Limnol. Oceanogr., 54(6, part 2), 2009, 2529–2541

Deciphering the effect of climate change and separating the influence of confounding factors in sediment core records using additive models

Gavin L. Simpson[a,*] and N. J. Anderson[b]

[a] Environmental Change Research Centre, Department of Geography, University College London, London, UK
[b] Department of Geography, Loughborough University, Loughborough, UK

Abstract

We describe a new approach to modeling sediment core records, one that uses additive models (AMs) incorporating a serial correlation structure to model residual autocorrelation. Species assemblages, for example, are reduced to ordination axis scores that capture major changes in the data through time. Each set of axis scores is then modeled using an AM, where covariates represent forcing variables (e.g., tree-ring–inferred temperature or proxies for atmospheric deposition) and/or trend and, where necessary, periodic components. The effect of forcing variables on species composition through time can be determined via the contribution each covariate makes to the fitted model, which can be used to separate effects due to competing forcing variables. We illustrate the approach using data from Kassjön, northern Sweden, and Loch Coire Fionnaraich, northwest Scotland.

There is a growing body of evidence from paleoecological studies showing that many remote arctic and boreal lakes have undergone substantial change in species composition over the past 100–200 yr. The cause of this ecological change, often unprecedented in nature and scale, has recently been attributed solely to global warming, especially in polar regions (Smol et al. 2005). The observed changes occur during a period of multiple stressors, and some studies have shown that the onset of biological change is concomitant with the first appearance of spheroidal carbonaceous particles (SCPs) in sediment cores (Battarbee et al. 2002). SCPs are produced via incomplete combustion of fossil fuels and are an independent measure of atmospheric deposition from anthropogenic sources (Rose 2001).

The similarity and coherency of the observed changes suggest the effects of multiple forcing factors (drivers of environmental and ecological change) acting throughout the Northern Hemisphere, which may vary in space and time (Smol et al. 2005; Rühland et al. 2008). However, quantification of the effects of the individual drivers on community composition is difficult using current statistical techniques and data sets, which, in turn, hampers our ability to determine the key driver or drivers of change (Anderson et al. 2006).

Variance partitioning (Borcard et al. 1992) is a statistical method used frequently in paleoecological studies to identify the relative magnitude of the effects of forcing variables on lake ecosystems through time. For example, Lotter and Birks (1997) used variance partitioning to show that prior to restoration efforts, 28% of the variance in varve thickness observed in a sediment core from Baldeggersee, Switzerland, was attributable to variation in temperature and precipitation, while just 6% was related to the eutrophication of the lake. The effects of climate, agriculture, and urbanization on chironomid communities from Canadian prairie lakes were investigated by Quinlan et al. (2002). Climate alone explained substantial proportions (~10% to 72%) of the variance in the chironomid communities of the seven studied lakes, compared with much smaller unique contributions of resource use and urbanization-related factors.

A major drawback of the variance partitioning approach, however, is that only the total effect of the various covariates over the time period of interest is given as an output of the analysis. This precludes asking questions about when and where the various covariates may be driving change in the response variable(s). To address these questions, time-series analysis techniques are required.

In one of the few examples of time-series analysis applied to paleolimnological data, Cottingham et al. (2000) used Bayesian Dynamic Linear Models (DLMs) to investigate changes in variability of algal communities in Lake 227, Experimental Lakes Area, Ontario, Canada, following experimental fertilization. The sediments of Lake 227 are annually laminated, allowing the application of DLMs, which require regular spacing of samples in time (this is one of the difficulties in applying classical time-series techniques to paleoecological data; Birks 1998).

In the majority of lakes, however, annually resolved sediments do not form, either due to bioturbation or a lack of strong seasonal inputs to the sediments. At best, such sediments can be sampled regularly with depth, but due to variations in sediment accumulation rates, each regular slice of sediment will encompass a variable amount of time, and therefore the "distance" in time between any two adjacent samples will also vary. Furthermore, often it is not possible to analyze every single sediment slice or annual varve, either due to financial constraints or impracticalities of counting the number of samples or applying analytical techniques. The end result is an irregularly sampled time series for which classical time-series techniques are inappropriate (Birks 1998).

Here, we present an alternative modeling approach based on the use of flexible additive models that solves some of these problems when fitting models to paleolim-

*Corresponding author: gavin.simpson@ucl.ac.uk

nological data. We demonstrate the way in which these models can be used to answer the critical questions of how much, and when, is climate a significant driver of trends and patterns in the observed data.

Methods

Additive models—Additive models are a nonparametric form of regression in which the sum of regression coefficients × explanatory variables of a linear regression is replaced by a sum of arbitrary smooth functions of the explanatory variables. This allows the shape of the relationship between the response and the explanatory variables to be determined from the data themselves, rather than being assigned a prescribed functional form (e.g., linear or quadratic). As a result, additive models are able to model local features of the relationship between the response and the covariates; the effect of the covariate is allowed to vary across its range.

Formally, an additive model of the type discussed here has the following form

$$y_i = \alpha + \sum_{j=1}^{k} f_j(x_{ji}) + \varepsilon_i, \qquad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda}). \tag{1}$$

where $f_j(x_{ji})$ is a centered smooth function of the jth explanatory variable, $i = 1, \ldots, n$, is the number of observations, and α is a constant representing the intercept. The errors, ε_i, are assumed to have a Gaussian distribution with mean 0 and variance $\sigma^2\mathbf{\Lambda}$. In the case of independent observations, it follows that the correlation matrix will equal the identity matrix, $\mathbf{\Lambda} = \mathbf{I}$. Where observations are not independent, a correlation structure may be assumed for $\mathbf{\Lambda}$, and any additional model parameters required are estimated alongside the other model parameters. The degree of smoothness (or wiggliness) of f_j is controlled through the use of penalized regression and, in the implementation used here, is determined automatically using a generalized cross-validation (GCV) routine (Wood 2006).

For regularly observed sediment time series, autoregressive (AR), moving average (MA), or a combination of the two (ARMA) models can be used to parameterize $\mathbf{\Lambda}$. The order 1 autoregressive model, denoted AR(1), is the simplest and often the most useful autoregressive model (Pinheiro and Bates 2000) for such situations. These correlation structures are not appropriate for the irregular time series most often encountered in paleolimnological studies. The AR(1) structure can be generalized to continuous time to allow for varying temporal differences between analyzed sediment samples, and in this case, it is known as the continuous time AR(1), or CAR(1), structure. It can be shown (Pinheiro and Bates 2000) that CAR(1) is equivalent to the exponential spatial correlation structure for a one-dimensional position vector. Other spatial correlation structures may also be employed to parameterize $\mathbf{\Lambda}$, such as the Gaussian, spherical, or rational quadratic structures, though these all have a superficially similar form, and the end result is often the same regardless the structure used.

Here, we use the CAR(1) correlation structure for its simplicity and because it directly relates to the more common AR(1) structure in its definition. In the CAR(1) structure, the correlation function, $h(\cdot)$, between two observations is

$$h(s, \phi) = \phi^s, \qquad s \geq 0, \qquad \phi \geq 0 \tag{2}$$

where s is the temporal distance between observations. The single parameter ϕ is estimated as part of the model fitting.

Fitting a nonlinear trend to paleolimnological data involves including a smooth function for the date of the sample in the additive model

$$y_i = \alpha + f(date_i) + \varepsilon_i, \qquad \varepsilon \sim N(0, \sigma^2 \mathbf{\Lambda}), \tag{3}$$

and the presence or absence of a trend in the data can be determined via a likelihood ratio test (LRT) between Eq. 3 and a model including the intercept, α, only. Information statistics, such as Akaike's information criterion (AIC) or the Bayesian information criterion (BIC), may also be used to assess whether the presence of a trend in the data is supported; however, the different models should be estimated using the full maximum likelihood to ensure correct testing of models with different fixed effects.

The need to include the correlation structure can likewise be determined using an LRT or via AIC (or BIC) by comparing a model fitted using the structure with a model using $\mathbf{\Lambda} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and indicates independence of observations. The latter model is a simpler model with fewer parameters, and, in the case of the AR(1) or CAR(1) structures, we are testing whether ϕ is significantly different from 0. Restricted maximum likelihood (REML) should be used to fit the models when performing these tests, and furthermore, as 0 is on the boundary of allowed values for ϕ, the p-value may be anticonservative.

We are not restricted to fitting models to identify trends in paleolimnological data using these techniques. Indeed, far more interesting questions may be addressed where independent data on forcing variables are available covering the same time period as the observed samples (Leavitt et al. 2009). For example, given independent temperature measurements for a recent sediment core sequence covering the past 1–200 yr, one could model the effect of temperature variations on species composition recorded in the sediment samples. The contribution to the fitted value of the explanatory variable (temperature in this example) for each sediment sample can be extracted, providing a time series of the effect of that explanatory variable on the response. Because the smooth functions of the explanatory variable ($f_j(x_{ji})$) need not be linear and are local, the magnitudes of the effect of the explanatory variables can vary through time, depending on the relationship between the variables and the response.

Where more than one independent forcing variable is available and included in the model, the fitted response can be partitioned into the contributions of the individual forcing variables for each sediment sample. This partitioning is possible due to the simple, additive nature of Eq. 1. The end result of this partitioning is two or more time series

of the varying magnitudes of the effects of the forcing variables, from which, the individual effects of the forcings may be elucidated.

Additive mixed models—The computational approach used here uses a formulation of Eq. 1 that is embedded in a linear mixed-effects framework (Ruppert et al. 2003; Wood 2006). A mixed-effects model is one that contains both fixed and random effects. Fixed effects are the standard representation of variables in a linear model, such as the effects of a treatment variable or explanatory variable on the response. Grouping structures may lead to additional sources of variability in the data. This additional variability may be described using a random effect.

The additive mixed model (AMM) version of Eq. 1 may take the following form

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^{k} f_j(x_{ji}) + \mathbf{Z}_i\mathbf{b} + \varepsilon_i \quad (4)$$

where y_i is the univariate response, $\boldsymbol{\beta}$ is a vector of fixed parameters, $\mathbf{X}_i$ is the ith row of the model matrix of fixed effects, which includes $\alpha = \beta_0$, the f_j values are smooth functions of the covariates as before, $\mathbf{Z}_i$ is the ith row of the model matrix of random effects, $\mathbf{b}$ is a vector of random effects coefficients, which are assumed to be normally distributed with mean zero and unknown covariance matrix $\boldsymbol{\psi}_\theta$ and parameter $\boldsymbol{\theta}$, and $\varepsilon \sim N(0, \sigma^2\boldsymbol{\Lambda})$ represent the model residuals (Wood 2006).

Equation 4, unlike the simple additive models described previously, does not require the concept of penalized regression for model fitting. Instead, it is possible to formulate the penalized smoothers of the additive model as components of the mixed model formulation (Eq. 4) (Ruppert et al. 2003; Wood 2006). Each smooth function (f_j) is separated into an unpenalized component (the fixed effect), which is included in $\mathbf{X}_i\boldsymbol{\beta}$, and a penalized component (the random effect), which is included in $\mathbf{Z}_i\mathbf{b}$. The random-effects component is assumed to be normally distributed, with mean 0 and one or more unknown variances to be estimated. Full details of the AMM formulation used here are given in Wood (2006, §6.6). Durbán et al. (2005) provide additional, practical advice on implementing models of this type.

In this AMM formulation, the model is actually a linear mixed model, albeit one that allows for smooth functions of the covariates, and therefore the usual methods of statistical inference for linear mixed models apply. As such, likelihood ratio tests and AIC (or BIC) can be used for model comparison in the same manner as that described previously (Pinheiro and Bates 2000; Ferguson et al. 2008).

The additive model approach described here is similar to the nonparametric regression-based approach used by Ferguson et al. (2008) to model changes and detect trends in several limnological parameters in Loch Leven, Scotland. In the approach described in this study, we use CAR(1) correlation structures to cope with irregularly sampled data and smoothers based on regression splines, not locally linear smooths (Ferguson et al. 2008).

Study sites

Kassjön—Kassjön is a small, dimictic lake in northern Sweden. It is ~0.23 km^2 in area, and it has a maximum depth of 12.2 m and a mean depth of 5.5 m (Anderson et al. 1995). From mid-November to late April or early May, the lake is ice covered and thermally stratifies following ice-out. Kassjön is a soft-water, mesotrophic lake that has experienced a long history of human disturbance in the catchment resulting in marked increases in planktonic diatom productivity between A.D. 1300 and the late 19th century (Anderson et al. 1995).

Only brief descriptions of sample preparation and analysis are provided here; more detailed descriptions can be found in Anderson et al. (1995). The sediment core was cut into 5-yr (varve) blocks, and an amalgamated sample was prepared for diatom analyses using standard methods. Between 350 and 400 diatom valves were counted per sample. The core sequence analyzed here covers the period 445 B.C. to A.D. 445, and dating was provided via varve counting (Anderson et al. 1995), represented by 177 contiguous samples. In total, 228 diatom taxa were found in the sediment core sequence studied here.

Tree-ring–inferred temperatures for the period of interest were obtained using the regional curve standardization (RCS) methodology applied to a 7400-yr tree-ring chronology for northern Swedish Lapland (Grudd et al. 2002). The RCS methodology emphasizes year-to-year and decadal to century timescale variability (Grudd et al. 2002).

Loch Coire Fionnaraich—Loch Coire Fionnaraich (LCFR) is a small, upland, polymictic loch located in northwest Scotland at an altitude of 236 m above sea level. The loch has a maximum depth of 14.5 m and covers an area of 0.093 km^2. The catchment of the loch is characterized by blanket peats and acid heathland. The loch is oligotrophic, with low alkalinity (16.4 μeq L^{-1}) and present-day pH of 5.9 (Pla et al. 2009).

A 25-cm-length core was retrieved from Loch Coire Fionnaraich in May 2001 using a gravity corer in ~14-m water depth (Pla et al. 2009). The core was extruded at 2-mm intervals and analyzed for diatom and spheroidal carbonaceous particles (SCPs) using standard methods (Rose 1990; Battarbee et al. 2001). Temperature data from the Central England Air Temperature series (CET) were used as an independent record of air-temperature variations in the Loch Coire Fionnaraich catchment, while SCP accumulation rate (No. particles cm^{-2} yr^{-1}) was used as a surrogate for atmospheric deposition loading from anthropogenic sources to the loch. The period covered by the CET series is represented by 74 samples from the LCFR sediment core. In total, 215 diatom taxa were present across the 74 samples from LCFR, and all taxa were used in the statistical analyses presented here. SCP data were interpolated linearly from the observed levels to provide a measure of SCP accumulation rate for each observation that CET and diatom data were available. The sediment core data and the applicability of the CET to LCFR are discussed in more detail in Pla et al. (2009). The data used here differ slightly from those of Pla et al. (2009) in the use

of Hellinger-transformed diatom data and in not pre-smoothing the raw CET time series.

Statistical analysis—The diatom data from both Kassjön and LCFR were Hellinger transformed prior to analysis with principal components analysis (PCA) (Legendre and Gallagher 2001). This preserves the Hellinger distances between samples in the resulting PCA, which is a dissimilarity coefficient that has been shown to have good properties for ecological data (Legendre and Gallagher 2001). In both data sets, two PCA axes were identified as explaining significant proportions of the variance in the species data when compared to the broken stick (null) distribution (Jackson 1993) and via inspection of scree plots of the axis eigenvalues. These axes were retained for subsequent analysis. PCA was performed using the vegan package (Oksanen et al. 2008) for the R statistical software (R Development Core Team 2009). AMMs of the form described previously were individually fitted to the PCA axis 1 and axis 2 scores for both sites using the mgcv package (Wood 2004, 2006) for R. All other statistical analyses were performed using standard R functions.

Preliminary exploratory data analysis (EDA) of the Kassjön diatom and RCS tree-ring temperature data suggested that the visible trends in the diatom PCA axis 1 and 2 scores were unlikely to be explained wholly by the RCS data. Furthermore, the results from application of the autocorrelation function (ACF) to the PCA axis scores revealed significant autocorrelations at lag 5, and higher multiples of 5, suggesting periodicity on the order of ~25 yr. In order to model variance in these series sufficiently well, so that statistical inference on the effect of temperature on diatom species composition was as robust as possible, we included sediment sample date and a dummy variable for periodicity as explanatory variables in the models. These two variables allowed the model to fit the observed trend and periodicity in the PCA axis scores so that the effect of RCS temperature could be determined.

For LCFR, preliminary EDA of the available data and the results of Pla et al. (2009) suggested that the diatom PCA axis 1 scores were strongly related to SCP accumulation rate. Therefore, initial model fitting proceeded using the CET and SCP data only, without an additional covariate for the trend through time. The diatom PCA axis 2 scores at LCFR showed no obvious sign of trend or periodicity, and thus model building proceeded as for the axis 1 scores, using only smooth functions of CET and SCP accumulation rate. It is likely that measurement error is present in the covariates in both the examples presented here. Like most other statistical modeling techniques, the models described do not take into account this source of error.

Likelihood ratio tests and AIC statistics were used to select the most parsimonious model for each of the diatom PCA axis score series. The contributions of the individual model covariates to the fitted values of the response were calculated by decomposing the fitted values into their component terms using the bases of the fitted smooth functions and the corresponding coefficients for each spline term, excluding the intercept term. In practice, these contributions are derived using function predict.gam argument type = "terms" in the function predict.gam of the mgcv package.

Results

Kassjön—Time-series plots of the diatom PCA axis 1 and 2 scores for Kassjön are shown in Fig. 1. These two axes account for ~9% of the variance in the transformed diatom data, a high proportion given the high number of taxa (n = 228) and the high degree of variability exhibited in the community data. Comparison with the broken stick distribution suggests that these two PCA axes represent the signal in the diatom data, and subsequent axes explain no more variance than expected under the null distribution.

The first half of the series of PCA axis 1 scores for Kassjön exhibits a moderate increasing trend to ca. 250 B.C., followed by decreasing scores, while scores in the latter half of the sequence remain constant, with a slight increasing trend observed in the last 100 yr of the record (Fig. 1a). The main feature of the data, however, is the marked periodicity that is present throughout the period, but that is most pronounced in the first half of the core, as shown by the regular spikes in the axis scores that occur every five samples. There is a large amount of variation in the size of these peaks, and the largest excursions are found in the first half of the core sequence.

The PCA axis 2 scores for Kassjön (Fig. 1) show a stronger trend but have lower variation than the axis 1 scores. A decline in axis scores is observed up to A.D. 300, and a slight rise in scores is apparent in the last 75 yr of the record. Although not as apparent as in the axis 1 scores, the regular five-sample periodicity is also present in the axis 2 scores (determined from an ACF plot of the record, not shown), although the variation about the trend is considerably smaller than for the axis 1 scores. It is also worth noting that variation in the diatom communities on this second PCA axis is smaller than on the first, as evidenced by the smaller range in axis scores (compare the range on the y-axes of the two panels in Fig. 1).

AMMs including components for the trend, periodicity, and RCS temperature were fitted to the Kassjön PCA axis 1 and 2 scores. Trend and periodicity components were required to adequately model the major patterns in the observed series as RCS temperature appeared to be most correlated with small-scale variations over and above the trends and periodicities observed in Fig. 1. Modeling the series in this manner allows the model to fit features of the data and explain residual variation that may affect statistical inference when assessing the effect of RCS temperature if left unaccounted for. The data for Kassjön were sampled on a regular 5-yr span, and, as such, we modeled autocorrelation in the series using the simpler AR(1) structure. Models were fitted with and without this structure and were compared using likelihood ratio tests and AIC.

The final fitted model for the axis 1 scores from Kassjön did not include a smooth function for RCS temperature because it was deemed insignificant on the basis of assessment by likelihood ratio tests and AIC. While

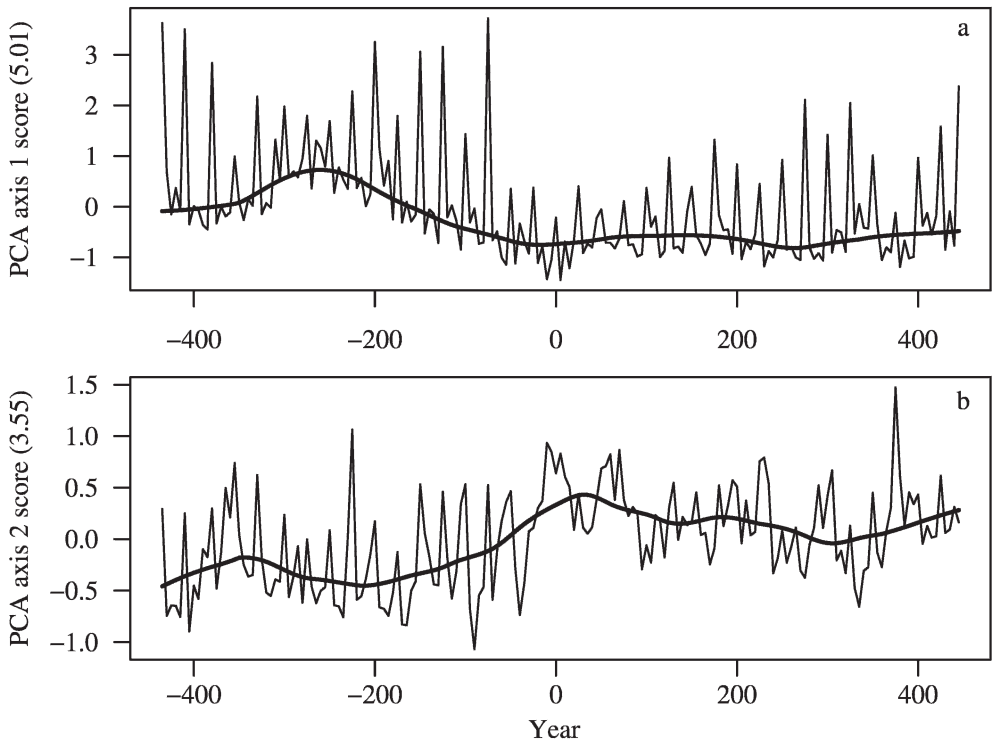

Fig. 1. Time series of (a) axis 1 and (b) axis 2 scores for the PCA of the Hellinger-transformed diatom data from Kassjön. The number in brackets on the *y*-axis label is the % variance explained by each axis. The thick lines are LOESS smoothers fitted through the observations to highlight trends in an exploratory manner.

interesting in their own right, because we are concerned with illustrating the additive modeling approach in deciphering the effects of forcing variables, including temperature, on aquatic ecosystems, we do not discuss this model further.

All three covariates contributed significantly to the model fitted to the PCA axis 2 scores for Kassjön (Table 1), although the fitted smooth function for RCS temperature is only marginally significant at the 95% level. As we are interested in the effect of temperature on the diatom communities in the Kassjön core, we retain this term in the model.

An AR(1) structure for the model errors was not supported by a likelihood ratio test ($p = 1$), and the two models provided effectively the same fit to the observed data. Therefore, the simpler model (without the AR structure) was retained. The fitted smooth functions for trend, periodicity, and RCS temperature are shown in Fig. 2. The three panels are drawn on the same scale, and it is immediately apparent that the smooth functions for the trend and periodic components exhibit more variation (greater range), having larger absolute contributions to the fitted response, than the smooth function for RCS temperature, which indicates a greater effect for these two components on diatom PCA axis 2 scores.

The fitted smooth function for the trend follows closely the exploratory locally weighted scatter plot smoothing (LOESS) smoother fitted to the observations in Fig. 1 (as one might expect). The smooth function for period indicates that the third sample in the five-sample periodicity leads to higher, positive PCA axis 2 scores, while either side of this peak, axis scores fall back to around zero or are slightly negative in general.

The smooth function for RCS temperature shows a slightly nonlinear relationship between RCS temperature and PCA axis 2 scores. There is much uncertainty in the shape of the curve as evidenced by the wide confidence bands (Fig. 2c), although at lower temperatures, the confidence bands do not include 0, indicating a significant effect of temperature on axis 2 scores. The effect includes 0 for the upper half of the observed temperature gradient during the period covered by the Kassjön core, and as such, the model indicates an effect of temperature on the diatoms only at the lowest observed temperatures.

This is clearly reflected in the time series of the contribution (or effect) of RCS temperature on the diatom PCA axis 2 scores (Fig. 3), which shows significant contributions during the two extreme cold events that occurred around 300 B.C. and ca. 100 B.C. (Fig. 3c). The effect of RCS temperature on the diatoms is generally insignificant during the more stable, warmer period in the latter half of the record. The fitted model (Fig. 4) deals well with the change in level in the PCA axis 2 scores after ca. 100 B.C., but substantial variability around this level remains unexplained.

Loch Coire Fionnaraich—Time-series plots of the diatom PCA axis 1 and 2 scores for LCFR are shown in Fig. 5.

Table 1. Model summary for the additive mixed model fitted to the Kassjön diatom PCA axis 2 scores. EDF = effective degrees of freedom for the regression spline. Ref. df = reference degrees of freedom used to compute the *p*-value.

Covariate	EDF	Ref. df	*F*	*p*-value
Trend	7.329	7.829	8.453	1.84×10^{-9}
Periodicity	2.714	3.21	6.43	0.000269
RCS temperature	2.042	2.542	2.900	0.041158

These two axes account for 17.2% of the variance in the transformed diatom data. Again, this is quite a high proportion given the high number of taxa (n = 215). Comparison with the broken stick distribution suggests that these two PCA axes represent the signal in the LCFR diatom data, and subsequent axes explain no more variance than expected under the null distribution.

A strong, declining trend in the LCFR diatom PCA axis 1 scores is present from approximately 1800 to 1850 (Fig. 5). A second, declining trend is apparent around 1950. Prior to 1800 and after the two periods of declining PCA axis 1 scores, relatively stable diatom communities were observed. The overall pattern is towards lower PCA axis 1 scores after 1800.

PCA axis 2 scores do not show such a striking trend compared to the axis 1 scores. The observed pattern suggests some quasi-periodic component to the diatom community, with negative PCA axis 2 scores prior to 1700 and around 1900, and a sharp decline to lower scores again

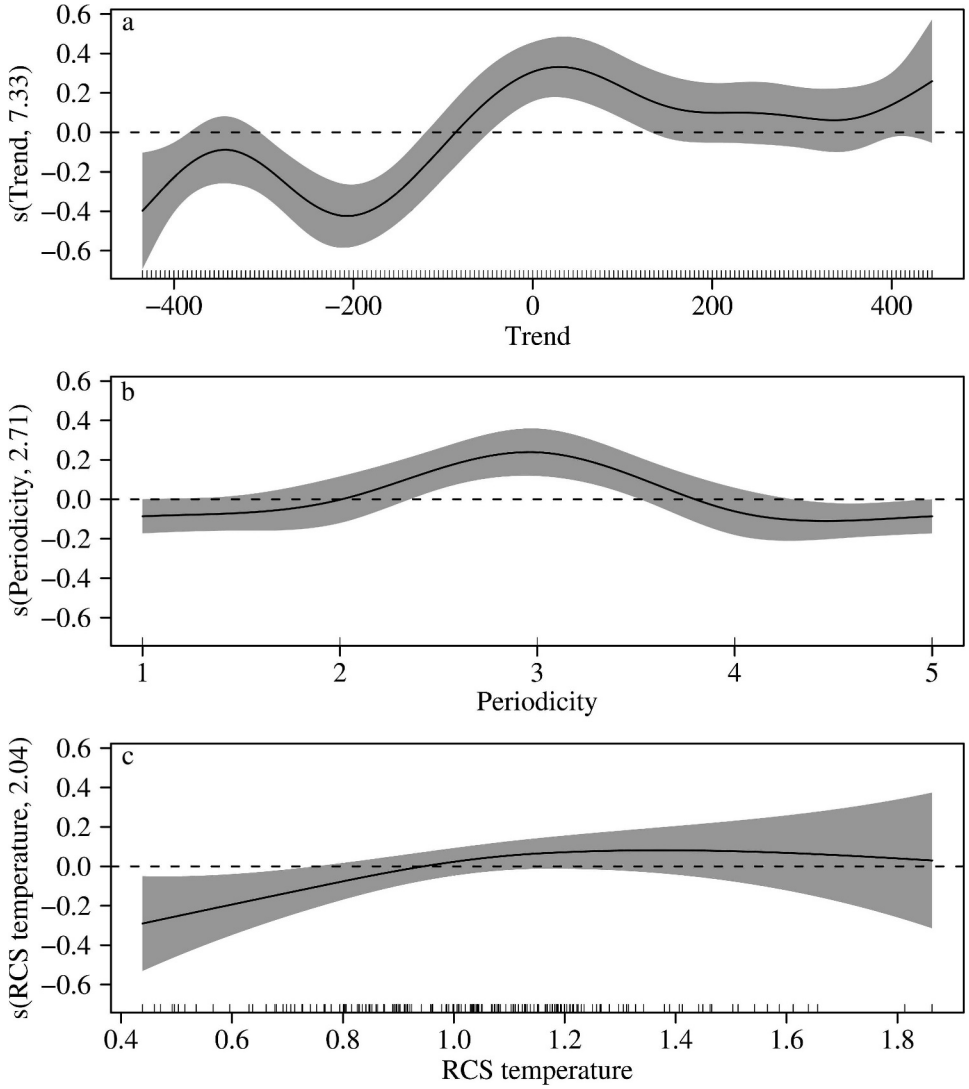

Fig. 2. The fitted smooth functions for (a) trend, (b) periodicity, and (c) RCS temperature from the final AMM for the Kassjön PCA axis 2 scores. The gray bands are approximate 95% pointwise confidence intervals on the fitted functions. The tick marks inside the panels on the *x*-axis show the distribution of observed values for the two covariates. The numbers in brackets on the *y*-axis (7.33, 2.71, and 2.042 for trend, periodicity, and RCS temperature, respectively) are the effective degrees of freedom for each smooth function.

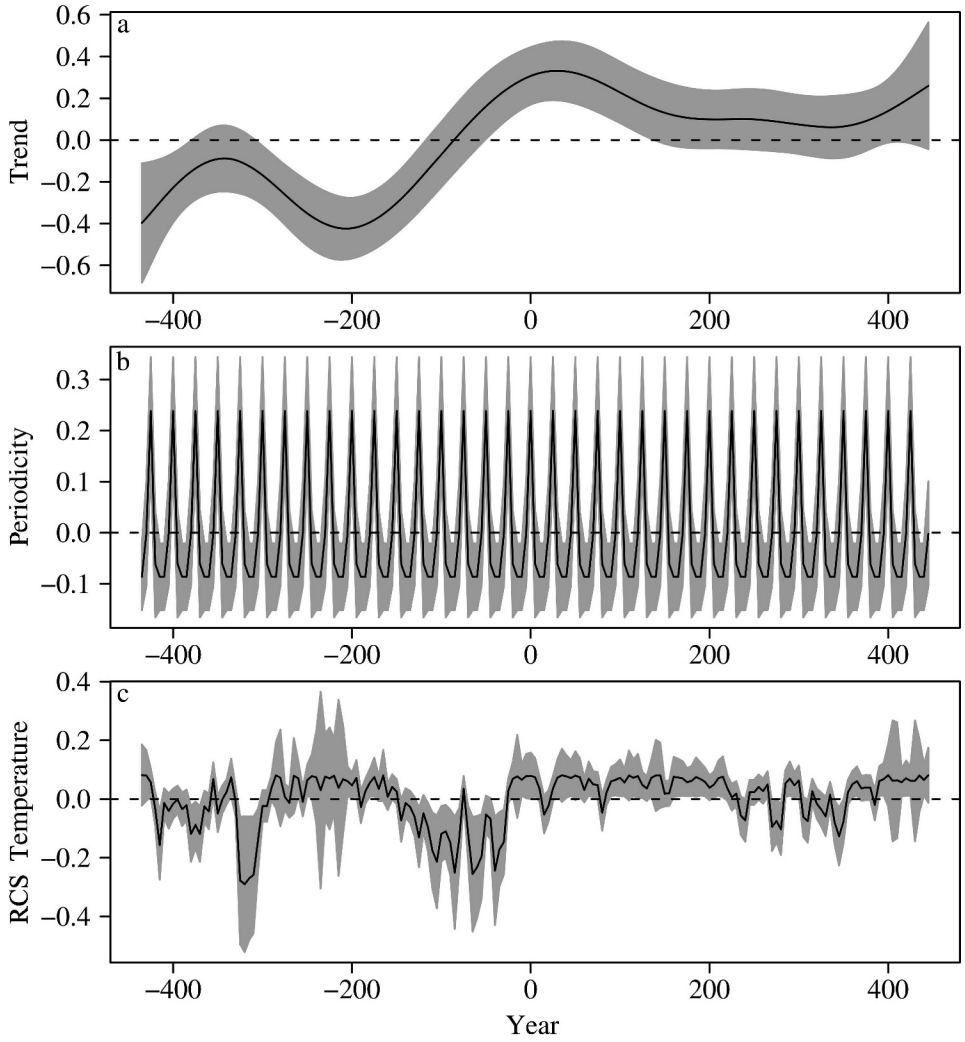

Fig. 3. The contribution of (a) trend, (b) periodicity, and (c) RCS temperature to the fitted diatom PCA axis 2 scores for the final Kassjön model. The gray band is an approximate 95% pointwise confidence interval on the contribution. Where the band includes the dashed zero line, the contribution of the covariate is not statistically significantly different from the intercept.

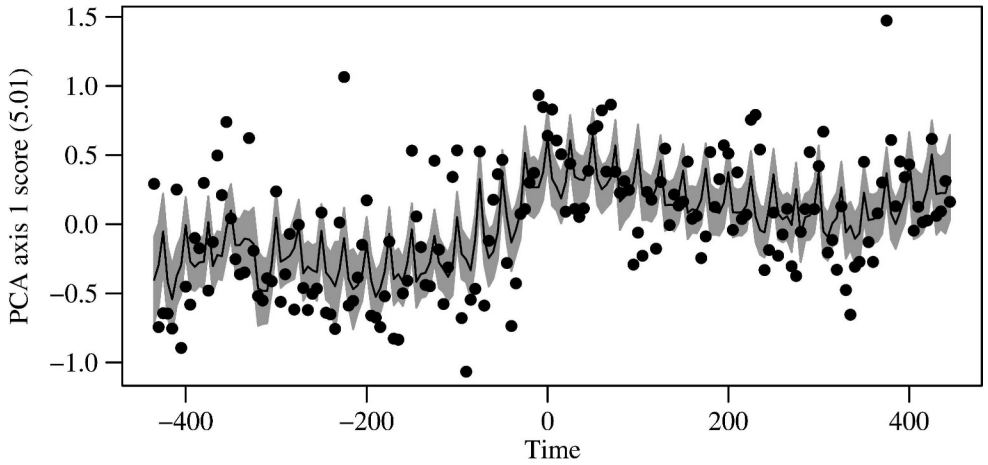

Fig. 4. Observed and AMM fitted values for PCA axis 2 scores for the Kassjön core. The gray band is an approximate 95% pointwise confidence interval on the fitted values.

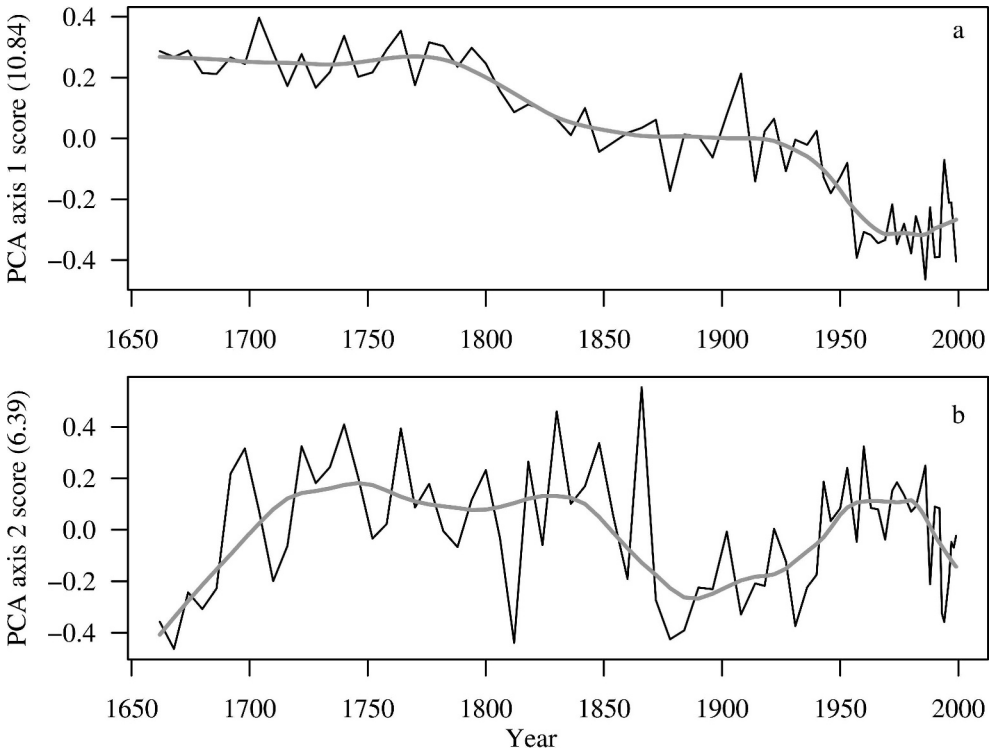

Fig. 5. Time series of (a) axis 1 and (b) axis 2 scores for the PCA of the Hellinger-transformed diatom data from Loch Coire Fionnaraich (LCFR). The number in brackets on the *y*-axis label is the % variance explained by each axis. The thick lines are LOESS smoothers fitted through the observations to highlight trends in an exploratory manner.

in the uppermost samples of the core. The scale of variability between samples is much greater on axis 2 than axis 1.

AMMs including smooth functions of the SCP accumulation rate and CET data as the sole covariates were fitted to both sets of axis scores, with and without a CAR(1) correlation structure for $\mathbf{\Lambda}$. The CAR(1) error structure was used for the LCFR data because the sediment core samples were irregularly distributed in time.

The AMM fitted to the PCA axis 1 scores included a significant smooth term for SCP accumulation rate ($p \leq 0.0001$) and a borderline significant smooth function for CET ($p = 0.057$). The CAR(1) error structure was not required, as assessed by a likelihood ratio test comparing AMMs with and without the structure ($p = 0.79$). AIC and BIC also favored the AMM without (AIC = −99.68, BIC = −85.85) rather than with (AIC = −97.75, BIC = −81.62) the CAR(1) structure. Dropping the CAR(1) error structure reduced the *p*-value for the CET smooth function by a small amount, though the interpretation remains the same; i.e., the effect of temperature on the diatom PCA axis 1 scores in LCFR is marginally significant. We retain this term in the model because our aim is to compare the effects of temperature and atmospheric deposition loading on the diatom communities of LCFR. Table 2 shows the model summary for the final AMM for the PCA axis 1 scores.

Table 2. Model summary for the additive mixed model fitted to the LCFR diatom PCA axis 1 scores. EDF = effective degrees of freedom for the regression spline. Ref. df = reference degrees of freedom used to compute the *p*-value.

Covariate	EDF	Ref. df	*F*	*p*-value
CET	1.00	1.50	3.47	0.0495
SCP	2.95	3.45	56.70	$\leq 2 \times 10^{-16}$

The fitted smooth functions for CET and SCP accumulation rate are shown in Fig. 6. The two panels are drawn to the same scale and show that variation in diatom PCA axis 1 scores with SCP accumulation rates is much greater than variation for CET. The smooth function for CET is linear, using one effective degree of freedom (EDF). The fitted relationship between diatom PCA axis 1 scores and SCP accumulation rate is nonlinear, using 2.97 EDF, and this suggests that the effect on the scores is smaller at high SCP accumulation rates than at low rates. For SCP accumulation rates between 30 and 40 n cm^{-2} yr^{-1}, PCA axis 1 scores do not vary as the SCP accumulation rate increases for example.

Figure 7 shows the fitted values from the final AMM overlain on a plot of the observed PCA axis 1 scores. This model explains ~74% of the variance in the diatom PCA axis 1 scores. The model captures well the changes in diatom community composition from 1900 onward, but it performs less well for the earlier decline in PCA axis 1 scores around 1800. The model also systematically underestimates the axis scores prior to 1800. Despite the lack of fit in these two periods, the model does exceptionally well

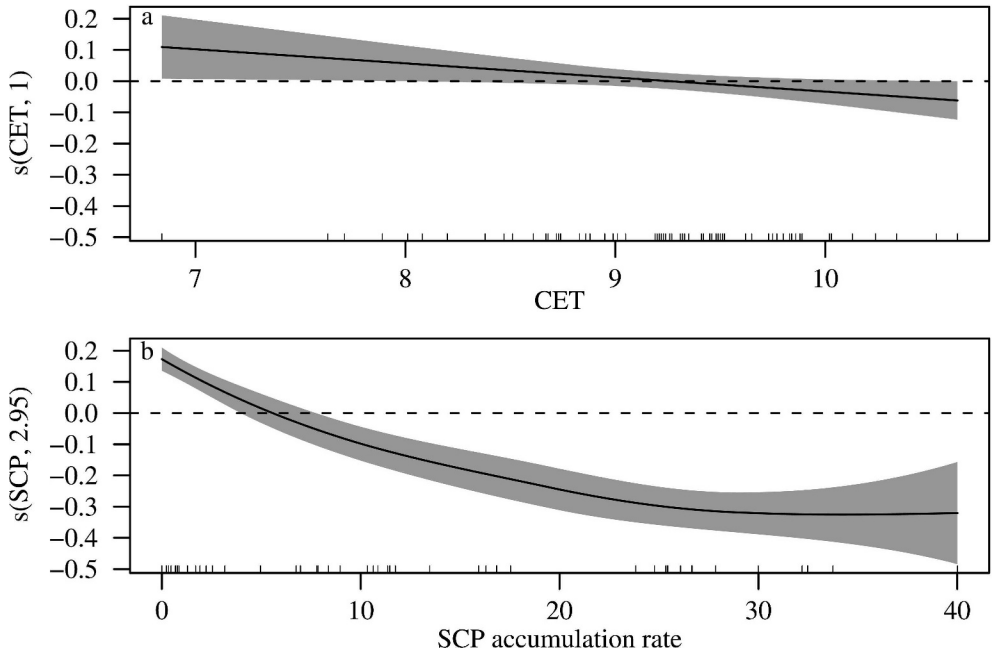

Fig. 6. The fitted smooth functions for (a) CET and (b) SCP accumulation rate from the final AMM for the LCFR PCA axis 2 scores. The gray bands are approximate 95% pointwise confidence intervals on the fitted functions. The tick marks inside the panels on the *x*-axis show the distribution of observed values for the two covariates. The numbers in brackets on the *y*-axis (1 and 2.95 for CET and SCP, respectively) are the effective degrees of freedom for each smooth. A value of 1 indicates a linear function.

in explaining the observed changes in community composition.

The contributions to the fitted values of the two covariates (CET and SCP) are shown in Fig. 8. The contribution to fitted axis scores is small for CET, as suggested by the borderline significance of the fitted smooth function. Between the start of the record and ~1800, the main contribution of CET appears to be to increase diatom axis 1 scores, and from the shape of the fitted model (Fig. 7), it is clear that the smooth function for CET explains the stochastic variation about the overall fitted trend in diatom PCA axis 1 scores. This overall trend in the axis scores is explained predominately by variation in the SCP accumulation rate (Fig. 8b). There is some evidence for a combined effect of CET and SCP accumulation rates on axis scores in the uppermost samples of the LCFR core, as indicated by the trend toward lower and negative diatom scores suggested by the CET effect, while SCP accumulation rate variation in this part of the core acts to increase scores as atmospheric deposition loads to LCFR decline.

The models fitted to the PCA axis 2 scores were not significant, and the approximate 95% confidence intervals

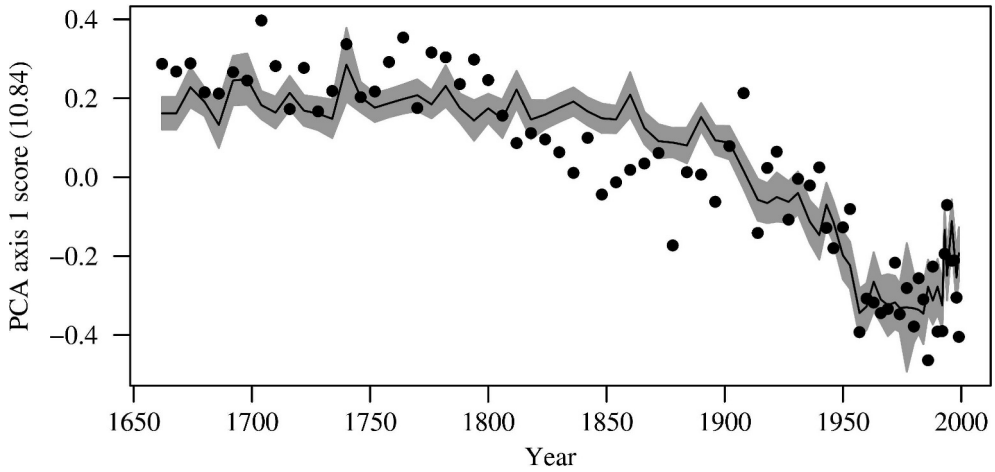

Fig. 7. Observed and AMM fitted values for PCA axis 2 scores for the LCFR core. The gray band is an approximate 95% pointwise confidence interval on the fitted values.

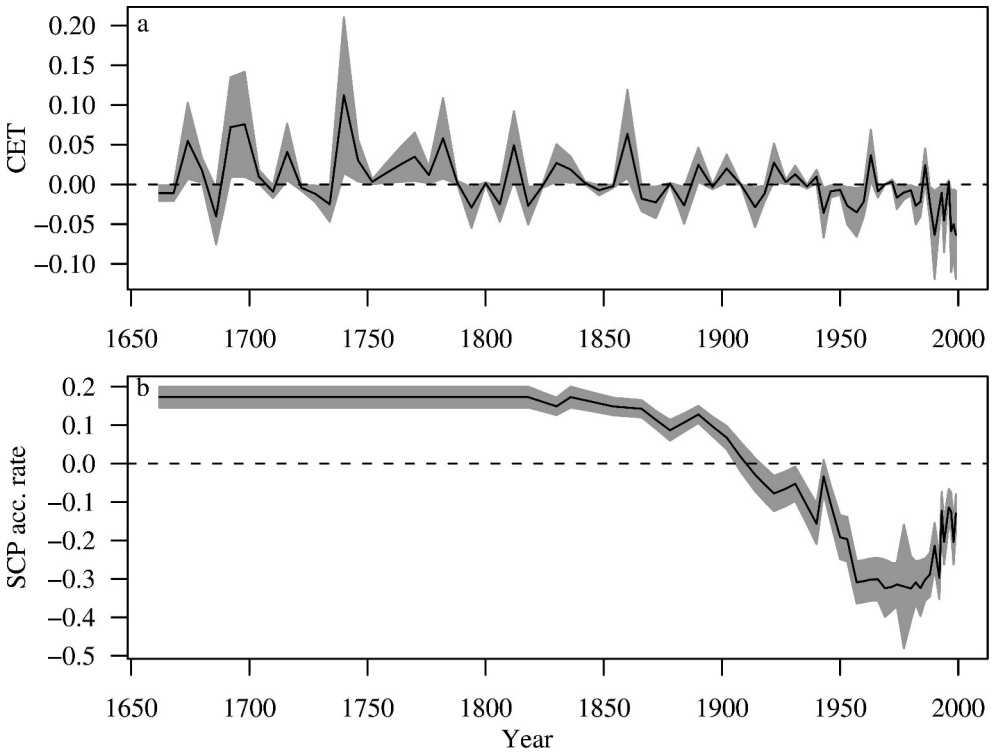

Fig. 8. The contribution of (a) CET and (b) SCP accumulation rate to the fitted diatom PCA axis 2 scores for the final LCFR model. The gray band is an approximate 95% pointwise confidence interval on the contribution. Where the band includes the dashed zero line, the contribution of the covariate is not statistically significantly different from the intercept.

for the smooth functions of SCP accumulation rate and CET included 0 throughout the range of each covariate, indicating no relationship between the covariates and the response. Therefore, we do not consider the PCA axis 2 scores for LCFR further.

Discussion

Kassjön—Additive modeling of the precultural diatom assemblage of Kassjön indicated small, but significant, direct climate effects on the littoral zone during short-term cold periods arising from changes in benthic diatom composition along PCA axis 2. While the direct effects of temperature on the diatom community of the lake are small, additive modeling provides useful additional information on where in time diatom species composition was changing, information that variance partitioning would not have provided. The results suggest that the diatom community of Kassjön was reasonably resilient to natural fluctuations in temperature. Modest warming events, if anything, appeared to support stable diatom assemblages, while species composition changed only during extreme cold periods.

Importantly, direct effects of climate (temperature) on diatom species composition were shown to be small and not associated with the main structural changes in community composition as expressed on the PCA axis 1 and 2 scores. This does not preclude indirect, possibly lagged, effects of climate on the diatoms in this period in Kassjön, however (Anderson et al. 1995). Indeed, the major diatom changes on PCA axis 1 appear to reflect fluctuating inputs of nutrients and levels of dissolved organic carbon driving changes in planktonic diatom species such as *Asterionella formosa* and *Aulacoseira tenella* (N. J. Anderson unpubl.).

Loch Coire Fionnaraich—The trend in the PCA axis 1 scores is associated with declines in the typically sediment-welling species *Fragilaria virescens* var. *exigua* and the pedunculate species *Cymbella* spp. and *Gomphonema* spp., and increases in several species in the genus *Eunotia*. These changes are indicative of a shift in the source of diatom productivity, from deep-water sediments to shallower rock-dominated habitats (Pla et al. 2009).

The major trend and patterns in the LCFR PCA axis 1 scores are strongly related to the increase and subsequent decline in atmospheric deposition (represented by SCP accumulation rate). Superimposed upon this pollutant trend, there is a small, but significant effect of climate throughout the observed sequence. These results concur with those of Pla et al. (2009), and yet the identification of subtle climate effects on diatom community composition and the ability to compare the magnitude of effects of the two driving factors are two important findings that will add substantial weight to the authors' conclusions, namely, that atmospheric deposition, possibly through nutrient enrichment, is a major driver of ecological change in LCFR.

The role of atmospheric deposition in driving change in remote ecosystems has been documented elsewhere (Wolfe

et al. 2001), and the contribution of atmospheric deposition to the changes observed elsewhere in the Northern Hemisphere (Smol et al. 2005) could be investigated more thoroughly using additive modeling techniques.

Additive models—The results presented here illustrate the power of the additive model framework for fitting statistical models to paleolimnological data, which allows the effects of covariates through time to be evaluated, separated, and compared. Appropriate correlation structures, such as AR(1) and CAR(1), allow models to be fitted to both regular and irregular sediment core time series, and they can account for temporal autocorrelation in model residuals when evaluating the effects of covariates.

In the two examples shown, the correlation structures were not required in the final models, presumably because autocorrelation within the times series was modeled by covariates already in the model. This will not be the case for all studies. When autocorrelation remains unmodeled by the covariates, a suitable correlation structure is required to correctly determine the statistical significance of the model covariates.

Care must be taken when interpreting the contribution of covariates plots such as those shown in Figs. 3 and 8, however. The *y*-axis on such plots describes the contribution of each covariate to the fitted response excluding the intercept; hence, they are centered on 0. Where the contribution for an individual covariate changes in time along the *x*-axis, the change in that covariate is associated with change in the response (here the diatom axis scores). In the LCFR example, we see that the contribution of SCP accumulation rate to the fitted values of the response is positive (Fig. 8b), but constant, between 1660 and 1850, with the 95% pointwise confidence interval bounded away from zero. In this period, low or zero SCP values are observed. As deposition to the loch increases post-1850, the SCP values begin to change rapidly, and this is associated with a change in the diatom PCA axis 1 scores. It is during this period that we observe a significant effect of atmospheric deposition on the diatom assemblage, where the contribution of the covariate changes rapidly. Contrast this with the contribution of CET to the fitted values of the model (Fig. 8a). For most of the period covered by the sediment core, the contribution of CET to the fitted values is small and not much different from the mean response (the intercept component), and there is little systematic change in the contribution of CET to the fitted values. However, the contribution of CET periodically takes large, positive values, suggesting a stochastic effect of temperature that is related to fluctuations in the axis 1 scores about the intercept. Overall, there is a trend in the contribution of CET toward negative contributions; however, this is small in magnitude compared the magnitude of change in the contribution of SCP, reflecting the lesser importance in describing the observed patterns in the diatoms in Loch Coire Fionnaraich.

Furthermore, it should be noted that the confidence interval in these plots is pointwise, not simultaneous, and is computed for each observation independent of the others. As with correcting for multiple comparisons in statistical testing, it is important not to overinterpret individual observations where the confidence interval only slightly excludes zero; for a sequence of 100 observations, we could expect five to have a 95% pointwise confidence interval bounded away from zero purely by chance. Large deviations of the contribution away from zero, relative to the magnitude of the confidence interval, for individual points are deserving of interpretation.

The AM framework has two main advantages over traditional time-series techniques (such as ARMA models or DLMs) and the variance partitioning approach that has proved popular in paleolimnology. The first advantage of AMs is that they provide a time series of the contribution of covariates to the fitted response instead of a single statistic for the amount of variance explained by each covariate, as is generally reported from variance partitioning analyses. Furthermore, this time series of contribution or effect has an associated standard error that allows confidence bands on the contribution to be created, leading to more formal statistical assessment of the model covariates than is possible with variance partitioning.

A time series of effect or contribution of a covariate is, however, still possible with variance partitioning. Using a single covariate in the constrained ordination model results in axis 1 scores that reflect the effect of the covariate on the species assemblages, and these can be plotted as a time series to produce a similar diagram to those shown in Figs. 3 and 8. Formal statistical analysis of this variance partitioning–based contribution is not possible, however, in the manner shown in Figs. 3 and 8.

A second advantage of the AM approach is that it is firmly rooted in the familiar regression framework. Additive models are now commonly used in ecological research to model species response curves to environmental variables or in spatial assessments of the effects of climate change on species presence or absence to give but two examples (Yee and Mitchell 1991). The utility of mixed models for ecological research is slowly entering the main stream (Zuur et al. 2007, 2009), and only a reasonable familiarity with the concepts and nomenclature is required in order to fit these models using modern, freely available computer software. However, as with any statistical approach, it is important for the implications of applying AMMs to data to be fully appreciated. In contrast, DLMs, while being mathematically very similar to the additive model, are unfamiliar to most ecologists, and the nomenclature (Kalman filtering, etc.) is like a foreign language to many.

One disadvantage of the AM approach is that it is not as simple to provide predictions outside the range of the observed covariates (forecasting) as it is with DLMs. This is because the splines used in the smooth functions that are central to additive models are chosen to fit the observed data and may behave erratically outside these bounds. This is a particular problem when attempting to forecast future observations from a model fitted to sediment core observations such as the study of Cottingham et al. (2000), where DLMs proved to be particularly useful in this regard.

A further disadvantage is that currently available species data need to be reduced a priori to ordination axes using unconstrained ordination techniques before additive mod-

els can be fitted to the response. Models are also fitted independently to each ordination axis in turn. In this regard, the approach described here is similar in spirit to the so-called indirect ordination approach popular before the development of constrained ordination techniques. Recent developments with vector generalized additive models (VGAMs) (Yee and Mackenzie 2002) and constrained additive ordination (CAO) (Yee 2004, 2006) may allow multivariate versions of the models fitted here to be applied in the future in a manner similar to variance partitioning with constrained ordination yet retain the additive model concept.

One could also choose to fit the models described in the results section here to constrained ordination axes instead of the PCA axes used here. However, without evidence to the contrary, the ability to extract the major patterns of variance in species assemblage data independently of the available explanatory variables for subsequent analysis is preferable to a situation where only the variation in the species data that can be ascribed to covariates is extracted and then modeled using the techniques described here. Extracting only the variance that can be explained by available covariates allows the effect of the covariates to be modeled through time on the part of the species data that they can explain, and this variance may not be included on the first couple of unconstrained ordination axes in the approach taken in this study. However, it is important to demonstrate that the effects of any covariates are leading to important shifts in species composition, and these should be related to the main patterns of variation in the species data. In practice, both the indirect approach used here and a direct approach using constrained ordination axes (or with CAO in the future) may be desirable, as is the case with ordination in general.

Additive modeling of paleoecological data represents a robust approach to investigate the causes of ecological change in lake ecosystems. The approach has wide applicability and represents a powerful technique that may be applied to decipher the relative magnitudes of forcing factors that are hypothesized to be the cause of the recent changes in species composition observed in many remote, Northern Hemisphere lakes.

Acknowledgments

We thank Veronica Gahlman and Ingemar Renberg for permission to use the diatom data from Kassjön; and Roger Flower, Neil Rose, Sergi Pla, and Don Monteith for the Loch Coire Fionnaraich diatom and spheroidal carbonaceous particle data. G.L.S. was supported while undertaking this study through the European Union Sixth Framework Integrated Project Eurolimpacs (GOCE-CT-2003-505540).

References

Anderson, N. J., H. Bugmann, J. A. Dearing, and M.-J. Gaillard. 2006. Linking palaeoenvironmental data and models to understand the past and to predict the future. Trends Ecol. Evol. **21:** 696–704.

———, I. Renberg, and U. Segerström. 1995. Diatom production response to the development of early agriculture in a boreal forest lake-catchment (Kassjön, northern Sweden). J. Ecol. **86:** 809–822.

Battarbee, R. W., J.-A. Grytnes, R. Thompson, P. G. Appleby, J. Catalan, A. Korhola, H. J. B. Birks, E. Heegaard, and A. Lami. 2002. Comparing palaeolimnological and instrumental evidence of climate change for remote mountain lakes over the last 200 years. J. Paleolimn. **28:** 161–179.

———, V. J. Jones, R. J. Flower, N. G. Cameron, H. Bennion, L. Carvalho, and S. Juggins. 2001. Diatoms, p. 155–202. *In* J. P. Smol, H. J. B. Birks and W. M. Last [eds.], Tracking environmental change using lake sediments. V. 3, Terrestrial, algal and siliceous indicators. Kluwer.

Birks, H. J. B. 1998. Numerical tools in palaeolimnology—progress, potentialities, and problems. J. Paleolimn. **20:** 307–332.

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

Cottingham, K. L., J. A. Rusak, and P. R. Leavitt. 2000. Increased ecosystem variability and reduced predictability following fertilisation: Evidence from palaeolimnology. Ecol. Lett. **3:** 340–348.

Durbán, M., J. Harezlak, M. P. Wand, and R. J. Carroll. 2005. Simple fitting of subject-specific curves for longitudinal data. Stat. Med. **24:** 1153–1167.

Ferguson, C., L. Carvalho, E. Scott, A. Bowman, and A. Kirika. 2008. Assessing ecological responses to environmental change using statistical models. J. App. Ecol. **45:** 193–203.

Grudd, H., K. R. Briffa, W. Karlén, T. S. Bartholin, P. D. Jones, and B. Kromer. 2002. A 7400-year tree-ring chronology in northern Swedish Lapland: Natural climatic variability expressed on annual to millennial timescales. Holocene **12:** 657–665.

Jackson, D. A. 1993. Stopping rules in principal components analysis: A comparison of heuristic and statistical approaches. Ecology **74:** 2204–2214.

Leavitt, P. R., and others. 2009. Paleolimnological evidence of the effects on lakes of energy and mass transfer from climate and humans. Limnol. Oceanogr. **54:** 2330–2348.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia **129:** 271–280.

Lotter, A. F., and H. J. B. Birks. 1997. The separation of the influence of nutrients and climate on the varve time-series of Baldeggersee, Switzerland. Aquat. Sci. **59:** 362–375.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, G. L. Simpson, M. H. H. Stevens, and H. Wagner. 2008. Vegan: Community ecology package. R package version 1.13-2 [Internet]. Available from http://vegan.r-forge.r-project.org/.

Pinheiro, J. C., and D. Bates. 2000. Mixed-effects models in S and S-PLUS. Springer.

Pla, S., D. Monteith, and R. Flower. 2009. Assessing the relative influence of climate warming and atmospheric contamination on the diatom and chrysophyte communities of a remote Scottish loch. Freshwater Biol. **54:** 505–523.

Quinlan, R., P. R. Leavitt, A. S. Dixit, R. I. Hall, and J. P. Smol. 2002. Landscape effects of climate, agriculture, and urbanization on benthic invertebrate communities of Canadian prairie lakes. Limnol. Oceanogr. **47:** 378–391.

R Development Core Team. 2009. R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing. Available from http://www.R-project.org.

Rose, N. 1990. A method for the extraction of carbonaceous particles from lake sediment. J. Paleolimnol. **3:** 45–53.

———. 2001. Fly-ash particles, p. 319–349. *In* J. P. Smol and W. M. Last [eds.], Physical and geochemical methods: Tracking environmental change using lake sediments. Kluwer.

Rühland, K., A. M. Paterson, and J. P. Smol. 2008. Hemispheric-scale patterns of climate-related shifts in planktonic diatoms from North American and European lakes. Glob. Change Biol. **14:** 1–15.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. Semiparametric regression. Cambridge Univ. Press.

Smol, P. J., and others. 2005. Climate-driven regime shifts in the biological communities of Arctic lakes. Proc. Natl. Acad. Sci. USA **102:** 4397–4402.

Wolfe, A. P., J. S. Baron, and R. J. Cornett. 2001. Anthropogenic nitrogen deposition induces rapid ecological changes in alpine lakes of the Colorado Front Range (USA). J. Paleolimnol. **25:** 1–27.

Wood, S. N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Amer. Stat. Assoc. **99:** 673–686.

———. 2006. Generalized additive models: An introduction with R. Chapman and Hall/CRC.

Yee, T. W. 2004. A new technique for maximum likelihood canonical Gaussian ordination. Ecol. Monogr. **74:** 685–701.

———. 2006. Constrained additive ordination. Ecology **87:** 203–213.

———, and M. Mackenzie. 2002. Vector generalized additive models in plant ecology. Ecol. Model. **157:** 141–156.

———, and N. D. Mitchell. 1991. Generalized additive models in plant ecology. J. Veg. Sci. **2:** 577–586.

Zuur, A., E. Ieno, and G. Smith. 2007. Analysing ecological data. Springer.

———, E. Ieno, N. Walker, A. Saveliev, and G. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer.

Associate editors: John P. Smol and Warwick F. Vincent

Received: 05 October 2008
Accepted: 20 April 2009
Amended: 01 June 2009