysis, redundancy analysis, and multivariate partial least squares regression

There are several techniques available for considering simultaneously the linear relationships between multivariate biological and multivariate environmental data, namely canonical correlation analysis (CCoA), redundancy analysis (RDA), and multivariate partial least squares regression (PLS-2). They have all been used occasionally for the simultaneous reconstruction of two or more environmental variables from biological data.

The best known of these is canonical correlation analysis (Gittins, 1985). It is the standard linear multivariate technique for relating two sets of variables simultaneously. Given a set of m biological variables (y_1, y_2, ..., y_m) and a set of p environmental variables (x_1, x_2, ..., x_q), the linear combinations of

$$g_j = \sum_{k=1}^{m} a_{jk} y_j \qquad \text{and} \qquad h_j = \sum_{k=1}^{q} b_{jk} x_j$$

are sought, where j = 1, 2, ..., for which the correlation between g_j and h_j is maximised, subject to g_j and h_j being uncorrelated with all other pairs of linear combinations. The canonical variable pairs (g_j, h_j (j = 1, 2, ...)) indicate which linear combinations of biological and environmental variable are most clearly related to one another. The jth environmental canonical variable is then linearly regressed onto the jth biological canonical variable. Since each environmental variable can be expressed in terms of the environmental canonical variables (h_j), the environmental variables are modelled in terms of the biological canonical variables. By restricting attention to the first few pairs

of canonical variates, it is hoped to derive a statistical model for the environmental variables based on the relevant variation in the biological data (e.g. Varis, 1991).

The mathematics of this combined canonical correlation and regression analysis are presented in the biological - environmental context by Webb and Bryson (1972) and Webb and Clark (1977) (see also Glahn, 1968). Because the scores for the biological variables are parameters estimated by multiple regression of the sample scores on the biological variables, there is an important practical limitation in CCoA, namely that the number of biological variables and the number of environmental variables must be less than the number of samples (Gittins, 1985; ter Braak, 1987c, 1987d). For many biological data-sets this is too restrictive and an alternative approach without this restriction is provided by redundancy analysis (ter Braak, 1994). Despite this restriction and its underlying assumption of linear responses between the two sets of variables (e.g. pollen and climate), CCoA has been used to reconstruct past climate from pollen-stratigraphical data by Webb and Bryson (1972), Webb and Clark (1977), Kay (1979), and Bryson and Swain (1981), and to "enhance the climatic signal" in pollen-stratigraphical data prior to applying a modern analogue technique for reconstruction purposes (Peng *et al.*, 1994).

The closely related but less well known technique of redundancy analysis (RDA) (van den Wollenberg, 1977; Israëls, 1984; ter Braak, 1987d, 1994) has no such restriction and there is no theoretical limit on the number of biological variables that can be analysed in RDA (ter Braak and Prentice, 1988). RDA provides an ordination of the biological data in which the ordination axes are constrained to be linear combinations of the environmental variables. It is, in effect, a constrained PCA, hence its alternative names of PCA of instrumental variables (Rao, 1964) and PCA of y with respect to x (Robert and Escoufier, 1976). It is also a constrained form of (multivariate) multiple regression, hence another of its alternative names is reduced-rank regression (Davies and Tso, 1982). For m biological variables and p environmental variables, RDA uses $2(p + m) + m$ parameters to describe the biological data whereas multiple regression uses $m(p + 1)$ parameters (ter Braak and Prentice, 1988). In RDA, linear combinations of independent environmental variables are formed that account for successively maximal proportions of the total sum of squares over the set of biological response variables, i.e. that give the maximum total sum of squares possible and hence the smallest total residual sum-of-squares, whereas in CCoA linear combinations are formed that maximise the multiple correlation coefficient between the two sets. At a more theoretical level, RDA also differs from CCoA in its assumption about error components. Independently and identically normally distributed errors with equal variance are assumed in the least-squares-based RDA, whereas correlated errors following a multivariate normal

distribution are assumed in the maximum-likelihood-based CCoA (Tso, 1981; ter Braak, 1987c; 1987d; van der Meer, 1991).

An important concept in RDA is the index of redundancy introduced by Stewart and Love (1968), namely the proportion of the total sum-of-squares in the biological data accounted for by the linear prediction of Y by X. For example, Bryson and Kutzbach (1974) used the index of redundancy to show that although four canonical variables account for 89% of the variance in the five-variable climate data-set, only 48% of the variance in the ten-variable modern pollen data-set was accounted for by these four canonical variables. The remaining 52% of the variance in the pollen data may reflect non-climate factors such as soils, natural disturbance, human activities, and local site taphonomic processes or it may result from the non-linear relationships between modern pollen and regional climate (Webb and Clark, 1977).

Although RDA has been used occasionally in palaeoecology (e.g. Odgaard, 1992; Lotter and Birks, 1993), it has not, as far as I know, been used for palaeoenvironmental reconstructions. Van Dobben and ter Braak (1993) have, however, developed a form of RDA to infer simultaneously temporal changes in several environmental variables from historical biological data from the same localities as the modern data or from an appropriate subset (see also ter Braak and Wiertz, 1994). This approach focuses on the magnitude of environmental change rather than quantitative reconstructions of particular environmental variables. A preliminary calibration function is derived initially from modern biological and environmental data. It is then applied to historical biological data to infer environmental variables for the past. These are then used with the modern data to "improve" the calibration function by imposing constraints on the environmental change, for example the same magnitude of change has occurred at all localities at a particular time or that the magnitude of change is constant within regions but may differ between regions. Van Dobben and ter Braak (1993) used this approach to infer the magnitude of change in atmospheric SO_2 and NH_3 concentrations in different regions from changes in the epiphytic lichen flora. Ter Braak and Wiertz (1994) reconstructed pH changes for different changes in water table from changes in fen-grassland vegetation. These approaches are clearly applicable in fine-resolution local-scale palaeoecological studies under the assumptions that at these scales the relationship between taxon abundance and the relevant environmental variables is linear, that the environment changes linearly with time, and that the relative environmental changes are constant in space. These assumptions can, however, be relaxed, in particular the assumptions of linearity of change with time and of spatial constancy with time (van Dobben and ter Braak, 1993). The approach is attractive because it is a classical multivariate calibration method - classical because it uses a specific vegetation-

environment response model and multivariate because changes in more than one environmental variable are estimated simultaneously (van Dobben and ter Braak, 1993).

The multivariate version of partial least squares regression, here abbreviated to PLS-2, uses a multivariate set of environmental variables as "response" variables, thereby permitting the modelling of two or more environmental variables simultaneously in relation to the biological data (Geladi, 1988; Martens and Næs, 1989). The algorithms for PLS-2 are iterative for each component, because the linear combinations of the two sets of variables have to be optimised with respect to each other (Manne, 1987; Höskuldsson, 1988). Martens and Næs (1989) discuss the situations when PLS-2 could be useful in calibration. These are (1) an exploratory data analysis including all p environmental variables rather than doing p separate PLS-1 analyses, and (2) if the environmental variables show strong intercorrelations, as PLS-2 uses this correlation structure to stabilise the solution against the inherent random noise in the variables. Non-linearities in the biological-environmental responses may, however, be a major problem, especially if different biological variables show different non-linear relationships with different environmental variables. In such instances PLS-2 may find a suboptimal compromise solution between the different environmental variables. When this occurs, Martens and Næs (1989) recommend using separate PLS-1 analyses for each environmental variable of interest in the final calibration. Korsman and Birks (1995) found little or no improvement in RMSEP for lake pH, alkalinity, or colour predicted from modern diatom assemblages when PLS-2 was used instead of three separate PLS-1 analyses for each chemical variable. Ter Braak *et al.* (1993) emphasise that PLS-2 is, in reality, a form of multivariate inverse regression with safeguards for multicollinearity in the predictor variables rather than being a technique for multivariate classical calibration. There is probably little or no statistical benefit in using PLS-2 or its close relative principal covariates regression (de Jong and Kiers, 1992) for calibration in palaeoenvironmental reconstructions.

Other linear approaches for inferring values for more than one environmental variable simultaneously based on the principle of maximum likelihood are considered by Brown (1982), Sundberg and Brown (1989), and Brown and Sundberg (1987).

NON-LINEAR UNIMODAL (GAUSSIAN)-BASED TECHNIQUES

Introduction

All the methods considered so far assume that the underlying response of taxa to the environmental variable of interest is linear or at least monotonic (Figure 6.2b). This

assumption is generally valid over short (< 2 SD units) environmental gradients (ter Braak and Prentice, 1988). Over longer gradients, taxa commonly exhibit a non-linear unimodal response (Figure 6.2a), with a maximum abundance at some value of the environmental variable and with declining abundances either side of this maximum. A variety of regression and calibration procedures are available that specifically assume such a unimodal response. These include maximum likelihood loglinear and logit regression and calibration, weighted averaging regression and calibration, weighted averaging partial least squares regression, and inverse regression procedures based on correspondence analysis.

Maximum likelihood regression and calibration

The basic idea is that the relationship between the abundance of a taxon (y) and an environmental variable (x) can be modelled by an ecological response curve consisting of systematic and random (error) components. Such a curve is fitted to the modern training set by non-linear regression. The response curves and their assumed error structure form a classical regression statistical model of biological composition in relation to the environmental variable of interest. The curves for all taxa determine jointly what biological composition and abundances are expected at a given value of x. This model of responses and their error structure can then be used to calculate the probability that a particular value of x would occur with a given biological assemblage over the range of possible values of x. The value of x that gives the highest probability is the maximum likelihood (ML) estimate. From a theoretical viewpoint, ML regression and calibration comprise the most statistically rigorous approach for environmental reconstruction. They form a classical approach to reconstruction (ter Braak, 1995). Such ML techniques require an explicit response model, consisting of a systematic part and a random part.

If we consider only one taxon (y) and one environmental variable (x), the simple linear response model of

$$Ey = b_0 + b_1 x$$

can be extended to a unimodal model by adding a quadratic term (x^2) to the regression equation, namely

$$Ey = b_0 + b_1 x + b_2 x^2$$

This has the effect of changing the response curve from a straight line to a parabola (ter Braak and Looman, 1987). Such a polynomial regression has the serious disadvantage that it can predict negative values for the taxon's abundance, whereas, in reality, all

abundance values are either zero or positive. If a parabola is fitted to log-transformed taxon abundances, namely

$$E(\log y) = b_0 + b_1 x + b_2 x^2$$

the model being fitted to the *original* taxon abundances is the Gaussian response model (Gauch and Whittaker, 1972, ter Braak and Looman, 1987) if $b_2 < 0$. If $b_2 > 0$, the fitted curve has a minimum rather than a maximum. This regression equation has little ecological meaning, but it can be more easily interpreted by transforming it in terms of u, t, and c (Figure 6.2a) (ter Braak and Looman, 1986; ter Braak and Prentice, 1988), namely

$$E(\log y) = a - \tfrac{1}{2}(x-u)^2/t^2$$

where a is a coefficient related to the height of the curve's peak (ter Braak and Looman, 1986). Estimates of u, t, and c can be obtained as

$u = -b_1/(2b_2)$ (u = taxon's optimum)

$t = 1/(-2b_2)^{1/2}$ (t = taxon's tolerance)

$c = \exp(b_0 + b_1 x + b_2 x^2)$ (c = curve maximum = exp (a))

Approximate standard errors of the estimated optimum and tolerance for the taxon can be derived from the variances and covariances of b_1 and b_2, and approximate 95% confidence intervals for the estimated optimum calculated (ter Braak and Looman, 1986, 1987).

In palaeoecology nearly all quantitative biological data are expressed as proportions or percentages. As the Gaussian response model considers absolute abundances, it is not appropriate for relative compositional data because it does not take account of the constraint that the abundances of all taxa sum to 1 or 100%. Instead a multinominal logit model (MLM) should be used (ter Braak and van Dam, 1989; ter Braak *et al.*, 1993).

The equation for the Gaussian response curve is the model for absolute abundances, namely

$$f_k(x) = c_k e^{-\frac{1}{2}(x-u_k)^2/t_k^2}$$

where $f_k(x)$ is the expected value of the abundance of taxon k as a function of the environmental variable x. We can obtain a model for proportions by dividing the

absolute abundances (for any given x) by their sum (Ihm and van Groenewoud, 1984). We obtain a multinomial logit model (MLM) (McCullagh and Nelder, 1989) of the form

$$y_k(x) = \frac{c_k e^{-\frac{1}{2}(x-u_k)^2/t_k^2}}{\sum_{i=1}^{m} c_i e^{-\frac{1}{2}(x-u_i)^2/t_i^2}}$$

where $y_k(x)$ is the expected proportional abundance of taxon k as a function of x (ter Braak and van Dam, 1989). Compared with the Gaussian model, the MLM is difficult to fit and its parameters are difficult to interpret because of indeterminancies (ter Braak, 1987c).

An alternative compromise approach is to fit a Gaussian logit model (GLM) (ter Braak and Looman, 1986, 1987), namely

$$f_k(x) = \frac{c_k e^{-\frac{1}{2}(x-u_k)^2/t_k^2}}{1 + c_k e^{-\frac{1}{2}(x-u_k)^2/t_k^2}}$$

This recognises that the each taxon's abundances must lie between 0 and 1 but it does not guarantee that the total is exactly 1 for each sample (ter Braak and van Dam, 1989). The GLM is usually applied to presence/absence data (ter Braak and Looman, 1986, 1987), but it can be used as a quasi-likelihood (*sensu* McCullagh and Nelder, 1989) model for proportional data (ter Braak and van Dam, 1989; Birks *et al.*, 1990a) and as approximation to the more complex MLM.

The fitting of multinomial logit and Gaussian logit curves is relatively straightforward as both the MLM and the GLM can be fitted by generalised linear modelling (McCullagh and Nelder, 1989; ter Braak and Looman, 1987; Brew and Maddy, this volume). In generalised linear models one can specify "link functions" and error distributions. No initial estimates are needed and convergence problems rarely occur (ter Braak and Prentice, 1988).

MLM can be fitted to a training set by loglinear regression (with a logarithmic link function and a multinomial error structure assumed for the random component of the response model) using GENSTAT (GENSTAT 5, 1987). This results in estimates of the taxon parameters a_k and b_k where $a_k = c_k - \frac{1}{2}u_k^2/t^2$ and $b_k = u_k/t^2$. Ter Braak *et al.* (1993) assume equal tolerances for the taxa to provide computational tractability, parsimony of parameters, and unimodality (Anderson, 1984). The hypothesis of equal tolerances can be tested by a quasi-likelihood F-test (Jørgensen, 1983). Ter Braak and van Dam (1989)

computed MLM twice under the assumption of equal tolerances and of unequal tolerances, and rejected the hypothesis of equal tolerances. However, MLM with unequal tolerances did not noticeably improve the prediction of lake-water pH from diatom assemblages based on a MLM with equal tolerances.

GLM can be fitted for each taxon by logit regression (with a logit link function and a binomial error structure assumed for the random component of the response model) using GENSTAT or GLIM (Payne, 1986). From the Gaussian logit regression coefficients (b_0, b_1, b_2), the optimum, tolerance, and height of the peak of the fitted Gaussian logit response curve (Figure 6.2a) can be calculated for each taxon, as

$$u = -b_1/(2b_2) \qquad \text{(u = taxon's optimum)}$$

$$t = 1/(-2b_2)^{1/2} \qquad \text{(t = taxon's tolerance)}$$

$$c = 1/\left(1+\exp\left(b_0+b_1x+b_2x^2\right)\right) \qquad \text{(c = curve maximum)}$$

along with approximate 95% confidence intervals for the estimated optimum and the standard error of the estimated optimum and tolerance (ter Braak and Looman, 1986, 1987).

In GLM the significance of the fit of each taxon to the environmental variable of interest can be tested against the simpler linear-logit (sigmoidal) model using a residual deviance test, and the significance of the regression coefficient b_2 tested against the null hypothesis of $b_2 \geq 0$ by means of a one-sided t-test (ter Braak and Looman, 1986, 1987). If the null hypothesis is rejected in favour of $b_2 < 0$, the taxon's optimum can be considered significant. If either the overall Gaussian unimodal model or the optimum are not significant, the linear logit model and its regression coefficient b_1 should be tested against the null model that the taxon shows no relationship to the environmental variable of interest by means of deviance and two-sided t-tests. In this way, the simplest acceptable response curve can be determined for each taxon (ter Braak and van Dam, 1989). One of the advantages of GLR is that statistical tests can be carried out on the relationship of each taxon to the environmental variable of interest. For example, Birks *et al.* (1990a) showed for 225 diatom taxa that 88 taxa have significant unimodal responses to pH, 78 taxa show significant sigmoidal curves, 53 taxa show no relationship, five taxa have unimodal curves with minima, and one taxon failed to converge.

It is not uncommon to find taxa with estimated optima lying beyond the range of sampled environmental values and with a statistically significant linear logit model. In

such cases the optimum has to be assumed to be the lowest value of x sampled for decreasing linear logit curves and the highest value of x sampled for increasing linear logit curves. Some taxa may have fitted curves with a minimum ($b_2 > 0$) instead of a maximum. For these taxa, it is advisable to fit a linear logit model and estimate their optima as the lowest or highest values of x in the training set for decreasing or increasing curves, respectively. Tolerances can only be estimated for taxa with unimodal response curves.

In ML calibration (ter Braak and Barendregt, 1986) the aim is to infer an unknown value of x_0 from a biological assemblage. The response curves of the taxa and the assumed error structure are used to find the value of x with the highest probability of producing the observed biological assemblage in fossil sample 0. In MLM (ter Braak *et al.*, 1993) x_0 is estimated by loglinear regression with b_k as the values for the predictor variable and a_k as offset (a predictor with a unit regression coefficient). Sampling errors in a_k and b_k are ignored (ter Braak *et al.*, 1993). In GLM, the expected abundance (U) of a single taxon at value x of the environmental variable is (Juggins, 1992)

$$U = 1/\left(1+e^{-\left(b_0+b_1x+b_2x^2\right)}\right)$$

where b_0, b_1, and b_2 are parameters of the GLM regression coefficients for the taxon. The probability of observing a taxon with a proportion y at a value x of the environmental variable when proportion U is expected is given by the likelihood function (Juggins, 1992)

$$\mathrm{L} = \mathrm{U}^{y}\,(1 - \mathrm{U})^{(1-y)}$$

This is the contribution of a single taxon. The total likelihood for sample 0 is given by the product of the likelihood for all taxa. The value of x that maximises the overall likelihood is the value of the environmental variable with the highest probability of producing the observed biological assemblage. In practice, it is usual to maximise the logarithm of this function, the log-likelihood function ℓ, summed over all taxa (Juggins, 1992).

$$\ell = \sum_{k=1}^{m}\left(y_k \log(\mathrm{U}_k) + (1-y_k)\log(1-\mathrm{U}_k)\right)$$

This equation can be maximised by an iterative procedure, for example a Gauss-Newton numerical optimisation procedure with Gallant's (1975) chopping rule for step-shortening (Line *et al.*, 1994). If the estimate fails to converge, a direct-search algorithm can be used that simply evaluates the log-likelihood function at intervals throughout the

likely range of the environmental variable being reconstructed (Line *et al.*, 1994). The point giving the largest value of the function is the estimate for x_0. As long as the function does not have any really abrupt changes of slope, the direct-search estimate will be a reasonable approach to the true result. This may, of course, fall between the points at which the function is evaluated, so a small step-size should be used in such direct searches. An alternative calibration procedure is outlined by Oksanen *et al.* (1990) as FORTRAN-style pseudocode for use with a non-linear regression program.

The predictive powers of classical calibrations such as MLM or GLM, as assessed by RMSE or RMSEP in cross-validation, are usually enhanced by a shrinkage regression, for example by regressing x_i on $\hat{x}_i$ (ter Braak and van Dam, 1989; ter Braak *et al.*, 1993). The effect of such a regression is to shrink the values inferred by ML calibration.

MLM has been rarely used in palaeoecology. Examples are ter Braak and van Dam (1989) who inferred lake-water pH from diatom assemblages, ter Braak *et al.* (1993) who estimated sea-surface temperatures from foraminiferal assemblages, and Barber *et al.* (1994) who modelled peat macrofossil abundances in relation to historical climatic data. GLM has been used more widely but mainly in palaeolimnology with diatoms or chrysophytes (e.g. ter Braak and van Dam, 1989; Birks *et al.*, 1990a; Cumming *et al.*, 1992a; Juggins, 1992; ter Braak *et al.*, 1993 Juggins *et al.*, 1994).

Oksanen *et al.* (1988,1990) fitted the related Gaussian model with Poisson error structure to modern diatom assemblages in relation to pH. They showed that separate Poisson, binomial (= GLM), and *multinomial* fitting all gave very similar fitted curves.

ML regression and calibration are very computer-intensive, particularly with large, taxon-rich training sets. Fortunately weighted averaging regression and calibration have essentially the same aims as the ML procedures, are mathematically much simpler, are computationally fast, and perform as well or even better than their ML counterparts.

Weighted averaging regression and calibration

The basic idea behind weighted averaging (WA) (ter Braak (1987b)) is that at a site with a particular environmental variable x, taxa with their optima for x close to the site's value of x will tend to be the most abundant taxa present, if the taxa show a unimodal relationship with x. A simple and ecologically reasonable and intuitive estimate of a taxon's optimum for x is the average of all the x values for sites in which the taxon occurs, weighted by the taxon's relative abundance, namely the optimum is a weighted average, abundance weighted mean, or centroid of x. Taxon absences carry zero weight. The taxon's tolerance or amplitude can be estimated as the weighted standard deviation

or abundance weighted standard deviation of x. Estimation of the optimum (u_k) and tolerance (t_k) of taxon k is done by WA regression.

An estimate of a site's value of x is the weighted average of the optima for x for all the taxa present (WA calibration). Taxa with a narrow tolerance for x can, if required, be given greater weight in WA than taxa with a wide tolerance for x.

The WA estimate of a taxon's optimum, u_k, (WA regression) is

$$\hat{u}_k = \left(\sum_{i=1}^{n} y_{ik} x_i\right) \Big/ \left(\sum_{i=1}^{n} y_{ik}\right) = \sum_{i=1}^{n} y_{ik} x_i / y_{k+}$$

and a taxon's tolerance, t_k, is

$$t_k = \left[\sum_{i=1}^{n} y_{ik} (x_i - \hat{u}_k)^2 \Big/ \sum_{i=1}^{n} y_{ik}\right]^{1/2}$$

Note that for a unbiased statistical comparison of tolerances, the effective number of occurrences of the taxon should be taken into account (ter Braak, 1990). An appropriate measure of the effective number of occurrences of a taxon is *N2* (Hill, 1979), analogous to Hill's (1973a) *N2* diversity measure for samples that is the inverse of Simpson's diversity index. A taxon with 5 actual occurrences with values of, say, 60%, 1%, 0.5%, 0.2%, and 0.1% will have its WA optimum effectively determined by the sample in which it occurs with an abundance of 60%. The *N2* for this taxon is thus close to 1. For presence-absence data, *N2* is simply the actual number of occurrences. For quantitative data, *N2* lies between 1 and the actual number of occurrences. WA tolerances can be corrected for bias by dividing t_k by $(1\text{-}1/N2)^{1/2}$ (Line *et al.*, 1994).

The estimated optima can be used to infer a site's value of x from its biological assemblage (WA calibration) by:

$$\hat{x}_i = \left(\sum_{k=1}^{m} y_{ik} \hat{u}_k\right) \Big/ \left(\sum_{k=1}^{m} y_{ik}\right)$$

whereas a tolerance - weighted estimate ($WA_{(tol)}$) of x would be:

$$\hat{x}_i = \left(\sum_{k=1}^{m} \frac{y_{ik} \hat{u}_k}{t_k^2}\right) \Big/ \left(\sum_{k=1}^{m} \frac{y_{ik}}{t_k^2}\right)$$

Weighted averaging regression and calibration represent an inverse approach to palaeoenvironmental reconstructions (ter Braak, 1995).

The theory of WA and the conditions under which WA approximates ML are presented and explored by ter Braak (1985, 1987a, 1987b), ter Braak and Looman (1986, 1987), ter Braak and Barendregt (1986), and ter Braak and Prentice (1988). For quantitative abundance data, as commonly occur in palaeoccology, WA regression efficiently estimates the optimum of a Gaussian response curve (Figure 6.2a) if the abundances are Poisson distributed and the sites are equally spaced over the whole range of the taxon's occurrences (ter Braak and Looman, 1987). A WA calibration estimate approximates a ML estimate if the taxon abundances are Poisson distributed, and the response curves are Gaussian with homogeneously distributed optima and equal tolerances and maxima (ter Braak and Barendregt, 1986). These conditions combine to define a species-packing model (ter Braak, 1987a), which is an ecological model based on the idea that taxa evolve to occupy maximally separate niches with respect to some limiting environmental resource. Such a model is unlikely to hold in real life. However, the theoretical basis of WA indicates the kinds of situation when WA will perform reasonably well, namely when taxa are **not** tightly clumped along an environmental gradient and there is a fairly even turnover of taxa along the gradient (ter Braak, 1987a).

In WA environmental reconstructions, averages are taken twice, once in WA regression and once in WA calibration. This results in shrinkage of the range of inferred values towards the mean of x. To correct for this, a simple linear "deshrinking" can be done in one of two ways (Birks *et al.*, 1990a; ter Braak and Juggins, 1993; Osborne, 1991).

(1) Inverse regression where the observed values, x_i, are regressed on the initial inferred values x_i for the modern training set using the linear regression model:

$$x_i = b_0 + b_1(\text{initial}x_i) + \varepsilon_i$$

and

$$\text{final } \hat{x}_i = b_0 + b_1(\text{initial}x_i)$$

where b_0 is the intercept and b_1 is the slope of the linear regression. Inverse regression minimises the root mean squared error in the training set (Krutchkoff, 1967; Lwin and Maritz, 1982; ter Braak and van Dam, 1989; Martens and Næs, 1989), but with the danger of introducing bias at the ends of the gradient (ter Braak and Juggins, 1993).

(2) Classical regression where the initial inferred values x_i for the training set are regressed on the observed values, x_i;

$$\text{initial } x_i = b_0 + b_1 x_i + \varepsilon_i$$

and

$$\text{final } \hat{x}_i = (\text{initial } x_i - b_0)/b_1$$

Classical regression deshrinks more than inverse regression (Martinelle, 1970) and it takes inferred values further away from the mean of the training set. The residuals are orthogonal to the inferred values ($\hat{x}_i$) and hence are uncorrelated with the original observed values (x_i). In contrast, in an inverse deshrinking the residuals are orthogonal to the original inferred values (initial x_i), not the observed values. Because inverse deshrinking takes the inferred values nearer to the mean of the training set than in classical deshrinking, the inferred values often tend to be overestimated at low values of x_i and underestimated at high values of x_i (ter Braak and Juggins, 1993). Ways of overcoming this bias in inverse deshrinking include non-linear deshrinkage (Wold, 1992; A. Marchetto, 1994; C.J.F. ter Braak, personal communication) and weighted averaging partial least squares (ter Braak and Juggins, 1993; ter Braak *et al.*, 1993).

Within WA regression and calibration, classical deshrinking may be appropriate for some variables and some data-sets (Birks *et al.*, 1990a), whereas inverse regression may be more appropriate for other variables and data-sets. The choice often depends on the part of the gradient of interest. If greatest accuracy is required at high or low values of the environmental variable being reconstructed, classical deshrinking is appropriate. If the emphasis is on inferences in the mid-range of x, inverse regression should be used (Gasse *et al.*, 1995).

Because the computations involved in WA regression, WA calibration, and inverse or classical deshrinking are simple and, if programmed efficiently, very fast, it is possible to use the computer-intensive procedure of bootstrapping (Efron, 1982; Efron and Gong, 1983; Diaconis and Efron, 1983; Efron and Tibshirani, 1993) to estimate RMSEP for inferred values of x_i for all modern training-set samples, for the training set as a whole, and for all individual fossil samples.

The idea of bootstrap error estimation is to do many bootstrap cycles, say 1000. In each cycle, a subset of training samples is selected randomly but with replacement from the original training set to form a bootstrap training set of the same size as the original training set. This mimics sampling variation in the training set. As sampling is with replacement, some samples may be selected more than once in a cycle. Any samples not selected for the training set form a bootstrap **test** set for that cycle. WA regression and calibration are then used with the bootstrap training set to infer the environment variable of interest (x_i) for the modern samples (with known observed environmental variables) in the bootstrap **test** set. In each cycle, WA calibration is also used to infer the environmental value, x_0, for each fossil sample. The standard deviation of the inferred

values for both modern and fossil samples is calculated. This comprises one component (s_1) of the prediction error, namely that part due to estimation error in the optima and tolerances of the taxa. The second component (s_2), the error due to variations in the abundance of taxa at a given environmental value, is estimated from the training set by the root mean square (across all training samples) of the difference between observed x_i and the mean bootstrap estimate of $\hat{x}_i$ when that modern sample is in the bootstrap test set. The first component varies from fossil sample to fossil sample, depending on the composition of the fossil assemblage, whereas the second component is constant for fossil samples. The estimated RMSEP for a fossil sample is the square root of the sum of squares for these two components. The same procedure can be applied to each modern training sample, except that s_2 varies among the modern samples, to derive sample-specific RMSEP. The underlying theory is presented by Birks *et al.* (1990a), and applications include Birks *et al.* (1990a, 1990b), Cumming *et al.* (1992a, 1992b), Kingston *et al.* (1992), Dixit *et al.* (1993), Cumming and Smol (1993a, 1993b), Bennion (1994), and Wilson *et al.* (1994).

"Hybrid" procedures can be used with taxon optima and tolerances estimated by ML Gaussian logit regression, followed by WA calibration (with and without tolerance downweighting) (e.g. Oksanen *et al.*, 1988; Birks *et al.* 1990a; Juggins, 1992). The use of ML regression to derive optima and tolerances allows one to identify those taxa with a statistically significant Gaussian or sigmoidal relation to the environmental variable of interest.

Despite the heuristic nature (ter Braak and Prentice, 1988) of WA regression and calibration, in nearly all comparative studies results from WA are as good as or even superior to results from the computationally intensive and difficult ML in terms of RMSE (e.g. ter Braak and van Dam, 1988; Birks *et al.*, 1990a; Kingston and Birks, 1990; Juggins, 1992; Cumming *et al.* 1992a; ter Braak *et al.*, 1993; Juggins *et al.*, 1994).

Although the underlying theory of WA has only recently been explored (ter Braak, 1987b), WA regression and/or calibration have a long history in ecology (e.g. Gause, 1930; Ellenberg, 1948), palaeoecology (Lynts and Judd, 1971; Berger and Bardner, 1975), and limnology (e.g. Descy, 1979; Chutter, 1972). WA regression and calibration are now widely used in palaeolimnology and other areas of palaeoecology for the quantitative reconstruction of past environmental variables from fossil assemblages. Examples include soil and land-use variables reconstructed from pollen (Gaillard *et al.*, 1992, 1995), lake-water pH reconstructed from diatoms or chrysophytes (e.g. Birks *et al.*, 1990a, 1990b, Dixit *et al.*, 1991; Kingston *et al.*, 1992; Cumming *et al.*, 1992b), lake salinity reconstructed from diatoms (Fritz *et al.*, 1991, Juggins *et al.*, 1994), lake-water total P or N reconstructed from diatoms (Anderson *et al.* 1993; Agbeti, 1992, Christie

and Smol, 1993; Hall and Smol, 1992; Fritz *et al.*, 1993; Anderson and Rippey, 1994), lake-water Al, Ni, Ca, and conductivity reconstructed from diatoms (Dixit *et al.*, 1991), estuarine salinity reconstructed from diatoms (Juggins, 1992), mire-water pH and mire water-table reconstructed from bryophytes (Janssens *et al.*, 1992), lake-water temperatures reconstructed from chironomid assemblages (Walker *et al.*, 1991; Wilson *et al.*, 1993), sea-surface temperatures reconstructed from foraminifera (Le, 1992; ter Braak *et al.*, 1993), and bog moisture reconstructed from testate amoebae (Warner and Charman, 1994). WA generally provides robust, reliable, and rapid regression and calibration procedures for situations when the response variables show a broadly unimodal response to the predictor environmental variable of interest.

Weighted averaging regression and calibration have gained considerable popularity in the last five years for various reasons (ter Braak and Juggins, 1993). First, they combine ecological plausibility (unimodal species responses and the species-packing model) with mathematical and computational simplicity, a rigorous underlying theory, and good empirical predictive power. Second, WA does not assume linear responses of species to their environment as inverse linear regression and related linear-based methods do, it is relatively insensitive to outliers, and it is not hindered by multicollinearity between variables or by the large number of taxa common in many training sets. In fact, WA appears to perform best, in terms of RMSEP estimated by bootstrapping, with large numbers of taxa (Birks, 1994). Third, because of the computational simplicity of WA, it is possible to use computer-intensive bootstrapping procedures to derive RMSEP for individual samples. These estimates are less prone to bias than in other cross-validation procedures. Fourth, WA appears to perform well in "no-analogue" situations where the fossil assemblages differ, to some degree, in composition to the modern assemblages, as the WA inferences are based on the weighted average of the individual optima of the taxa in common between the modern and fossil assemblages. It is, in many ways, a combination of both the "indicator" species and the "assemblage" approaches to palaeoenvironmental reconstructions (Birks and Birks, 1980). Fifth, WA consistently performs as well as or even better than the theoretically more rigorous ML regression and calibration, as assessed in terms of RMSEP in cross-validation. Why this is remains unclear (Birks *et al.*, 1990a), but possible reasons are that (a) ML uses more of the data, especially the taxon absences that WA ignores and the precise percentage values, and (b) ML is more sensitive than WA to the high variability and skewness in the relative abundances of taxa. Sixth, WA appears to perform best with noisy, species-rich, compositional data including many taxa that may be absent from many of the sites and extending over a long environmental gradient (> 3 SD units) (ter Braak and Juggins, 1993). Weighted averaging does, however, have three important weaknesses (ter Braak and Juggins, 1993). First, WA is sensitive to the distribution of sites within the training set along the environmental gradient of interest (ter Braak and Looman, 1986). Second,

WA considers each environmental variable separately. Third, WA disregards the residual correlations in the biological data, namely correlations that remain in the biological data after fitting the environmental variable and that are often caused by environmental variables that are not considered in WA. The incorporation of partial least squares regression into WA regression (ter Braak and Juggins, 1993; ter Braak *et al.*, 1993) helps to overcome some of these weaknesses by considering residual correlations in the biological data in an attempt to improve estimation of the optima for the taxa.

Correspondence analysis regression

Correspondence analysis (CA) (= reciprocal averaging; Hill, 1973b; ter Braak, 1987d; Kovach this volume) is the unimodal-based equivalent of PCA. It is appropriate with biological data that have a compositional gradient of 2 or more standard deviation units (ter Braak and Prentice, 1988). CA is an extension of weighted averaging except that no known environmental variables are used in the analysis. It is thus, like PCA, an indirect gradient analysis procedure (ter Braak and Prentice, 1988), in contrast to direct gradient analysis methods such as regression and constrained ordination techniques such as RDA, CCoA, PLS, and canonical correspondence analysis.

A reduction in the number of taxa used as predictors in an inverse multiple linear regression can be achieved by CA of the modern biological data and then using a limited number of orthogonal CA axes as the predictor variables in the regression. The regression and reconstruction steps in CA regression (CAR) are identical to PCR (see above) except that CA is used instead of PCA and the fossil biological data are transformed into CA sample scores prior to reconstruction. CAR is an inverse regression procedure.

In a re-analysis of the Imbrie and Kipp (1971) modern foraminiferal data from the Atlantic used by Imbrie and Kipp to develop their variant of PCR, Roux (1979) used CA instead of non-centred PCA to allow for the unimodal response of species to their environment (ter Braak, 1985) and for the compositional nature of the data. Subsequent inverse regression using 3 CA axes produced better inferences (lower RMSE) of sea-surface temperatures (RMSE = 1.72°C for summer temperature, 1.37°C for winter temperature), at least in the training set, than Imbrie and Kipp's modified PCR (2.55°C and 2.57°C, respectively).

Despite Roux's (1979) work, CAR has hardly been used for reconstruction purposes in palaeoecology. Examples include Rousseau (1991) and Rousseau *et al.* (1994) working with terrestrial mollusca in loess profiles and Juggins *et al.* (1994) working with diatoms in saline lakes. Huttunen and Merilāinen (1986), Turkia and Huttunen (1988), and Davis

et al. (1990) used the first axis of a DCA (Hill and Gauch, 1980) of modern diatom data in an inverse regression with pH or alkalinity as the "response" variable. Surprisingly they did not use other DCA axes. The first DCA axis is identical to the first CA axis, so these studies are, in fact, examples of CAR with only one CA axis.

Correspondence analysis regression shares many of the same problems as PCR, in particular deciding how many CA axes to include and the fact that CA axes are selected to maximise the dispersion of the taxa scores within the biological data (ter Braak, 1987d) and not to maximise the covariance with the environmental variable of interest. Clearly a procedure is required that selects axes to maximise the covariance with the environmental variable and decides how many axes to include by a cross-validation procedure, and yet has the underlying unimodal response model of CA and WA. The recently developed weighted averaging partial least squares regression provides such a technique.

Weighted averaging partial least squares regression

As discussed above, one important limitation of WA is that it ignores residual correlations among the biological data, namely correlations that remain after fitting the environmental variable of interest, x. Such correlations may result from environmental variables that are not taken into account in WA (ter Braak and Juggins, 1993). Ter Braak (ter Braak and Juggins, 1993; ter Braak *et al.*, 1993) has recently invented weighted averaging partial least squares regression (WA-PLS) to take account of any such residual correlations. WA-PLS uses the residual correlation structure in the data to improve the fit between the biological data and x in the modern training set. WA-PLS can be regarded as the unimodal-based equivalent of the linear-based PLS, just as WA of site scores is the unimodal-based equivalent of multiple linear regression, WA of taxon scores is the unimodal-based equivalent of inverse linear regression, CA (= reciprocal averaging) is the unimodal-based equivalent of PCA, canonical correspondence analysis is the unimodal-based equivalent of RDA (= canonical or constrained PCA) (ter Braak and Prentice, 1988), and CAR is the unimodal-based equivalent of PCR. Just as PLS is an inverse regression procedure for environmental reconstruction, WA-PLS is also an inverse procedure (ter Braak, 1995).

Principal components analysis finds an optimal linear combination of variables with maximum variance. CA finds an optimal WA so that the vector of weighted averages has maximum variance. Neither PCA or CA consider the environmental variable of interest to be predicted, x, so the linear combinations or optimal weighted averages do not necessarily have any strong predictive power. In PLS the first component is selected to maximise the covariance between the linear combination and x. Thus in WA-PLS the

first component is selected to maximise the covariance between the vector of weighted averages and x. Subsequent components in PLS and WA-PLS are chosen to maximise the same criterion but with the restriction that they be orthogonal and hence uncorrelated to earlier components (ter Braak *et al.*, 1993).

Environmental reconstructions by WA consist of three steps - WA regression, WA calibration, and a deshrinking regression. The final prediction formula for inferring the value of an environmental variable, x_0, from a fossil assemblage using WA and an inverse deshrinking regression is

$$\hat{x}_0 = b_0 + b_1 x_0^* = b_0 + b_1 \sum_{k=1}^{m} y_{0k}\hat{u}_k^* \Big/ \sum_{k=1}^{m} y_{0k} = \sum_{k=1}^{m} y_{0k}\hat{u}_k \Big/ \sum_{k=1}^{m} y_{0k}$$

where b_0 and b_1 are the coefficients of the deshrinking regression and $\hat{u}_k = b_0 + b_1\hat{u}_k^*$ where $\hat{u}_k^*$ are the initial estimates of the taxa optima. The final formula is a weighted average but with updated optima (ter Braak and Juggins, 1993).

Ter Braak and Juggins (1993) show that, with a small modification, WA is equivalent to the first PLS component of transformed data. The modification changes the inverse deshrinking regression to a weighted inverse regression with weights proportional to the site total (y_{i+}). For percentage compositional data with a constant total, this modification is unimportant. In WA-PLS further orthogonal components are obtained as two-way weighted averages of the residuals for the environmental variable of interest, in other words the residuals of the regression of x on the components extracted to date are used as new site scores in the two-way WA algorithm (ter Braak and Juggins, 1993). In ordinary PLS these residuals are used as new site scores in the two-way weighted summation algorithm of PCA (ter Braak, 1987d). In WA-PLS a joint estimate of x_i is a linear combination of the components of the WA-PLS, each of which is a WA of the taxon scores, hence the name WA-PLS. The final prediction equation is thus a WA of the updated optima, but in contrast to WA (see above) the optima are updated not only by an inverse deshrinking regression but also by considering the residual correlations in the biological data. In practice, taxa that are abundant in sites that have large residuals are most likely to have updated optima (ter Braak and Juggins, 1993). As in PLS (Martens and Næs, 1989), the number of components to be retained is determined by cross-validation (leave-one-out jack-knifing) on the basis of prediction error sum-of-squares (PRESS). If only the first component is retained, WA-PLS reduces to WA with an inverse deshrinking regression. WA-PLS will thus equal or perform better than WA, depending on whether the optimal number (giving the minimal RMSEP in cross-validation) is 1 or more than 1 (ter Braak and Juggins, 1993).

Recent experience with using WA-PLS and a variety of training sets (freshwater diatoms, marine diatoms, freshwater invertebrates, terrestrial pollen, marine foraminifera) shows that there is little or no improvement in reducing RMSEP over simple WA for some data-sets, whereas with other data-sets there can be considerable improvement (up to a 60% reduction in RMSEP) over WA when 2 or 3 WA-PLS components are included (e.g. ter Braak and Juggins, 1993; Juggins *et al.*, 1994; Korsman, 1994; Korsman and Birks, 1995; Birks, unpublished results).

Ter Braak and Juggins (1993) and ter Braak *et al.* (1993) have used simulated training and independent test sets generated using Minchin's (1987) COMmunity PAttern Simulator (COMPAS) to try to discover under what conditions WA-PLS outperforms WA and linear PLS. Ter Braak and Juggins (1993) used simulated data with a single underlying environmental gradient and varied the length of the compositional gradient (in SD units) and the amount of unstructured qualitative noise (percentage absences) in the data. WA-PLS greatly reduces the RMSEP for the test set (up to 72%) for data with low noise, but the reduction declined with increasing noise. At very high noise levels, WA-PLS performs slightly worse (-3%) than WA. If the gradient length is varied (2.8, 5.6, 11 SD units), WA-PLS outperforms both WA and linear PLS in terms of RMSEP and maximum bias, with the greatest improvement (33%) at the intermediate gradient length.

Ter Braak *et al.* (1993) simulated training and test data with two underlying gradients, one major gradient (x) and a secondary gradient (z), and varied the magnitude of z and the gradient length of x. RMSEP is reduced by about 50% when WA-PLS is used compared with WA. Gradient length of x (2-8 SD units) has little influence on relative performance, as does the secondary gradient, except when the gradient length of x is short (2 SD units). WA-PLS was also compared with PLS and multinomial logit models (MLM). In all cases WA-PLS outperforms PLS and MLM, but the differences with MLM are small when z is unimportant. Although WA tends to perform less well than WA-PLS in all these simulation experiments, both methods perform reasonably well with simulated data despite the complexity and variability of the underlying response curves (ter Braak and Juggins, 1993), illustrating the robustness of WA-based methods to deviations from a Gaussian unimodal response model. In all experiments, the correlation between the WA and WA-PLS estimates and the true values in the independent test sets is 0.95 or more.

There are two main reasons why WA-PLS outperforms WA in certain instances. First, WA is an approximation to maximum likelihood regression and calibration using a Gaussian response model (ter Braak and Prentice, 1988; ter Braak and van Dam, 1989). It is not an ideal approximation because WA suffers from "edge effects" (Mohler, 1983)

with poor estimation of optima due to truncation of the response curves at the ends of the gradient (e.g. Oksanen *et al.*, 1988; Birks *et al.*, 1990a; Juggins, 1992). As a result, WA overestimates optima at the low end and underestimates optima at the high end, and non-linear distortions appear that are well known in correspondence analysis (Hill and Gauch, 1980). A linear deshrinking regression in WA does not fully remove these edge effects but WA-PLS does remove them in a predominantly unidimensional data set (ter Braak and Juggins, 1993). Second, in real life there are almost always additional environmental variables that influence the composition and abundance of biological assemblages. Patterns resulting from these variables are ignored in WA as it assumes that environmental variables other than the one of interest have negligible influence (Birks *et al.*, 1990a). WA-PLS uses this additional structure to improve estimates of taxa optima. For optimal performance in prediction, the joint distribution of these additional environmental variables in the fossil assemblages should be the same as in the modern training set (Brown, 1979; Birks *et al.*, 1990a). Although WA-PLS shows little improvement over WA when unstructured noise is added to simulated data, it shows a large improvement if the noise is structured in the form of secondary environmental gradients, thereby providing additional structure in the data for WA-PLS to exploit.

Ter Braak *et al.* (1993, p. 556) conclude that "until the time that such sophisticated methods mature and demonstrate their power for species-environment calibration, WA-PLS is recommended as a simple and robust alternative". For data that span an environmental gradient of 2 or more SD units, WA-PLS is an appropriate and robust reconstruction procedure.

Canonical correspondence analysis and multivariate weighted averaging partial least squares regression

Canonical correspondence analysis (CCA) (ter Braak 1986, 1987d, 1987e) is the unimodal-based equivalent of the linear-based RDA and is the constrained or canonical version of CA. CCA performs a direct gradient analysis or constrained ordination of biological "response" data in relation to two or more environmental "predictor" variables. It selects ordination axes that are linear combinations of the environmental variables that maximise the dispersion of the taxon scores. A CCA ordination diagram simultaneously displays the main patterns of biological variation, as far as these reflect environmental variation, and the major patterns in the weighted averages (cf. the correlations in RDA) of each taxon in relation to the environmental variables (ter Braak, 1986, 1987e). CCA is thus intermediate between CA and separate WA regressions for each taxon (ter Braak and Prentice, 1988).

A CCA ordination plot can be a useful graphical aid in environmental reconstruction because "passive" fossil samples with no known environmental variables can be positioned into the plane of CCA axes 1 and 2 on the basis of similarities in their biological composition to the active modern samples with known environmental variables. The location of the fossil samples and, when joined up into stratigraphical order, their "time-track" in relation to the modern environmental variables can strikingly illustrate patterns of past change in one or more environmental variables; - see Birks *et al.* (1990b), Allott *et al.* (1992), Grønlund (1991), Battarbee (1991), Kingston *et al.* (1992), Battarbee *et al.* (1989), Wilson *et al.* (1993), Ollikainen *et al.* (1993), Gaillard *et al.* (1992, 1994), and Warner and Charman (1994) for examples of such plots. In positioning fossil samples into the plane of the modern training-set samples, it is important to check if the positioning of the fossil samples is geometrically reliable (Birks and Gordon, 1985). The square residual chi-square distance of each sample from the plane formed by the CCA axes (ter Braak 1990) provides an appropriate measure of fit in CCA.

Canonical correspondence analysis can also be used for quantitative reconstruction of a single environmental variable from fossil assemblages given a modern training set (ter Braak, 1987c). The modern training set of biological data and the environmental variable of interest act as "active" samples (*sensu* ter Braak, 1987c) and the fossil samples whose values of the environmental variable are unknown and are to be inferred are treated as "passive" samples. In this case CCA reduces to WA regression and calibration with equal tolerances and a classical deshrinking linear regression (Line and Birks, 1990; ter Braak *et al.*, 1993). Examples of the use of CCA for quantitative environmental reconstruction include Stevenson *et al*, (1989), Dixit *et al.* (1989) Fritz (1990), Fritz *et al.* (1994), Marsicano and Siver (1993), and Siver (1993). It is unclear what advantage there is in using CCA in this way in preference to WA with inverse or classical deshrinking and WA-PLS, especially as easy-to-use software exists for WA and WA-PLS (see below).

A special form of CCA was developed for quantitative palaeoenvironmental reconstruction by Gasse and Tekaia (1979, 1983) as an improvement on Roux's (1979) CAR. Gasse and Tekaia (1979, 1983; see also Gasse, 1986) divided the environmental variable of interest, in their case pH, into classes and then did a CA of the taxon-by-class table, each cell of which contained the total abundance of a taxon in the sample with pH values that fall within the corresponding pH class. The final inference equation is obtained by multiple regression of pH on the CA axes. Ter Braak (1986) (see also ter Braak and van Dam, 1989) noted that this method is a special case of CCA with a single nominal predictor variable. This special case is called "analysis of concentration" by Feoli and Orlóci (1979) and provides an ordination constrained to show maximum

separation among pre-defined groups of samples. As environmental variables in CCA can be quantitative, the division into classes is not needed, and there is no advantage in the Gasse and Tekaia procedure over CCA (= WA with classical deshrinking), WA, or WA-PLS. Despite this, the approach of CCA of environmental classes has recently been used to reconstruct temperature and a range of chemical variables (e.g. Ca, Cl, Na, K, Mg, Si, So, alkalinity) from diatom assemblages in Bolivia by Roux *et al.* (1991) and Servant-Vildary and Roux (1990), and to reconstruct salinity, water depth, and the Mg/Ca ratio from ostracod assemblages in Bolivia by Mourguiart and Carbonel (1994).

Ter Braak *et al.* (1993) discuss if simultaneous calibration of several environmental variables by classical multivariate calibration (Brown, 1982) would reduce the prediction error for the variable of interest as there would be some information about additional variables. Classical calibration by (full-rank) CCA reduces to a form of multivariate linear regression of the biological data on the transformed environmental variables. This can, in theory, be inverted to produce an inverse regression of the environmental variables on the biological data. It would fail, in practice, because of multicollinearity, whereas WA-PLS is the inverse regression procedure that guards against multicollinearity. Ter Braak *et al.* (1993) suggest, however, that there is no advantage in using a multivariate version of WA-PLS, namely WA-PLS2, for joint calibration of environmental variables because it is a form of inverse regression, admittedly multivariate, and not a form of multivariate classical calibration (see Martens and Næs, 1989; Korsman and Birks, 1995).

To complete the series of unimodal-based methods that parallel the linear-based methods (e.g. CA and PCA, CAR and PCR, WA-PLS and PLS, WA-PLS2 and PLS2, CCA and RDA), it is of interest to note, for the sake of completion, that ter Braak (personal communication) has shown that, in theory, there is a unimodal-based equivalent of CCoA. This has not, as far as is known, been implemented. It is unclear what advantages, if any, it would have over CCA.

OTHER METHODS

In this section I consider quantitative environmental reconstruction methods that, although numerical and requiring a computer for implementation, do not have an underlying statistical model such as a linear or unimodal species-environment response model. These methods fall into two main types, modern analogue techniques and response-surface methods, including the mutual climatic range method.

Modern analogue techniques

The basic idea of modern analogue techniques (MAT) is to compare numerically, using an appropriate dissimilarity or similarity measure, the biological assemblage in a fossil sample with the biological assemblages in all available modern samples that have associated environmental data. Having found the modern sample(s) that is most similar to the fossil sample, the past environment for that sample is inferred to be the modern environmental variable(s) for the analogous modern sample(s). The procedure is repeated for all fossil samples and a simultaneous reconstruction for several environmental variables is made for the stratigraphical sequence on the basis of modern analogues. Examples of this simple MAT include Hutson (1980), Prell (1985), Morley (1989b), D. Anderson *et al.* (1989), Le (1992), Bartlein and Whitlock (1993), and Thunell *et al.* (1994). Ter Braak (1995) considers MAT to be a k-nearest neighbours methods of regression via smoothing (Stone, 1977; Hastie and Tibshirani, 1990). It is an inverse procedure because it estimates the environmental variable x_0 given the fossil assemblage y_0 (ter Braak, 1995).

The approach clearly requires an extensive set of modern surface samples and associated environmental data that covers the likely range of biotic assemblages and environmental conditions in the past if the reconstructions are to be reliable and informative. The modern training set should be of consistent taxonomy and quality, and all the samples should be from the same type of sedimentary environment as the fossil data sets used for environmental inference in order to minimise taxonomic, methodological, and taphonomic differences (Birks, 1994). These requirements can create serious limitations, especially when attempting reconstructions for glacial stages. A second problem is how to measure dissimilarity between fossil and modern assemblages. Squared chord distance (Prentice, 1980; Overpeck *et al.*, 1985) is widely used, although other coefficients may be more appropriate for particular data-sets and problems, for example the squared chi-square distance (Birks *et al.*, 1990a, 1990b). This measure is implicit in CA, DCA, and CCA. Although chord and chi-square distances maximise the "signal-to-noise" ratio when used with pollen percentage data (Prentice, 1980; Overpeck *et al.*, 1985), the intuitive concept of dissimilarity (or similarity) between assemblages is complex. It includes not only the relative abundances of the taxa but also the total taxonomic composition and the few numerically dominant taxa. A composite dissimilarity measure that incorporates three or four different dissimilarity coefficients, each of which emphasises different properties of the assemblage (cf. Hill, 1989) can be useful in MAT (Birks, unpublished results). A third problem is how to assess if the lowest dissimilarity found between a fossil sample and a modern sample represents a convincing match and analogue. It can be misleading to assume that modern and fossil assemblages with the lowest dissimilarity are necessarily good analogues (e.g. Huntley, 1990a), simply

because in any comparison, some modern sample has to have the lowest dissimilarity irrespective of whether the dissimilarity is statistically different from all the other dissimilarities. A common solution is to take the extreme 5% or 10% of the dissimilarities calculated between all modern samples as an approximate threshold value to indicate a "good analogue" (P. Anderson *et al.*, 1989; Bartlein and Whitlock, 1993). Another approach is to use permutation tests (Manly, 1991) to assess the approximate significance of an observed dissimilarity value (Birks *et al.*, 1990a, 1990b; Birks, 1993). The fossil and modern assemblages are permuted many times, subject to certain constraints, and an approximate empirical statistical distribution of dissimilarity values is derived. This enables approximate limits to be set for a "close" analogue and a "good" analogue in terms of the extreme 2.5% and 5% tails of the empirical distribution of dissimilarity coefficients. It permits the identification of fossil assemblages that appear to have no modern analogues in the available modern training set. Such no analogues are a major problem in MAT because if no modern analogues can be identified, no realistic environmental reconstruction is possible as MAT cannot extrapolate when applied to fossil data (Bartlein and Whitlock, 1993). This is a particular problem when, in many situations, modern biological assemblages appear to have been influenced, to varying degrees, by different types of human impact in the recent and historical past. Guiot (1990, p. 49) cautions that "the present relationship between climate and vegetation is disturbed by human action". Analogue matching, with associated permutation tests, is an important means of evaluating the likely reliability of environmental reconstructions based on other methods such as WA (see Birks *et al.* (1990a, 1990b), Hall and Smol (1993), and Gaillard *et al.* (1995) for examples of this type of evaluation).

Having found the modern sample or group of samples that most closely match the fossil sample, the environmental reconstruction can be based on the single modern sample that most closely resembles the fossil assemblage, it can be based on a mean of, say, the 5 or 10 most similar modern samples (D. Anderson *et al.*, 1989; Hutson, 1980; Thunell *et al.* 1994), or it can be based on a weighted mean of the 5 or 10 most similar modern samples, with the weights being the inverse of the dissimilarity values so that modern samples that have the lowest dissimilarity (i.e. are most similar) will have the greatest weight in the reconstruction (e.g. Prell, 1985; Morley, 1989b; Bartlein and Whitlock, 1993).

MacDonald and Reid (1989) present a modified graphical MAT. After comparing modern and fossil pollen assemblages using the squared chord distance, contour maps of the chord distances are constructed in relation to the modern annual precipitation and annual degree days greater than 5°C of the modern samples. These maps show where the closest modern analogues to the fossils assemblages of a particular age occur in modern climate space. This procedure avoids the problem of deciding on the one, five, or ten

most similar analogues, as all chord distances are plotted and contoured. The method provides a clear display of fossil pollen analogues that lack convincing modern analogues (see P. Anderson *et al.*, 1989 for other graphical methods for detecting no-analogue fossil assemblages).

It is difficult to estimate reliable reconstruction errors for individual fossil samples using MAT. Bartlein and Whitlock (1993) estimate an approximate error of reconstruction as the standard deviation of the mean of the weighted environmental values associated with the ten most similar modern samples. The overall predictive power of a training set in MAT can be estimated by cross-validation involving jack-knifing. Modern environment values at individual sites are estimated by MAT using the training set but excluding, in turn, the individual modern site whose environmental values are being inferred. The excluded sample forms an independent test set, and r^2 between observed and predicted values (Bartlein and Whitlock, 1993) and $RMSEP_{(jack)}$ can be calculated (Juggins *et al.*, 1994).

Bartlein and Whitlock (1993) suggest that an advantage of MAT is that it is the "least statistical" reconstruction procedure as it does not involve any underlying model of taxon-environment relationships. An important disadvantage is that MAT depends entirely on the range and the composition of the modern training-set. As a result, MAT environmental reconstructions often show greater short-term variation than reconstructions based on other, more statistically based techniques. A second practical disadvantage is that MAT requires very large and comprehensive training sets of comparable quality, taxonomy, and taphonomy covering large geographical areas and biotic and environmental gradients. Such high-quality training sets do not yet exist for many biological groups or environments, particularly terrestrial and freshwater environments.

A more mathematically complex variant of MAT has been developed by Guiot (1990; Guiot *et al.* 1993a, 1993b) based, in part, on his earlier techniques of palaeoenvironmental reconstructions (Guiot, 1985, 1987; Guiot and Pons, 1986). The approach has been extensively used to reconstruct a range of past climatic variables from fossil pollen assemblages (e.g. Guiot *et al.* 1989, 1992, 1993a, 1993b; Bonnefille *et al.*, 1990, 1992; Guiot and Couteaux, 1992; Vincens *et al.*, 1993), including January, April, July, and October temperature and precipitation anomalies for the last interglacial-glacial cycle (Guiot *et al.*, 1992). In its simplest form, Guiot's (1990) approach differs from simple MAT in that the dissimilarities between fossil and modern assemblages are assessed by a weighted log-transformed Euclidean distance (the so-called distance operator) to find the closest analogue for each fossil spectrum. The weights, the so-called palaeobioclimatic operators (PBO), are computed from either a time-series analysis

(with a lag of one sample) of the fossil pollen-stratigraphical sequence used for reconstruction purposes (Guiot, 1990) or a PCA of fossil pollen data from a large spatial array of sites used for reconstruction purposes (Guiot *et al.*, 1993a). The weights are selected in an attempt to emphasise the climatic signal within the fossil data and are chosen to highlight those pollen taxa that show the "most coherent behaviour in the vegetational dynamics" (Guiot, 1990, p. 50). As the weights are based solely on fossil data, they are considered to "minimise the human action which has significantly disturbed the modern spectra" (Guiot, 1990, p. 50) and to reduce noise (i.e. non-climate signals) within the fossil data resulting from local site factors, disturbance, etc. In the case of stratigraphical sequences, palaeoclimatic time-series for the sequences are computed from the eigenvectors of the multiple auto-correlation matrix in which the diagonal elements are the first-order auto-correlations of each taxon and the off-diagonal elements are the first-order cross-correlations between taxa. Any palaeoclimatic time-series representing < 10% of the variance in the autocorrelation matrix is discarded as "noise". The weights (PBO) of each taxon in the retained eigenvectors are then used in calculating the weighted log-transformed Euclidean distances to find the closest analogues. In the case of spatial fossil pollen data, the PBO weights are simply the taxon loadings on the first principal components of the log-transformed pollen data (Guiot *et al.*, 1993a).

The reconstructed values for the climatic variables are weighted means of the estimates based on the 20, 40, or 50 most similar modern assemblages. The weights are inverses of the dissimilarity values (Guiot, 1990). Standard deviations of these estimates give an approximate error for the reconstruction. Alternatively a modified bootstrap approach is used to derive "approximate 70% confidence intervals" for the reconstruction (Guiot, 1990; Guiot *et al.*, 1993a).

In this modified MAT, large confidence intervals and apparently anomalous climatic reconstructions may occur, for example in glacial stages and in transitional periods between glacial and interglacial conditions. These occur because of ambiguous and poorly defined palynological analogues. In an attempt to improve the precision of MAT and to overcome the lack of unambiguous modern palynological analogues, Guiot and associates often introduce constraints to narrow the choice of possible modern analogues. These constraints on the possible palynological matches include geographical constraints (in effect climatic constraints at the broad spatial scale of Guiot's training set) (Guiot *et al.*, 1989, 1992; Guiot, 1990), inferred biome type (Guiot *et al.*, 1993b) as defined in the global biome model of Prentice *et al.* (1992), sedimentary organic content (Seret *et al.*, 1992), fossil beetle assemblages (Guiot *et al.*, 1993b), and inferred lake-level changes (Guiot *et al.*, 1993a). Such constrained analogue-based methods appear to have little statistical or theoretical basis. They do, however, provide the palaeoecologist

with great scope at imposing whatever constraints are considered appropriate. Such methods should perhaps be used with caution.

The MAT has been further modified by Peng *et al.* (1993) to reconstruct vegetational "ecosystems" in Europe at 6000 B.P. They delimit 23 "ecosystem" units today and characterise each by three "typical" modern spectra. They used the linear technique of canonical correlation analysis to "optimize the climatic signal in the pollen data" and then transformed all modern and fossil pollen assemblages to 4 canonical axes. The MAT is thus done not with pollen assemblages but with their canonical correlation representation. Fossil pollen data were then compared with the 69 "typical" modern spectra to find modern "ecosystem" analogues and to reconstruct past "ecosystem" distribution and abundance. From this reconstruction, the carbon density (in gigatons) in plant biomass at 6000 B.P. was estimated for Europe. The rationale for many of the modifications to the basic MAT is unclear in this study and highlights the desirability of using reconstruction techniques with an explicit and stated theoretical basis rather than making *ad hoc* modifications to existing methods.

Response surfaces

Response surfaces are two- or three-dimensional graphical representations of the occurrence and/or abundance of taxa considered individually in modern environmental space. The x and y axes represent explanatory environmental variables and the z axis represents the response variable such as the presence or absence of a taxon or its relative abundance. In this section, emphasis is on the use of response surfaces for the quantitative reconstruction of environmental variables from fossil data rather than on the details of constructing response surfaces to display the behaviour and responses of taxa today along environmental gradients. The basic idea in the use of response surfaces for reconstruction or calibration purposes, given a fossil assemblage whose past environment is to be reconstructed, is to find the combination of modern environmental variables today that supports a biological assemblage of similar composition and abundance to the fossil assemblage. The values of these environmental variables are the inferred values of the past. How this search is done depends on the type of response surface and on how it is constructed.

Response surfaces can be of two main types depending on the nature of the response variable, namely presence-or-absence (+/-) data or quantitative abundances. The surfaces can be constructed graphically without any computer contouring or smoothing; by computer contouring and/or smoothing but without any underlying statistical response model; by the fitting of a specific statistical model within the general framework of generalised linear modelling; or by the semi-parametric approach of generalised additive

modelling. Techniques for environmental reconstruction (calibration) from response surfaces are, however, less well developed than for constructing response surfaces.

Iversen (1944) presented simple presence/absence response plots to show the presence or absence of *Ilex*, *Hedera*, and *Viscum* in relation to the mean temperature of the coldest month and the mean temperature of the warmest month near the northern geographical range-limits of these shrubs. Having established thermal-limit curves for the three taxa, Iversen then positioned fossil samples containing *Ilex*, *Viscum*, and *Hedera* pollen, *Hedera* and *Ilex* pollen, and *Hedera* and *Viscum* pollen only to infer past changes in mean January and July temperatures. This approach was extended by Grichuk (1969) by plotting the entire modern geographical range of several taxa in relation to two climatic variables. Reconstruction of past climate from a fossil pollen assemblage was done by finding where today the taxa whose pollen occurred together in the fossil assemblage coincide in modern climate space.

The approach of Grichuk (1969) forms the basis for the mutual climate range method (MCRM) developed by Atkinson *et al.* (1986, 1987) to reconstruct past climates from fossil coleopteran data. In MCRM the limits of geographical range of selected taxa today are plotted together in modern climate space, for example mean temperature of the warmest month, mean temperature of the coldest month, or the temperature range between the warmest and coldest months. Past climate is inferred from a fossil assemblage to be where the occurrences of the taxa in the fossil spectrum all overlap in modern climate space. Contours are drawn to show the percentage overlap of the modern climatic ranges of the taxa present in the fossil assemblage, and the most likely estimate of past climate lies within the mutual climatic ranges, as defined by the 100% overlap contour. As the zone of overlap for a fossil assemblage may be quite considerable, the reconstructed values may show a very large range (e.g. Atkinson *et al.*, 1987; Ponel and Coope, 1990; Lemdahl, 1991; Walker *et al.*, 1993; Lowe *et al.*, 1994; Gaillard and Lemdahl, 1994; Hammerlund and Lemdahl, 1994).

A limitation of all these presence-absence response surfaces is that they take no account of the probability of occurrence of a taxon within modern climate space. As the probability of occurrence of an individual taxon is not equal within its geographical range (Hengeveld, 1990), being highest near the centre or optimum of its range and lowest at the range margins, it would be important to take into account the probability of occurrence. This can be done by using data on the presence and absence of the taxon throughout its range and fitting Gaussian logit response surfaces (ter Braak and Looman, 1987), logistic regression surfaces (Lenihan, 1993), or more flexible surfaces based on local logistic regression or generalised additive models that may more closely follow the data. It is likely that if probabilities of occurrence are considered, more precise

reconstructions with narrower ranges and associated probabilities could be obtained using maximum likelihood estimation.

Pollen analysts have constructed modern pollen-climate response surfaces to illustrate how the relative abundances of pollen in modern surface samples vary in relation to major climatic gradients (e.g. Bartlein *et al.*, 1986; Anderson *et al.*, 1991; Prentice *et al.*, 1991; Huntley *et al.*, 1989; Webb *et al.*, 1987; Bartlein and Whitlock, 1993; R. Webb *et al.*, 1993; T. Webb *et al.*, 1993). Related surfaces have been fitted for modern ostracods in relation to alkalinity and sulphate (Smith, 1993) and for bryophytes in relation to ecological factors and climate (Gignac *et al.*, 1991a, 1991b). The surfaces have been fitted either by multiple regression or by a robust locally weighted regression.

In the case of response surfaces estimated by multiple regression, Bartlein *et al.* (1986) fitted a second- or third-order polynomial model for two predictor power-transformed climatic variables and arcsine-transformed pollen percentages as the response variable. A third-order model was fitted first, and if there were non-significant higher order regression coefficients in the model, a second-order model was examined. The goodness-of-fit of the final fitted surface was evaluated for each taxon. This varied from 9% to 78%. This approach is a form of generalised linear modelling (GLM) (McCullagh and Nelder, 1989) with particular systematic and random components. Other types of GLM-based response models can be constructed for percentage response data using a logit or logarithmic link function and a binomial or multinomial error structure (e.g. Barber *et al.*, 1994; ter Braak and Looman, 1987). For example, the Gaussian logit model with two predictor environmental variables, x_1 and x_2, is (ter Braak and Looman, 1987)

$$\log y = b_0 + b_1 x_1 + b_2 x_1^2 + b_3 x_2 + b_4 x_2^2 + b_5 x_1 x_2$$

If b_2 and b_4 are both negative and $4b_2b_4 - b_{52}$ is positive, the equation describes a unimodal response surface. If b_5 is not equal to 0, the optimum with respect to x_1 depends on the value of x_2, and the two predictor variables show an interaction effect on the taxon. If $b_5 = 0$, the optimum of taxon y with respect to x_1 does not depend on the value of x_2 (i.e. there is no interaction). The GLM-based approach has only rarely been used in palaeoecology (e.g. Bartlein *et al.*, 1986; Smith, 1993; Barber *et al.*, 1994). Instead a robust non-parametric locally weighted averaging approach has been developed that is an approximation to the smoothing of a multidimensional scatter plot by robust locally weighted regression (LOWESS; Cleveland and Devlin, 1988; Cleveland, 1993, 1994).

In the locally weighted averaging approach of Huntley *et al.* (1989), Prentice *et al.* (1991), Anderson *et al.* (1991), Huntley (1992, 1993), T. Webb *et al.* (1993), R. Webb *et*

al. (1993), and Bartlein and Whitlock (1993), a window is moved over a 20 × 20 point grid in the climate space defined by the predictor variables. The expected or fitted values of the pollen taxon is calculated for the centre of the window as a locally weighted average using a tricubic weighting function of the pollen percentages for the samples falling within the window at each grid point. The resulting response surface is constructed by interpolation between these fitted values (T. Webb *et al.*, 1993). The width of the window and the spacing of the grid points can be varied so as to produce surfaces that are relatively smooth and that portray the major patterns of variation in the pollen data within modern climate space (Prentice *et al.*, 1991). Details of the actual computations are given by Gignac *et al.* (1991b). Like the GLM-based approach, there are simple statistical measures of goodness-of-fit (Prentice *et al.*, 1991). One possible problem in this methodology is deciding on the size of the grid and the window widths and on the extent of data-smoothing. Despite this problem, the approach is favoured because it "neither assumes nor provides any simple mathematical description of the fitted surface; thus the value at any location in climate space can be determined only using the surrounding data values" (Huntley, 1993, p. 216) and because "response surfaces can be extrapolated to some extent by calculating weighted averages of the fitted values for regions just outside the portion of climate space covered by the data" (Bartlein and Whitlock, 1993, p. 277). However, Shane and Andersen (1993) encountered serious problems in using response surfaces to reconstruct past climate in the late-glacial of east-central USA, as the response-surface training set was not sufficiently similar to provide adequate modern analogues for climatic reconstruction. Examples of pollen-climate response surfaces derived by local weighted averaging are given by Anderson *et al.* (1991), Prentice *et al.* (1991), Huntley *et al.* (1989), Webb *et al.* (1987), Bartlein and Whitlock (1993), R. Webb *et al.* (1993), and T. Webb *et al.* (1993). Huntley (1990b) and Huntley and Prentice (1993) present approximate optima of 22 major European pollen taxa in relation to January and July mean temperatures.

An attractive new approach that combines the advantages of the GLM approach of Bartlein *et al.* (1986) with the flexibility of the locally weighted regression approach of Bartlein and Whitlock (1993) is generalised additive models (GAM) (Hastie and Tibshirani, 1986, 1990; Yee and Mitchell, 1991; Smilauer and Birks, 1995). GAMs are a semi-parametric smooth maximum-likelihood extension of GLMs and are primarily data-driven rather than model-driven. The two types can be contrasted as follows. In GLM, the expected abundance of taxon *y*, (E*y*), is modelled in relation to *x* predictor environmental variables by

$$g(\mathrm{E}\,y) = \alpha + \sum_{j=1}^{p} \beta_j x_j$$

where g is a prespecified link function, α is the intercept or constant term, and ß is the regression coefficient for variable *j*.

In GAM, the expected abundance of taxon *y* is modelled as

$$g(\mathrm{E}\,y) = \alpha + \sum_{j=1}^{p} f_j(x_j)$$

where f_j are semiparametric smoothing functions estimated, in practice, from the data using LOWESS techniques developed for smoothing scatter plots. These transform the predictor variables so as to achieve the maximum explanatory power. GAMs allow the data to determine the shape of the response surface, rather than being limited by the shapes implicit in parametric models (Yee and Mitchell, 1991). Features such as bimodality and pronounced asymmetry in the data can be easily detected. GAMs provide a more powerful tool for data exploration than GLMs. As in GLM, a forward-selection procedure with cross-validation can be used to find a minimal set of predictor variables that best fit, in a statistical sense, the taxon abundances (Smilauer and Birks, 1995). GAMs thus combine the desirable statistical aspects of GLM such as goodness-of-fit measures, model selection, and regression diagnostics with the flexibility of non-parametric LOWESS-type smoothing. They are potentially very powerful semi-parametric techniques for response-surface estimation.

Despite their current popularity in palynology, response surfaces have only occasionally been used in palaeoenvironmental reconstructions (e.g. Kelly and Huntley, 1991; Prentice *et al.*, 1991; Shane and Anderson, 1993; R. Webb *et al.*, 1993. T. Webb *et al.*, 1993; Bartlein and Whitlock, 1993). The methodology used involves "stacking" the modern pollen-climate response surfaces to produce synthetic pollen spectra for each of the grid nodes, usually 20 × 20 nodes, in modern climate space. These synthetic spectra are then compared to the fossil pollen assemblages being used for reconstruction by means of the squared chord distance (Prentice, 1980; Overpeck *et al.*, 1985) dissimilarity measure, as in MAT. The climate values for the 10 grid nodes with pollen spectra most similar (i.e. with the lowest dissimilarities) to the fossil assemblages are used to infer the past climate. The final inferred value is a mean of the climate values weighted by the inverse of the squared chord distances (Prentice *et al.*, 1991). In other words environmental reconstructions are done by MAT but here the modern training set consists of fitted pollen values in relation to modern climate and not the original pollen values. The fitted values naturally smooth the pollen data to a varying degree depending on the parameters of the smoothing procedure used (Cleveland and Devlin, 1988), and much of the inherent local site variability in modern pollen data that is assumed to be unrelated to broad-scale climate is removed (Bartlein and Whitlock, 1993; Huntley and

Prentice, 1993). As in MAT, apparent RMSE and r^2 can be estimated by inferring modern climate from modern pollen data and comparing inferred and observed values. Unless an independent test set of modern samples is used or unless some statistical cross-validation is performed, the estimated RMSE and r^2 will be over-optimistic and the RMSE will provide an under-estimate of the predictive power of the method and of particular data-sets (Bartlein and Whitlock, 1993).

Huntley (1993, 1994) discusses constraining the climatic variables that are to be inferred by the response-surface approach. Such constraints may be based on the fossil pollen spectra themselves in the form of a so-called "rule-based model to assign pollen spectra to 'biomes'" (Huntley, 1993, p. 220) and thus to give "palaeoclimatic reconstructions that are less variable between samples and have smaller uncertainties" (Huntley, 1993, p. 220). Although no indication is given by Huntley (1993) as to how uncertainty is assessed in this type of constrained response-surface MAT reconstruction, according to Huntley (1994) uncertainty appears to be assessed as the standard error of the environment variable for the ten closest analogues found. Alternatively the constraints may be based on plant macrofossils (Huntley, 1994). Such constraints can enable "a stable reconstruction to be obtained whereas without constraint the reconstruction was unstable and has large standard errors" (Huntley, 1994, p. 54). Such constraints are imposed by "visual inspection of the maps" (Huntley, 1993, p. 220) of the modern distributions of the species whose macroscopic remains are found fossil. Range limits are compared with maps of the climatic variables used in the response surfaces and a restricted climatic space is defined on the basis of the most limited geographical range of the species represented by macrofossils. The approach appears to resemble closely the MCRM method of Grichuk (1969) and Atkinson *et al.* (1986, 1987).

When using response surfaces fitted by GLM or GAM, reconstructions of past environment values from fossil assemblages can be made, in theory at least, by maximum likelihood calibration (ter Braak, 1987a). GAMs can be used not only for description and modelling but also for predicting abundances of taxa for a particular combination of environmental variables (Yee and Mitchell, 1991) and for estimating values of environmental variables for particular combinations and abundances of taxa in, for example, fossil assemblages (Smilauer, personal communication).

Limitations in the current use of response surfaces for palaeoenvironmental reconstructions are the need, as in MAT, for large high-quality training sets from a wide environmental range, the absence of any realistic error estimation for reconstructed values, and the problems inherent in MAT, namely delimiting threshold values for close, good, poor, and no analogues, the choice of dissimilarity measure, and the lack of modern analogues (see Shane and Anderson, 1993).

The response-surface approach is a classical approach to environmental reconstruction (ter Braak, 1995). It is favoured by Prentice *et al.* (1991) and Bartlein and Whitlock (1993) because the methodology allows some extrapolation of the estimated surface into "no-analogue" situations whereas MAT and restricted inverse linear regression approaches encounter serious problems with no-analogues (cf. Shane and Anderson, 1993; T. Webb *et al.*, 1993). As the response surfaces are fitted locally rather than globally (as in a GLM-estimated surface), some extrapolation is possible to extend the surface beyond the range of the training set, thereby permitting inferences of environmental variables from fossil assemblages that differ from samples in the modern training-set. Ter Braak *et al.* (1993) used simulated training sets to predict environmental variables from independent test data that are simulated to lie outside the environmental space of the training sets, thereby studying the effects of hidden extrapolation in no-analogue situations. They showed that WA, WA-PLS, and MLM all performed poorly and no one method performed consistently better than other methods. However, for very strong extrapolation, WA consistently performed best. Hutson (1977) similarly found that WA produced the most accurate estimates under no-analogue conditions compared with inverse regression and PCR. Clearly there is a need to compare the predictive abilities of response surfaces, WA, and WA-PLS on test sets for simulated no-analogue conditions.

EVALUATION OF RECONSTRUCTED ENVIRONMENTAL VALUES

Introduction

All quantitative palaeoenvironmental reconstruction procedures will produce a result. There is, however, no simple means of evaluating how reliable the result is (Imbrie and Webb, 1981). The lack of such evaluation procedures is rather surprising considering the range of sophisticated reconstruction procedures that are currently available. In addition to an overall performance measure for the training set like RMSEP estimated from an independent test set (e.g. Kipp, 1986; ter Braak and van Dam, 1989) or by statistical cross-validation (e.g. Birks *et al.*, 1990a, ter Braak and Juggins, 1993; Juggins *et al.*, 1994), it is important to assess the reliability of individual reconstructed values for each fossil sample. In this section I discuss possible means of conducting such evaluations.

Lack-of-fit measures

It is often useful to screen the modern training data prior to any regression and calibration (Birks *et al.*, 1990a). The squared residual distance (ter Braak, 1990) of

individual modern samples to the environmental variable of interest in a redundancy analysis (if the gradient length is < 2 SD units) or in a canonical correspondence analysis (if the gradient length is > 2 SD units) provides a useful measure of lack-of-fit of a sample to the environmental variable of interest. In such constrained ordinations, the only constraining external variable used is the environmental variable of interest (Birks *et al.*, 1990a). Some modern samples may have a poor fit to the environmental variable because other environmental factors may be dominant in influencing the biological composition and relative abundances in that sample or the composition of the assemblage may have been altered by differential preservation. Samples with a poor fit will have a high residual distance from the first ordination axis constrained by the single environmental variable of interest.

The same procedure can be used to evaluate individual fossil samples, except they are used as "passive" samples in the constrained ordination. This means that, as they lack external environmental data, they are positioned on the ordination axes by transition formulae (ter Braak, 1987c) on the basis of their biological composition as if they had participated in the derivation of the ordination axes. Any fossil sample whose residual distance is equal to or larger than the residual distance of the extreme 5% of the modern training set's distances can be considered to have a "very poor" fit to that particular environmental variable, whereas those fossil samples with residual distances equal to or larger than the extreme 10% can be deemed to have a "poor" fit (Birks *et al.*, 1990a).

A reconstructed environmental value is likely to have some reliability if the fossil sample in question has close modern analogues within the training set used for regression and calibration (ter Braak, 1995). Modern analogue techniques (MAT) (see above) are useful to detect fossil samples that lack "close" or "good" analogues within the available modern training set. Clearly environmental reconstructions for fossil samples that lack good or close analogues may be less reliable than inferred values for fossil samples with close or good analogues within the modern training set (ter Braak, 1995; Bartlein and Whitlock, 1993). Examples of the use of MAT as an aid in the evaluation of inferred environmental values include Birks *et al.* (1990a, 1990b), Hall and Smol (1993), Gaillard *et al.* (1994, 1995), and Juggins *et al.* (1994).

Root mean squared errors for reconstructed values

Bootstrapping (see above) provides a means of estimating sample-specific root mean squared errors of prediction for individual fossil samples (see Birks *et al.*, 1990a; Line *et al.*, 1994). As one of the components of RMSEP, the standard deviation of the reconstructed values for the fossil sample (s_1) in the bootstrap cycles, can vary from sample to sample depending on the fossil composition of the samples, RMSEP for fossil

samples can often vary within a core. Component s_1 is likely to be relatively small for fossil assemblages consisting of taxa that are frequent and abundant in the modern training set and to be relatively large for fossil assemblages consisting of taxa that are infrequent and rare in the training set. Birks *et al.* (1990a) showed that RMSEP for pH reconstructions from fossil diatom assemblages in a lake in south-west Scotland varied from 0.314 to 0.322 pH units during the Holocene.

Other approaches

Imbrie and Kipp (1971) discuss the evaluation of reconstructed environmental values derived from their modified principal components regression procedure. In this type of procedure it is possible to calculate how much of the variance in a fossil assemblage is captured by the 3 or 4 principal components used in PCR (or CAR). This is estimated by the sum of squares or communality and would be 1 if the fossil data are completely represented by the 3 or 4 components used in PCR. Imbrie and Kipp (1971) show interestingly that the communality of the fossil samples decreases down-core. Imbrie and Kipp also discuss what fraction of the down-core reconstruction pattern represents random effects and to what extent can the reconstruction be validated by independent evidence. Wang (1994) has presented a new reconstruction procedure involving target factor analysis (Malinovski and Hovery, 1980), an oblique dimension-reduction technique, to help overcome the no-analogue problems that can arise from evolutionary changes over long geological periods. Wang's approach is an extension of Imbrie and Kipp's (1971) PCR technique and has been used to reconstruct sea-surface temperatures in the late Tertiary and early Quaternary.

In the restricted inverse linear regression approach, Bartlein and Webb (1985) and Bartlein and Whitlock (1993) use the Mahalanobis distance (Prentice, 1980) as a measure of dissimilarity between a particular fossil sample and the mean of the modern spectra within a particular geographical region. This distance measure is also an approximate measure of the extent of extrapolation in the reconstruction that can result from "no-analogue" conditions (Bartlein and Whitlock, 1993). Such no-analogues can produce unreliable reconstructions in inverse regression because of "hidden extrapolation" (Weisberg, 1985; Bartlein and Whitlock, 1993; Huntley, 1993).

A useful but informal evaluation procedure is to reconstruct the same environmental variable by several numerical methods (e.g. Hutson, 1977; Webb and Clark, 1977; Le, 1992; Bartlein and Whitlock, 1993), to look for differences in the reconstructions, and to attempt to produce a consensus reconstruction based on several methods (e.g. Bartlein and Whitlock, 1993).

The most powerful evaluation procedure (Imbrie and Webb, 1981), unfortunately only rarely possible, is to reconstruct the same environmental variable from different fossil groups (e.g. Molfino *et al.*, 1982) or from other independent palaeoenvironmental records (e.g. Imbrie *et al.*, 1973; Engstrom and Nelson, 1991; Fritz *et al.*, 1994) and to validate the reconstruction, at least for the recent past, against known historical records (e.g. Fritz, 1990, Fritz *et al.*, 1994; Renberg and Hultberg, 1992).

AN EXAMPLE OF A QUANTITATIVE PALAEOENVIRONMENTAL RECONSTRUCTION

Introduction

As an example of the analysis of a modern training set and the use of such a training set for palaeoenvironmental reconstructions, I use the data from Imbrie and Kipp (1971). These consist of 61 core-top samples from the Atlantic containing a total of 27 foraminiferal taxa along with three environmental variables, namely summer sea-surface temperature (SSST), winter sea-surface temperature (WSST), and salinity. For reconstruction purposes, 110 core samples from Caribbean core V12-122 covering at least the last 450,000 years are used.

The modern training set is summarised in Table 6.1.

Gradient lengths

The gradient lengths of 3.86 and 4.22 SD units for SSST and WSST, respectively, indicate the unimodal nature of the foraminiferal abundance data. For salinity the gradient length is smaller (2.83 SD units). SSST and WSST are highly correlated (r = 0.97) but less strongly correlated with salinity (*ca.* 0.7). A detrended canonical correspondence analysis (ter Braak, 1986, 1987c) of the 61 modern samples with all three environmental variables as explanatory variables gives eigenvalues of 0.75, 0.04, and 0.01. The fourth, unconstrained axis has a small eigenvalue (= 0.05). These DCCA results show the essentially one-dimensional nature of these data. WSST and SSST explain about 38% of the total variance in the species data and either environmental variable could therefore usefully be reconstructed (ter Braak *et al.*, 1993).

	SSST°C	WSST°C	Salinity ‰
Range	2 - 29	-1.0 - 26.5	33.0 - 37.2
Standard deviation	7.02	8.00	0.99
Mean	21.39	16.92	33.70
Median	24.50	19.00	35.69
DCCA axis 1			
λ	0.72	0.73	0.52
Gradient length (SD units)	3.86	4.22	2.83
% variance explained	37.7	38.1	27.4
DCA axis 2			
λ	0.14	0.08	0.41
Gradient length (SD units)	2.08	1.69	2.51
% variance explained	7.3	4.2	21.6

	Minimum	Median	Maximum
Effective number of occurrences per taxon (*N2*)	5	20	45
Effective number of occurrences per sample (*N2*)	1	4	11

Table 6.1 *Summary of the Imbrie and Kipp (1971) training set. The range, standard deviation, mean, and median of summer sea-surface temperature (SSST), winter sea-surface temperature (WSST), and salinity are tabulated. The eigenvalues (λ), gradient lengths (in standard deviation (SD) units), and percentage variance of the species data explained by the environmental variable in a detrended canonical correspondence analysis (detrending by segments) with the environmental variable of interest as the sole constraining external variable are listed. The minimum, median, and maximum effective number of occurrences (N2) per taxon and per sample are also given.*

Modern Taxon-Environment Relationships

In an attempt to assess how many of the taxa show statistically significant relationships to the environmental variables today, Gaussian logit curves were fitted for all taxa to each environmental variable (Table 6.2). For WSST, 21 of the taxa show a significant fit

to a Gaussian unimodal model, three taxa show a significant increasing sigmoid logit curve, one taxon shows a significant decreasing sigmoidal curve, and two taxa have no significant relationship. As Table 6.2 shows, nearly all the 27 taxa have statistically significant relationships to one or more of the environmental variables.

	SSST	WSST	Salinity
Significant Gaussian logit unimodal model	19	21	21
Significant increasing sigmoidal logit model	6	3	4
Significant decreasing sigmoidal logit model	1	1	0
No statistically significant relationship	1	2	2

Table 6.2 *Result of Gaussian logit-regression of the Imbrie and Kipp (1971) training set (61 samples, 27 taxa) in relation to summer sea-surface temperature (SSST), winter sea-surface temperature (WSST), and salinity.*

Reconstruction Results

Thirteen numerical reconstruction procedures have been applied to the Imbrie and Kipp (1971) training set, and the results, as apparent RMSE, are presented in Tables 6.3-6.5. For six of these methods no $RMSEP_{(jack)}$ estimates are available, whereas RMSEP based on bootstrapping is only available for two methods (WA, $WA_{(tol)}$). Comparison between methods in terms of predictive abilities is thus not straight forward. In terms of apparent RMSE, the best performance (i.e. lowest RMSE) is PLS using 7 components for SSST, PLS with 8 components for WSST, and PLS with 3 components for salinity. However, as ter Braak and Juggins (1993) show, PLS or WA-PLS models with many components will naturally fit the training-set data well but may have little predictive value. This is shown by the high $RMSEP_{(jack)}$ for these PLS models. In all cases WA-PLS had the second lowest apparent RMSE and, when available, one of the lowest $RMSEP_{(jack)}$, closely followed by CAR. Being a predecessor of WA-PLS the good performance of CAR is not surprising, especially for a data set in which the first correspondence axis is so highly correlated with the environmental variables (ter Braak *et al.*, 1993). What is more surprising is the consistently good performance of MAT using squared chord distance and a weighted mean of the 5 closest analogues (cf. ter Braak, 1995).

	Apparent RMSE	$RMSEP_{(jack)}$	$RMSEP_{(boot)}$
Imbrie and Kipp (1971)			
PCR (4 components and linear model)	2.55	-	-
PCR (quadratic model)	2.15	-	-
Roux (1979) and Juggins (unpublished)			
CAR (3 components - Roux)	1.72	-	-
CAR (3 components - Juggins)	1.67	1.85	-
ter Braak *et al.* (1993) and unpublished			
WA	2.02	2.22	2.31
$WA_{(tol)}$	2.05	2.29	2.37
WA-PLS (3 components)	1.53	1.81	-
PLS (7 components)	1.29	2.03	-
MLM	1.82	1.95	-
MLM + inverse regression	1.70	-	-
GLM + ML calibration	2.09	-	-
GLM + WA calibration	1.94	-	-
GLM + $WA_{(tol)}$ calibration	1.80	-	-
MAT (squared chord distance, weighted mean of 5 closest matches)	-	1.76	-
MAT (squared chord distance, weighted mean of optimal number of closest matches = 6)	-	1.72	-

Table 6.3 *Apparent root mean squared error (RMSE) and root mean squared error of prediction based on leave-one-out jackknifing (RMSEP(jack)) or bootstrapping (RMSEP(boot)) for summer sea-surface temperature (0C) using the Imbrie and Kipp (1971) training set in several reconstruction procedures. (PCR = principal components regression, CAR = correspondence analysis regression, WA = weighted averaging, WA(tol) = weighted averaging with tolerance downweighting, WA-PLS = weighted averaging partial least squares, PLS = partial least squares; MLM = multinomial logit model, GLM = Gaussian logit model, ML = maximum likelihood, MAT = modern analogue technique).*

The consistently superior performance of WA-PLS over WA is shown in Figures 6.3-6.5 where the predicted values (based on leave-one-out jackknifing) and prediction residuals (predicted-observed) are plotted against observed values for SSST, WSST, and salinity for a 1-component WA-PLS (= simple WA) model and a 3-component WA-PLS model. A three-component WA-PLS model for all the variables is the optimum model, as assessed by $RMSEP_{(jack)}$ (ter Braak *et al.*, 1993).

For SSST the first WA-PLS component (i.e. WA) tends to underestimate the low observed values of SSST (Figure 6.3a). Three WA-PLS components remove some of this bias (Figure 6.3b), and there is little trend remaining in the residuals for the 3-component model (Figure 6.3d). In the case of WSST, the curvature in Figure 6.4a (WA) is largely removed after extracting three WA-PLS components (Figure 6.4b). WA tends

to underestimate seriously the high values of salinity (Figure 6.5a) but not to produce biased estimates at the low end. A three-component WA-PLS model largely removes the bias at the high end (Figure 6.5b), although a slight bias persists in the residuals for this model (Figure 6.5d).

	Apparent RMSE	$RMSEP_{(jack)}$	$RMSEP_{(boot)}$
Imbrie and Kipp (1971)			
PCR (4 components and linear model)	2.57	-	-
PCR (quadratic model)	1.54	-	-
Roux (1979) and Juggins (unpublished)			
CAR (3 components - Roux)	1.37	-	-
CAR (4 components - Juggins)	1.30	1.39	-
ter Braak *et al.* (1993) and unpublished			
WA	1.97	2.14	2.23
$WA_{(tol)}$	1.93	2.09	2.19
WA-PLS (3 components)	1.17	1.45	-
PLS (8 components)	0.99	2.05	-
MLM	1.51	1.65	-
MLM + inverse regression	1.38	-	-
GLM + ML calibration	3.21	-	-
GLM + $WA_{(tol)}$ calibration	1.25	-	-
MAT (squared chord distance, weighted mean of 5 closest matches)	-	1.22	-
MAT (squared chord distance, weighted mean of optimal number of closest matches = 4)	-	1.20	-

Table 6.4 *Apparent root mean squared error (RMSE) and root mean squared error of prediction based on leave-one-out jackknifing ($RMSEP_{(jack)}$) or bootstrapping ($RMSEP_{(boot)}$ for winter sea-surface temperature (°C) using the Imbrie and Kipp (1971) training set in several reconstruction procedures. For abbreviations, see Table 6.3.*

	Apparent RMSE	$RMSEP_{(jack)}$	$RMSEP_{(boot)}$
Imbrie and Kipp (1971)			
PCR (4 components and linear model)	0.57	-	-
PCR (quadratic model)	0.57	-	-
Roux (1979) and Juggins (unpublished)			
CAR (3 components - Roux)	0.50	-	-
CAR (4 components - Juggins)	0.49	0.52	-
ter Braak *et al.* (1993) and unpublished			
WA	0.57	0.60	0.61
$WA_{(tol)}$	0.56	0.60	0.60
WA-PLS (3 components)	0.45	0.53	-
PLS (3 components)	0.41	0.55	-
MLM	0.81	0.85	-
MLM + inverse regression	0.60	-	-
GLM + ML calibration	0.71	-	-
GLM + WA calibration	0.56	-	-
GLM + $WA_{(tol)}$ calibration	0.53	-	-
MAT (squared chord distance, weighted mean of 5 closest matches)	-	0.56	-
MAT (squared chord distance, weighted mean of optimal number of closest matches = 10)	-	0.50	-

Table 6.5 *Apparent root mean squared error (RMSE) and root mean squared error of prediction based on leave-one-out jackknifing ($RMSEP_{(jack)}$) or bootstrapping ($RMSEP_{(boot)}$) for salinity (‰) using the Imbrie and Kipp (1971) training set in several reconstruction procedures. For abbreviations, see Table 6.3.*

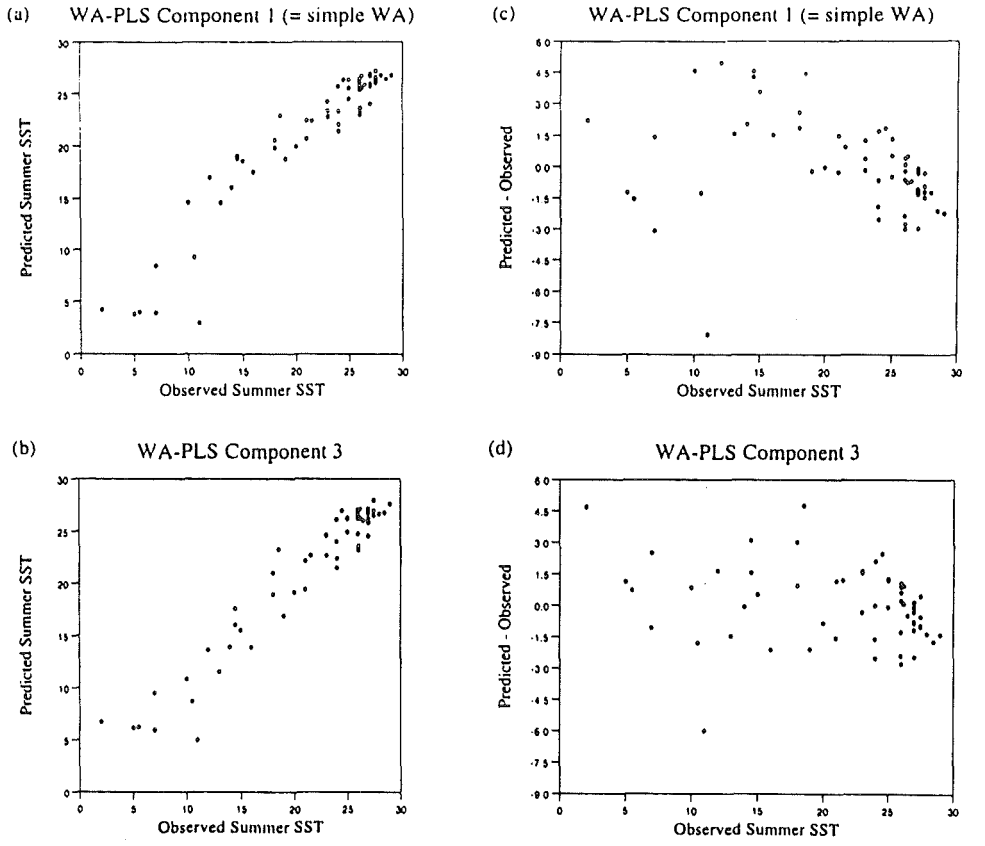

Figure 6.3 *(a) Plot of predicted summer sea-surface temperature (SST) (°C) as predicted by leave-one-out from the species composition of foraminifera after extraction of 1 WA-PLS (= WA) component plotted against observed summer SST. (b) Plot of predicted summer SST after extraction of 3 WA-PLS components against observed summer SST. (c) Plot of (predicted-observed) residuals of summer SST after extraction of 1 WA-PLS (= WA) component against observed summer SST. (d) Plot of (predicted-observed) residuals of summer SST after extraction of 3 WA-PLS components against observed summer SST. The optimal number of components in terms of RMSEP(jack) is 3. Data are from Imbrie and Kipp (1971).*

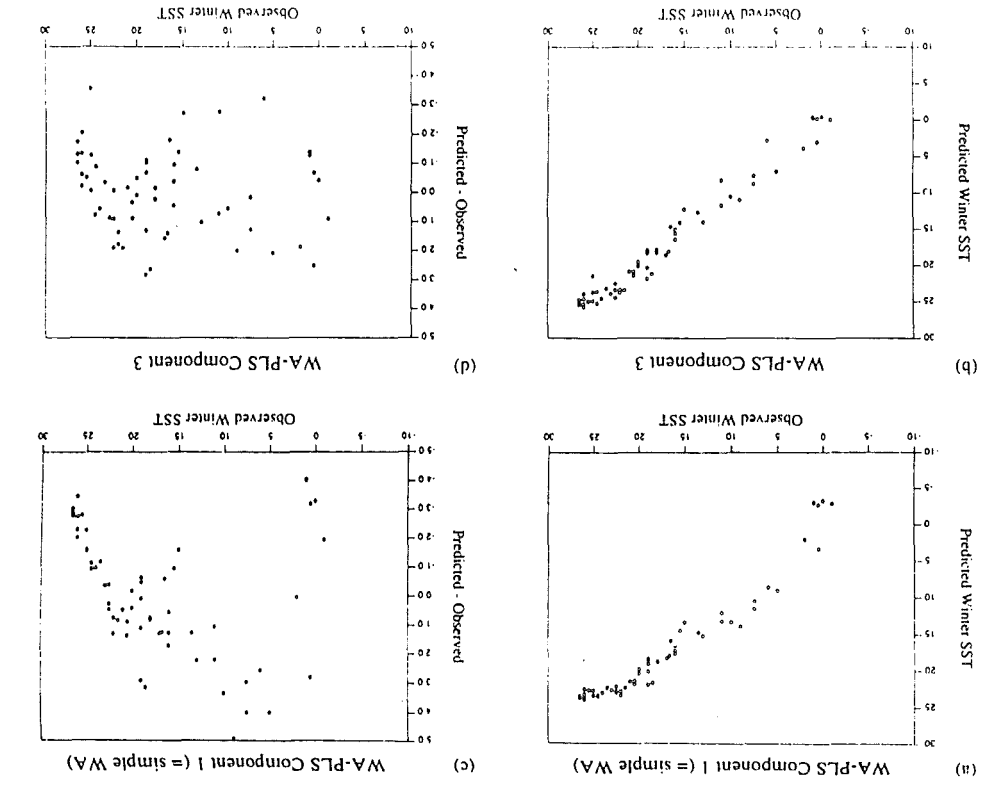

Figure 6.4 (a) Plot of predicted winter sea-surface temperature (SST) (°C) as predicted by leave-one-out from the species composition of foraminifera after extraction of 1 WA-PLS (= WA) component plotted against observed winter SST. (b) Plot of predicted winter SST after extraction of 3 WA-PLS components against observed winter SST. (c) Plot of (predicted-observed) residuals of winter temperature after extraction of 1 WA-PLS (= WA) component against observed winter SST. (d) Plot of (predicted-observed) residuals of winter SST after extraction of 3 WA-PLS components against observed winter SST. The optimal number of components in terms of RMSEP(jack) is 3. Data are from Imbrie and Kipp (1971).

Figure 6.5 *(a) Plot of predicted salinity (‰) as predicted by leave-one-out from the species composition of foraminifera after extraction of 1 WA-PLS (= WA) component plotted against observed salinity. (b) Plot of predicted salinity after extraction of 3 WA-PLS components against observed salinity. (c) Plot of (predicted-observed) residuals of salinity after extraction of 1 WA-PLS (≡ WA) component against observed salinity. (d) Plot of (predicted-observed) residuals of salinity after extraction of 3 WA-PLS components against observed salinity. The optimal number of components in terms of RMSEP(jack) is 3. Data are from Imbrie and Kipp (1971).*

The superior performance of unimodal-based models (excluding PLS) over the linear-based PCR for SSST and WSST is not surprising, given the length of the gradients (3.86 and 4.22 SD units, respectively). In the case of salinity (gradient length = 2.83 SD units), there is little difference in apparent RMSE between linear- and unimodal-based models. Despite their statistical sophistication, the generally slightly inferior performances of MLM and GLM compared to WA-PLS are surprising. However, in almost all instances where maximum likelihood- and weighted-averaging-based methods have been compared, the WA-based techniques have performed better (e.g. Birks *et al.*, 1990a; ter Braak and van Dam, 1989; Kingston and Birks, 1990; Cumming *et al.* 1992a; ter Braak *et al.*, 1993; Juggins *et al.*, 1994). In comparing WA with WA-PLS, WA-PLS must, by definition, perform as well or better than WA. As WA-PLS with one component is WA (ter Braak and Juggins, 1993) the performance of WA-PLS relative to WA depends on whether the optimal number of components is 1 (i.e. WA) or more than 1, as determined by cross-validation on the basic of $RMSEP_{(jack)}$.

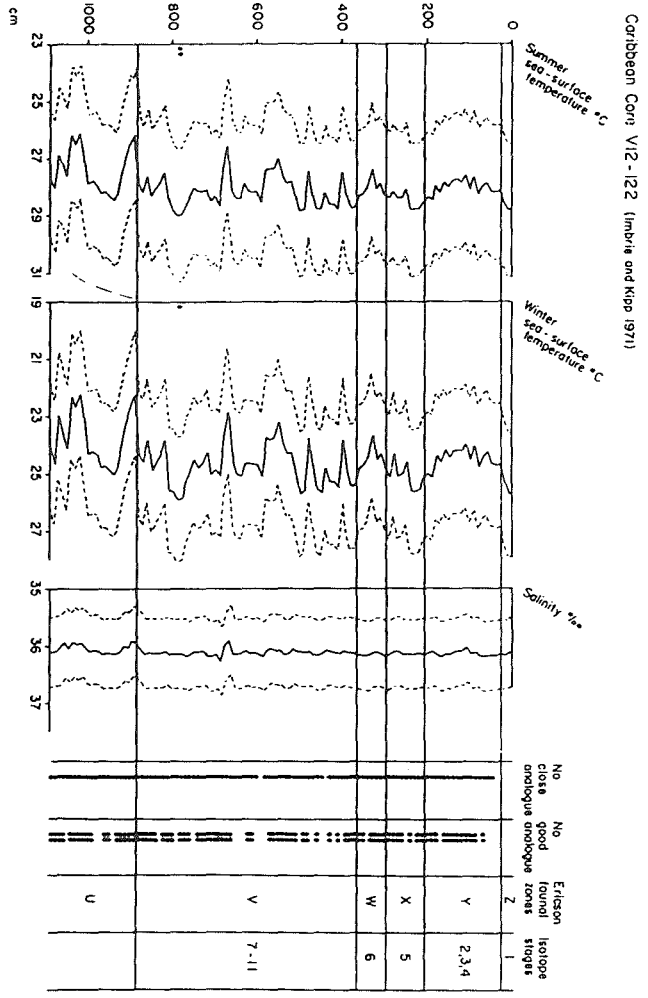

Figure 6.6 *Reconstructed summer sea-surface temperature, winter sea-surface temperature, and salinity values for Caribbean core V12-122 plotted against depth (solid line) based on simple weighted averaging. The RMSE of prediction for these estimates, derived by bootstrapping (1000 cycles) are shown as broken lines. Samples with a poor (•) or very poor (••) fit to particular environmental variables and those lacking a close or a good modern analogue in the modern training set of Imbrie and Kipp (1971) are shown. The Ericson faunal zones and isotope stages are also indicated. Data are from Imbrie and Kipp (1971).*

A simple WA reconstruction, with bootstrap error estimates, is presented in Figure 6.6 for SSST, WSST, and salinity for Caribbean Core V12-122. In terms of evaluation statistics, all the samples except for one at 790 cm show low squared residual distances to the individual environmental variables. The 790 cm sample has a "very poor" fit for SSST and a "poor" fit for WSST. In terms of modern analogues, almost all the core samples (except for Ericson faunal zone Z, isotope stage 1) lack "close" or "good" analogues in the modern training set. Imbrie and Kipp (1971) report communalities of 78-90% (mean = 84%) for zone Z and 47-86% (mean = 68%) for zones Y-U. The lack of analogues and the low communalities for zones Y-U suggest that the reconstructions for these zones may be subject to some "hidden extrapolation" and thus should be viewed with some caution.

Juggins *et al.* (1994) present the results of comparing the performance of seven calibration procedures (CAR, CCA of classes, WA, $WA_{(tol)}$, WA-PLS, MAT, and GLR (called "maximum likelihood" by Juggins *et al.*, 1994)) used with a training set of 55 modern diatom samples from saline lakes in the Northern Great Plains of North America. They show that in terms of $RMSE_{(jack)}$, WA, $WA_{(tol)}$, and CCA of classes give the best performance. They favour WA because of its computational and model simplicity and the direct ecological interpretation of the regression coefficients as species optima.

Further comparative studies using different training sets are clearly required to evaluate the relative abilities and performances of different reconstruction procedures (Bartlein and Whitlock, 1993; ter Braak, 1995).

COMPUTER SOFTWARE

As all the methods discussed in this chapter are totally computer dependent, I list in Table 6.6 computer programs for PCs that implement most of the palaeoenvironmental reconstruction methods discussed above. The list makes no attempt to be exhaustive. It is based solely on my own experience and/or preferences. Sources of further information regarding the programs is given in Table 6.7.

WACALIB	Weighted averaging regression and calibration; bootstrapping; Gaussian logit calibration.
CALIB	Weighted averaging regression and calibration; partial least squares regression and calibration; weighted averaging partial least squares regression and calibration; jack-knifing.
WAPLS	Weighted averaging partial least squares regression and calibration and Monte Carlo simulation of sample-specific error estimation.
ANALOG	Modern analogue techniques; associated permutation tests.
MAT	Modern analogue techniques
RSURF	Modern analogue techniques.
SIRIUS	Partial least squares regression (PLS-1); multivariate partial least squares regression (PLS-2).
CANCOR	Canonical correlation analysis.
GLIM	Generalised linear modelling including Gaussian logit regression.
GLR	Gaussian logit regression.
GENSTAT	Generalised linear modelling including Gaussian logit regression, multinomial logit regression and calibration, classical and inverse linear regression, and multiple linear regression; principal components regression; correspondence analysis regression; canonical correlation analysis.
CANOCO	Detrended correspondence analysis; canonical correspondence analysis; detrended canonical correspondence analysis, redundancy analysis; principal components regression; correspondence analysis regression.
CABFAC, THREAD, REGRESS	Imbrie and Kipp (1971) variant of principal components regression (Q-mode factor analysis with varimax rotation).
RMSEP	Prediction error and bias statistics (Power, 1993; Mayer and Butler, 1993; Altman and Bland, 1983).
S-PLUS	Generalised additive models.
GAIM	Generalised additive models.

PPPHALOS	Package of statistical programs for palaeoenvironmental reconstructions developed by Joel Guiot, and including modern analogue techniques, inverse regression, principal components analysis, etc.

Table 6.6 *Computer Software for Palaeoenvironmental Reconstructions*

CANOCO is available from Microcomputer Power, 111 Clover Lane, Ithaca, New York 14850, USA.

WACALIB, ANALOG, RSURF, CANCOR, and RMSEP are available from H.J.B. Birks, Botanical Institute, University of Bergen, Allégaten 41, N-5007 Bergen, Norway.

CALIB, WAPLS, GLR, and MAT are available from Steve Juggins, Department of Geography, University of Newcastle, Newcastle-upon-Tyne NE1 7RH, UK.

GENSTAT and GLIM are available from NAG Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 8DR, UK.

S-PLUS is available from StatSci Europe, 52 Sandfield Road, Oxford OX3 7RJ, UK.

SIRIUS is available from Pattern Recognition Systems Ltd, Rosenbergsgt. 8 N-5010 Bergen, Norway.

CABFAC, THREAD, and REGRESS are available from Hans Schrader or Tim Schrader, Department of Geology, University of Bergen, Allégaten 41, N-5007 Bergen, Norway.

GAIM is available from S.N. Tibshirani Enterprises Inc., 5334 Yonge Street, Suite 1714, Toronto, Ontario M2N 6M2, Canada.

PPPHALOS is available from J. Guiot, Laboratoire de Botanique Historique & Palynologie, UA 1152 du CNRD, Faculté de St Jérôme, F-13397 Marseille cedex 13, France.

Table 6.7 Sources of Computer Software listed in Table 6.6.

DISCUSSION

There is currently a bewildering range of numerical procedures available for quantitative palaeoenvironmental reconstructions. Some of these have been developed for one group or organisms in relation to particular environmental variables over a specific spatial scale, whereas others have been used with several groups of organisms and environmental variables at a range of scales.

The most important question in trying to decide what reconstruction methods to use for a particular data-set is whether to use linear- or unimodal-based methods. The amount of biological compositional turnover along the environmental gradient of interest, as

estimated by DCCA, is a useful guide. After deciding whether to use linear- or unimodal-based methods, it is impossible to say *a priori* that a particular method is recommended to use with a given data-set. Little is currently known about the statistical properties of the different reconstruction methods with data-sets that may differ in their sample size, taxon richness, gradient length, amounts of non-structured and structured noise, deviations from linearity or unimodality, etc. There is clearly a need for critical comparative studies.

Comparisons of the predictive power of different methods using "real" data are not simple. Ter Braak (1995) discusses the problems involved and shows that RMSEP estimated by statistical cross-validation is not always an unbiased estimate of the true RMSEP, as assessed by an independent test set. RMSEP based on cross-validation can show bias depending on the reconstruction method used. For example, RMSEP is strongly overestimated by cross-validation in the MAT (ter Braak, 1995). A rigorous and reliable comparison of the predictive abilities of different methods requires an independent test or evaluation set (ter Braak and van Dam, 1989; Juggins, 1992; Borggaard and Thodberg, 1992). As very few independent evaluation data-sets are currently available, an alternative approach is to use simulated data of known properties for both training and test sets (e.g. Thomas and Haaland, 1990; Haaland and Thomas, 1988a, 1988b; ter Braak and Juggins, 1993; ter Braak *et al.*, 1993; Le and Shackleton, 1994; Borggaard and Thodberg, 1992) to discover the strengths and weaknesses of current reconstruction procedures with data of different properties. The problem is how to design the simulated training and test sets so that they are as ecologically realistic as possible, when we have little idea of what is ecologically realistic! Minchin's (1987) COMPAS simulation procedure provides a valuable basis for simulating a large number of biological data-sets based on a range of gradient lengths, species-environment response models, taxon richness, etc. Much work remains to be done using simulated training and test sets and different reconstruction procedures, along the lines of ter Braak and Juggins (1993), ter Braak *et al.* (1993), and ter Braak (1995), to evaluate different procedures.

For the palaeoecologist, it is desirable, given the current state of reconstruction methods, to reconstruct the same environmental variable using different reconstruction procedures, to compare the resulting reconstructions, and even to develop a consensus reconstruction (e.g. Bartlein and Whitlock, 1993). Such comparative studies are remarkably few, despite the pioneering comparisons by Hutson (1977) and Webb and Clark (1977). Limited comparisons of different reconstruction techniques include ter Braak and van Dam (1989), Birks *et al.* (1990a), Le (1992), ter Braak and Juggins (1993), ter Braak *et al.* (1993), ter Braak (1995), Bartlein and Whitlock (1993), Thunell

et al. (1994), Shane and Anderson (1993), Juggins *et al.* (1994), and Korsman and Birks (1995).

A further topic that may repay critical study is the question of estimating optimal transformations for particular reconstruction procedures and data-sets. Questions of transformations of biological data and environmental data are only rarely considered by palaeoecologists (e.g. Cumming and Smol, 1993b; Legendre and Dutilleul, 1991; Bartlein *et al.*, 1986; Bartlein and Webb, 1985). The alternating conditional expectation (ACE) algorithm is a recently developed non-parametric tool for finding optimal transformations that will maximise r^2 of the response variable Y with the predictors X in multiple regression (Breiman and Freidman, 1985; De Veaux, 1990). It could usefully be expanded for other regression-based procedures to find the appropriate optimal transformations.

It is perhaps surprising to palaeoecologists that much new research in statistical calibration that is potentially very relevant to palaeoenvironmental reconstruction techniques is being made in the fields of analytical chemistry and chemometrics (e.g. Martens and Næs, 1989). For example, various non-linear extensions of PLS are being developed (Frank, 1990; Wold, 1992; Sekulic *et al.*, 1993) by chemometricians. Neural-network techniques are becoming available for calibration purposes in chemometrics (e.g. Borggaard and Thodberg, 1992; Carlin *et al.*, 1994; Smits *et al.*, 1994; Cense *et al.*, 1994), with impressive reductions (50%-75%) of the RMSEP for simulated test data-sets (Borggaard and Thodberg, 1992). Techniques are being developed to linearise non-linear near infra-red absorbance data by multiplicative scatter correction (Næs *et al.*, 1990) and to decrease the RMSEP by up to 70% using an interactive variable selection procedure in PLS (Lindgren *et al.*, 1994). All these developments and several others in chemometrics are of potential value to palaeoecologists attempting to reconstruct palaeoenvironmental variables from complex assemblages of fossils that result from both linear and non-linear responses to the past environment.

Despite these methodological developments and the need for comparative studies using real and simulated data, the biggest limitation in quantitative palaeoenvironmental reconstructions is the need for high-quality modern training sets of consistent and detailed taxonomy and comparable methodology and sedimentary environment (Imbrie and Kipp, 1971; Kipp, 1976; Birks *et al.*, 1994). There is a need to try to minimise "noise" in these data as much as possible, as techniques such as WA-PLS and MLM perform best with data containing low noise. Considerable care is needed in the sampling design of training sets along the primary environmental gradients of interest (Mohler, 1981, 1983; ter Braak and Looman, 1986). The problems of secondary gradients need critical evaluation in light of ter Braak *et al.'s* (1993) analysis of

simulated data with a primary gradient of interest and a secondary, "nuisance" gradient. A further problem is how to allow for the inherent (e.g seasonal, annual) variability in the modern environmental variables of interest (e.g. Palmer and Dixon, 1990) in the numerical reconstruction procedures.

In conclusion there are now several numerical techniques available for quantitative palaeoenvironmental reconstructions. Some have a clear underlying theory and rationale, some are computationally simpler than others, and some are more ecologically realistic than others. At present WA-PLS is to be recommended for species-rich data over long (> 2 standard deviation units) environmental gradients, whereas PLS is recommended for data over shorter gradients. PLS and the unimodal-based equivalent WA-PLS combine the desirable features of inverse linear regression and WA regression (high correlation) and of PCR and CAR (stable predictors of high variance). PLS and WA-PLS are attractive and robust reconstruction techniques for a wide range of palaeoecological data, although care is taken to select the optimal number of components by statistical cross-validation.

ACKNOWLEDGEMENTS

I am indebted to the many colleagues who have stimulated my interest in quantitative palaeoenvironmental reconstruction techniques, who have collaborated in such reconstruction work, and who have patiently discussed many aspects of the theory and practice of quantitative reconstructions. I am particularly grateful to Rick Battarbee, Hilary Birks, John Boyle, Brian Cumming, John Imbrie, Steve Juggins, Tom Korsman, Olav Kvalheim, John Line, Petr Smilauer, John Smol, Cajo ter Braak, and Tom Webb for their help, discussions, and stimulation. Many of these people have contributed directly or indirectly to the ideas presented in this chapter. I am also greatly indebted to Anne Birgit Ruud Hage for her skilled and devoted word-processing of the manuscript and to Sylvia Peglar for help with the illustrations. Steve Juggins has generously provided me with copies of his unpublished C++ programs for PLS, WA-PLS, GLR, and MAT.

REFERENCES

Agbeti, M.D. 1992. Relationship between diatom assemblages and trophic variables: a comparison of old and new approaches. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 1171-1175.

Allott, T.E.H., Harriman, R. and Battarbee, R.W. 1992. Reversibility of lake acidification at the Round Loch of Glenhead, Galloway, Scotland. *Environmental Pollution*, 77, 219-225.

Altman, D.G. and Bland, J.M. 1983. Measurement in medicine: the analysis of method comparison studies. *The Statistician*, 32, 307-317.

Anderson, D.M., Prell, W.L. and Barratt, N.J. 1989. Estimates of sea surface temperature in the Coral Sea at the Last Glacial Maximum. *Paleoceanography*, 4, 615-627.

Anderson, J.A. 1984. Regression and ordered categorical variables. *Journal of the Royal Statistical Society B*, 46, 1-30.

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

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

Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K., and Ritchie, J.C. 1989. Modern analogues of late-Quaternary pollen spectra from the western Interior of North America. *Journal of Biogeography*, 16, 573-596.

Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K. and Ritchie, J.C. 1991. Vegetation-pollen-climate relationships for the arcto-boreal region of North America and Greenland. *Journal of Biogeography*, 18, 565-582.

Andrews, J.T., Davis, P.T., Mode, W.N., Nichols, H., and Short, S.K. 1981. Relative departures in July temperatures in northern Canada for the past 6,000 yr. *Nature*, 289, 164-167.

Andrews, J.T. Mode, W.N., and Davis, P.T. 1980. Holocene climate based on pollen transfer functions, eastern Canadian Arctic. *Arctic and Alpine Research*, 12, 41-64.

Andrews, J.T. and Nichols, H. 1981. Modern pollen deposition and Holocene paleotemperature reconstructions, central Northern Canada. *Arctic and Alpine Research*, 13, 387-408.

Atkinson, T.C., Briffa, K.R., and Coope, G.R. 1987. Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature*, 325, 587-592.

Atkinson, T.C., Briffa, K.R., Coope, G.R., Joachim, M.J., and Perry, D.W. 1986. Climatic calibration of coleopteran data. In *Handbook of Holocene Palaeoecology and Palaeohydrology* (ed. B.E. Berglund), pp. 851-858. J. Wiley & Sons, Chichester.

Austin, M.P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio*, 42, 11-21.

Austin, M.P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio*, 69, 33-45.

Austin, M.P. 1992. Modelling the environmental niche of plants: implications for plant community response to elevated CO_2 levels. *Australian Journal of Botany*, 40, 615-630.

Austin, M.P. and Gaywood, M.J. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science*, 5, 473-482.

Austin, M.P., Nicholls, A.O., Doherty, M.D., and Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a ß-function. *Journal of Vegetation Science*, 5, 214-228.

Barber, K.E., Chambers, F.M., Maddy, D., Stoneman, R., and Brew, J.S. 1994. A sensitive high-resolution record of late Holocene climatic change from a raised bog in northern England. *The Holocene*, 4, 198-205.

Bartlein, P.J., Prentice, I.C., and Webb, T. 1986. Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, 13, 35-57.

Bartlein, P.J. and Webb, T. 1985. Mean July temperatures at 6000 yr B.P. in eastern North America: Regression equations for estimates from fossil-pollen data. *Syllogeus* 55, 301-342.

Bartlein, P.J., Webb, T., and Fleri, E. 1984. Holocene climatic change in the northern Midwest: pollen-derived estimates. *Quaternary Research*, 22, 361-374.

Bartlein, P.J. and Whitlock, C. 1993. Paleoclimatic interpretation of the Elk Lake pollen record. *Geological Society of America Special Paper*, 276, 275-293.

Battarbee, R.W. 1991. Recent paleolimnology and diatom-based environmental reconstruction. In *Quaternary Landscapes* (eds. L.C.K. Shane and E.J. Cushing), pp. 129-174. University of Minnesota Press, Minneapolis.

Battarbee, R.W., Flower, R.J., Stevenson, A.C., and Harriman, R.W. 1989. Rapid reversibility of lake acidification? Reply. *Nature*, 337, 217-228.

Beebe, K.R. and Kowalski, B.R. 1987. An introduction to multivariate calibration and analysis. *Analytical Chemistry*, 59, 1007-1016.

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

Berger, W.H. and Gardner, J.V. 1975. On the determination of Pleistocene temperatures form planktonic foraminifera. *Journal of Foraminiferal Research*, 5, 102-113.

Bernabo, J.C. 1981. Quantitative estimates of temperature changes over the last 2700 years in Michigan based on pollen data. *Quaternary Research*, 15, 143-159.

Binford, M.W., Brenner, M., Whitmore, J.J., Higuera-Gundy, A., Deevey, E.S., and Leyden, B. 1987. Ecosystems, paleoecology and human disturbance in subtropical and tropical America. *Quaternary Science Reviews*, 6, 115-128.

Birks, H.J.B. 1993. Quaternary palaeoecology and vegetation science - current contributions and possible future developments. *Review of Palaeobotany and Palynology*, 79, 153-177.

Birks, H.J.B. 1994. The importance of pollen and diatom taxonomic precision in quantitative palaeoenvironmental reconstruction. *Review of Palaeobotany and Palynology*, 83, 107-117.

Birks, H.J.B. and Birks, H.H. 1980. *Quaternary Palaeoecology*. Edward Arnold, London, 289 pp.

Birks, H.J.B. and Gordon, A.D. 1985. *Numerical Methods in Quaternary Pollen Analysis*. Academic Press, London, 317 pp.

Birks, H.J.B., Juggins, S., and Line, J.M. 1990b. Lake surface-water chemistry reconstructions from palaeolimnological data. In *The Surface Waters Acidification Programme* (ed. B.J. Mason), pp. 301-313. Cambridge University Press, Cambridge.

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

Bonnefille, R., Chalié, F., Guiot, J., and Vincens, A. 1992. Quantitative estimates of full glacial temperatures in equatorial Africa from palynological data. *Climate Dynamics*, 6, 251-257.

Bonnefille, R., Roeland, J.C., and Guiot, J. 1990. Temperature and rainfall estimates for the past 40,000 years in equatorial Africa. *Nature*, 346, 347-349.

Borggaard, C. and Thodberg, H.H. 1992. Optimal minimal neural interpretation of spectra. *Analytical Chemistry*, 64, 545-551.

Breiman, L. and Friedman, J.H. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-619.

Brenner, M., Whitmore, T.J., Flannery, M.S., and Binford, M.W. 1993. Paleolimnological methods for defining target conditions in lake restoration: Florida case studies. *Lake and Reservoir Management*, 7, 209-217.

Brown, G.H. 1979. An optimization criterion for linear inverse estimation. *Technometrics*, 21, 575-579.

Brown, P.J. 1982. Multivariate calibration. *Journal of the Royal Statistical Society B*, 44, 287-321.

Brown, P.J. and Sundberg, R. 1987. Confidence and conflict in multivariate calibration. *Journal of the Royal Statistical Society B*, 49, 46-57.

Bryson, R.A. and Kutzbach, J.E. 1974. On the analysis of pollen-climate canonical transfer functions. *Quaternary Research*, 4, 162-174.

Bryson, R.A. and Swain, A.M. 1981. Holocene variations of monsoon rainfall in Rajasthan. *Quaternary Research*, 16, 135-145.

Carlin, M., Kavli, T., and Lillekjendlie, B. 1994. A comparison of four methods for non-linear data modelling. *Chemometrics and Intelligent Laboratory Systems*, 23, 163-177.

Cense, J.M., Diawara, B., Legendre, J.J., and Roullet, G. 1994. Neural networks prediction of partition coefficients. *Chemometrics and Intelligent Laboratory Systems*, 23, 301-398.

Charles, D.F. 1985. Relationships between surface sediment diatom assemblages and lake water characteristics in Adirondack lakes. *Ecology*, 66, 994-1011.

Charles, D.F. and Smol, J.P. 1988. New methods for using diatoms and chrysophytes to infer past Ph of low-alkalinity lakes. *Limnology and Oceanography*, 33, 1451-1462.

Christie, C.E. and Smol, J.P. 1993. Diatom assemblages as indicators of lake trophic status in southeastern Ontario lakes. *Journal of Phycology*, 29, 575-586.

Chutter, F.M. 1972. An empirical biotic index of the quality of water in South African streams and rivers. *Water Research*, 6, 19-30

Cleveland, W.S. 1993. *Visualizing data*. Hobart Press, Summit, New Jersey, 360 pp.

Cleveland, W.S. 1994. *The Elements of Graphing Data* (second edition). Hobart Press, Summit, New Jersey, 297 pp.

Cleveland, W.S. and Devlin, S.J. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, 596-610.

Cole, H.S. 1969. Objective reconstruction of the paleoclimatic record through the application of eigenvectors of present-day pollen spectra and climate to the late-Quaternary pollen stratigraphy. Doctoral thesis, University of Wisconsin, 110 pp.

Cumming, B.F. and Smol, J.P. 1993a. Development of diatom-based salinity models for paleoclimatic research from lakes in British Columbia (Canada), *Hydrobiologia*, 269/270, 179-196

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

Cumming, B.F., Smol, J.P., and Birks, H.J.B. 1992a. Scaled chrysophytes (Chrysophyceae and Synurophyceae) from Adirondack drainage lakes and their relationship to environmental variables. *Journal of Phycology*, 28, 162-178.

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., and Selle, A.R. 1992b. How much acidification has occurred in Adirondack lakes (New York, USA) since preindustrial times? *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 128-141.

Cook, R.D. and Weisberg, S. 1982. *Residuals and influence in regression*. Chapman and Hall, London, 230 pp.

D'Antoni, H.L. 1993. Paleotemperatures of La Malinche: a palynological hypothesis. *Grana*, 32, 354-358.

D'Antoni, H.L. and Spanner, M.A. 1993. Remote sensing and modern pollen dispersal in Southern Patagonia and Tierra del Fuego (Argentina): Models for palaeoecology. *Grana*, 32, 29-39.

Davies, P.T. and Tso, M.K.-S. 1982. Procedures for reduced-rank regression. *Applied Statistics*, 31, 244-255.

Davis, J.C. 1986. *Statistics and data analysis in geology* (Second edition). J. Wiley & Sons, New York, 646 pp.

Davis, R.B. and Anderson, D.S. 1985. Methods of pH calibration of sedimentary diatom remains for reconstructing history of pH in lakes. *Hydrobiologia*, 120, 69-87.

Davis, R.B., Anderson, D.S., Whiting, M.C., Smol, J.P., and Dixit, S.S. 1990. Alkalinity and pH of three lakes in northern New England, U.S.A., over the past 300 years. *Philosophical Transactions of the Royal Society of London* B, 327, 413-421.

Descy, J.P. 1979. A new approach to water quality estimation using diatoms. *Nova Hedwigia Beihefte*, 64, 305-323.

De Veaux, R.D. 1990. Finding transformations for regression using the ACE algorithm. In *Modern Methods of Data Analysis* (eds. J. Fox and J. Scott Long), pp. 177-208. Sage Publications, Newsbury Park.

Diaconis, P. and Efron, B. 1983. Computer-intensive methods in statistics. *Scientific American*, 248 (5), 96-100.

de Jong, S. 1991. Chemometrical applications in an industrial food research laboratory. *Mikrochimica Acta*, 2, 93-101.

de Jong, S. 1993. PLS fits closer than PCR. *Journal of Chemometrics*, 7, 551-557.

de Jong, S. and Kiers, H.A.L. 1992. Principal covariates regression Part I. Theory. *Chemometrics and Intelligent Laboratory Systems*, 14, 155-164.

Dixit, S.S. 1986. Diatom-inferred pH calibration of lakes near Wawa, Ontario. *Canadian Journal of Botany*, 64, 1129-1133.

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

Dixit, S.S., Dixit, A.A., and Smol, J.P. 1990. Relationship between chrysophyte assemblages and environmental variables in seventy-two Sudbury lakes as examined by canonical correspondence analysis (CCA). *Canadian Journal of Fisheries and Aquatic Sciences*, 46, 1167-1176.

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

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics CMBS-NSF Regional Conference Series in Applied Mathematics, 38, 1-92.

Efron, G. and Gong, G. 1983. A leisurely look at the bootstrap, the jackknife, and cross validation. *American Statistician*, 37, 36-48.

Efron, B. and Tibshirani, R.J. 1993. *Introduction to the bootstrap*. Chapman and Hall, London, 450 pp.

Ellenberg, H. 1948. Unkrautgesellschaften als Mass für den Säuregrad, die Verdichtung und andere Eigenschaften des Ackerbodens. *Berichte über Landtechnik, Kuratorium für Technik und Bauwesenin der Landwirtschaft*, 4, 130-146.

Engstrom, D.R. and Nelson, S.R. 1991 Paleosalinity from trace metals in fossil ostracodes compared with observational records at Devils Lake, North Dakota, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 82, 47-73.

Feoli, E. and Orlóci, L. 1979. Analysis of concentration and detection of underlying factors in structured tables. *Vegetatio*, 40, 49-54.

Flower, R.J. 1986. The relationships between surface sediment diatom assemblages and pH in 33 Galloway lakes: some regression models for reconstructing pH and their application to sediment cores. *Hydrobiologia*, 143, 93-103.

Frank, I.E. 1990. A nonlinear PLS model. *Chemometrics and Intelligent Laboratory Systems*, 8, 109-119

Frank, I.E. and Friedman, J.H. 1993. A statistical view of some chemometrics regression tools. *Technometrics*, 35, 109-148.

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, 1771-1781.

Fritz, S.C., Engstrom, D.R., and Haskell, B.J. 1994. "Little Ice Age" aridity in the North American Great Plains: a high-resolution reconstruction of salinity fluctuations from Devils Lake, North Dakota, USA. *The Holocene*, 4, 69-73.

Fritz, S.C., Juggins, S., Battarbee, R.W., and Engstrom, D.R. 1991. Reconstruction of past changes in salinity and climate using a diatom-based transfer function. *Nature*, 352, 706-708.

Fritz, S.C., Kingston, J.C. and Engstrom, D.R. 1993. Quantitative trophic reconstruction from sedimentary diatom assemblages: a cautionary tale. *Freshwater Biology*, 30, 1-23.

Gaillard, M.-J., Birks, H.J.B., Emanuelsson, U., and Berglund, B.E. 1992. Modern pollen/land-use relationships as an aid in the reconstruction of past land-uses and cultural landscapes: an example from south Sweden. *Vegetation History and Archaeobotany*, 1, 3-17.

Gaillard, M.-J., Birks, H.J.B., Emanuelsson, U., Karlsson, K., Lagerås, P., and Olausson, D. 1994. Application of modern pollen/land-use relationships to the interpretation of pollen diagrams - reconstructions of land-use history in south Sweden, 3000-0 BP. *Review of Palaeobotany and Palynology*, 82, 47-73.

Gaillard, M.-J., Birks, H.J.B. Karlsson, K. and Lagerås, P. 1995. Quantitative reconstruction of past land-use and soil conditions using the modern analogue approach - a case study in south Sweden. *PACT Volume for Urve Miller* (submitted).

Gaillard, M.-J. and Lemdahl, G. 1994. Late-glacial insect assemblages from Grand Marais, south-western Switzerland - climatic implications and comparison with pollen and plant macrofossil data. *Dissertationes Botanicae*, 234, 287-308.

Gallant, A.R. 1975. Nonlinear regression. *American Statistician*, 29, 73-81.

Gasse, F. 1986. East African diatoms and water pH. In *Diatoms and Lake Acidity* (eds. J.P. Smol, R.W. Battarbee, R.B. Davis, and J. Merilãinen), pp. 149-168. Dr. W. Junk, Dordrecht.

Gasse, F., Juggins, S., and Ben Khelifa, L. 1995. Diatom-based transfer functions for inferring hydrochemical characteristics of African palaeolakes. *Palaeogeography, Palaeoclimatology, Palaeoecology* (in press).

Gasse, F. and Tekaia, F. 1979. La paléoécologie des diatomées évolution des lacs de l'Afar Central. *Les cahiers de l'Analyse des Données*, 4, 81-94.

Gasse, F. and Tekaia, F. 1983. Transfer functions for estimating paleoecological conditions (pH) from East African diatoms. *Hydrobiologia*, 103, 85-90.

Gauch, H.G. 1982. Noise reduction by eigenvector ordination. *Ecology*, 63, 1643-1649.

Gauch, H.G., and Whittaker, R.H. 1972. Coenocline simulation. *Ecology*, 53 446-451.

Gause, G.F. 1930. Studies on the ecology of the Orthoptera. *Ecology*, 11, 307-325.

Geladi, P. 1988. Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics*, 2, 231-246.

Geladi, P. and Kowalski, B.R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1-17.

GENSTAT 5 Committee. 1987. *GENSTAT 5 Reference Manual*. Clarendon Press, Oxford, 749 pp.

Gignac, L.D., Vitt, D.H., and Bayley, S.E. 1991b. Bryophyte response surfaces along ecological and climatic gradients. *Vegetatio*, 93, 29-45.

Gignac, L.D., Vitt, D.H., Zoltai, S.C., and Bayley, S.E. 1991a. Bryophyte response surfaces along climatic, chemical, and physical gradients in peatlands of western Canada. *Nova Hedwigia*, 53, 27-71.

Gittins, R. 1985. Canonical analysis. A review with applications in ecology. Springer-Verlag, Berlin, 351 pp.

Glahn, H. 1968. Canonical correlation and its relationship to discriminant analysis and multiple regression. *Journal of Atmospheric Science,s* 25, 23-31.

Grichuk, V.P. 1969. An attempt to reconstruct certain elements of the climate of the northern hemisphere in the Atlantic Period of the Holocene. In *Golotsen* (ed. M.I. Neishtadt). Moscow, Izd-vo Nauka.

Grimm, E.C. 1988. Data analysis and display. In *Vegetation History* (ed. B. Huntley and T. Webb), pp. 43-76. Kluwer, Dordrecht.

Grönlund, E. 1991. Sediment characteristics in relation to cultivation history in two varved lake sediments from East Finland. *Hydrobiologia*, 214, 137-142.

Guiot, J. 1985. A method for palaeoclimatic reconstruction in palynology based on multivariate time-series analysis. *Géographie physique et Quaternaire*, 39, 115-125

Guiot, J. 1987. Late Quaternary climatic change in France estimated from multivariate pollen time series. *Quaternary Research*, 28, 100-118.

Guiot, J. 1990. Methodology of the last climatic cycle reconstruction in France from pollen data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 80, 49-69.

Guiot, J. and Couteaux, M. 1992. Quantitative climate reconstruction form pollen data in the Grand Duchy of Luxembourg since 15000 yr BP. *Journal of Quaternary Science*, 7, 303-309.

Guiot, J. and Pons, A., 1986. Une méthode de reconstruction quantitative du climat à partir de chroniques pollenanalytiques: le climat de la France depuis 15 000 ans. *Comptes Rendus de l'Académie des Sciences (Paris) Série II*, 302, 911-916.

Guiot, J., Pons, A., de Beaulieu, J.L., and Reille, M. 1989. A 140,000-year continental climate reconstruction from two European pollen records. *Nature*, 338, 309-313.

Guiot, J., Reille, M., de Beaulieu, J.L., and Pons, A. 1992. Calibration of the climatic signal in a new pollen sequence from La Grand Pile. *Climate Dynamics*, 6, 259-264.

Guiot, J., de Beaulieu, J.L., Cheddadi, R., David, F., Ponel, P., and Reille, M. 1993b. The climate in western Europe during the last glacial/interglacial cycle derived from pollen and insect remains. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 103, 73-93.

Guiot, J., Harrison, S.P., and Prentice, I.C. 1993a. Reconstruction of Holocene precipitation patterns in Europe using pollen and lake-level data. *Quaternary Research*, 40, 139-149.

Haaland, D.M. and Thomas. E.V. 1988a. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and extraction of qualitative information. *Analytical Chemistry*, 60, 1193-1202.

Haaland, D.M. and Thomas, E.V. 1988b. Partial least-squares methods for spectral analyses. 2 Application to simulated and glass spectral data. *Analytical Chemistry*, 60, 1202-1208.

Hall, R.I. and 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. and Smol, J.P. 1993. The influence of catchment size on lake trophic status during the hemlock decline and recovery (4800 to 3500 BP) in southern Ontario lakes. *Hydrobiologia*, 269/270, 371-390.

Hammerlund, D. and Lemdahl, G. 1994. A Late Weichselian stable isotope stratigraphy compared with biostratigraphical data: a case study from southern Sweden. *Journal of Quaternary Science*, 9, 13-31.

Hastie, T.J. and Tibshirani, R.J. 1986. Generalized additive models. *Statistical Science* 1, 297-318.

Hastie, T.J. and Tibshirani, R.J. 1990 *Generalized additive models*. Chapman and Hall, London, 335 pp.

Hawkins, D.M. 1973. On the investigation of alternative regressions by principal component analysis. *Applied Statistics*, 22, 275-286.

Helland, I.S. 1988. On the structure of partial least squares regression. *Communications in Statistics, Simulation and Computing*, 17, 581-607.

Helland, I.S. 1990. Partial least squares regression and statistical methods. *Scandinavian Journal of Statistics*, 17, 97-114.

Hengeveld, R. 1990. *Dynamic biogeography*. Cambridge University Press, Cambridge, 249 pp.

Heusser, C.J. and Heusser, L.E. 1980. Sequence of pumiceous tephra layers and the late Quaternary environmental record near Mount St. Helens. *Science*, **210**, 1007-1009.

Heusser, C.J. and Heusser, L.E. 1981. Palynology and paleotemperature analysis of the Whidbey Formation, Puget Lowland, Washington. *Canadian Journal of Earth Sciences*, **18**, 136-149.

Heusser, C.J., Heusser, L.E., and Streeter, S.S. 1980 Quaternary temperatures and precipitation for the north-west coast of North America. *Nature*, **286**, 702-704.

Heusser, C.J. and Streeter, S.S. 1980. A temperature and precipitation record of the past 16,000 years in southern Chile. *Science*, **210**, 1345-1347.

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

Hill, M.O. 1973b. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology*, **61**, 237-249.

Hill, M.O. 1979. DECORANA - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Cornell University, Ithaca, New York, USA.

Hill, M.O. 1989. Computerized matching of relevés and association tables, with an application to the British National Vegetation Classification. *Vegetatio*, **83**, 187-194.

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

Hocking, R.R. 1983. Developments in linear regression methodology: 1959-1982. *Technometrics*, **25**, 219-230.

Howe, S. and Webb, T. 1977. Testing the statistical assumptions of paleoclimatic calibration functions. Preprint Volume: Fifth Conference on Probability and Statistics in Atmospheric Sciences, American Meteorological Society, Boston, pp.152-157.

Howe, S. and Webb, T. 1983. Calibrating pollen data in climatic terms: improving the methods. *Quaternary Science Reviews*, **2**, 17-51.

Huntley, B. 1990a. Dissimilarity mapping between fossil and contemporary pollen spectra in Europe for the past 13,000 years. *Quaternary Research*, **33**, 360-376.

Huntley, B. 1990b. European post-glacial forests: compositional changes in response to climatic change. *Journal of Vegetation Science*, **1**, 507-518.

Huntley, B. 1992. Pollen-climate response surfaces and the study of climate change. In *Applications of Quaternary Research* (ed. J.M. Gray), pp. 91-99. *Quaternary Proceedings*, **2**, Quaternary Research Association, Cambridge.

Huntley, B. 1993. The use of climate response surfaces to reconstruct palaeoclimate from Quaternary pollen and plant macrofossil data. *Philosophical Transactions of the Royal Society of London B*, **341**, 215-224.

Huntley, B. 1994. Devensian late-glacial palaeoclimate: the Durham contribution. In *Palaeoclimate of the Last Glacial/Interglacial Cycle* (eds. B.M. Funnell and R.L.F Kay), pp. 53-55. Special Publication 94/2 of the NERC Earth Sciences Directorate.

Huntley, B., Bartlein, P.J., and Prentice, I.C. 1989. Climatic control of the distribution and abundance of beech (*Fagus* L.) in Europe and North America. *Journal of Biogeography*, **16**, 551-560.

Huntley, B. and Prentice, I.C. 1988. July temperatures in Europe from pollen data, 6000 years before present. *Science* **241**, 687-690.

Huntley, B. and Prentice, I.C. 1993. Holocene vegetation and climates of Europe. In *Global Climates since the Last Glacial Maximum* (eds. H.E. Wright, J.E. Kutzbach, T. Webb, W.F. Ruddiman, F.A. Street-Perrott, and P.J. Bartlein), pp. 136-168. University of Minnesota Press, Minneapolis.

Hutson, W.H. 1977. Transfer functions under no-analog conditions: experiments with Indian Ocean planktonic foraminifera. *Quaternary Research*, **8**, 355-367.

Hutson, W.H. 1978. Application of transfer functions to Indian Ocean planktonic foraminifera. *Quaternary Research*, **9**, 87-112.

Hutson, W.H. 1980. The Agulhas Current during the last Pleistocene: analysis of modern faunal analogs. *Science*, **207**, 64-66.

Huttunen, P. and Meriläinen, J. 1986. A comparison of different pH indices derived from diatom assemblages. *University of Joensuu Publications of the Karelian Institute*, **79**, 41-46.

Höskuldsson, A. 1988. PLS regression methods. *Journal of Chemometrics*, **2**, 211-228.

Ihm, P. and van Groenewoud, H. 1984. Correspondence analysis and Gaussian ordination. *COMPSTAT Lectures*, **3**, 5-60.

Imbrie, J. and Kipp, N.G. 1971. A new micropaleontological method for quantitative paleoclimatology: application to a late Pleistocene Caribbean core. In *The Late Cenozoic Glacial Ages* (ed. K.K. Turekian), pp. 71-181. Yale University Press, New Haven and London.

Imbrie, J., van Donk, J., and Kipp, N.G. 1973. Paleoclimatic investigation of a Late Pleistocene Caribbean deep-sea core: comparison of isotopic and faunal methods. *Quaternary Research*, **3**, 10-38.

Imbrie, J. and Webb, T. 1981. Transfer functions: calibrating micropaleontological data in climatic terms. In *Climate Variations and Variability: Facts and Theories* (ed. A. Berger), pp. 125-134. D. Reidel, Dordrecht.

Israëls, A.Z. 1984. Redundancy analysis for qualitative variables. *Psychometrika*, **49**, 331-346.

Iversen, J. 1944. Viscum, Hedera and Ilex as climate indicators. Geologiska Föreningens i Stockholm Förhandlingar, **66**, 463-483.

Jackson, J.E. 1991. *A user's guide to principal components*. J. Wiley & Sons, New York, 569 pp.

Jager, J.C. and Looman, C.W.N. 1987. Data collection. In *Data analysis in community and landscape ecology* (ed. R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren), pp. 10-28. Pudoc, Wageningen.

Janssens, J.A., Hansen, B.C.S., Glaser, P.H., and Whitlock, C. 1992. Development of a raised-bog complex. In *The Patterned Peatlands of Minnesota* (eds. H.E. Wright, B.A. Coffin, and N.E. Aaseng), pp. 189-221. University of Minnesota Press, Minneapolis.

Jolliffe, I.T. 1982. A note on the use of principal components in regression. *Applied Statistics*, **31** 300-303.

Jolliffe, I.T. 1986. *Principal component analysis*. Springer-Verlag, New York, 271 pp.

Juggins, S. 1992. Diatoms in the Thames Estuary, England: Ecology, palaeoecology, and salinity transfer function. *Bibliotheca Diatomologica*, **25**, 216 pp.

Juggins, S., Battarbee, R.W., and Fritz, S.C. 1994. Diatom/ salinity transfer functions and climate change: an assessment of methods and application to two Holocene sequences from the northern Great Plains, North America. In *Palaeoclimate of the Last Glacial/Interglacial Cycle* (eds. B.M. Funnell and R.L.F. Kay), pp. 37-41. Special Publication 94/2 of the NERC Earth Sciences Directorate.

Jørgensen, B. 1983. Maximum likelihood estimation and large-sample inference for generalized linear and non-linear regression models. *Biometrika*, 70, 19-28.

Karpuz, N.K. and Schrader, H. 1990. Surface sediment diatom distribution and Holocene paleotemperature variations in the Greenland, Iceland and Norwegian Sea. *Paleoceanography*, 5, 557-580.

Kay, P.A. 1979. Multivariate statistical estimates of Holocene vegetation and climate change, forest-tundra transition zone, NWT, Canada. *Quaternary Research*, 11, 125-140.

Kay, P.A. and Andrews, J.T. 1983. Re-evaluation of pollen-climate transfer functions in Keewatin, northern Canada. *Annals of the Association Geographers*, 73, 550-559.

Keen, K.L. and Shane, L.C.K. 1990. A continuous record of Holocene eolian activity and vegetation change at Lake Ann, east-central Minnesota. *Geological Society of American Bulletin*, 102, 1646-1657.

Kelly, M.G. and Huntley, B. 1991. An 11,000-year record of vegetation and environment from Lago di Martignano, Latium, Italy. *Journal of Quaternary Science*, 6, 209-224.

Kingston, J.C. and Birks, H.J.B. 1990. Dissolved organic carbon reconstructions from diatom assemblages in PIRLA project lakes, North America. *Philosophical Transactions of the Royal Society of London* B, 327, 279-288.

Kingston, J.C., Birks, H.J.B., Uutala, A.J., Cumming, B.F., and Smol, J.P. 1992. Assessing trends in fishery resources and lake water aluminum from paleolimnological analyses of siliceous algae. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 116-127.

Kipp, N.G. 1976. New transfer functions for estimating past sea-surface conditions from sea-bed distribution of planktonic foraminiferal assemblages in the North Atlantic. *Geological Society of America Memoir*, 145, 3-41.

Klovan, J.E. and Imbrie, J. 1971. An algorithm and FORTRAN-IV program for large-scale Q-mode factor analysis and calculation of factor scores. *Mathematical Geology*, 3, 61-77.

Korsman, T. 1994. Acidification trends in Sweden. An assessment of past water chemistry conditions using lake sediments. Doctoral thesis, University of Umeå, 72 pp.

Korsman, T. and Birks, H.J.B. 1995. Diatom-based water chemistry reconstructions: a comparison of reconstruction techniques. *Journal of Paleolimnology* (in press).

Korsman, T., Nilsson, M., Öhman, J., and Renberg, I. 1992. Near-Infrared reflectance spectroscopy of sediments: a potential method to infer the past pH of lakes. *Environmental Science and Technology*, 26, 2122-2126.

Kowalski, B.R. and Seasholtz, M.B. 1991. Recent developments in multivariate calibration. *Journal of Chemometrics*, 5, 129-145.

Krutchkoff, R.G. 1967. Classical and inverse regression methods of calibration. *Technometrics*, 9, 425-439.

Le, J. 1992. Palaeotemperature estimation methods: sensitivity test on two western equatorial Pacific cores. *Quaternary Science Reviews*, 11, 801-820.

Le, J. and Shackleton, N.J. 1994. Estimation of paleoenvironment by transfer functions: simulation with hypothetical data. *Marine Micropaleontology* (in press).

Legendre, P. and Dutilleul, P. 1991. Comments on Boyle's acidity and organic carbon in lake water: variability and estimation of means. *Journal of Paleolimnolgy*, 6, 103-110.

Lemdahl, G. 1991. A rapid climatic change at the end of the Younger Dryas in south Sweden - palaeoclimatic and palaeoenvironmental reconstructions based on fossil insect assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 83, 313-331.

Lenihan, J.M. 1993. Ecological response surfaces for North American boreal tree species and their use in forest classification. *Journal of Vegetation Science*, 4, 667-680.

Lindgren, F., Geladi, P., Rännar, S., and Wold, S. 1994. Interactive variable selection (IVS) for PLS. Part 1: Theory and Algorithms. *Journal of Chemometrics*, 8, 349-363.

Line, J.M. and Birks, H.J.B. 1990. WACALIB version 2.1 - a computer program to reconstruct environmental variables from fossil assemblages by weighted averaging. *Journal of Paleolimnology*, 3, 170-173.

Line, J.M., ter Braak, C.J.F., and Birks, H.J.B. 1994. WACALIB version 3.3 - a computer program to reconstruct environmental variables from fossil assemblages by weighted averaging and to derive sample-specific errors of prediction. *Journal of Paleolimnology*, 10, 147-152.

Lorber, A., Wangen, L.E., and Kowalski, B.R. 1987. A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*, 1, 19-31.

Lotter, A.F. and Birks, H.J.B. 1993. The impact of the Laacher See Tephra on terrestrial and aquatic ecosystems in the Black Forest (Southern Germany). *Journal of Quaternary Science*, 8, 263-276.

Lowe, J.J., Coope, G.R., Keen, D., and Walker, M.J.C. 1994. High resolution stratigraphy of the last glacial-interglacial transition (LGIT) and inferred climatic gradients. In *Palaeoclimate of the Last Glacial/Interglacial Cycle* (eds. B.M. Funnell and R.L.F. Kay), pp. 47-52. Special Publication 94/2 of the NERC Earth Sciences Directorate.

Lwin, T. and Maritz, J.S. 1982. An analysis of the linear-calibration controversy from the perspective of compound estimation. *Technometrics*, 24, 235-242 (minor correction in *Technometrics*, 27, 44).

Lynts, G.W. and Judd, J.B. 1971. Lake Pleistocene paleotemperatures at Tongue of the Ocean, Bahamas. *Science* 171, 1143-1144.

McAndrews, J.H. and Campbell, I.D. 1993. 6 ka July temperature in eastern Canada from Bartlein and Webb's (1985) pollen transfer functions: comments and illustrations. In *Proxy Climate Data and Models of the Six Thousand Years Before Present Time Interval: The Canadian Perspective*. (ed. A. Telka), pp. 22-25. Royal Society of Canada, Ottawa.

McCullagh, P. and Nelder, J.A. 1989. *Generalized linear models* (Second edition). Chapman and Hall, London, 511 pp.

MacDonald, G.M. and Reid, R.T. 1989. Pollen-climate distance surfaces and the interpretation of fossil pollen assemblages from the western interior of Canada. *Journal of Biogeography*, 16, 403-412.

Malinovski, E.R. and Howery, D.G. 1980. *Factor analysis in chemistry*, J. Wiley & Sons, New York, 251 pp.

Malmgren, B.A. and Healy-Williams, N. 1978. Variation in test diameter of *Orbulina universa* in the paleoclimatology of the Late Quaternary of the Gulf of Mexico. *Palaeogeography Palaeoclimatology, Palaeoecology*, 25, 235-240.

Malmgren, B.A. and Kennett, J.P. 1978a. Test size variations in *Globigerina bulloides* in response to Quaternary palaeoceanographic changes. *Nature*, 275, 123-124.

Malmgren, B.A. and Kennett, J.P. 1978b. Late Quaternary paleoclimatic applications of mean size variations in *Globigerina bulloides* D'Orbigny in the Southern Indian Ocean. *Journal of Paleontology*, 52, 1195-1207.

Manly, B.F.J. 1991. *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London, 281 pp.

Manne, R. 1987. Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, 2, 187-197.

Mansfield, E.R., Webster, J.T., and Gunst, R.F. 1977. An analytic variable selection technique for principal component regression. *Applied Statistics*, 26, 34-40.

Marchetto, A. 1994. Rescaling species optima estimated by weighted averaging. *Journal of Paleolimnology*, 12, 155-162.

Marsicano, L.J. and Siver, P.A. 1993. A paleolimnological assessment of lake acidification in five Connecticut lakes. *Journal of Paleolimnology*, 9, 209-221.

Martens, H. and Næs, T. 1984. Multivariate calibration. I. Concepts and distinctions. *Trends in Analytical Chemistry*, 3, 204-210.

Martens, H. and Næs, T. 1989. *Multivariate calibration*. J. Wiley & Sons, Chichester, 419 pp.

Martinelle, S. 1970. On the choice of regression in linear calibration. Comments on a paper by R.G. Krutchkoff. *Technometrics*, 12, 157-161.

Mathewes, R.W. and Heusser, L.E. 1981. A 12000 year palynological record of temperature and precipitation trends in southwestern British Columbia. *Canadian Journal of Botany*, 59, 707-710.

Meyer, D.G. and Butler, D.G. 1993. Statistical validation. *Ecological Modelling*, 68, 21-32.

Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio*, 71, 145-156.

Mohler, C.L. 1981. Effects of sample distribution along gradients on eigenvector ordination. *Vegetatio*, 45, 141-145.

Mohler, C.L. 1983. Effect of sampling pattern on estimation of species distribution along gradients. *Vegetatio*, 54, 97-102.

Molfino, B., Kipp, N.G., and Morley, J.J. 1982. Comparison of foraminiferal, coccolithophorid, and radiolarian paleotemperature equations: assemblage coherency and estimate concordancy. *Quaternary Research*, 17, 279-313.

Montgomery, D.C. and Peck, E.A. 1982. *Introduction to linear regression analysis*. J.Wiley & Sons, New York, 504 pp.

Morley, J.J. 1979. A transfer function for estimating paleoceanographic conditions based on deep-sea surface sediment distribution of radiolarian assemblages in the South Atlantic. *Quaternary Research*, 12, 381-395.

Morley, J.J. 1989a. Radiolarian-based transfer functions for estimating paleoceanographic conditions in the South Indian Ocean. *Marine Micropaleontology*, 13, 293-307.

Morley, J.J. 1989b. Variations in high-latitude oceanographic fronts in the southern Indian Ocean: An estimation based on faunal changes. *Paleoceanography*, 4, 547-554.

Morzuch, B.J. and Ruark, G.A. 1991. Principal components regression to mitigate the effects of multicollinearity. *Forest Science*, 37, 191-199.

Mourguirart, P., and Carbonel, P. 1994. A quantitative method of palaeolake-level reconstruction using ostracod assemblages: an example from the Bolivian Altiplano. *Hydrobiologia*, 288, 183-194.

Nilsson, M. and Bohlin, E. 1993. Methane and carbon dioxide concentrations in bogs and fens - with special reference to the effects of the botanical composition of the peat. *Journal of Ecology*, 81, 615-625.

Norton, D.A., McGlone, M.S., and Wigley, T.M.L. 1986. Quantitative analysis of modern pollen-climate relationships in New Zealand indigenous forests. *New Zealand Journal of Botany*, 24, 331-342.

Næs, T., Irgens, C., and Martens, H. 1986. Comparison of linear statistical methods for calibration of NIR instruments. *Applied Statistics*, 35, 195-206.

Næs, T., Isaksson, T., and Kowalski, B. 1990. Locally weighted regression and scatter correction for near-infrared reflectance data. *Analytical Chemistry*, 62, 664-673.

Næs, T. and Martens, H. 1984. Multivariate calibration. II. Chemometric methods. *Trends in Analytical Chemistry* 3, 266-271.

Næs, T. and Martens, H. 1985. Comparison of prediction methods for multicollinear data. *Communications in Statistics, Simulation and Computing*, 14, 545-576.

Næs, T. and Martens, H. 1988. Principal component regression in NIR analysis: viewpoints, background details and selection of components. *Journal of Chemometrics*, 2, 155-167.

Odgaard, B.V. 1992. The fire history of Danish heathland areas as reflected by pollen and charred particles in lake sediments. *The Holocene*, 2, 218-226.

Oksanen, J., Läärä, E., Huttunen, P., and Meriläinen, J. 1988. Estimation of pH optima and tolerances of diatoms in lake sediments by the methods of weighted averaging, least squares and maximum likelihood, and their use for the prediction of lake acidity. *Journal of Paleolimnology*, 1, 39-49.

Oksanen, J., Läärä, E., Huttunen, P., and Meriläinen, J. 1990. Maximum likelihood prediction of lake acidity based on sedimented diatoms. *Journal of Vegetation Science*, 1, 49-56.

Ollikainen, M., Simola, H., and Niinioja, R. 1993. Changes in diatom assemblages in the profundal sediments of two large oligohumic lakes in eastern Finland. *Hydrobiologia*, 269/270, 405-413.

Osborne, C. 1991. Statistical calibration: a review. *International Statistical Review*, 59, 309-336.

Osten, D.W. 1988. Selection of optimal regression models via cross-validation. *Journal of Chemometrics*, 2, 39-48.

Ott, W.R. 1990. A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air Waste Management Association*, 40, 1378-1383.

Overpeck, J.T., Webb, T., and Prentice, I.c. 1985. Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. *Quaternary Research*, 23, 87-108.

Palmer, M.W. and Dixon, P.M. 1990. Small-scale environmental heterogeneity and the analysis of species distributions along gradients. *Journal of Vegetation Science*, 1, 57-65.

Payne, C.D. (Ed.) 1986. *The GLIM System Release 3.77*. Numerical Algorithms Group, Oxford.

Peet, R.K., Knox, R.G., Case, J.S., and Allen, R.B. 1988. Putting things in order: the advantages of detrended correspondence analysis. *American Naturalist*, 131, 924-934.

Picard, R.R. and Cook, R.D. 1984. Cross-validation of regression models. *Journal of the American Statistical Association*, **79**, 575-583.

Ponel, P. and Coope, G.R. 1990. Late glacial and early Flandrian coleoptera from La Taphanel, Massif Central, France: climatic and ecological implications. *Journal of Quaternary Science*, **5**, 234-249.

Power, M. 1993. The predictive validation of ecological and environmental models. *Ecological Modelling*, **68**, 33-50.

Prell, W.L. 1985. The stability of low-latitude sea-surface temperatures: an evaluation of the CLIMAP reconstruction with emphasis on the positive SST anomalies. *Technical Report* **TR025**, Department of Energy, Washington DC, 60 pp.

Prentice, I.C. 1980. Multidimensional scaling as a research tool in Quaternary palynology: a review of theory and methods. *Review of Palaeobotany and Palynology*, **31**, 71-104.

Prentice, I.C., Bartlein, P.J., and Webb, T. 1991. Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology*, **72**, 2038-2056.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., and Solomon, A.M. 1992. A global biome based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117-134.

Rao, C.R. 1964. The use and interpretation of principal components analysis in applied research. *Sankhya A*, **26** 329-358.

Renberg, I. and Hultberg, H. 1992. A paleolimnological assessment of acidification and liming effects on diatom assemblages in a Swedish lake. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 65-72.

Robert, P. and Escoufier, Y. 1976. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied Statistics*, **25**, 257-265.

Rousseau, D.-D. 1991. Climatic transfer function from Quaternary molluscs in European loess deposits. *Quaternary Research*, **36**, 195-209.

Rousseau, D.-D., Limondin, N., Magnin, F., and Puissegur, J.-J. 1994. Temperature oscillations over the last 10,000 years in western Europe estimated from terrestrial mollusc assemblages. *Boreas*, **23**, 66-73.

Roux, M. 1979. Estimation des paléoclimats d'apres l'ecologie des foraminifères. *Cahiers de l'Analyse des Données*, **4**, 61-79.

Roux, M., Servant-Vildary, S., and Servant, M. 1991. Inferred ionic composition and salinity of a Bolivian Quaternary lake, as estimated from fossil diatoms in the sediments. *Hydrobiologia*, **210**, 3-18.

Sachs, H.M., Webb, T., and Clark, D.R. 1977. Paleoecological transfer functions. *Annual Review of Earth and Planetary Sciences*, **5**, 159-178.

Scheffé, H. 1973. A statistical theory of calibration. *Annals of Statistics*, **1**, 1-37.

Schrader, H., Swanberg, N., Lycke, A.K., Paetzel, M., Schrader, T., and Schrader, T. 1993. Diatom-inferred productivity changes in the eastern equatorial Pacific: The Quaternary record of ODP Leg 111, Site 677. *Hydrobiologia*, **269/270**, 137-151.

Schwartz, L.M. 1977. Nonlinear calibration. *Analytical Chemistry*, **49**, 2062-2068.

Sekulic, S., Seasholtz, M.B., Wang, Z., Kowalski, B.R., Lee, S.E., and Holt, B.R. 1993. Nonlinear multivariate calibration methods in analytical chemistry. *Analytical Chemistry*, **65**, 835-845.

Seret, G., Guiot, J., Wansard, G., de Beaulieu, J.L., and Reille, M. 1992. Tentative palaeoclimatic reconstruction linking pollen and sedimentology in La Grande Pile (Vosges, France). *Quaternary Science Reviews*, **11**, 425-430.

Servant-Vildary, S. and Roux, M. 1990. Variations de température estimées à partir du déplacement en altitude des associations de diatomées dans une séquence holocène de la Cordillère Orientale de Bolivie. *Comptes Rendus de l'Académie des Sciences (Paris) Série II*, **311**, 429-436.

Shane, L.C.K. and Anderson, K.H. 1993. Intensity, gradients and reversals in late glacial environmental change in east-central North-America. *Quaternary Science Reviews*, **12**, 307-320.

Siver, P.A. 1993. Inferring the specific conductivity of lake water with scaled chrysophytes. *Limnology and Oceanography*, **38**, 1480-1492.

Sjöström, M., Wold, S., Lindberg, W., Persson, J-Å., and Martens, H. 1983. A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Analytica Chimica Acta*, **150**, 61-70.

Smilauer, P. and Birks, H.J.B. 1995. The use of generalised additive models in the description of diatom-environment response surfaces. *Proceedings of the Salten Skov Conference on Lake Eutrophication*, May 1994 (in press).

Smith, A.J. 1993. Lacustrine ostracodes as hydrochemical indicators in lakes of the north-central United States. *Journal of Paleolimnology*, **8**, 121-134.

Smits, J.R.M., Melssen, W.J., Buydens, L.M.C., and Kateman, G. 1994. Using artificial neural networks for solving chemical problems Part I. Multi-layer feed-forward networks. *Chemometrics and Intelligent Laboratory Systems*, **22**, 165-189.

Snee, R.D. 1977. Validation of regression models: methods and examples. *Technometrics*, **19**, 415-425.

Stevenson, A.C., Birks, H.J.B., Flower, R.J., and Battarbee, R.W. 1989. Diatom-based pH reconstruction of lake acidification using canonical correspondence analysis. *Ambio*, **18**, 229-233.

Stewart, D. and Love, W.A. 1968. A general canonical correlation index. *Psychological Bulletin*, **70**, 160-163.

Stone, C.J. 1977. Consistent non-parametric regression. *Annals of Statistics*, **5**, 595-645.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, **36**, 111-147.

Stone, M and Brooks, R.J. 1990. Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society B*, **52**, 237-269.

Sundberg, R. 1985. When is the inverse regression estimator MSE-superior to the standard regression estimator in multivariate controlled situations? *Statistics and Probability Letters*, **3**, 75-79.

Sundberg, R. and Brown, P.J. 1989. Multivariate calibration with more variables than observations. *Technometrics*, **31**, 365-371.

Swain, A.M., Kutzbach, J.E., and Hastenrath, S. 1983. Estimates of Holocene precipitation for Rajasthan, India based on pollen and lake-level data. *Quaternary Research*, **19**, 1-17.

ter Braak, C.J.F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, **41**, 859-873.

ter Braak, C.J.F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167-1179.

ter Braak, C.J.F. 1987a. Calibration. In *Data analysis in community and landscape ecology* (ed. R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren), pp. 78-90. Pudoc, Wageningen.

ter Braak, C.J.F. 1987b. *Unimodal models to relate species to environment*. Doctoral thesis. University of Wageningen, 152 pp.

ter Braak, C.J.F. 1987c. CANOCO - a FORTRAN program for *cano*nical community *o*rdination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1). TWO Institute of Applied Computer Science, 6700 AC Wageningen, The Netherlands, 95 pp.

ter Braak, C.J.F. 1987d. Ordination. In *Data analysis in community and landscape ecology* (ed. R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren), pp. 91-173. Pudoc, Wageningen.

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

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

ter Braak, C.J.F. 1994. Canonical community ordination. Part I: Basic theory and linear methods. *Ecoscience*, 1, 127-140.

ter Braak, C.J.F. 1995. Non-linear methods for multivariate calibration and their use in paleoecology: a comparison of inverse (K-Nearest Neighbours, PLS and WA-PLS) and classical approaches. *Chemometrics and Intelligent Laboratory Systems* (in press).

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

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

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

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

ter Braak, C.J.F. and Looman, C.W.N. 1987. Regression. In *Data analysis in community and landscape ecology* (ed. R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren), pp. 29-90. Pudoc, Wageningen.

ter Braak, C.J.F. and Prentice, I.C. 1988. A theory of gradient analysis. *Advances in Ecological Research*, 18, 271-317.

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

ter Braak, C.J.F. and Wiertz, J. 1994. On the statistical analysis of vegetation change: a wetland affected by water extraction and soil acidification. *Journal of Vegetation Science*, 5, 361-372.

Thomas, E.V. and Haaland, D.M. 1990. Comparison of multivariate calibration methods for quantitative spectral analysis. *Analytical Chemistry*, 62, 1091-1099.

Thunell, R., Anderson, D., Gellar, D., and Miao, Q. 1994. Sea-surface temperature estimates for the Tropical Western Pacific during the Last Glaciation and their implications for the Pacific Warm Pool. *Quaternary Research*, 41, 255-264.

Tso, M.K.-S. 1981. Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society B*, 43, 183-189.

Turkia, J. and Huttunen, P. 1988. Diatoms of surface sediments in relation to lake acidity in Finland. *University of Stockholm Department of Quaternary Research Report*, 12, 99-102.

van den Wollenberg, A.L. 1977. Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika*, 42, 207-219.

van der Meer, J. 1991. Exploring macrobenthos-environment relationship by canonical correlation analysis. *Journal of Experimental Marine Biology and Ecology*, 148. 105-120.

van Dobben, H.F. and ter Braak, C.J.F. 1993. Monitoring of epiphytic lichens in The Netherlands In *Vegetation as a monitor for deposition of nitrogen and acidity* (by H.F. van Dobben), pp. 79-102. Doctoral thesis, University of Utrecht.

Varis, O. 1991. Associations between lake phytoplankton community and growth factors - a canonical correlation analysis. *Hydrobiologia*, 210, 209-216.

Vincens, A., Chalié, F., Bonnefille, R., Guiot, J., and Tiercelin, J.-J. 1993. Pollen-derived rainfall and temperature estimates from Lake Tanganyika and their implication for Late Pleistocene water levels. *Quaternary Research*, 40, 343-350.

Walker, I.R., Mott, R.J., and Smol, J.P. 1991. Allerød-Younger Dryas lake temperatures from midge fossils in Atlantic Canada. *Science*, 253, 1010-1012.

Walker, M.J.C., Coope, G.R., and Lowe, J.J. 1993. The Devensian (Weichselian) lateglacial palaeoenvironmental record from Gransmoor, east Yorkshire, England. *Quaternary Science Reviews*, 12, 659-680.

Wallach, D. and Goffinet, B. 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling*, 44, 299-306.

Wang, L. 1994. Sea surface temperature history of the low latitude western Pacific during the last 5.3 million years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 108, 379-436.

Warner, B.G. and Charman, D.J. 1994. Holocene changes on a peatland in northwestern Ontario interpreted from testate amoebae (Protozoa) analysis. *Boreas*, 23, 270-279.

Wartenberg, D. 1985a. Spatial autocorrelation as a criterion for retaining factors in ordinations of geographic data. *Mathematical Geology*, 17, 665-682.

Wartenberg, D. 1985b. Multivariate spatial correlation: a method for exploratory geographical analysis. *Geographical Analysis*, 17, 263-283.

Wartenberg, D. 1985c. Canonical trend surface analysis: a method for describing geographic patterns. *Systematic Zoology*, 34, 259-279.

Webb, R.S., Anderson, K.H. and Webb, T. 1993. Pollen response-surface estimates of late-Quaternary changes in the moisture balance of northeastern United States. *Quaternary Research*, 40, 213-227

Webb, T., Bartlein, P.J., Harrison, S.P., and Anderson, K.H. 1993. Vegetation, lake levels, and climate in eastern north America for the past 18,000 years. In *Global climates since the Last Glacial Maximum* (eds. H.E. Wright, J.E. Kutzbach, T. Webb, W.F. Ruddiman, F.A. Street-Perrott, and P.J. Bartlein), pp. 415-467. University of Minnesota Press, Minneapolis.

Webb, T., Bartlein, P.J., and Kutzbach, J.E. 1987. Climate change in eastern North America during the past 18,000 years; Comparisons of pollen data with model results. In *North America and adjacent oceans during the last deglaciation* (ed. W.F. Ruddiman and H.E. Wright), pp. 447-462. Geological Society of America, Boulder.

Webb, T. and Bryson, R.A. 1972. Late- and postglacial climatic change in the northern Midwest, USA: Quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. *Quaternary Research*, 2, 70-115.

Webb, T. and Clark, D.R. 1977. Calibrating micropaleontological data in climatic terms: a critical review. *Annals of the New York Academy of Sciences*, 288, 93-118.

Weisberg, S. 1985. *Applied linear regression*. J. Wiley & Sons, New York, 324 pp.

Whitmore, T.J. 1989. Florida diatom assemblages as indicators of trophic state and pH. *Limnology and Oceanography*, 34, 882-895.

Whitmore, T.J. 1991. Sedimentary diatom concentrations and accumulation rates as predictors of lake trophic state. *Hydrobiologia*, 214, 163-169.

Wilson, S.E., Cumming, B.F. and Smol, J.P. 1994. Diatom-salinity relationships in 111 lakes from the Interior Plateau of British Columbia, Canada: the development of diatom-based models for paleosalinity reconstructions. *Journal of Paleolimnology*, 12, 197-221.

Wilson, S.E., Walker, I.R., Mott, R.J., and Smol, J.P. 1993. Climatic and limnological changes associated with the Younger Dryas in Atlantic Canada. *Climate Dynamics*, 8, 177.187.

Wold, S. 1992. Nonlinear partial least squares modelling. II. Spline inner relation. *Chemometrics and Intelligent Laboratory Systems*, 14, 71-84.

Wold, S., Ruhe, A., Wold, H., and Dunn, W.J. 1984. The collinearity problem in linear regression: the partial least squares (PLS) approach to generalized inverses. *Society for Industrial and Applied Mathematics Journal of Scientific Statistical Computing*, 5, 735-743.

Yee, T.W. and Mitchell, N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2, 587-602.

Økland, R.H. 1986. Rescaling of ecological gradients. I. Calculation of ecological distance between vegetation stands by means of their floristic composition. *Nordic Journal of Botany*, 6, 651-660.

Corrections, additions, and up-dates to H.J.B. Birks (1995) 'Quantitative Environmental Reconstructions' In *Statistical Modelling of Quaternary Science Data* (eds. D. Maddy & J.S. Brew), *Quaternary Research Association Technical Guide* **5**, Quaternary Research Association, Cambridge.

As I never saw the proofs of this article, there are inevitably some corrections, additions, and up-dates. These are as follows:

page iii — My affiliation should also include Environmental Research Centre, Department of Geography, University College London, 26 Bedford Way, London WC1H OAP, UK.

page 163 — line 4 from bottom. Jongman *et al.* (1987) should now be Jongman *et al.* (1995).

page 168 — line 10. Korsman and Birks, 1995, should be Korsman and Birks, 1996.

page 171 — line 4 from bottom. Power, 1993 should be Power (1993).

page 174 — line 6 from bottom. Korsman and Birks, 1995, should be Korsman and Birks 1996.

page 177 — line 12. Equation should be $\hat{x}_o = b_o + b_1 y_o$ for fossil sample o.

page 177 — line 10 from bottom. Equation should be

$$x = b_o + b_1 y_1 + b_2 y_2 + b_3 y_3 + \ldots + b_m y_m$$ for the modern training set.

line 4 from bottom. Equation should be

$$\hat{x} = b_o + b_1 y_1 + b_2 y_2 + b_3 y_3 + \ldots + b_m y_m$$ for fossil sample o.

page 184 — line 25. Korsman and Birks, 1995 should be Korsman and Birks 1996.

page 188 — line 19. Korsman and Birks, 1995 should be Korsman and Birks 1996.

page 190 — line 14. Equation should be

$$c = e\,(b_o + b_1 x + b_2 x^2)$$ (c = curve maximum = e(a)).

page 192 — line 12. Equation should be

$$c = 1 / (1 + e\,(b_o + b_1 x + b_2 x^2))$$ (c = curve maximum).

page 194 — line 21. multinomial, not *multinonial*.

page 203 — line 7. Korsman and Birks, 1995, should be Korsman and Birks 1996.

page 206 — line 21. Korsman and Birks, 1995, should be Korsman and Birks 1996.

page 211 — line 3. Peng *et al.* (1993) should be Peng *et al.* (1994).

page 223 — Table 6.3, Caption line 3. *(°C)*, not *(OC)*.

page 235 — line 2. Korsman and Birks, 1995, should be Korsman and Birks 1996.

page 242 — Gaillard *et al.* 1995. Reference is *Landscapes and Life. Studies in Honour of Urve Miller* (ed. A.-M. Robertsson *et al.*) PACT **50**, 431-442.

page 242 — Gasse *et al.* 1995. Reference is **117**, 31-54.

page 245 — Add Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R. (Eds.) (1995) *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, 324 pp.

page 246 — Korsman and Birks 1995. Should be 1996, *Journal of Paleolimnology* **15**, 65-77.

page 246 — Le and Shackleton 1994. Reference is **24**, 187-199.

page 249 — Add Peng, C.H., Guiot, J., van Campo, E., and Cheddadi, R. 1994. The vegetation carbon storage variation in Europe since 6000 BP: reconstruction from pollen. *Journal of Biogeography* **21**, 19-31.

page 251 — Smilauer and Birks 1995. Reference is *Geological Survey of Denmark (DGU) Service Report* **7**, 42-47.

page 252 — ter Braak 1995. Correct reference is 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.

HJBB
6.iv.96