Ecological Applications, 11(2), 2001, pp. 618–630
© 2001 by the Ecological Society of America

APPLYING BAYESIAN STATISTICS TO ORGANISM-BASED ENVIRONMENTAL RECONSTRUCTION

HANNU T. T. TOIVONEN,[1,5] HEIKKI MANNILA,[2,5] ATTE KORHOLA,[3] AND HEIKKI OLANDER[4]

[1]*Rolf Nevanlinna Institute, P.O. Box 4, FIN-00014 University of Helsinki, Finland*
[2]*Laboratory for Computer and Information Science, Helsinki University of Technology, P.O. Box 5400, FIN-02015 HUT, Finland*
[3]*Department of Ecology and Systematics, P.O. Box 17, FIN-00014 University of Helsinki, Finland*
[4]*Geological Survey of Finland, P.O. Box 96, FIN-02151 Espoo, Finland*

Abstract. Reconstruction of environmental conditions from biological indicators, such as chironomids, can provide valuable insight into the variance of natural systems and the extent of current environmental problems. Various numerical techniques have been developed for this challenging task, based on different models and assumptions, but ignoring some of the uncertainties inherent in the reconstruction.

We study the use of Bayesian modeling and inference in organism-based palaeoenvironmental reconstruction. We propose a Bayesian model, BUM, and compare it empirically with eight other methods, including the widely used weighted averaging (WA) technique. The methods are evaluated on a surface-sediment chironomid training data set from 53 subarctic lakes in northern Fennoscandia by comparing the prediction statistics of these data. The resulting calibration models are also applied to fossil chironomid assemblages in order to evaluate the differences in Holocene temperature reconstructions. The empirical results indicate that BUM is competitive compared with the state-of-the-art methods.

We also describe a generic Bayesian framework for reconstruction, to demonstrate Bayesian tools for reasoning with a variety of ecological response models. Bayesian statistics support the "classical" approach to regression and calibration whereas the state-of-the-art methods, including WA, are based on the conceptually more controversial "inverse" approach. Further, the use of probability distributions rather than point estimates for species responses gives a principled method for handling uncertainty in response modeling.

Key words: *Bayesian modeling and inference; Bayesian statistics; Chironomidae; environmental reconstruction, organism-based; Fennoscandia, northern; Holocene palaeotemperatures; palaeoclimate; regression and calibration techniques; reconstruction methods.*

INTRODUCTION

The importance of long-term monitoring data for the assessment of both the variance of natural systems and the extent of recent disturbance is widely acknowledged. Unfortunately, for most abiotic and biotic systems there is a lack of direct observational data of sufficient accuracy for the natural state to be defined. Parts of the past climatological and environmental systems have been recorded in detail in natural archives, such as peat deposits, lake sediments, and tree rings. A significant amount of this information is in the form of biological indicators, such as diatoms, chironomids, chrysophytes, and pollen, which are well preserved in sedimentary deposits. Such records are highly valuable for a temporal perspective of current environmental problems and for the validation and further development of global systems models, such as general circulation models (COHMAP Members 1988).

The development of quantitative transfer functions by means of modern calibration data sets has been a major advance in recent quantitative palaeoecology (Anderson 1993, Charles and Smol 1994). Palaeoecologists now have the possibility to quantitatively estimate the rates and magnitudes of environmental change and to address and test competing hypotheses (Birks 1995, 1998).

There is a whole spectrum of numerical techniques that are relevant in the context of organism-based palaeoenvironmental reconstruction (Birks 1995, 1998). Each is based on slightly different statistical models or underlying ecological assumptions. Popular methods include multiple linear regression (or inverse regression) (Davis and Anderson 1985, Huntley and Prentice 1988) and modern analogue technique (Overbeck et al. 1985, Bartlein and Whitlock 1993), both of which are widely used in palaeoecological climate studies. Canonical correspondence analysis (CCA) and its single-predictor variant, weighted-averaging (WA) regression and calibration, have proven to be the most suitable statistical techniques for the development of quantitative transfer functions (ter Braak 1988, ter Braak and van Dam 1989, Birks et al. 1990). Weighted-averaging methods have found favor particularly with palaeolim-

Manuscript received 23 November 1998; revised 17 January 2000; accepted 26 January 2000; final version received 8 March 2000.

[5] Present address: Nokia Research Center, Helsinki, Finland. E-mail: Hannu.TT.Toivonen@nokia.com

nologists, because they perform well with noisy, species-rich, compositional data, with many zero values or absences and some "rogue" observations (outliers). Most of these techniques involve two steps: (1) the regression step where the responses of taxa to the contemporary environment are modeled; and (2) a calibration step where the environmental variable of interest is predicted from the fossil compositional data (Birks 1995).

The reconstructions contain many uncertainties. The regression and calibration models are usually stochastic, to reflect the stochasticity of nature, and this gives variation in the reconstructions. Further, the regression step is based on observed data, which itself is stochastic, so the uncertainty in the modeled responses should be taken into account. Finally, no statistical model can be considered to be a true image of the nature, so there is uncertainty about how well the model works as a model of reality. All three of these levels of uncertainty should be accounted for in any reconstruction.

In this paper, we consider a new statistical approach to the reconstruction problem, namely the application of Bayesian inferential approaches to regression and calibration. The issue of uncertainty in modeled responses is one that is handled explicitly within the paradigm of Bayesian analysis, unlike in most current techniques. In Bayesian statistics, all unknown quantities—parameters as well as the environmental variables to be reconstructed—are treated as random variables with probability distributions. The smaller the variance, the more we think we know about the quantity. Within this paradigm, the knowledge gained from analyzing modern training data (also called "calibration data") is explicitly conditioned on the statistical model used, our prior knowledge, and the data. Further, reconstructions are also conditioned on these factors.

The regression step of environmental reconstruction is an example of a difficult statistical inversion problem: species abundances are known, and the task is to estimate the parameters that describe how the environment influences the abundances. Bayesian statistical analysis is well suited for the task, since Bayesian statistics is essentially about solving such probabilistic inversion problems. The flexibility of Bayesian statistics allows the use of complex models, if necessary, to describe the statistical response of species to the environment.

We describe a Bayesian framework for reconstruction. Using this framework, a wide variety of statistical models can be used for the response of the species to the environment. We show how the framework specifies the probability distributions of past environments, based on the response model used and the observed training and fossil data, and how the regression and calibration steps are contained in the process. Although the framework as such is a fairly simple application of Bayesian principles to the reconstruction problem, it offers a remarkably flexible setting for regression and calibration.

TABLE 1. Notation used to describe the reconstruction problem.

Symbol	Interpretation
n	number of modern sites
m	number of taxa
x_i	value of the environmental variables of interest at site i
$\mathbf{X} = (x_i)$	matrix of values for the environmental variables of interest
y_{ik}	abundance of taxon k at site i
$\mathbf{Y} = (y_{ik})$	matrix of abundances
$\mathbf{y}_0 = (y_{01}, y_{02}, \ldots, y_{0m})$	the fossil assemblage for which the environmental variables are to be inferred
x_0	value of the environmental variables of interest for the fossil assemblage; to be inferred

We introduce one specific Bayesian response model, called "BUM," that illustrates the use of Bayesian modeling and reasoning. BUM is based on a unimodal response model, and it can be seen as a Bayesian version of some previous reconstruction models, such as weighted-averaging models and the Gaussian logit model (ter Braak and van Dam 1989, Birks et al. 1990). A significant difference is, however, how the statistical inversion problem of response regression is solved in these methods.

We empirically compare BUM to the weighted-averaging technique, including the recently developed weighted-averaging partial least-squares method (WA-PLS) (ter Braak and Juggins 1993, ter Braak et al. 1993), and to other popular reconstruction techniques. In the comparison, our focus is on palaeoclimate reconstruction using modern and fossil data sets of nonbiting midges (Chironomidae). We concentrate on statistical evaluation and do not try to validate the accuracy of reconstruction in terms of Holocene climate history.

METHODS AND MATERIALS

Suppose we have n sites and m taxa, and let y_{ik} be the abundance of taxon k at site i. Let $\mathbf{Y} = (y_{ik})$ be the $n \times m$ matrix of abundances (Table 1). Suppose that for the n sites we have measured p physicochemical variables of the environment; denote by x_{ij} the value of variable j at site i, and by $\mathbf{X} = (x_{ij})$ the $n \times p$ matrix of these variables. The environment at site i is thus described by the vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$. The problem in quantitative reconstruction is to obtain good estimates of the environmental variables in the past, such as temperature, based on fossil assemblages.

Statement of the quantitative reconstruction problem

Let $\mathbf{Y} = (y_{ik})$ be the matrix of abundances and $\mathbf{X} = (x_{ij})$ the matrix of environmental variables. Let $\mathbf{y}_0 = (y_{01}, y_{02}, \ldots, y_{0m})$ be a fossil assemblage, and let $\mathbf{x}_0 =$

$(x_{01}, x_{02}, \ldots, x_{0p})$ be the set of values of environmental variables for the fossil assemblage. Given $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{y}_0$, the task is to reconstruct the environmental variables $\mathbf{x}_0$.

For simplicity, we consider only one environmental variable, for instance temperature. This corresponds to a setting where $\mathbf{x}_i$, the environment at site i, is a scalar value (denoted by x_i) rather than a vector of values. With a slight abuse of notation, we denote variables and their values by the same names and, if not necessary, we do not distinguish between probabilities, densities, and distributions.

Different ways to solve the quantitative reconstruction problem can be classified into two broad categories (Birks 1995, 1998, ter Braak 1995). In the *classical* approach to regression and calibration one first estimates the ecological response function $f(x_i) = \mathbf{y}_i$, where x_i is the environment at site i and $\mathbf{y}_i$ the organism assemblage at the site. For reconstruction, in principle, the inverse of the response function is then used to estimate the past environment based on the fossil assemblage: $x_0 \approx f^{-1}(\mathbf{y}_0)$. (Typically the inverse function does not exist; consequently one can estimate x_0 by minimizing the difference between $f(x_0)$ and $\mathbf{y}_0$.) The *inverse* approach is based on directly estimating the inverse function $x_i = f^{-1}(\mathbf{y}_i)$ and then applying the same function in reconstruction to obtain $x_0 \approx f^{-1}(\mathbf{y}_0)$.

The classical approach is often considered more attractive (ter Braak 1995); the response function f is explicit, it can be constrained to take some ecologically realistic form, and it can be verified for ecological plausibility. However, the classical approach has not gained wide popularity due to computational problems in solving the statistical inversion problem and a comparatively weak performance. While the inverse approach is mathematically and computationally much simpler to use, it is not always obvious what the form of the response function $f(x_i) = (f^{-1})^{-1}(x_i)$ is and how plausible it is. Ter Braak (1995) concludes that, in general, the inverse approach performs slightly better in interpolation and the classical approach at the extremes and with slight extrapolation.

Reconstruction in the Bayesian framework

We adopt the classical approach to regression and calibration. Our aim is to show how Bayesian statistics can be used to solve the inversion problems in the classical approach and to demonstrate with a specific model, BUM, that competitive performance can in fact be achieved. The methods we mostly contrast our approach to, the weighted-averaging models, assume responses that are similar to BUM, but are inverse methods.

To keep the discussion simple, we restrict the argumentation to models with the following properties. First, we consider only a single environmental variable. Second, we model the abundance of each taxon separately, i.e., we assume that the species abundances at a site are conditionally independent given the environment at the site. Third, we take the classical approach to calibration and approximate the abundance y_{ik} by a function of the environment x_i and the taxon k.

Under the above assumptions, the density of abundance y_{ik} is conditionally independent of other information, given the parameters of the taxon k and the environment at site i. A central part of the model is the *likelihood function* $p(y_{ik} | x_i, \mathbf{w}_k)$, which defines the probability of observing abundance y_{ik} given the environment x_i and the vector $\mathbf{w}_k = (w_{k1}, w_{k2}, \ldots, w_{kh})$ of parameters for taxon k. The likelihood function determines the forms of the responses assumed in the model, and defining this function is a major decision in the construction of a response model. For now, we make no assumptions about the form of the response.

In the reconstruction problem we are interested in obtaining the *posterior density,*

$$p(x_0 | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0) \tag{1}$$

of the unknown x_0 (past environment) given the data $\mathbf{X}$ (modern environment), $\mathbf{Y}$ (modern abundances), and $\mathbf{y}_0$ (fossil abundances). This can be written as

$$p(x_0 | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0) = \int p(x_0, \mathbf{W} | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0)\, d\mathbf{W} \tag{2}$$

where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ is the matrix of all parameters and the integral is taken over all possible values of $\mathbf{W}$. Under reasonable conditional independence assumptions the posterior density can be rewritten as $\int p(x_0, \mathbf{W} | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0)\, d\mathbf{W} = \int p(\mathbf{W} | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0) p(x_0 | \mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{y}_0)\, d\mathbf{W} = \int p(\mathbf{W} | \mathbf{X}, \mathbf{Y}) p(x_0 | \mathbf{W}, \mathbf{y}_0)\, d\mathbf{W}$, to illustrate the role of the terms corresponding to regression and calibration, $p(\mathbf{W} | \mathbf{X}, \mathbf{Y})$ and $p(x_0 | \mathbf{W}, \mathbf{y}_0)$, respectively.

Eq. 2 shows how the conditional density of x_0, the past environment, is obtained by integrating over all possible parameter values $\mathbf{W}$ for the response functions. The equation indicates an important property of the Bayesian approach: we do not just use the "best" response function for each taxon; rather, all possible response functions are considered (in principle) and their impact on the final results is determined by their likelihoods and prior probabilities.

According to Bayes' theorem, we can write

$$p(x_0, \mathbf{W} | \mathbf{X}, \mathbf{Y}, \mathbf{y}_0)$$

$$= \frac{p(x_0, \mathbf{W} | \mathbf{X}) p(\mathbf{Y}, \mathbf{y}_0 | \mathbf{X}, x_0, \mathbf{W})}{p(\mathbf{Y}, \mathbf{y}_0 | \mathbf{X})} \tag{3}$$

$$\propto p(x_0, \mathbf{W} | \mathbf{X}) p(\mathbf{Y}, \mathbf{y}_0 | \mathbf{X}, x_0, \mathbf{W}). \tag{4}$$

Assuming that x_0 and $\mathbf{W}$ are independent of $\mathbf{X}$ we have $p(x_0, \mathbf{W} | \mathbf{X}) = p(x_0, \mathbf{W})$. Assuming that the abundances at different sites are conditionally independent given the environment at each site and the taxon parameters, we can write the term $p(\mathbf{Y}, \mathbf{y}_0 | \mathbf{X}, x_0, \mathbf{W})$ as a product

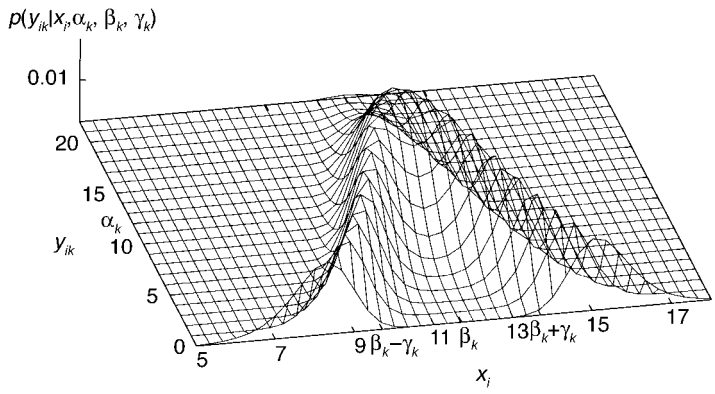

FIG. 1. An example of a density surface for the abundance y_{ik} of a species k and the environment x_i. The species has parameters $\alpha_k = 12$, $\beta_k = 11.7$, and $\gamma_k = 2$ (see Eq. 6), and the surface has been plotted assuming that variable x_i is normally distributed with a mean of β_k and a variance of 3.

over all n sites. Assuming further that taxa within a site are conditionally independent given the environment and the taxon-specific parameters, we can write another nested product over all m taxa:

$$p(x_0, \mathbf{W}|\mathbf{X})p(\mathbf{Y}, \mathbf{y}_0|\mathbf{X}, x_0, \mathbf{W}) = p(x_0, \mathbf{W}) \prod_{i=0}^{n} \prod_{k=1}^{m} p(y_{ik}|x_i, \mathbf{w}_k). \quad (5)$$

Here $p(x_0, \mathbf{W})$ is the *prior density* of x_0 and $\mathbf{W}$, and $p(y_{ik}|x_i, \mathbf{w}_k)$ is the likelihood function.

The conclusion from Eqs. 2–5 is that the posterior distribution of x_0, the temperature to be reconstructed, is proportional to the prior density of x_0 and the parameters $\mathbf{W}$, and the likelihood over all observations y_{ik}, integrated over the parameter space. Different response models can be plugged in by defining the likelihood term $p(y_{ik}|x_i, \mathbf{w}_k)$ in different ways; existing knowledge of the parameters and environment can be expressed in the prior density.

BUM model specification

In the case of one environmental variable, such as temperature, a plausible ecological response function is a smooth unimodal function such as a bell-shaped curve. We follow the weighted-averaging and Gaussian logit models (ter Braak and van Dam 1989, Birks et al. 1990): in the BUM model the expected response of each taxon k is specified by the unimodal curve

$$E(y_{ik}|x_i, \mathbf{w}_k) = r(x_i, \alpha_k, \beta_k, \gamma_k) = \alpha_k \exp\left[-\left(\frac{x_i - \beta_k}{\gamma_k}\right)^2\right] \quad (6)$$

where $\mathbf{w}_k = (\alpha_k, \beta_k, \gamma_k)$ are the model parameters for taxon k. Given a value of the environmental variable x_i, the function specifies the relative suitability of the environment for taxon k. Parameter β_k represents the optimum temperature, γ_k a measure of tolerance for temperature, and α_k the expected abundance of taxon k in the optimal environment.

We assume here that the species abundances are represented as actual counts and that a (large) constant number of individuals has been counted at each site, i.e., y_{ik} indicates how many representatives of taxon k were observed at site i and $\Sigma_k y_{ik} = c$ for all i. In the model we assume that the abundances y_{ik} are Poisson distributed with parameter $r(x_i, \alpha_k, \beta_k, \gamma_k)$, i.e.,

$$p(y_{ik}|x_i, \alpha_k, \beta_k, \gamma_k) = (y_{ik}!)^{-1} r(x_i, \alpha_k, \beta_k, \gamma_k)^{y_{ik}} \exp[-r(x_i, \alpha_k, \beta_k, \gamma_k)]. \quad (7)$$

The value of $r(\cdot)$ is both the mean and the variance of the Poisson distribution. In this model zero abundances are not handled separately. Instead, the Poisson distribution assigns them a density—which is typically small—that depends on the expected value of the distribution.

The modeled response of a species k can be illustrated as a density surface where the first axis is the environment x_i, the second axis is the abundance y_{ik}, and the height of the surface corresponds to the probability of observing abundance y_{ik} in environment x_i (Fig. 1). Given a value for the environmental variable x_i, the conditional distribution of abundance y_{ik} is specified by the corresponding slice through the surface parallel to the y_{ik} axis (after rescaling the area to unity). More interestingly, the slices parallel to the x_i axis specify conditional distributions for the environmental variable x_i, given the abundance y_{ik} (after rescaling to unity).

The form of the surface for species k is defined by the parameters α_k, β_k, and γ_k. In the regression step the surface is fitted to the observations by considering, in a sense, the slices parallel to the y_{ik} axis. In calibration, the slices parallel to the x_i axis are used to estimate the probability distribution of the environmental variable x_i. In reconstruction with a number of species, the distribution of the environmental variable x_i is specified by the product over all the species k.

To complete the model specification, we need to give the prior densities $p(x_0, \mathbf{W})$. For the prior distributions of the parameters we write

$$p(x_0, \mathbf{W}) = p(x_0) \prod_{k=1}^{m} p(\alpha_k)p(\beta_k)p(\gamma_k) \quad (8)$$

i.e., the unknown temperature and the individual parameters for the k response functions are assumed to be a priori independent.

For our experiments we have selected one set of prior distributions. In order to reduce the effect of subjective prior distributions and to keep experiments fair, the distributions are taken from our data set and have been constant for both statistical evaluation and reconstruction comparison. Note, however, that for reconstruction studies—which are outside the scope of this paper—it would be useful to use more informative prior distributions for the environmental variables.

In our experiments the temperature x_0 to be reconstructed is a priori normally distributed with a mean of 11.7 and a variance of 0.30. The mean has been chosen to be the same as the mean of the modern training set. The variance is one fifth of the variance of the modern training set. The prior variance of temperatures in the fossil data is smaller since data points in the modern training set represent intervals of several years or a decade, whereas in the fossil data set one data point corresponds to the average over several decades.

For the parameters α_k, β_k, and γ_k we use the following prior distributions: scaling factor α_k has a uniform distribution in [0.1, 100], the optimum temperature β_k has a normal distribution with parameters 11.7 and 1.21, and the tolerance parameter γ_k is assumed to be gamma distributed with parameters 9 and 3. The prior distribution of α_k allows any possible value with equal density. The optimum temperatures β_k are assumed to have roughly the same distribution as the modern temperatures. Finally, the prior distribution of γ_k is defined to encourage relatively large tolerances and therefore gently sloping response curves. Such a prior distribution is selected in order to avoid overfitting, by assuming conservative rather than sharp responses, unless strongly evidenced by the data.

BUM serves as an illustration and as a first application of a Bayesian reconstruction model: we plan to develop more realistic models that also take further advantage of the benefits of the Bayesian approach.

Data set

The methodology was applied to the training data set of Olander et al. (1999), collected to reconstruct Holocene midsummer air and surface-water palaeotemperatures from remains of Chironomidae in northern Fennoscandia. The value of chironomids as providing one of the most sensitive records of past temperature variations is widely recognized (Walker et al. 1991, 1997, Levesque et al. 1994, 1997, Lotter et al. 1997).

The training set consists of surface-sediment chironomid head capsules and site-specific air-temperature and epilimnetic water-temperature data collected from 53 lakes located mainly in northwestern Finnish Lapland (66°92–69°18 N, 20°67–25°40 E). Vegetation, climate, and bedrock characteristics of the study area are described in detail by Olander et al. (1997). Mean surface-water temperatures in July varied between 6.1° and 15.4°C and altitude-corrected mean air temperatures in July between 8.5° and 14.9°C.

The fossil assemblages originate from an AMS-dated sediment core from lake Tsuolbmajavri (68°41′ N, 22°05′ E), northwestern Finnish Lapland, and they are used to reconstruct Holocene temperature trends. Tsuolbmajavri is a relatively small (surface area 13.9 ha), shallow (maximum depth 5.35 m), and oligotrophic clear-water lake with a mean water temperature of 10.9°C and a mean air temperature of 11.0°C during July. A 291-cm-long sediment core was obtained from the lake in 1997 and analyzed for chironomids at every 2-cm interval (equivalent to a resolution of 50–70 yr).

Rare taxa were excluded from the data set by retaining only those occurring in at least two lakes with a relative abundance of >2% in at least one lake. This left a total of 38 chironomid taxa for calibration purposes. Further details of the data set can be found in Olander et al. (1997, 1999).

Evaluation methodology

Chironomid–temperature calibration models have previously been developed from the data set using eight different numerical techniques, each based on slightly different underlying models or ecological assumptions (Olander et al. 1999). The methods used were: (1) simple weighted averaging (WA) with an "inverse" deshrinking regression (WA inverse) (Birks et al. 1990, Birks 1995); (2) weighted averaging with taxon-tolerance weighting and an "inverse" deshrinking regression (WAtol inverse) (Birks et al. 1990, Birks 1995); (3) simple weighted averaging with a "classical" deshrinking regression (WA classical) (Birks et al. 1990, Birks 1995); (4) weighted averaging with taxon-tolerance weighting and a "classical" deshrinking regression (WAtol classical) (Birks et al. 1990, Birks 1995); (5) partial least squares (PLS) (Martens and Naes 1989, ter Braak and Juggins 1993, Birks 1995); (6) weighted-averaging partial least squares (WA-PLS) (ter Braak and Juggins 1993, Birks 1995, ter Braak 1995); (7) Gaussian logit model (GLM) (maximum-likelihood method) (ter Braak and van Dam 1989, Birks et al. 1990, Birks 1995); and (8) modern analogue technique (MAT) (Bartlein and Whitlock 1993, Birks 1995, ter Braak 1995). (Note that all WA methods are inverse methods. Terms "inverse" and "classical" in the names refer to a deshrinking step, performed in order to reduce estimation bias at the extremes.)

We compare here the calibration models developed by using the above-listed eight methods plus our BUM model. In particular, we compare BUM to the WA technique, which is one of the most popular calibration techniques among palaeoenvironmental scientists. It is a mathematically and computationally simple method, based on the concepts of niche-space partitioning and ecological optima and tolerances of taxa (ter Braak and Juggins 1993). The basic model in WA assumes that

the response curve for an environmental variable x has a unimodal shape. The optimum value of x for a taxon is estimated as a weighted average of the observed values of x. The tolerance for the taxon with respect to the environmental variable is similarly estimated as a weighted standard deviation of the observed values. WA and WA-PLS are perhaps the most successful and widely used calibration methods in current palaeoecological and paleoclimate research (Bennion et al. 1996, Korsman and Birks 1996, Lotter et al. 1997, Walker et al. 1997, Korhola et al. 2000). Moreover, simulation studies, theoretical analyses, and palaeoecological applications demonstrate that WA techniques generally perform well. For example, ter Braak et al. (1993) recommend WA-PLS as a simple and robust alternative to more complicated models.

The performance of each method in the modern training set is assessed on the basis of the following standard evaluation methods in palaeoecological reconstruction: (1) the root mean square error of the difference between the observed and the inferred (RMSE) and the observed and the predicted temperature values (RMSEP); (2) the squared correlations (r^2) between measured and chironomid-inferred temperature values; (3) the "mean bias" (difference between observed and predicted values); and (4) the "maximum bias" along the temperature gradient (ter Braak and Juggins 1993). These statistics are computed for both the estimation error (or "apparent" error) and the leave-one-out cross-validation error (or "jackknifing" error). Obviously, cross-validation gives more realistic prediction errors and is more useful in validating the results (Birks et al. 1990, ter Braak and Juggins 1993). For calculating bias statistics, we followed the procedure given in ter Braak and Juggins (1993) and Olander et al. (1999) where the temperature gradient was subdivided into 10 equal intervals, the mean bias (mean of the differences between observed and inferred) per interval was calculated, and the largest absolute value of mean bias for an interval was used as a measure of maximum bias. In order to compare the reconstructions given by different methods, BUM and the most widely used calibration models are applied to the fossil chironomid assemblages.

As BUM predictions in these evaluations we use the mean model prediction integrated over the posterior distribution. The BUM model was implemented using stochastic Markov chain Monte Carlo (MCMC) methods (see, e.g., Gilks et al. 1996). The particular implementation is based on Bassist (Toivonen et al. 1999), a general-purpose MCMC software tool under development at the University of Helsinki.

Results

Prediction accuracy

The Bayesian model BUM compares well with the rest of the models, yielding relatively low prediction error for water temperature both in terms of "apparent" as well as cross-validation (jackknifed) statistics (Table 2, top). For air temperatures (Table 2, bottom), the Bayesian technique also performs comparably to other methods as assessed by the error estimates. However, for the current data set the BUM model offers no improvement in prediction accuracy over WA (weighted-averaging) or corresponding methods. In terms of RMSEP only, there is not much difference between WA (inverse), PLS, WA-PLS, GLM, MAT, and BUM (see *Evaluation methodology*, above, for explanation of these model and technique abbreviations).

Cross-validation tests with water- and air-temperature calibration functions (Fig. 2A and B) and the respective residuals (Fig. 2C and D) show that, in the structure of the residuals, little difference exists between WA (inverse) and BUM. However, BUM has somewhat smaller temperature-dependent bias (not reported in Table 2): the least-squares linear fit to the residuals gives slope coefficients of −0.59 (water) and −0.56 (air) vs. −0.49 (water) and −0.39 (air) for WA and BUM, respectively. Recall that the prior distribution for temperature in our experiments was actually chosen for the fossil data set; the prior variance is only one fifth of the variance in the modern training set. Using larger variance leads to smaller bias, but eventually also to a degradation of prediction accuracy in terms of RMSEP. (See Robertson et al. [1999] for a discussion on regression errors and bias in environmental reconstruction.)

The overall results (Table 2) indicate that WA (inverse), PLS, WA-PLS, GLM, MAT, and BUM could all be used to develop a transfer function for surface-water July temperature with a cross-validation prediction error of about 1.5–1.6°C and for July air temperature with a prediction error of 0.8–1.1°C. These results generally compare well with other existing chironomid–summer air- and water-temperature inference models (Lotter et al. 1997, Walker et al. 1997, Brooks and Birks 2000).

BUM response modeling

An examination of the posterior response functions of BUM shows that responses are modeled in a meaningful manner; sparse observations result in response curves with large variance, and the curves are plausible for both interpolation and slight extrapolation. For example, *Heterotrissocladius subpilosus* (Fig. 3A) shows a more distinct response to temperature, whereas *Psectrocladius sordidellus* (Fig. 3B) seems to have abundances of up to 10% all over the temperature gradient. Accordingly, the response curves are fairly different: *H. subpilosus* (Fig. 3A) has a distinctive peak at the lower end of the temperature gradient, although with a lot of variance, whereas the response curve of *P. sordidellus* (Fig. 3B) is fairly flat, indicating a much weaker response to temperature.

The response curves extrapolate smoothly outside the observed temperature gradient. Since we prefer

TABLE 2. Performance statistics of nine different inference models.

Model and method†	Apparent				Jackknifing			
	RMSE (°C)	r^2	Mean bias (°C)	Maximum bias (°C)	RMSEP (°C)	r^2	Mean bias (°C)	Maximum bias (°C)
Summer surface-water temperature								
WA (inverse	1.34	0.52	0.00	3.56	1.56	0.34	−0.04	3.73
WAtol (inverse)	1.31	0.54	0.00	3.45	1.76	0.23	−0.03	3.73
WA (classical)	1.86	0.52	0.00	4.09	2.06	0.41	−0.03	3.82
WAtol (classical)	1.79	0.54	0.00	3.13	2.49	0.27	−0.03	3.76
PLS (1 component)	1.33	0.52	0.00	3.08	1.57	0.35	0.00	3.56
WA-PLS (1 component)	1.34	0.52	0.00	3.58	1.53	0.37	0.00	3.88
GLM‡	1.17	0.63	0.00	0.87	1.37	0.59	0.00	1.28
MAT (6 analogues)	n.a.§	n.a.	n.a.	n.a.	1.48	0.43	0.00	1.90
BUM	1.27	0.56	−0.02	2.57	1.49	0.42	−0.04	2.88
Mean July air temperature								
WA (inverse)	0.85	0.51	0.00	2.70	0.95	0.39	−0.02	3.01
WAtol (inverse)	0.85	0.50	0.00	2.82	1.00	0.34	−0.02	3.10
WA (classical)	1.18	0.51	0.00	2.23	1.29	0.40	−0.04	2.79
WAtol (classical)	1.20	0.50	0.00	2.46	1.39	0.35	−0.05	2.94
PLS (3 components)	0.42	0.88	0.00	0.68	0.75	0.63	0.00	2.23
WA-PLS (2 components)	0.59	0.77	0.00	1.44	0.87	0.50	0.00	2.93
GLM‡	0.62	0.74	0.00	0.49	1.13	0.69	0.01	1.02
MAT (4 analogues)	n.a.	n.a.	n.a.	n.a.	0.84	0.56	0.00	0.81
BUM	0.72	0.65	0.04	2.24	0.97	0.43	−0.03	2.54

Notes: The results are from Olander et al. (1999) except for BUM. Statistics are computed for both the estimation error ("Apparent") and the leave-one-out cross-validation error ("Jackknifing"). RMSE = root mean square error of the difference between the observed and inferred values; RMSEP = root mean square error of the difference between the observed and predicted values; r^2 = squared correlation between the measured and chironomid-inferred values; mean bias = the mean of the difference between observed and predicted values; maximum bias = largest absolute value of mean bias in a tenth of the temperature gradient.

† For explanation of the models and techniques, see *Methods and materials: Evaluation methodology.*

‡ The results for GLM in Olander et al. (1999) are wrong due to a software error; here we show the new, correct results (H. J. B. Birks, *personal communication*).

§ Not applicable.

fairly wide tolerances in the prior distributions, the response curves tend to be wide. The preference for wide tolerances may sometime "push" the optimum temperature to either side; this is why the temperature optimum for *H. subpilosus* (Fig. 3A) is colder than the peak in the data. This is indeed the desired behavior: moving the optimum to the right would result in a bad fit with most of the data points if the curve is not narrowed, but decreasing the tolerance would, on the other hand, give sharper responses than are ecologically plausible.

Fig. 3C and D demonstrates one effect of prior distributions on the posterior response curves. The plots are the same as in Fig. 3A and B except that here the prior distribution of tolerance has been changed to favor larger tolerances. The posterior response of *H. subpilosus* (Fig. 3C) is now somewhat wider, whereas for *P. sordidellus* (Fig. 3D) it remains practically the same. This illustrates, again, the Bayesian treatment of uncertainty. The bump in the left panels is in an area with only few data points, and accordingly, the responses reflect our prior distributions more than the response curves in the right panels, where there are more nonzero data points along the bump.

Reconstructions

A comparison of the BUM air-temperature reconstruction with the reconstructions produced by other commonly used models shows that there is a marked correspondence particularly between the reconstructions derived by our Bayesian model and the WA (inverse) and WAtol (inverse) models (Fig. 4A).

The Bayesian reconstruction generally compares well with the results achieved by WA-PLS for the early part of the Holocene, yet for the latter part of Holocene WA-PLS gives ~1°C lower temperature readings (Fig. 4B). The discrepancy is still more pronounced in the case of PLS, which suggests temperature variations ranging from 5°C (1 component) to >10°C (3 components) in amplitude during the Holocene. It should be noted that a variability of this magnitude is highly unlikely for the Holocene climate on the basis of other available palaeoclimatic information. It seems that PLS overfits to the data.

Despite the difference in absolute values between the reconstructions by BUM, WA (inverse), and WAtol (inverse), and those produced by WA-PLS and PLS, the general patterns and trends are more or less the same in the temperature curves between the most plausible models, i.e., BUM, WA (inverse), WAtol (inverse), WA-PLS (2 components), and PLS (with 1 component). The tendency of WA-PLS and PLS to yield more extreme reconstructions in comparison to simple WA is known from previous studies (Walker et al. 1997), and should be more carefully analyzed in future

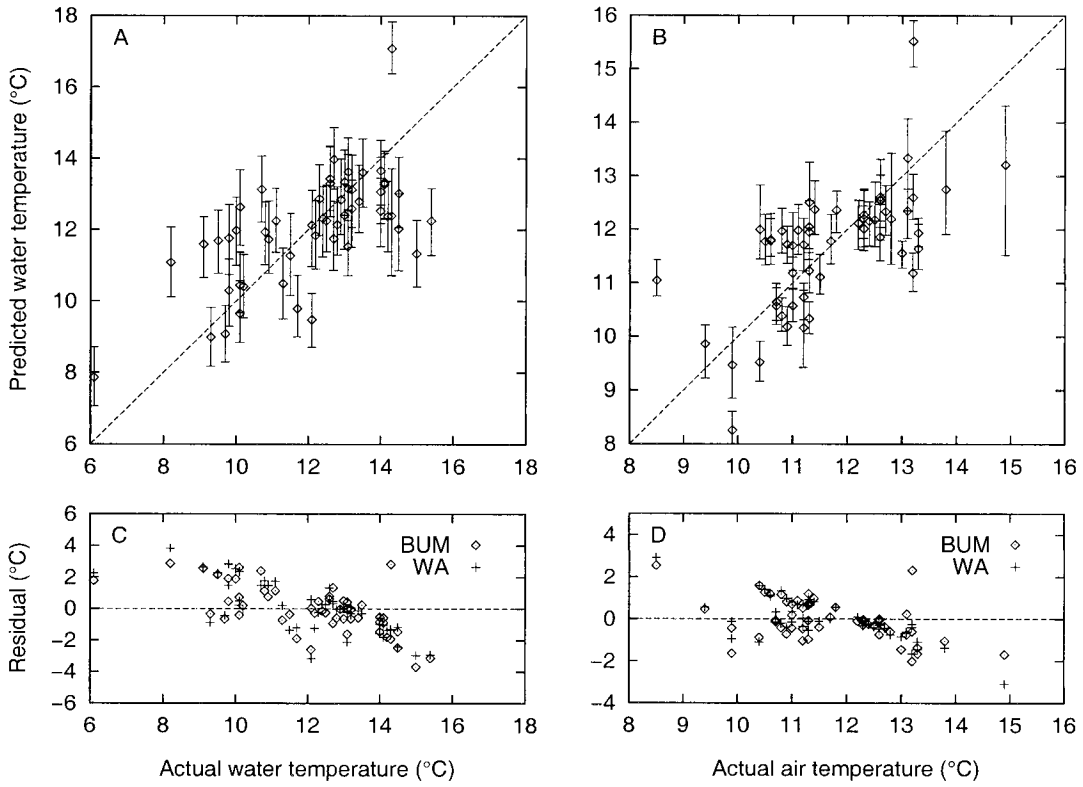

FIG. 2. Actual vs. predicted (A) summer surface-water temperatures and (B) mean air temperatures in July, from cross-validation experiments with BUM. The plots show the posterior mean value as the predicted value, and the error bars show 95% posterior intervals for the prediction. Residuals are plotted for (C) water temperatures and (D) air temperatures, for both the BUM and the weighted-average (WA) (inverse) methods.

modeling and calibration studies. This tendency is particularly evident when using more than one component in the PLS and WA-PLS calibration.

Temperatures reconstructed by MAT (Fig. 4C) are in the same range with BUM, WA (inverse), and WAtol (inverse). MAT agrees with the other methods on a number of the peaks, but differs substantially in terms of trends, suggesting that the approach is not suitable for the current material.

Finally, we briefly assess the stability of the BUM reconstruction with respect to different choices of prior distributions. Temperature reconstructions obtained with the BUM model and four different prior distribution settings are very similar, and the model seems to be quite robust to small changes in the prior distributions in that respect (Fig. 5). Experiments with prediction accuracy (not shown) also confirm the conclusion. The most obvious effect is that favoring larger tolerances eventually leads to a decline in the prediction accuracy. This is due to the fact that flat response curves are not capable of expressing those responses that are critical for reconstruction.

DISCUSSION

The weighted-average (WA) methodology is maybe the most successful current calibration technique. It—as well as a majority of widely used reconstruction techniques—is based on the inverse approach to regression and calibration. The major disadvantages associated with the WA methodology, both ecologically and statistically, can be summarized as follows; most inverse methods share the same problems. (1) The basic assumption behind the formulation of a prediction equation in WA methodology is that there are true, fixed values for the parameters of interest. From an ecological perspective, there are many difficulties with this assumption (Ellison 1996). For example, no two organisms are likely to respond to a change in their living environment in precisely the same way. (2) WA techniques are based on the philosophy of maximum likelihood; when several response curves have approximately the same likelihood, only one is chosen. (See Ludwig [1996] for a discussion of how point estimates lose information.) (3) Weighted averaging is sensitive to the distribution of the environmental variable in the calibration data set. (4) All WA-based methods using inverse deshrinking regression suffer from problems of "edge effects": there is a tendency for predicted values to be overestimated at the low end of the sampled gradient and to be underestimated at the high end of the gradient (ter Braak and Juggins 1993, ter Braak et al. 1993, Birks 1995).

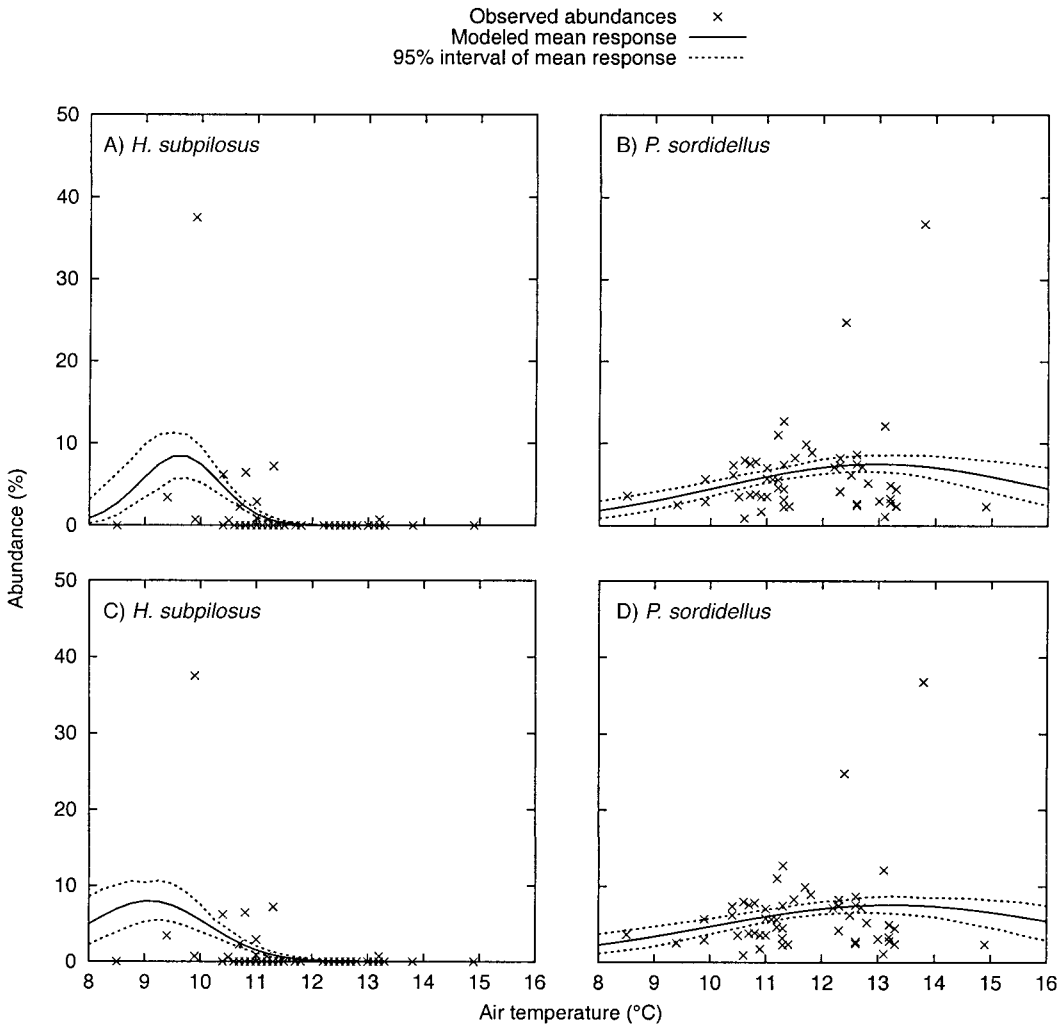

FIG. 3. Point-wise expected responses of two chironomid taxa (*Heterotrissocladius subpilosus* and *Psectrocladius sordidellus*) as a function of mean air temperature in July, according to the Bayesian model BUM. Small crosses show the actual abundances in the modern calibration set. The prior mean of tolerance is either 3 (A and B) or 5 (C and D); the variance is 1. Note that the panels depict intervals for the function $r(\cdot)$, the *expected* abundance, not for the abundance. The expected response and the 95% intervals are computed point-wise for each temperature by integrating over the posterior distribution.

The main benefits of weighted averaging over many other calibration techniques are that the method does not assume a linear response of species with their environment, it is not restricted by internal correlations (multicollinearity) or by the large number of taxa, it is relatively insensitive to outliers, and it also performs well in situations where the analogy between modern and fossil assemblages is not perfect (ter Braak and Juggins 1993, Birks 1995, ter Braak 1995, Birks 1998). The method usually has a good empirical predictive power and it produces calibration models that have low prediction errors and low average and maximum bias, as assessed by cross-validation techniques. In WA-PLS, there is, further, an option to utilize the residual correlations in the taxon data to improve estimates of species optima in the final WA predictor. This has been shown in many experiments to improve the predictive power of the model (ter Braak and Juggins 1993, ter Braak et al. 1993, Birks 1995).

While the classical approach to regression and calibration is conceptually attractive (ter Braak 1995), a major benefit of WA and other inverse-approach techniques is their superior performance (ter Braak et al. 1993, Bennion et al. 1996, Korsman and Birks 1996, Lotter et al. 1997, Walker et al. 1997, Korhola et al. 2000). In our experiments the BUM model seems to perform considerably better than other calibration techniques based on the classical approach and competitively with the best inverse methods. It is thus possible to achieve the reliability of inverse methods with the classical approach.

The above-mentioned issues associated with the WA-based methods are less problematic with BUM. (1) Instead of assuming that there are true values for

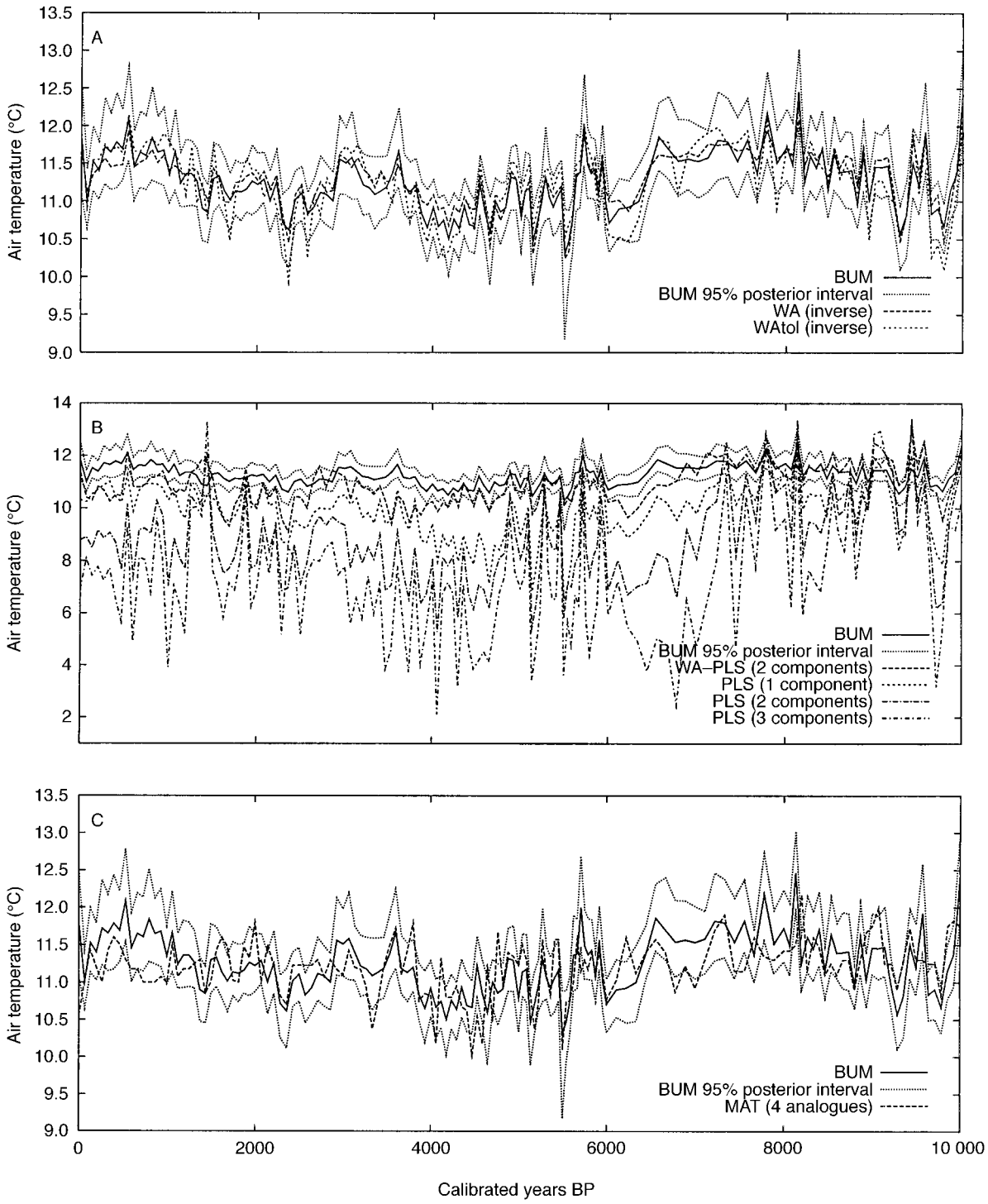

FIG. 4. Comparison of mean air temperatures in July as reconstructed by different models. Note the differences in temperature scales. As the BUM reconstruction we give the posterior mean; the mean and the 95% intervals are computed point-wise for each fossil assemblage by integrating over the joint posterior distribution of all parameters. WA = weighted average, PLS = partial least squares, MAT = modern analogue technique; for further information see *Methods and materials: Evaluation methodology.*

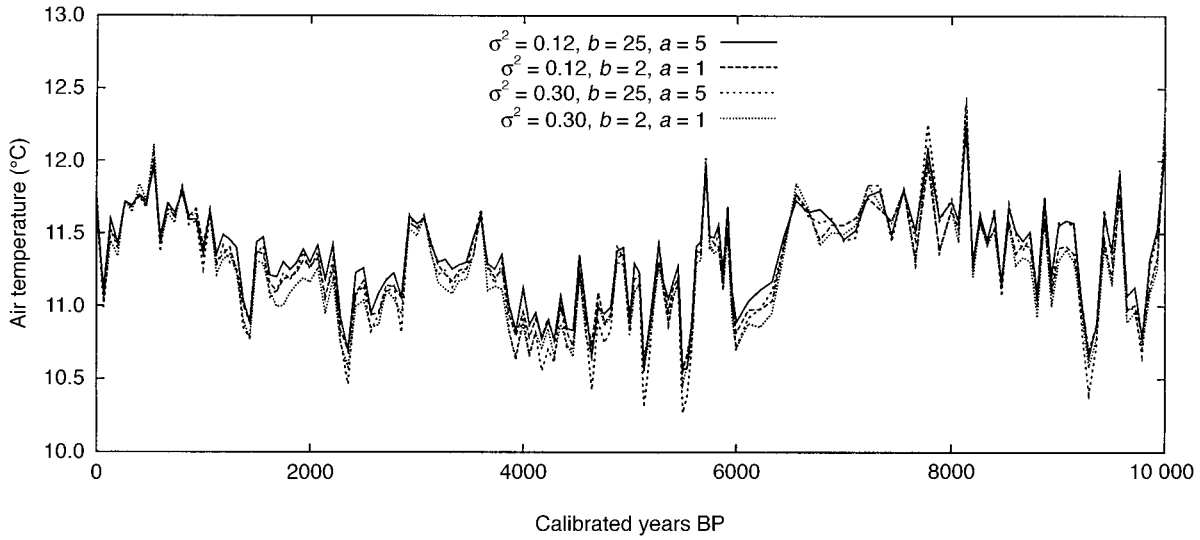

FIG. 5. Examples of the effect of prior distributions on the BUM-reconstructed mean air temperature in July: σ^2 is the variance of the prior distribution of temperature (mean = 11.7); b and a are parameters of the gamma prior distribution for the tolerance with respect to temperature.

parameters, BUM considers the probability distributions of parameters. (2) Maximum-likelihood estimates for response curves are replaced by (posterior) distributions, which handle explicitly the uncertainty in response modeling. (3) BUM is less sensitive to the distribution of the environmental variable in the calibration data set. (4) In the BUM model, there is no systematic tendency for edge effects.

The close correspondence between the results achieved by BUM and WA-based methods gives credence to the whole idea of organism-based environmental reconstruction, as the approaches used are based on very different statistical reasoning. The main differences between our Bayesian modeling approach and the more traditional reconstruction methods are summarized in Table 3.

Explicit prior distributions are useful, e.g., for ensuring that the parameters cannot have meaningless values. For instance, consider a parameter that defines the tolerance of a taxon to an environmental variable. Probably no species has an extremely small tolerance; this prior belief about tolerance can be expressed in the prior distribution of the tolerance parameter. All reconstruction models, whether Bayesian or not, make assumptions on variables and have various biases. One goal of Bayesian modeling is to make such assumptions and biases explicit.

Recently, several papers using the Bayesian approach have appeared in the ecological literature (Gazey and Staley 1986, Carpenter 1990, Reckhow 1990, Solow 1994, Stow et al. 1995, West 1996, Robertson et al. 1999). Robertson et al. (1999) make the only attempt we are aware of to apply Bayesian approaches to palaeoenvironmental reconstruction. However, instead of explicitly modeling the possible responses, they use kernel-based density estimation. Kernel-based estimation is essentially an improved form of the modern analogue technique (MAT) (Bartlein and Whitlock 1993, Birks 1995, ter Braak 1995). Such estimation methods can be effective in interpolation with large amounts of data, but they are likely to perform weakly with small amounts of observations or at the extremes and in extrapolation.

Finally, although we aim at a plausible model, we

TABLE 3. Summary of differences between the more traditional frequentist reconstruction methods and the Bayesian modeling approach to reconstruction.

Traditional frequentist methods	Bayesian full-probability models
"True" parameter values	Parameter probabilities
Does not take advantage of existing knowledge	Prior knowledge (prior probability distributions) used
Maximum likelihood	Posterior distribution
Often needs deshrinking	No deshrinking
Stable process	Iterative process
Response variables continuous	No requirements on variable types
Restricted to a single source of information	Allows the combination of various types of information

realize that BUM is not completely ecologically realistic; it assumes the abundances of different taxa are a priori independent, although they are measured as proportions and therefore depend on each other. BUM is our first attempt at a Bayesian reconstruction model, and we are currently working on more realistic models.

Conclusion

We propose the use of Bayesian techniques in organism-based palaeoenvironmental reconstruction. We have outlined a Bayesian approach to reconstruction and presented a simple unimodal Bayesian model, BUM, in more detail. We have compared empirically the performance of the model to eight other techniques using the chironomid data set of Olander et al. (1999). Our results indicate that BUM can be applied for reconstruction purposes with at least the same degree of reliability as the commonly used methods.

However, we think that full probability models and Bayesian inference have two clear advantages over the more traditional analysis methods. First, in Bayesian statistics, it is natural to use the classical approach to regression and calibration. This allows explicit modeling of species responses to the environment. There is always uncertainty of how well a statistical model corresponds to nature, but transparent models give ecologists the opportunity to study and judge models, and to develop more realistic ones. Realistic models can be expected to extrapolate better than less plausible ones, which is important for reconstructions based on fossil data (ter Braak 1995).

Second, Bayesian analysis provides methods for solving statistical inverse problems such as regression of species responses. Further, using probability distributions throughout the inference rather than point estimates provides a principled way of handling uncertainty in the regression step and passing it on to the reconstructions. The use of probability distributions for variables can also be viewed as multiple working hypotheses (Ellison 1996).

Our future work on Bayesian reconstruction models will address the following issues. The BUM model will be improved so that it is less specific to the current data set and its properties. The assumption of independence between species abundances will be relaxed. The use of a taxonomical species hierarchy in the model is also worth considering, as is the application of nonsymmetric expected response curves and the employment of mixture models for more realistic modeling of zero abundances. As with any models, the Bayesian models will need to be validated on a number of data sets with different characteristics. Software for Bayesian reconstruction is also under development.

Acknowledgments

We are grateful to Arvid Odland for deriving the modern climate values for the 53 lakes and to Jan Weckström for help in data collection. Elja Arjas, John Birks, David M. Lodge, and anonymous reviewers provided constructive comments on manuscripts and helped improve this paper a lot.

The study was funded by the Academy of Finland, the European Community Environment and Climate Programme (CHILL-10,000 Project; Contract Number ENV4-CT97-0642), the Nordic Council of Ministers (Contract Number FS/HFj-X-96005); and the University of Helsinki. This paper is CHILL-10,000 contribution Number 8 and NORD-CHILL contribution Number 9.

Literature cited

Anderson, N. J. 1993. Natural versus anthropogenic change in lakes: the role of the sediment record. Trends in Ecology and Evolution **8**:356–361.

Bartlein, P. J., and C. Whitlock. 1993. Pages 275–293 *in* J. P. Bradbury and W. E. Dean, editors. Paleoclimatic interpretation of the Elk Lake pollen record. Geological Society of America Special Paper **276**.

Bennion, H., S. Juggins, and N. J. Anderson. 1996. Predicting epilimnetic phosphorus concentrations using an improved diatom-based transfer function and its application to lake eutrophication management. Environmental Science and Technology **30**:2004–2007.

Birks, H. J. B. 1995. Quantitative palaeoenvironmental reconstructions. Pages 161–254 *in* D. Maddy and J. S. Brew, editors. Statistical modelling of Quaternary science data. Technical Guide 5. Quaternary Research Association, Cambridge, UK.

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

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

Brooks, S. J., and H. J. B. Birks. 2000. Chironomid-inferred late-glacial and early-Holocene mean July air temperatures for Krøakenes Lake, western Norway. Journal of Paleolimnology **23**:77–89.

Carpenter, S. R. 1990. Large-scale perturbations: opportunities for innovation. Ecology **71**:2038–2043.

Charles, D. F., and J. P. Smol. 1994. Long-term chemical changes in lakes: quantitative inferences from biotic remains in the sediment record. Pages 3–31 *in* L. Baker, editor. Environmental chemistry of lakes and reservoirs. Advances in Chemistry Series 237. American Chemical Society, Washington, D.C., USA.

COHMAP [Cooperative Holocene Mapping Project] members. 1988. Climatic changes of the last 18,000 years: observations and simulations. Science **241**:1043–1052.

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

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications **6**:1036–1046.

Gazey, W. J., and M. J. Staley. 1986. Population estimation from mark–recapture experiments using a sequential Bayes algorithm. Ecology **67**:941–951.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman & Hall, London, UK.

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

Korhola, A., J. Weckström, L. Holmström, and P. Erästö. 2000. A quantitative Holocene climatic record from diatoms in northern Fennoscandia. Quaternary Research **54**: 284–294.

Korsman, T., and H. J. B. Birks. 1996. Diatom-based water chemistry reconstructions from northern Sweden: a com-

parison of reconstruction techniques. Journal of Paleolimnology **15**:65–77.

Levesque, A. J., L. C. Cwynar, and I. R. Walker. 1994. A multi-proxy investigation of late-glacial climate and vegetation change at Pine Ridge Point, southwest New Brunswick, Canada. Quaternary Research **42**:316–327.

Levesque, A. J., L. C. Cwynar, and I. R. Walker. 1997. Exceptionally steep north–south gradients in lake temperatures during the last deglaciation. Nature **385**:423–426.

Lotter, A. F., H. J. B. Birks, W. Hofmann, and A. Marchetto. 1997. Modern diatom, cladocera, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps. I. Climate. Journal of Paleolimnology **18**:395–420.

Ludwid, D. 1996. Uncertainty and the assessment of extinction probabilities. Ecological Applications **6**:1067–1076.

Martens, H., and T. Naes. 1989. Multivariate calibration. John Wiley & Sons, Chichester, UK.

Olander, H., H. J. B. Birks, A. Korhola, and T. Blom. 1999. An expanded calibration model for inferring lakewater and air temperatures from fossil chironomid assemblages in northern Fennoscandia. The Holocene **9**:279–294.

Olander, H., A. Korhola, and T. Blom. 1997. Surface sediment Chironomidae (Insecta: Diptera) distributions along an ecotonal transect in subarctic Fennoscandia: developing a tool for palaeotemperature reconstructions. Journal of Paleolimnology **18**:45–59.

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

Reckhow, K. H. 1990. Bayesian inference in nonreplicated ecological studies. Ecology **71**:2053–2059.

Robertson, I., D. Lucy, L. Baxter, A. M. Pollard, R. G. Aykroyd, A. C. Barker, A. H. C. Carter, V. R. Switsur, and J. S. Waterhouse. 1999. A kernel-based Bayesian approach to climatic reconstruction. The Holocene **9**:495–500.

Solow, A. R. 1994. On the Bayesian estimation of the number of species in a community. Ecology **75**:2139–2142.

Stow, C. A., S. R. Carpenter, and K. L. Cottinham. 1995. Resource vs. ratio-dependent consumer resource models: a Bayesian perspective. Ecology **76**:1986–1990.

ter Braak, C. J. F. 1988. CANOCO—a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1). Technical Report LWA-88-02. Agricultural Mathematics Group, Wageningen, The Netherlands.

ter Braak, C. J. F. 1995. Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (*k*-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. Chemometrics and Intelligent Laboratory Systems **28**:165–180.

ter Braak, C. J. F., and S. Juggins. 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., S. Juggins, H. J. B. Birks, and H. van der Voet. 1993. Weighted averaging partial least squares regression (WA-PLS): definition and comparison with other methods for species–environment calibration. Pages 525–560 *in* G.P. Patil and C.R. Rao, editors. Multivariate environmental statistics. Elsevier Science, Amsterdam, The Netherlands.

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

Toivonen, H., H. Mannila, J. Seppänen, and K. Vasko. 1999. Bassist user's guide. Technical Report C-1999-36. Department of Computer Science, University of Helsinki, Helsinki, Finland.

Walker, I. R., A. J. Levesque, L. C. Cwynar, and A. F. Lotter. 1997. An expanded surface-water palaeotemperature inference model for use with fossil midges from eastern Canada. Journal of Paleolimnology **18**:165–178.

Walker, I. R., J. P. Smol, D. R. Engstrom, and H. J. B. Birks. 1991. An assessment of Chironomidae as quantitative indicators of past climatic change. Canadian Journal of Fisheries and Aquatic Sciences **48**:975–987.

West, M. 1996. Some statistical issues in palaeoclimatology. Pages 461–484 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian Statistics 5. Oxford University Press, Oxford, UK.