A Bayesian palaeoenvironmental transfer function model for acidified lakes

Philip B. Holden, Anson W. Mackay and Gavin L. Simpson

Environmental Change Research Centre,

Department of Geography,

University College London,

Pearson Building,

Gower Street,

London WC1E 6BT

UK

pbholden@ntlworld.com, amackay@geog.ucl.ac.uk, gavin.simpson@ucl.ac.uk

Key words: Environmental reconstruction, Transfer functions, Bayesian model selection, Diatoms, Acidification.

Abstract

A Bayesian approach to palaeoecological environmental reconstruction deriving from the unimodal responses generally exhibited by organisms to an environmental gradient is described. The approach uses Bayesian model selection to calculate a collection of probability-weighted, species-specific response curves (SRCs) for each taxon within a training set, with an explicit treatment for zero abundances. These SRCs are used to reconstruct the environmental variable from sub-fossilised assemblages. The approach enables a substantial increase in computational efficiency (several orders of magnitude) over existing Bayesian methodologies.

The model is developed from the Surface Water Acidification Programme (SWAP) training set and is demonstrated to exhibit comparable predictive power to existing Weighted Averaging and Maximum Likelihood methodologies, though with improvements in bias; the additional explanatory power of the Bayesian approach lies in an explicit calculation of uncertainty for each individual reconstruction. The model is applied to reconstruct the Holocene acidification history of the Round Loch of Glenhead, including a reconstruction of recent recovery derived from sediment trap data.

The Bayesian reconstructions display similar trends to conventional (Weighted Averaging Partial Least Squares) reconstructions but provide a better reconstruction of extreme pH and are more sensitive to small changes in diatom assemblages. The validity of the posteriors as an apparently meaningful representation of assemblage-specific uncertainty and the high computational efficiency of the approach open up the possibility of highly constrained multiproxy reconstructions.

Introduction

In order to understand future environmental and climatic changes, it is necessary to understand how they may have changed in the past; the demand for such data is increasing to define the boundary conditions and validate the predictions of Earth System models (Birks 2003). Species require particular environmental conditions for reproduction and growth, so a species assemblage is likely to reflect the local environment. By applying the principal of Uniformitarinism (Rymer 1978), the analysis of sub-fossilised assemblages preserved in e.g. lake and ocean sediments can be a powerful tool to derive past conditions. Since the development of the first quantitative approach of Imbrie and Kipp (1971) who applied principal component regression to reconstruct sea surface temperatures and ocean salinity from foraminifera assemblages, a number of statistical techniques have been developed in order to more accurately quantify uncertainty and to take account of ecologically realistic responses of organisms to their environment (ter Braak et al. 1993).

There are two philosophically distinct approaches to statistical analysis. Conventional, or frequentist, statistics assumes that the parameters being estimated (the "model") are fixed and that measured data are random observations distributed about these values; conversely, Bayesian statistics assumes that the model is the unknown and it is the measured data which are fixed (Box and Tiao 1992, Dennis 1996). Almost all of the reconstruction methodologies routinely used by palaeolimnologists apply frequentist statistics. Although very powerful tools, the major weakness of these approaches is that they do not explicitly model the uncertainty associated with individual reconstructions,

but rather assume a dataset-specific RMSEP error calculated by cross-validation techniques (see e.g. Birks 1995). Sample-specific calculations of RMSEP are occasionally performed in an attempt to address this issue (e.g. Birks et al 1990). A Bayesian approach, in contrast, considers all possible solutions and ascribes a probability to each of them, thus calculating not only the most likely reconstruction but also the uncertainty associated with that reconstruction. Although natural variability limits reconstruction accuracy, so that alternative transfer functions generally provide broadly similar performance statistics for a given training set (see e.g. Birks et al. 1990), the additional explanatory power of a Bayesian approach lies in this explicit calculation of uncertainty.

Individual reconstructions may be associated with an uncertainty that differs greatly from that implied by RMSEP. If for example there is a strong representation of a few species in a "pioneering" late-glacial assemblage, the assemblage may have no modern analogues, leading to reduced confidence in the reconstruction (Birks 1998). Conversely, an assemblage in equilibrium with the local environment and dominated, for example, by species with narrow tolerances might be expected to provide a relatively precise solution. Although multivariate techniques and best analogue coefficients can provide a useful assessment of reliability (Birks et al. 1990), they do not quantify this uncertainty.

A number of groups have developed detailed Bayesian models to reconstruct climate from ecological proxies. The BUMMER model (Vasko et al., 2000, Korhola et al., 2002) has been developed for chironomid-based temperature reconstructions. The model assumes that the probability that a random individual from a site is of a particular

taxon is a random variable with a Dirichlet distribution derived from a multinomially distributed Gaussian response: a Bayesian analogue for the multinomial logit model (MLM) (ter Braak et al. 1993). The posterior is integrated with a Markov Chain Monte Carlo (MCMC) methodology. Haslett et al. (2006) extended these ideas and developed an approach similar in spirit to the response surface method (Huntley 1993) via the MCMC modelling of pollen response functions in two-dimensional climate space.

The method described here applies Bayesian model selection to derive a collection of probability-weighted, species-specific response curves for each taxon within a training set, with an explicit treatment for zero abundances. This enables an analytical solution to be derived from species abundance data, avoiding the need for MCMC integration, resulting in a very substantial (several orders of magnitude) increase in computational efficiency over the approaches of Vasko et al. (2002) and Haslett et al. (2006). The model is applied to the diatom-based reconstruction of the Holocene acidification history of the Round Loch of Glenhead, Scotland, including a reconstruction of recent recovery derived from sediment trap data. This site has been chosen because a comprehensive set of analyses already exist with which to compare our model development. The concepts described are far more widely applicable, both in terms of organism types and the environmental variables that can be reconstructed.

Methods

The n training set sites and the m taxa found within the training set sites are represented by the ($n \times m$) matrix of percentage abundances $\mathbf{Y}$, where y_{ik} represents the abundance of

taxon *k, k = 1,...,m* at site *i, i=1,...,n*. The measured environmental variable is represented by the matrix **X**, where x_i is the value in site *i* and $\hat{x}_i$ the reconstructed value. The percentage abundance of taxon *k* in the fossil sample is y_{k0} and the reconstructed value $\hat{x}_0$.

Probability distribution of species counts

The probability p_{ik} that taxon *k* is present in site *i* is assumed to follow a Gaussian distribution about some optimum value u_k of the environmental variable x_i (*c.f.* Kühl et al. 2002):

$$p_{ik} = p_k \exp\left\{-\left(x_i - u_k\right)^2 / 2\tau_k{}^2\right\} \qquad (1)$$

where p_k is the probability that the species is present at its optimum and tolerance τ_k is a measure of how far away from its optimum a species can survive. The probability that the species is absent is given by $(1 - p_{ik})$. Although alternative unimodal response curves could be fitted, a Gaussian model represents a compromise between ecological realism and simplicity (ter Braak and van Dam 1989). The Gaussian model does not preclude a uniform probability of presence; a hypothetical species found in all training set lakes would, for instance, be described with $p_k=1$ and $\tau_k \rightarrow \infty$, so that the probability of presence is 1 for all values of x_i.

If a species is present, the expected abundance N_{ik} is also assumed to follow a Gaussian distribution (*c.f.* Vasko et al. 2000) about the same optimum, although not necessarily with the same tolerance:

$$N_{ik} = N_k \exp\left\{-(x_i - u_k)^2 / 2t_k^2\right\} \quad (2)$$

The second measure of tolerance t_k describes of how far away from its optimum a species can exist at high abundance. N_k is the expected abundance (given presence) at the species optimum. Although it is not a requirement of the methodology to assume that p_{ik} and N_{ik} are both maximised at u_k, it seems reasonable to assume that any pH related process which maximises the probability of species presence is also likely to maximise the expected species count; analysis of the Surface Water Acidification Program training set (SWAP; Stevenson et al 1991) does not suggest this is a poor assumption.

As the expressions for p_{ik} and N_{ik} are of the same analytical form, Equation 1 can be written (using Equation 2) in terms of the squared ratio of the tolerances, $P_k = t_k^2/\tau_k^2$:

$$p_{ik} = p_k \left(N_{ik} / N_k\right)^{P_k} \quad (3)$$

This enables a convenient representation to couch the models in terms of parameters which are presumably independent (so that the joint prior distribution can be determined from the individual priors).

The variable P_k allows different distributions for N_{ik} and p_{ik}. Low values ($P_k < 1$) are required to model species which are common across the environmental gradient but

exhibit clear abundance peaks, suggesting that although they need near optimum conditions in order to flourish and dominate an assemblage, they can survive even when conditions are far removed from this optimum. In contrast, the need to allow high values ($P_k > 1$) is less clear, on the assumption that any pH related process which reduces the ability of a species to survive would also affect its ability to flourish in high numbers. Data for *Achnanthes marginulata* is plotted in Figure 1 as an example of a distribution described by $P_k < 1$. With a modelled optimum u_k=5.07 pH units, the taxon is found in 77% of the 117 SWAP lakes of pH <6 at an average abundance of 5.5% (maximum 46.9%). It is still found in 56% of the 50 lakes of pH >6, although at a much reduced average abundance of 0.9% (maximum 3.6%) i.e. while the probability of presence is not substantially reduced in the more alkaline lakes, the expected count is greatly reduced, implying $P_k < 1$ (so that $\tau_k > t_k$).

The probability of a non-zero count y_{ik} of species k in site i at a given value of x_i is assumed to follow an exponential decay, with decay constant $1/N_{ik}$ (from Equation 2), normalised so that the total probability of all non-zero counts is equal to the probability of presence p_{ik} (from Equation 3):

$$prob(y_{ik} \mid x_i) = (p_{ik} / N_{ik}) \exp(- y_{ik} / N_{ik}) \qquad y_{ik} > 0 \text{ (present)} \qquad (4a)$$

with the probability of a zero count given by;

$$prob(y_{ik} \mid x_i) = (1 - p_{ik}) \qquad y_{ik} = 0 \text{ (absent)} \qquad (4b)$$

Alternative distributions for non-zero counts are possible; the exponential profile is essentially empirical, having been found to provide the best fit SWAP as determined by Bayesian model selection (i.e. maximising the posterior ratio of alternative distributions). Alternative probability distributions considered were a Gaussian distribution about the expected abundance and a uniform distribution. The exponential distribution reflects the observation that most species counts are lower than the expected abundance, presumably because other variables and/or inter-species competition often limit species abundance, even near the pH optimum for the species (Lancaster and Belyea 2006). The exponential distribution may not necessarily provide the best model for all organism types and/or environmental variables.

The model does not enforce the constraint $\Sigma_{k=1,m}\, y_{ik}=1$ and the individual species counts are assumed to be independent. The average expected total, $1/n\ \Sigma_{i=1,n}\ \Sigma_{k=1,m}\, E(y_{ik})$, where $E(y_{ik})=p_{ik}N_{ik}$, in the SWAP lakes is 0.922 (i.e. 92.2% of the SWAP diatom counts are expected to be from the 225 included species). The actual percentage of SWAP counts that are from these 225 species is 91.6%, suggesting that the model performs realistically in this respect.

Calculation of Species Responses Curves (SRCs)

Each of the five SRC variables are discretised and form a collection of s Species Response Curves SRC_{jk} for each taxon k, where all combinations of the species-specific variables are represented by the index j, $j=1,\ldots,s$. The model here considers s=8,000 SRCs for each taxon, derived from a matrix of dimensions (20,4,5,5,4) for (u_k, N_k, t_k, p_k

and P_k) respectively. The *a priori* probabilities of each SRC are assumed equal and these are progressively refined using Bayes' Equation for each of the *n* species samples (or (*n-1*) samples in a jack-knifed calculation):

$$prob(SRC_{jk} \mid y_{ik}, x_i) \propto prob(y_{ik} \mid SRC_{jk}, x_i) \times prob(SRC_{jk} \mid x_i) \quad (5)$$

where $prob\ (y_{ik} \mid SRC_{jk}, x_i)$ is given by Equation 4a or 4b (depending upon whether the count is non-zero or zero), introducing the conditionality on SRC_{jk}.

SRC probabilities are normalised from the constraint $\Sigma_{j=1,s}\ prob\ (SRC_{jk} \mid \mathbf{Y}, \mathbf{X}) = 1$ so that a series of SRCs are ascribed to each species, each of different probability, which we write as $prob(SRC_{jk})$. The mechanics of this procedure are described in some detail in the Appendix.

Common species are well constrained by the training set and are associated with a few SRCs of high probability whereas rare species, especially those which do not exhibit a clear unimodal response, can be associated with many significant SRCs. We define "significant" for these purposes, somewhat arbitrarily, by a probability > 10% of the most likely SRC, but note that all SRCs are included in the reconstruction so this choice does not affect the calculation. Although conventional statistics quantify uncertainty through the standard errors of the regression coefficients for each taxon, these errors are not incorporated into the reconstruction. In contrast, by assigning a probability to each SRC and using them all in the reconstruction, the uncertainty, in particular the uncertainty associated with rare taxa, is explicitly incorporated into the Bayesian reconstruction.

Reconstruction

A likelihood function for x_0, $L_y(x_0|y_{k0})$, given a count of y_{k0} of species k is constructed from all SRCs, weighted according to their relative probability:

$$L_y(x_0 \mid y_{k0}) \propto \sum_j \left\{ prob(SRC_{jk}) \times prob(y_{ko} \mid SRC_{jk}, x_0) \right\} \quad (6)$$

An alternative likelihood function $L_p(x_0|presence_{k0})$ is also calculated. This ignores the species count, assuming the species presence alone provides a valid, although less constrained, solution:

$$L_p(x_0 \mid presence_{k0}) \propto \sum_j \left\{ prob(SRC_{jk}) \times prob(presence_{k0} \mid SRC_{jk}, x_0) \right\} \quad (7)$$

where $prob(presence_{k0}|SRC_{jk}, x_0)$ is given by Equation 3, introducing the conditionality on SRC_{jk}.

Although the likelihood function of Equation 6 may be justifiable, a more conservative form for the likelihood function is derived by combining Equations 6 and 7:

$$L(x_0 \mid y_{k0}) = (1-\eta) L_y(x_0 \mid y_{k0}) + \eta L_p(x_0 \mid y_{k0}) \quad (8)$$

where $0 \le \eta \le 1$. A value of η=0.5 is assumed for the base case analysis presented here. This allows the reconstruction to be dominated by L_y, but broadens the likelihood function to allow for the possibility of outlying species counts. RMSEP was found to be only weakly

dependent upon η, suggesting that presence/absence data alone contains sufficient information to derive a useful predictive model.

The likelihood function that derives from the species count (Equation 6) can produce bimodal likelihood functions (when both $y_{ik}<N_k$ *and* $P_k<1$, where the low count suggests environmental conditions are likely to be away from the species optimum). This is illustrated in Figure 1c for *Achnanthes marginulata* and reflects the increased frequency of low counts towards the tails of the distribution (Figure 1b). In contrast, the presence-absence likelihood function (Equation 7) is always unimodal; presence always implies optimum conditions are most likely in this simpler model. Although the likelihood function derived from a single SRC (whether unimodal or bimodal) is always symmetrical, asymmetric likelihood functions arise in the full calculation which combines all possible SRCs. This is apparent in Figure 1c; species response for pH<4.5 is not well defined by the training set and an asymmetric likelihood function which allows for the possibility of a low pH results, even when the species is observed at high abundance.

The likelihood functions of the species comprising an assemblage are combined to give the posterior probability distribution for the reconstructed variable:

$$prob(x_0 \mid assemblage) \propto prob(x_0) \times \prod_{k=1,m} L(x_0 \mid y_{k0}) \qquad (9)$$

This is calculated across the environmental range and normalised. The term $prob(x_0)$ is the *a priori* probability distribution for the environmental variable. A uniform prior between the limits 3→9pH units is applied in the calculations described here; these limits

are sufficiently distant that the prior for lake pH has no impact upon the solution. Only species with an abundance ≥2% are included in the reconstructions presented here (although all species counts, including zero counts, are incorporated into the SRC calculation). The inclusion of species below 2% does not improve performance statistics in this data-set, presumably because very low abundance species exhibit broad likelihood functions (see e.g. Figure 1c) which contribute relatively little to the solution, and additionally their inclusion significantly increases computational demands. Furthermore, the absence of an explicit error structure in the model is likely to limit the reliability of likelihood functions derived from very low counts. The inclusion of very low count and possibly absent species, which would necessitate the incorporation of a binomial error structure, may be more useful in species-poor assemblages

A calculation on a modern PC requires approximately 10 minutes to derive the 8,000 SRC probabilities for each of the 225 species using the SWAP training set (167 lakes) and perform a 101 data point core reconstruction. Although computationally expensive, the model compares favourably with BUMMER (Vasko et al. 2000) which requires 1-2 days for a 150 data point reconstruction with a 63 lake, 52 taxa training set. The computational demands of BUMMER scale linearly with the number of sites and quadratically with the number of taxa, suggesting a 100 data-point core reconstruction using SWAP would require approximately 1 month of CPU time on a (2000) PC. The additional demands of BUMMER can presumably be attributed largely to the additional complexity associated with the multinomial assumption and MCMC integration; the model described here assumes (less robustly) that species counts are independent.

Alternative reconstruction methodologies

The Bayesian reconstructions are compared with several alternative methodologies (see e.g. Birks 1995): weighted averaging with classical deshrinking (WA Cla) and inverse deshrinking (WA Inv), weighted average partial least squares (WA-PLS) and Gaussian logit regression / maximum likelihood (GLR/ML). These models were developed using the C^2 software (Juggins 2003). The 1st component WA-PLS model (equivalent to WA Inv) was selected as the minimum adequate model (ter Braak and Juggins 1993).

Data Set

A substantial benefit of European acid rain research has been the development of the high quality SWAP training set (Stevenson et al. 1991), in part achieved through the use of taxonomic workshops which resolved a number of problems associated with differing nomenclature, splitting/ amalgamation of species and identification criteria (Munro et al. 1990). The full SWAP training set consists of 267 taxa present in surface sediment samples from 178 European lakes. Approximately 500 counts per sample were made. This data set was screened to derive a pruned training set of 167 lakes (Birks et al. 1990), outliers being removed using multivariate techniques or if the error of prediction was greater than 0.75pH units in weighted averaging both with and without tolerance downweighting. pH explains 8.1% of the total variance in the diatom assemblages (Birks 1994). The pruned SWAP training set was used to generate the model and derive performance statistics.

The model is applied to reconstructions of acidification in the Round Loch of Glenhead which has played a key role in acidification studies for 25 years (Battarbee et al. 2005). Since 1988, when the Loch was included in the Acid Waters Monitoring Network (AWMN; Monteith and Evans 2005), it has been closely monitored for both biology and chemistry. The loch is naturally acidic (Jones et al. 1989) but suffered post-industrial acidification (Flower and Battarbee 1983) to a minimum pH of ~4.7 in the 1980's (Flower et al 1987). Since EU directives limiting S and N emissions, the loch has exhibited some recovery to its current annual pH of ~5.2; in the last 2-3 years the pH has varied between about 5.5 in September and 5.1 during winter months (Monteith pers comm.). A long term reconstruction is performed on core RLGH3 (Jones et al. 1989), spanning the entire Holocene, dated from a combination of ^{210}Pb and ^{14}C measurements (Jones et al. 1989). Reconstructions are also performed on sediment trap data taken since 1991 (Battarbee et al. 2005) and on surface sediment assemblages from the K05 core (Allott et al. 1992) which was taken to investigate recovery from acidification.

Model Evaluation

The five SRC variables are each assigned uniform priors with limits provided in Table 1. The most important prior is that of the SRC optima; mathematically valid but ecologically unrealistic SRCs exist well beyond the limits of the training set. Vasko et al. (2000) assumed a normally distributed prior for u_k centred upon the observed modern mean with the observed modern variance. The *a priori* assumption made here is a uniform probability for u_k within the range $(x_{min}–x_r)$ to $(x_{max}+x_r)$, where x_{min} and x_{max} are

the extremes of the training set data and x_r is some constant. Although species-specific priors derived from consideration of large diatom data-sets such as the European Diatom Database (EDDI) (Battarbee et al. 2001) might enable a more robust form for the u_k priors, a global value of x_r=0.5 pH units was selected as the model exhibits negligible systematic bias with this prior. This value is thus assumed to approximate the transition from a regime in which species optima are ecologically realistic but over-constrained by the environmental range of the training set to a regime in which the solutions are not constrained but may be ecologically unrealistic. Detrended Canonical Correspondence Analysis (DCCA) with pH as the sole constraint reveals a gradient length of 2.56 over the pH range (2.92) of the training set, indicating a species turnover of ~1.1 pH units. Optima >~0.5 beyond the extremes of the environmental gradient are therefore presumably unlikely for species that exhibit clear maxima and high abundances within the training set.

The model exhibits little dependence on the tolerance prior, provided tolerances t_k as low as ~0.6 pH units are allowed. A conservative minimum value for t_k=*0.4* was applied here; although lower values reproduce the SWAP training set equally well, they allow low tolerances for rare species which may not be justified and hence may produce erroneous reconstructions on fossilised data. The performance of tolerance downweighted models shows substantial improvements when species with narrow tolerances are ascribed a minimum value of 0.1×environmental gradient (~0.3pH units), preventing very high tolerance weights for rare taxa (Köster et al. 2004).

The performance characteristics of the base model, which allows 5 values for P_k = *0.4→1.0* (allowing tolerance τ_k = *0.4→2.7*), are very similar to those of a model which

allows 40 values in the range $P_k = 0.0 \rightarrow 4.0$ (allowing tolerance $\tau_k = 0.2 \rightarrow \infty$). This suggests that there is little, if any, merit to the more demanding model and that the decision to restrict $P_k \leq 1$ on ecological grounds has not impaired model performance. In fact, a model which restricts P_k to a single value also exhibits similar performance statistics when optimised at $P_k \sim 0.7$. Although the model does not appear to be sensitive to the exact form of the P_k prior, further investigation of different proxies and/or environmental variables may enable a fuller understanding of the role of this variable.

The Bayesian approach does not define a specific predicted value, but rather ascribes a probability to all possible values. For comparative purposes it is useful to define the point prediction as the expectation of the posterior:

$$\hat{x} = \int x \; prob(x) dx \tag{10}$$

The model is evaluated in terms of five performance criteria: (i) root mean squared error (RMSE), (ii) root mean squared error of prediction (RMSEP), obtained by leave-one-out (jack-knifed) cross validation (Efron 1983) and a more realistic measure of performance than RMSE, (iii) coefficient of determination (r^2) between measured and jack-knifed reconstructed pH, (iv) maximum bias, calculated as the maximum of the average jack-knifed bias within 10 equal intervals across the environmental gradient (ter Braak and Juggins 1993) and (v) the linear least squares error slope parameter (LLSESP) between the jack-knifed residuals and the observed values (Vasko et al. 2000). More computer intensive error measures, such as bootstrapping (Efron 1983), have not been considered due to the high computational demands of the approach; the precise choice of

cross-validation method may be of less importance in a Bayesian approach as the statistic is not used to define reconstruction error. Performance characteristics are generally similar to existing methods (see Table 2), although an improvement in systematic bias is apparent. Although this reduction in systematic bias derives in part from the choice of the SRC optima prior, all reasonable priors produce a model with low LLSESP (below ~0.05); the approach is an example of classical calibration (as are WA Cla and GLR/ML) and as such is expected to exhibit reduced systematic bias at the expense of slightly higher RMSEP (ter Braak 1995). The jack-knifed Bayesian point predictions and residuals are plotted against measured pH in Figure 2.

A measure of the uncertainty implied by the posterior is given by:

$$\Delta = \sqrt{\int (x - \bar{x})^2 \, prob(x) dx} \qquad (11)$$

analogous to the calculation of RMSEP. The posteriors (though not the individual likelihood functions) approximate well to a Gaussian in these calculations, so that $\pm\Delta$ approximates the 68% confidence level and $\pm 2\Delta$ the 95% confidence interval; this is largely a consequence of the species-rich diatom assemblages and should not be assumed to be generally the case.

Taking $\eta = 0.5$ (Equation 8), the average jack-knifed posterior width $\overline{\Delta}$ of the SWAP lakes is 0.311 pH units (compared to RMSEP = 0.328 pH units) with individual posterior widths ranging from 0.175 to 0.554pH units (*c.f.* WA Cla sample-specific bootstrapped error which varies from 0.314 to 0.376, Birks et al 1990). With this value for η, 68.3% of the training set lakes have a measured pH that lies within the 68%

confidence interval of their respective jack-knifed posterior distributions; 92.2% of the training set lakes have a measured pH that lies within the 95% confidence interval. These figures suggest that the posteriors do meaningfully quantify the uncertainty of the reconstruction. 65.9% of the lakes have a measured pH in the range $\bar{x} \pm \Delta$ and 92.2% in the range $\bar{x} \pm 2\Delta$, demonstrating that (reconstruction-specific) Δ is a useful measure of the uncertainty. Although the calculation of uncertainty appears to slightly underestimate the predictive error at the 95% level, the comparison is between predicted pH and measured pH and both of these terms are associated with uncertainty; it is well known that RMSEP is likely to overstate predictive error for this reason (ter Braak and van Dam 1989).

Taking $\eta = 0.0$ reduces $\bar{\Delta}$ to 0.232 pH units. With this value for η, only 52.1% of the training set lakes have a measured pH that lies within the 68% confidence interval of their respective jack-knifed posterior distributions, with 79.6% predicted to the 95% confidence interval. Although these figures may appear to suggest that the model performs better with $\eta = 0.5$, again this conclusion may not be valid due to pH measurement errors. An assumed average error $\bar{\Delta}_m = \pm 0.23$ pH units in the measurement of pH would be sufficient to reconcile the apparent failure of the model with $\eta = 0.0$ (assuming $RMSEP^2 = \bar{\Delta}^2 + \bar{\Delta}_m^2$). The choice of $\eta = 0.5$ for the calculations described here is conservative, ascribing almost all of the error implied by RMSEP to the reconstruction.

The jack-knifed reconstructions were separated into two subsets with $\Delta<0.3$ (87 lakes) and $\Delta>0.3$ (80 lakes). Figure 3 compares the residual histograms for the two subsets with the expected residual distribution (as defined by the average Δ). These

comparisons clearly demonstrate the relationship between broad posteriors and large residuals, although some caution should be exercised, especially with more tightly constrained solutions; 3 of the 87 lakes with $\Delta<0.3$ pH units are not predicted to the 99% probability interval.

The Round Loch of Glenhead Reconstructions

Figures 4a and 4b compare WA-PLS1 and Bayesian point predictions for the RLGH3 core, with upper and lower bounds defined by WA-PLS1 sample-specific RMSEP and Bayesian posterior width Δ. The trends in the two reconstructions are strikingly similar. The gradual acidification of the loch during the early Holocene (due to the development of organic soils) is apparent in both, as are the rapid fluctuations between 4,000 and 2,000 cal BP, thought to be a result of the spread of blanket mire and declining tree cover at ~4,000 cal BP and the erosion of peat at ~3,000 cal BP (Jones et al. 1989). Both reconstructions exhibit a rapid post-industrial pH decline to levels unprecedented in the Holocene.

The Bayesian reconstruction of pre-acidification baseline pH (~5.6) is comparable to the WA based reconstructions of Battarbee et al. (2005) and does not help reconcile differences with numerical simulations of historical surface water chemistry using MAGIC (Cosby et al 2001), although the improved quantification of the discrepancy may be of use: the Bayesian posterior ascribes a probability of 7.4% to a pH equal to or greater than the MAGIC prediction of 6.1pH units.

The RLGH3 assemblage-specific Bayesian Δ is plotted illustrated in Figure 4c, and suggests far greater variability in uncertainty (0.231 to 0.475pH units) than sample-

specific WA-PLS bootstrapping (0.312 to 0.319pH units). A sample-specific range (0.314 to 0.322pH units) was derived in the WA Cla RLGH3 reconstructions of Birks et al (1990). Although it may not be appropriate to assume a causal relationship, it is interesting to note that the two periods of greatest Bayesian uncertainty are also periods associated with environmental and consequent catchment instability; the late glacial/early Holocene (rapid climate change) and ~3,000BP (changing vegetation). The Bayesian temperature reconstructions of Korhola et al. (2002) also indicated large uncertainties during the early Holocene which they associated with chironomid assemblages that were not in equilibrium with the local environment as a consequence of catchment instability and high erosion rates. The period of post-industrial acidification is not associated with increased uncertainty. The uncertainty does not exhibit any appreciable increasing trend at depth, despite the fact that assemblages below 48.5cm (~1,200 cal BP) all lack close modern analogues (Birks et al. 1990).

Surface sediment and sediment trap reconstructions are illustrated in Figure 5. The Bayesian surface sediment reconstruction of 4.66 pH units is consistent with pre-AWMN measurements which ranged from 4.50 to 4.86 (Flower et al. 1987) and improves on both the WA inverse calculations (~4.9) of Battarbee et al. (2005) and, to a lesser extent, the WA classical calculations (~4.75) of Birks et al. (1990). This is presumably a reflection of lower systematic bias and consequently improved performance at the extremes of the environmental gradient. The earliest diatom assemblages in the sediment traps are almost identical to those in the uppermost sediment, and are generally representative of the epilithic flora (Jones and Flower 1986). Although the changes in species composition of sediment trap assemblages from 1991 to 2004 are not statistically

significant as a whole (Monteith et al. 2005), they are sufficient to produce a significant (>Δ) shift in the Bayesian reconstructed pH. The principal changes are a progressive reduction in the relative abundances of *Tabellaria quadriseptata* and *Navicula cumbriensis* and progressive increases in *Navicula leptostriata*, *Frustulia rhomboides var. viridula* and *Eunotia vanheurckii var. intermedia*. These changes are all indicative of increasing pH and result in Bayesian reconstructed values increasing by 0.36pH units (*c.f.* Δ_{2004}=0.27pH units) from 4.66 to 5.02 in 2004; this recovery is only weakly observed in WA reconstructions to 2002 (Battarbee et al. 2005). The recovery displayed by the Bayesian reconstructions is delayed with respect to the measurements; it is not possible to say whether this results from the finite response time of diatom assemblages to their environment or simply reflects inaccuracies in the reconstruction as there is in each case a substantial overlap between the posterior and the measured pH range.

Discussion

Although the improvements in bias are welcome, the primary motivation for the development of a Bayesian model lies in the quantification of reconstruction uncertainty, and the potential applications that this may enable. The validity of the posterior width as a measure of reconstruction uncertainty has been demonstrated in Figure 3. The principal requirements for an assemblage that minimises reconstruction uncertainty can be summarised:

i) The ecological response of common species are the most well defined as the model selection process is highly constrained, resulting in fewer significant SRCs. Species with

few significant SRCs provide more precise reconstruction information, manifested by narrow likelihood functions. Conversely, rare species are characterised by many possible SRCs, particularly when they do not exhibit clear unimodal responses, and produce broader likelihood functions which contribute less information to the posterior.

ii) Species that are present at *relatively* high abundances are suggestive of near-optimum conditions and strongly constrain the solution. Some species are only ever found at low abundances in the training set and, as such, even a low count of such species may produce a narrow likelihood function. WA-based approaches may not adequately capture this effect for these "low count" species.

iii) Species-rich assemblages provide the most precise solutions, especially when low tolerance species are present. This is not in conflict with Racca et al. (2002) who concluded that 85% of species can be removed without a reduction in predictive power and that tolerance is not a good criterion for species inclusion. Intelligent pruning might well give the same result for this model, although it would inevitably result in poorly constrained posteriors which would not provide a realistic measure of the uncertainty implied by the entire assemblage.

Two of the training set lakes have reconstructed values that differ by substantially more than RMSEP when evaluated using alternative methods: Loch Doon (WA Inv 5.46, GLR/ML 4.91) and Hagasjon (WA-PLS1 6.95, GLR/ML 7.40). These lakes both exhibit broad posteriors (Δ = 0.497 and 0.493pH units respectively) again suggesting that broad posteriors are associated with sites that are less reliably reconstructed. The uncertainty associated with the core reconstructions does not appear to increase for poor analogue assemblages. This is a surprising result and suggests the need for further analysis of the

reconstructions for poor analogue sites within the training set. In contrast, periods of environmental instability do appear to be associated with high uncertainty although it may not be appropriate to assume a causal relationship at this stage.

The power that derives from an explicit calculation of uncertainty is the ability to combine reconstructions from independent data, in particular from independent proxies to derive a multiproxy reconstruction. Although smoothing techniques such as LOESS (Cleveland 1979) can be applied to combine frequentist reconstructions derived from different proxies (Birks and Birks 2003), the absence of an explicit uncertainty limits the reliability of such an approach. Haslett et al. (2006) utilise this power of a Bayesian methodology in a different way by modelling temporal autocorrelation through a core. These authors note the potential, in theory at least, for a Bayesian approach to perform a self-consistent, multiproxy, multi-core analysis, modelling climate as a stochastically structured space-time process, whilst simultaneously resolving dating uncertainties.

A secondary benefit of the Bayesian approach is the transparency of the solutions. Each species can be removed from the reconstruction in turn and the contribution of that species to the posterior quantified in terms of both its contribution to the expected pH and to the uncertainty. Figure 6 plots the data for the uncommon (observed in 12 of the 167 SWAP lakes) species *Aulacoseira ambigua* which dominates the assemblage of Hagasjon; the poorly characterised response of the dominant taxon in the assemblage explains the large uncertainty associated with Hagasjon and the very different reconstruction values derived from alternative methodologies. Fig 6a plots the training set species counts, the black point being the count in Hagasjon, and the curve the N_{ik} distribution defined by the most probable jack-knifed SRC. Very many significant SRCs

(126, as defined by a probability >10% of the most probable SRC) produce the relatively broad likelihood function (given the very high abundance) that is plotted as the dashed line in Fig 6b; there are few counts of the species so the SRCs are poorly constrained. The solid black line in Fig 6b plots the jack-knifed posterior for the lake, and the solid grey line the posterior that would be derived if *Aulacoseira ambigua* were not included; the difference between these curves provides a quantifiable representation of the contribution of the species to both the value and the precision of the reconstruction.

The Bayesian reconstructions are generally similar to those derived from WA-based approaches. In particular, the trends displayed by the full Holocene reconstruction are very similar to the WA-PLS1 reconstruction, although there is some evidence that the Bayesian approach is better able to reconstruct extreme values as a consequence of improved bias. The recent pH recovery (Figure 5) is only weakly apparent in the WA reconstructions (Battarbee et al. 2005); the improvements here can be attributed to the increased sensitivity of the Bayesian approach to "low count" species. WA estimates are dominated by the highest abundance species, in this case *Eunotia incisa* and *Frustulia rhomboides* var. *saxonica*. The species that most tightly constrain the Bayesian reconstruction are those which have high abundances relative to *expected* abundance. In addition to the dominant species, the "low count" species *Eunotia vanheurckii* var. *intermedia* and *Frustulia rhomboides* var. *viridula* are also present at *relatively* high abundances (5.5% and 3.0% respectively in 2004) and both significantly shift the posterior to higher pH.

Conclusions

A Bayesian model for the reconstruction of surface water acidification has been developed which displays comparable performance to existing methodologies, although with improvements in systematic bias. The Bayesian reconstructions of the Round Loch of Glenhead display similar trends to WA-PLS reconstructions but the Bayesian approach reconstructs extreme pH more accurately and is more sensitive to small changes in the diatom assemblage. The transparency of the Bayesian solutions reveal that this increased sensitivity derives from "low-count" species which are only ever found at low abundances (and hence have little effect on WA-based reconstructions) but can result in narrow Bayesian likelihood functions when *relatively* high counts are made.

The posteriors appear to meaningfully quantify assemblage-specific uncertainty, although some caution should be exercised with very tightly constrained solutions; 3 from 87 lakes with $\Delta<0.3$ pH units are not predicted to the 99% probability interval in jack-knifed reconstructions. The Bayesian calculations here exhibit substantially greater variability in reconstruction uncertainty than is implied by sample-specific RMSEP.

The incorporation of a binomial error structure into the model may improve performance by enabling the incorporation of very low abundance, and possibly absent, species into the reconstruction, although substantial improvements are presumably unlikely for species-rich (e.g. diatom) assemblages. Any performance improvements would need to be weighed against the inevitable reduction in computational efficiency.

The model is relatively simple, although apparently adequate, and substantial increases in computational efficiency are displayed over existing Bayesian approaches,

opening up the possibility for application to multiproxy studies. Independent proxies provide independent reconstructions of an environmental variable; the Bayesian approach enables a quantitative evaluation of the consistency of these different reconstructions, potentially assisting the validation of palaeoreconstructions in general and helping to identify potential problems that may be associated with migrational lags, disequilibriated assemblages, taphonomy, taxonomy or evolution of species response. Independent reconstructions can be combined into a single multiproxy reconstruction; this is, in theory at least, more robustly achievable in a Bayesian framework due to the assemblage-specific probability distributions associated with each proxy. Several (potentially species-poor) assemblages could be combined to yield a single "organism-rich" assemblage and provide a very well constrained solution. Holocene temperature reconstructions are notoriously difficult as variability generally lies within reconstruction errors (Korhola et al. 2002); the ability to further constrain reconstructions through the combination of several proxies suggests a potentially fruitful way forward.

Acknowledgements

We are grateful to Devinder Sivia, John Birks and Richard Telford for useful discussions. We are additionally grateful for the constructive comments of both referees which have substantially improved the paper.

Appendix

The calculation of SRC probabilities and reconstruction methodology is demonstrated in the following example, which considers a single species with two possible SRCs derived from a training set of three lakes. The calculation is directly analogous to the full calculation which considers 8,000 SRCs derived from 167 SWAP training set counts for each of 225 diatom species.

Training set pH is measured at x_1 = 4.5, x_2 = 5.5 and x_3 = 6.5pH units and the taxon k has fractional abundances of y_{1k} = 5, y_{2k} = 20 and y_{3k} = 0. For the purpose of the illustration, it is assumed that the only variable not known with certainty is the species optimum u_k. SRC_{1k} variables are arbitrarily assigned the values u^1_k = 5.0pH units, N^1_k = 15, t^1_k =1.2pH units, p^1_k=50% and P^1_k=0.7 (introducing the superscript to represent the SRC index j). SRC_{2k} variables are identical with the exception u^2_k = 6.0pH units. The two SRCs are assumed to have equal *apriori* probabilities of 0.5. We might conclude qualitatively that the SRC with the lower optimum (SRC_{1k}) is more likely due to the presence of the taxon in the 4.5pH lake; the approach enables a quantification of the relative probabilities of the two optima.

The expected fractional abundance (given presence) of taxon k in lake 1 given SRC_{1k} is given by (Equation 2):

$$N^1_{1k} = N^1_k \exp\left\{-\left(x_1 - u^1_k\right)^2 / 2t^{1^2}_k\right\} = 15 \times \exp\left\{-\left(4.5 - 5.0\right)^2 / \left(2 \times 1.2^2\right)\right\} = 13.75$$

The probability of occurrence of taxon k in site 1 given SRC_{1k} is given by (Equation 3):

$$p_{1k}^1 = p_k^1 \left(N_{1k}^1 / N_k^1 \right)^{p_k^1} = 50\% \times \left(13.75/15 \right)^{0.7} = 47.05\%$$

The probability of the measured abundance y_{1k} (given SRC_{1k} and x_1) is thus given by (Equation 4a):

$$prob(y_{1k} \mid SRC_{1k}, x_1) = \left(p_{1k}^1 / N_{1k}^1 \right) \exp\left\{ - y_{1k} / N_{1k}^1 \right\} = \left(47.05/13.75 \right) \times \exp\left\{ - (5/13.75) \right\} = 2.38\%$$

For the second lake, a similar calculation gives $prob(y_{2k} \mid SRC_{1k}, x_2) = 0.80\%$. The first two steps of the calculation are the same for the third lake, but the zero count requires Equation 4b to be used, giving $prob(y_{3k} \mid SRC_{1k}, x_3) = (1-p^1_{3k}) = 71.05\%$. Bayes' Equation (Equation 5) is used to combine the data from the three lakes to derive an expression for the probability of SRC_{1k} given the training set:

$$prob(SRC_{1k} \mid \mathbf{Y}, \mathbf{X}) = C \times prob(y_{1k} \mid SRC_{1k}, x_1) \times prob(y_{2k} \mid SRC_{1k}, x_2) \times prob(y_{3k} \mid SRC_{1k}, x_1)$$

$$= C \times 2.38\% \times 0.80\% \times 71.05\% = 1.35 \times 10^{-4} C$$

where the normalisation constant C has been introduced (and incorporates the uniform prior of 0.5).

An identically analogous calculation gives $prob(SRC_{2k} \mid \mathbf{Y}, \mathbf{X}) = 8.61 \times 10^{-5} C$. As the two SRCs are the only two possibilities allowed in this hypothetical example, their

probabilities sum to unity and the normalisation constant C is defined (i.e. $1.35 \times 10^{-4} C + 8.61 \times 10^{-5} C = 1$) giving posterior probabilities for the two optima of 61.1% and 38.9% respectively. Note: the probabilities of the two optima are a function of the values assigned to the other four SRC variables; for instance, a tolerance assumption t_k = 0.75pH units reduces the probability of u_k=6.0pH units to 12.6% as presence in the acidic lake suggests u_k=6.0pH is far less probable under an assumption of narrow tolerance. The full calculation allows all five SRC variables to vary simultaneously.

The reconstruction is illustrated considering a fossilised sediment sample with y_{0k}=10. The likelihood function $L_y(x_0|y_{k0})$ for x_0=5.0pH units is given by Equation 6:

$$\begin{aligned} L_y(x_0 = 5.0 \mid y_{0k} = 10.0) &= C' \{ prob(SRC_{1k}) \times prob(x_0 = 5.0 \mid SRC_{1k}, y_{0k} = 10.0) + \\ &\quad prob(SRC_{2k}) \times prob(x_0 = 5.0 \mid SRC_{2k}, y_{0k} = 10.0) \} \\ &= C' \{ 0.611 \times 0.0171 + 0.389 \times 0.0144 \} = 0.016 C' \end{aligned}$$

This calculation is performed for all possible pH values and the normalisation constant C' calculated to define $L_y(x_0|y_{k0})$ across the environmental gradient. Under these highly illustrative assumptions, the likelihood function peaks at 5.41pH units. The assemblage reconstruction is performed by combining the likelihood functions of all species in the fossilised sample (which may include both L_y and L_p terms, combined according to Equation 8), together with the assumed prior for lake pH according to Equation 9, and normalising the solution.

Please contact the authors for a copy of the FORTRAN source code.

References

Allott TEH, Harriman R, Battarbee RW (1992) Reversibility of lake acidification at the Round Loch of Glenhead, Galloway, Scotland. Environmental Pollution 77: 219-225

Battarbee RW, Juggins S, Gasse F, Anderson NJ, Bennion H, Cameron NG, Ryves DB, Pailles C, Chalie F, Telford R (2001) European Diatom Database (EDDI). An information system for palaeoenvironmental reconstruction. ECRC Research Report No. 81, 94 pp.

Battarbee RW, Monteith DT, Juggins S, Evans CD, Jenkins A, Simpson GL (2005) Reconstructing pre-acidification pH for an acidified Scottish loch: A comparison of palaeolimnological and modelling approaches. Environmental Pollution 137: 135-150

Birks HJB (1994) The importance of pollen and diatom taxonomic precision in quantitative palaeoenvironmental reconstructions. Review of Palaeobotany and Palynology 83: 107-117

Birks HJB (1995) Quantitative palaeoenvironmental reconstructions. In: Maddy D and Brew JS (eds) Statistical Modelling of Quaternary Science Data. Technical guide 5. Quaternary Research Association, Cambridge, pp 161-254

Birks HJB (1998) Numerical tools in palaeolimnology: progress, potentialities and problems. Journal of Paleolimnology 20: 307-332

Birks HJB (2003) Quantitative palaeoenvironmental reconstructions from Holocene biological data. In: Mackay AW, Battarbee RW, Birks HJB and Oldfield F (eds) Global change in the Holocene. Hodder Arnold, New York, pp 107-123.

Birks HH, Birks HJB (2003) Reconstructing Holocene climates from pollen and plant macrofossils. In: Mackay AW, Battarbee RW, Birks HJB and Oldfield F (eds) Global change in the Holocene. Hodder Arnold, New York, pp 342-357.

Birks HJB, Line JM, Juggins S, Stevenson AC, ter Braak CJF (1990) Diatoms and pH reconstruction. Philosophical Transactions of the Royal Society of London 327: 263-278

Box CEP, Tiao GC (1992) Bayesian Inference in Statistical Analysis. Wiley-Interscience, New York, 608pp.

Cosby BJ, Ferrier RC, Jenkins A, Wright RF (2001) Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. Hydrology and Earth System Sciences 5: 499-517

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74: 829-836

Dennis B (1996) Discussion: should ecologists become Bayesians? Ecological Applications 6:1095-1103

Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association 78: 316-330

Flower RJ, Battarbee RW (1983) Diatom evidence for the acidification of two Scottish Lochs. Nature 305: 130-133

Flower RJ, Battarbee RW, Appleby PB (1987) The recent palaeolimnology of acid lakes in Galloway, south-west Scotland. Diatom analysis, pH trends, and the role of afforestation. Journal of Ecology 75: 797-824

Haslett J, Whiley M, Bhattacharya S, Salter-Townshend M, Wilson SP, Allen JRM, Huntley B, Mitchell FJG (2006) Bayesian palaeoclimate reconstruction. Journal of the Royal Statistical Society A 169: 395-438

Huntley B (1993) The use of climate response surfaces to reconstruct palaeoclimate from quaternary pollen and plant macrofossil data. Philosophical Transactions of the Royal Society of London B 341: 215-223

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

Jones VJ, Flower RJ (1986) Spatial and temporal variability in periphytic diatom communities: palaeoecological significance in an acidified lake. In: Smol JP, Battarbee RW, Davis RB, Meriläinen J (eds) Diatoms and Lake Acidity. Dr W. Junk Publishers, Dordrecht, pp 87-94

Jones VJ, Stevenson AC, Battarbee RW (1989) Acidification of lakes in Galloway, south west Scotland – a diatom and pollen study of the post-glacial history of the Round Loch of Glenhead. Journal of Ecology 77: 1-23

Juggins S (2003) C^2 user guide. Software for ecological and palaeoecological data analysis and visualisation. University of Newcastle, Newcastle upon Tyne, UK, 69pp

Korhola A, Vasko K, Toivonen HTT, Olander H (2002) Holocene temperature changes in northern Fennoscadia reconstructed from chironomids using Bayesian modeling. Quaternary Science Reviews 21: 1841-1860

Köster D, Racca JMJ, Pienitz R (2004) Diatom-based inference models and reconstruction revisited: methods and transformation. Journal of Paleolimnology 32: 233-246

Kühl N, Gebhardt C, Litt T, Hense A (2002) Probability Density Functions as botanical-climatological transfer functions for climate reconstruction. Quaternary Research 58: 381-392

Lancaster J and Belyea LR (2006) Defining the limits to local density: alternative views of abundance-environment relationships. Freshwater Biology 51:783-796

Monteith DT, Evans CD (2005) The United Kingdom Acid Waters Monitoring Network: a review of the first 15 years and introduction to the special issue. Environmental Pollution 137: 3-13

Monteith DT, Hildrew AG, Flower RJ, Raven PJ, Beaumont WRB, Collen P, Kreiser AM, Shilland EM, Winterbottom JH (2005) Biological responses to the chemical recovery of acidified freshwaters in the UK. Environmental Pollution 137: 83-102

Munro MAR, Kreiser AM, Battarbee RW, Juggins S, Stevenson AC, Anderson DS, Anderson NJ, Berge F, Birks HJB, Davis RB, Flower RJ, Fritz SC, Haworth EY, Jones VJ, Kingston JC and Renberg I (1990) Diatom quality control and data handling. Philosophical Transactions of the Royal Society of London B 327: 257-261

Racca JMJ, Wild M, Birks HJB, Prairie YT (2002) Separating wheat from chaff: Diatom taxon selection using an artificial neural network pruning algorithm. Journal of Paleolimnology 29: 123-133

Rymer L (1978) The use of uniformitariansim and analogy in palaeoecology, particularly pollen analysis. In: D Walker and JC Guppy (eds) Biology and Quaternary Environments. Australian Academy of Sciences, Canberra, pp 245-258

Stevenson AC, Juggins S, Birks HJB, Anderson DS, Anderson NJ, Battarbee RW, Berge F, Davis RB, Flower RJ, Haworth EY, Jones VJ, Kingston VJ, Kreiser AM, Line JM, Munro MAR, Renberg I (1991) The Surface Waters Acidification Project palaeolimnology program: modern diatom / lake-water chemistry set. ENSIS, London, 86 pp.

ter Braak CJF (1995) Non linear models for multivariate statistical calibration and their use in palaeoecology; a comparison of inverse k-nearest neighhbours, partial least squares and weighted averaging partial least squares, and classical approaches. Chemometrics and Intelligent Laboratory Systems. 28: 165-180

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

ter Braak CJF and van Dam H (1989) Inferring pH from diatoms: a comparison of old and new calibration methods. Hydrobiologia 178: 209-233

ter Braak CJF, Juggins S, Birks HJB and van der Voet H (1993) Weighted averaging partial least squares regression (WA-PLS): Definition and comparison with other methods for species-environment calibration. In: GP Patil and CR Roa (eds) Multivariate Environmental Statistics. Elsevier Science Publishers, Amsterdam, pp 525-560.

Vasko K, Toivonen HTT, Korhola A (2000) A Bayesian multinomial response model for organism-based environmental reconstruction. Journal of Paleolimnology 24: 243-250

Warren Spring Laboratory (1983) Acid Deposition in the United Kingdom. Warren Spring Laboratory, Stevenage, 72 pp.

Wright RF, Emmett BA, Jenkins A (1998) Acid deposition, land use change and global change: MAGIC 7 model applied to Aber, UK (NITREX project) and Risdalsheia, Norway (RAIN and CLIMEX projects). Hydrology and Earth System Sciences 2: 385-397

Figure Captions

Figure 1 Example species *Achnanthes marginulata* in the SWAP training set. a) Calculated probability of presence p_{ik} compared to observed percentage presence. b) Expected count N_{ik} (given presence) compared to observed counts. c) Example likelihood functions L_y. illustrating that low counts produce broader likelihood functions and can be bimodal (when $P_k<1$).

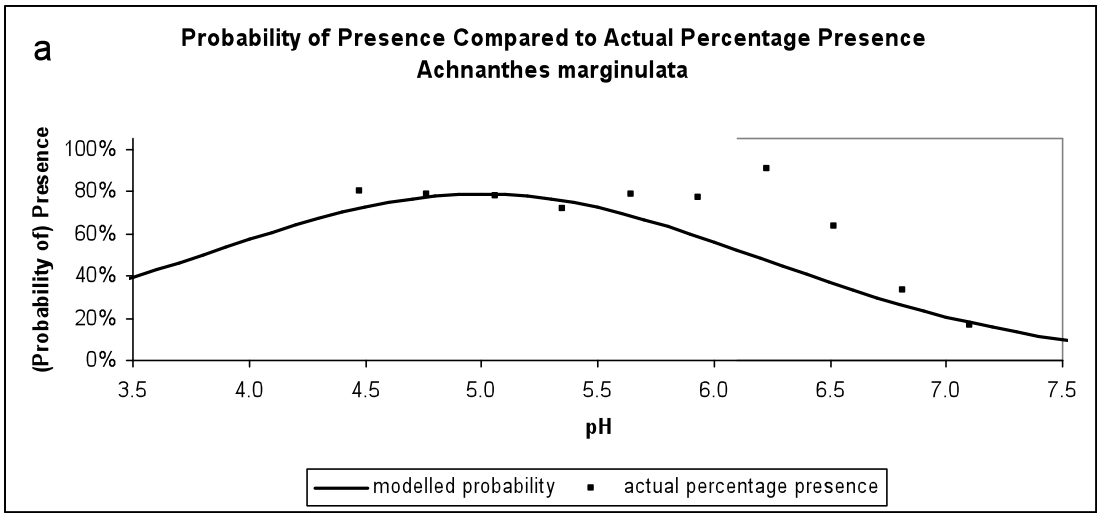
Figure 2 Plots of a) the relationship between observed and jack-knifed diatom-inferred lake pH and b) the jack-knifed residuals and their linear least squares fit.

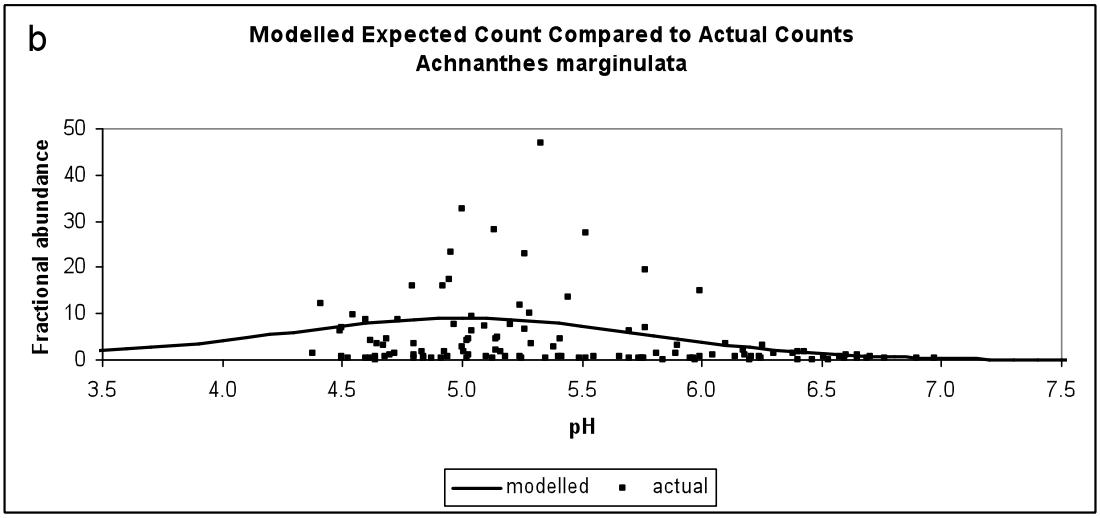
Figure 3 Residual histograms for lakes with a) $\Delta<0.3$ pH units (87 lakes) and b) $\Delta>0.3$ pH units (80 lakes). The curves are the expected residual distributions as defined by the average Δ of lakes within the subset. A clear relationship between broad posteriors and large residuals is apparent.

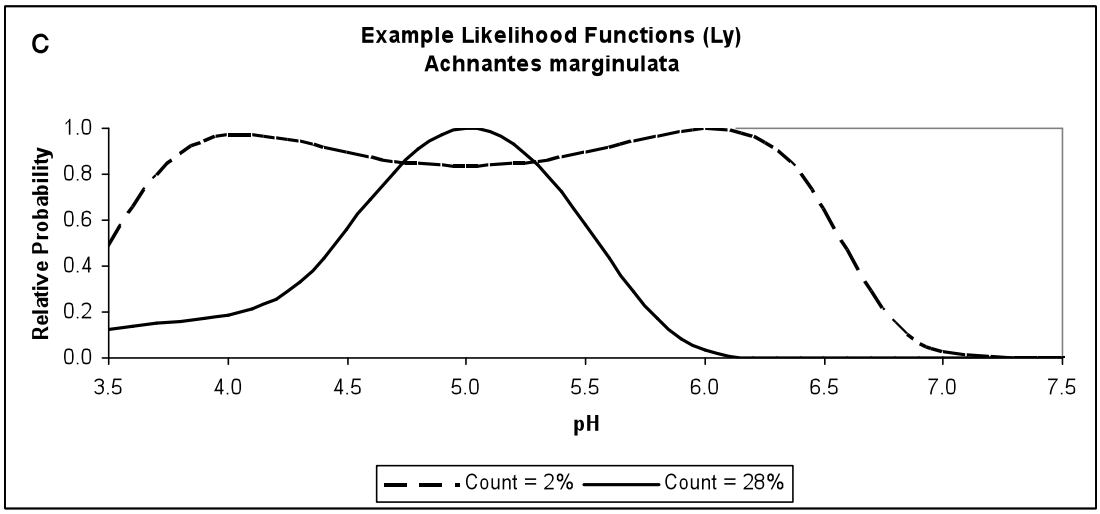
Figure 4 Holocene pH reconstructions of the Round Loch of Glenhead from the RLGH3 core (Jones et al. 1989). a) WA-PLS1 reconstructions (upper and lower bounds defined by sample-specific RMSEP), b) Bayesian point predictions (upper and lower bounds defined by posterior width Δ), and c) Bayesian posterior width Δ. Variations in the WA-PLS1 sample-specific error are almost negligible, with values ranging from 0.312 to 0.319pH units.

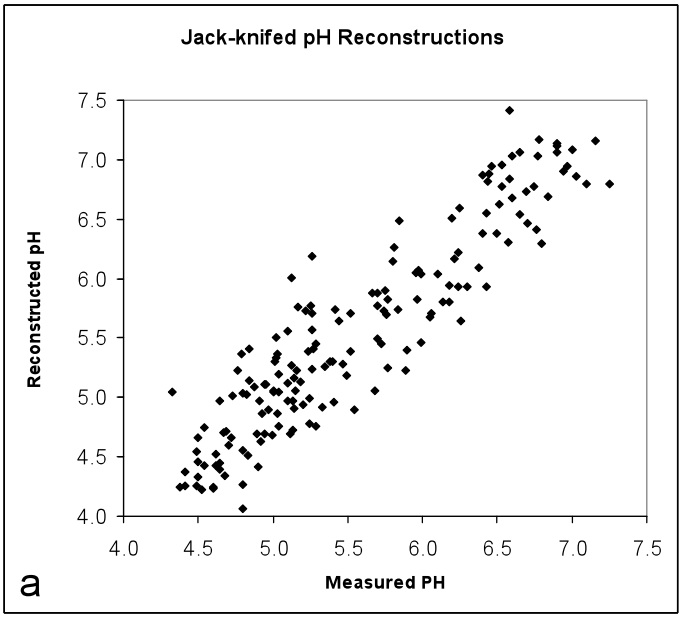
Figure 5 Comparison between measured and Bayesian reconstructed pH since 1979. Curves represent time integrated Bayesian posteriors from a) 1979-1989 K05 surface sediment (Allott at al. 1992) and b,c,d,e) 1991-2004 sediment trap data (Battarbee et al. 2005, Monteith pers. comm.). Horizontal lines represent the range of measured values from a) 1980-1981 (Flower et al. 1987) (data not available for entire period) and b,c,d,e) 1991-2004 (Battarbee et al. 2005, Monteith pers. comm.). The recovery from acidification is apparent in the reconstructions, though apparently delayed with respect to the measured recovery.

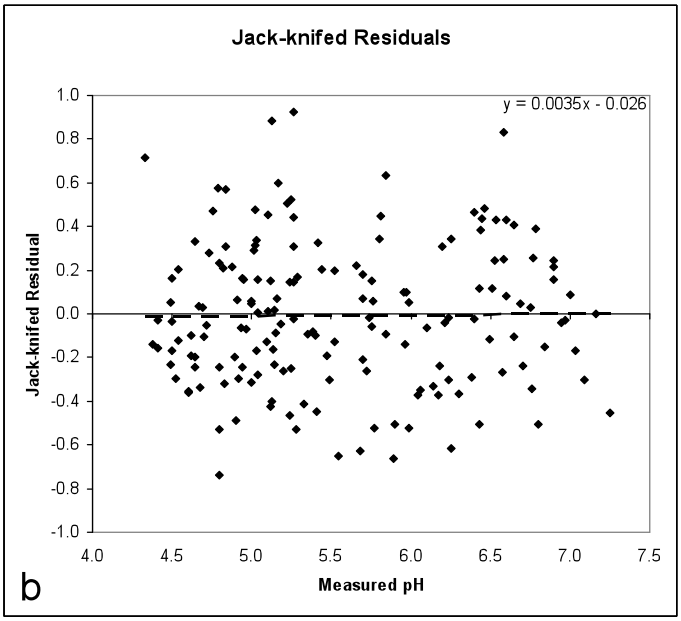
Figure 6 Illustrative pH reconstruction for Hagasjon, focussing on the uncommon species *Aulacoseira ambigua* (present in 12 of 167 SWAP sites) which dominates the assemblage. a) SWAP data for *Aulacoseira ambigua* in Hagasjon (filled square) and in other training set sites (open squares). The N_{ik} distribution for the most probable *jack-knifed* SRC is plotted as the black curve. b) Vertical black line: measured pH in Hagasjon. Dashed curve: likelihood function for *Aulacoseira ambigua* in Hagasjon. Black curve: pH posterior derived from the Hagasjon diatom assemblage. Grey curve: pH posterior if *Aulacoseira ambigua* is not included in the reconstruction.

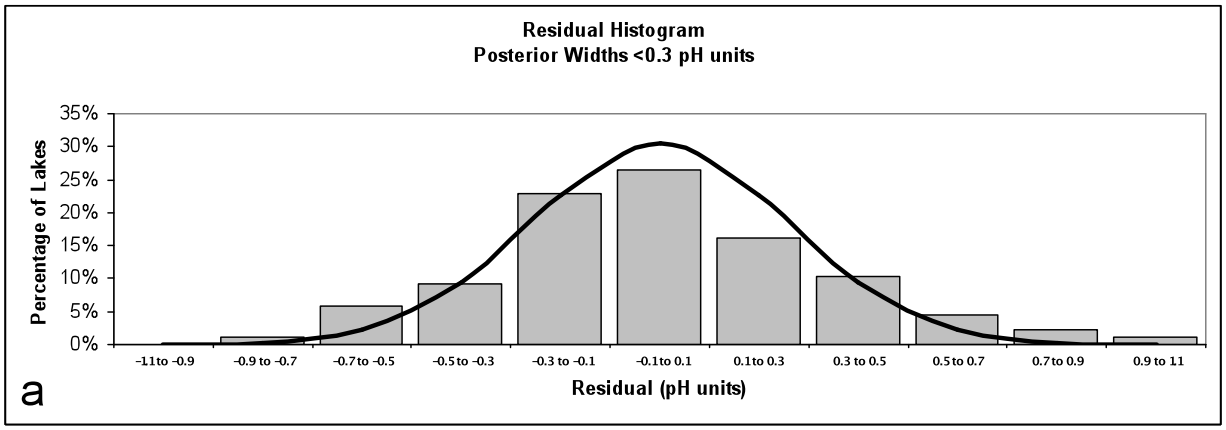

a
Probability of Presence Compared to Actual Percentage Presence
Achnanthes marginulata
(Probability of) Presence
100%
80%
60%
40%
20%
0%
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
pH
modelled probability
actual percentage presence

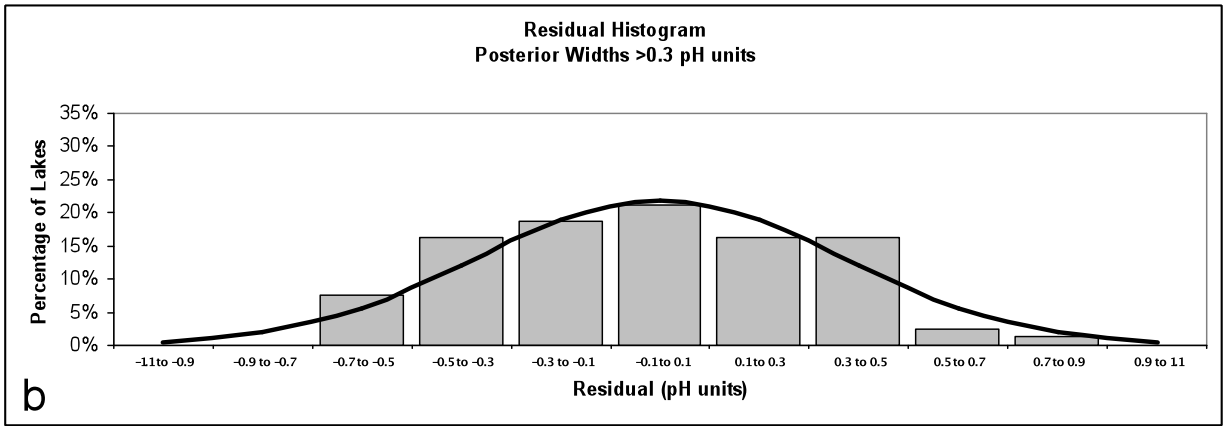

b
Modelled Expected Count Compared to Actual Counts
Achnanthes marginulata
Fractional abundance
50
40
30
20
10
0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
pH
modelled
actual

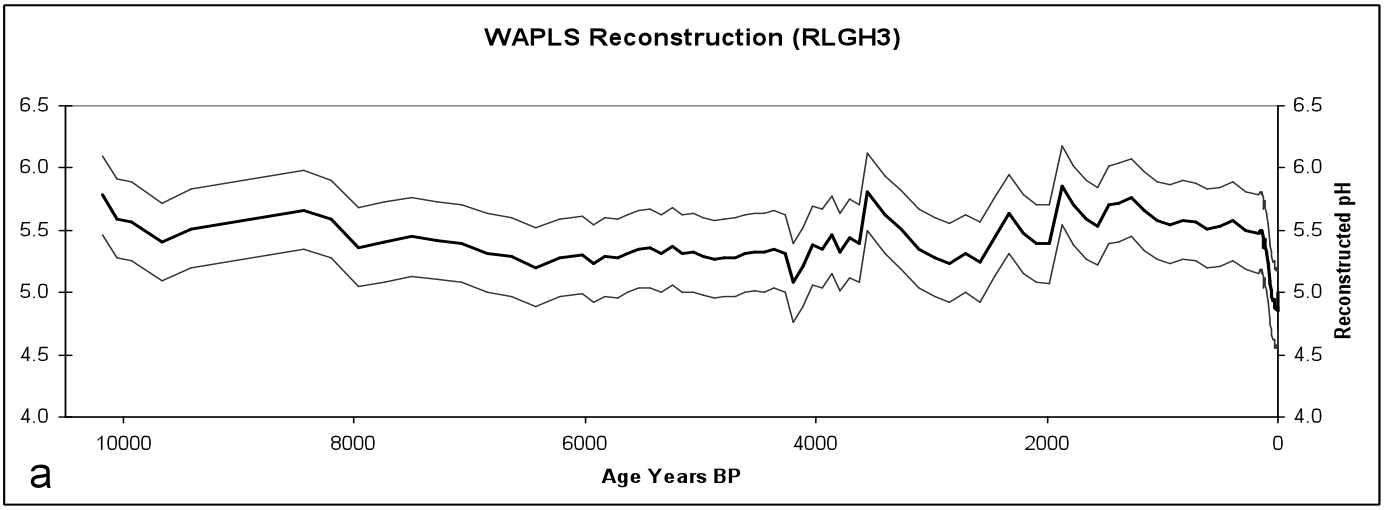

c
Example Likelihood Functions (Ly)
Achnantes marginulata
Relative Probability
1.0
0.8
0.6
0.4
0.2
0.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
pH
Count = 2%
Count = 28%

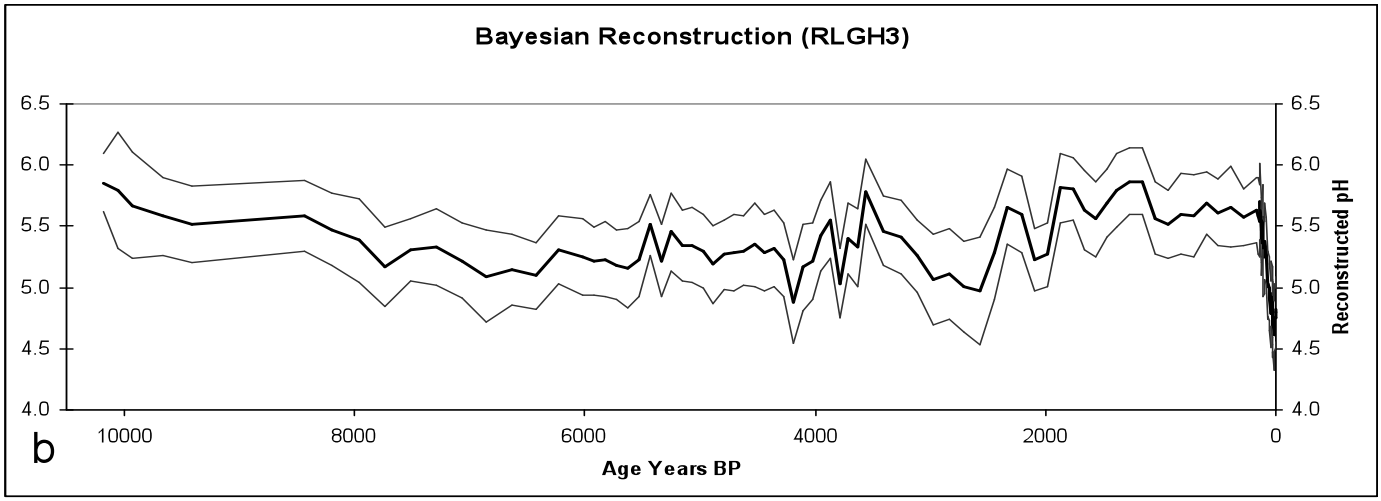

Jack-knifed pH Reconstructions
Reconstructed pH
Measured PH
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
a

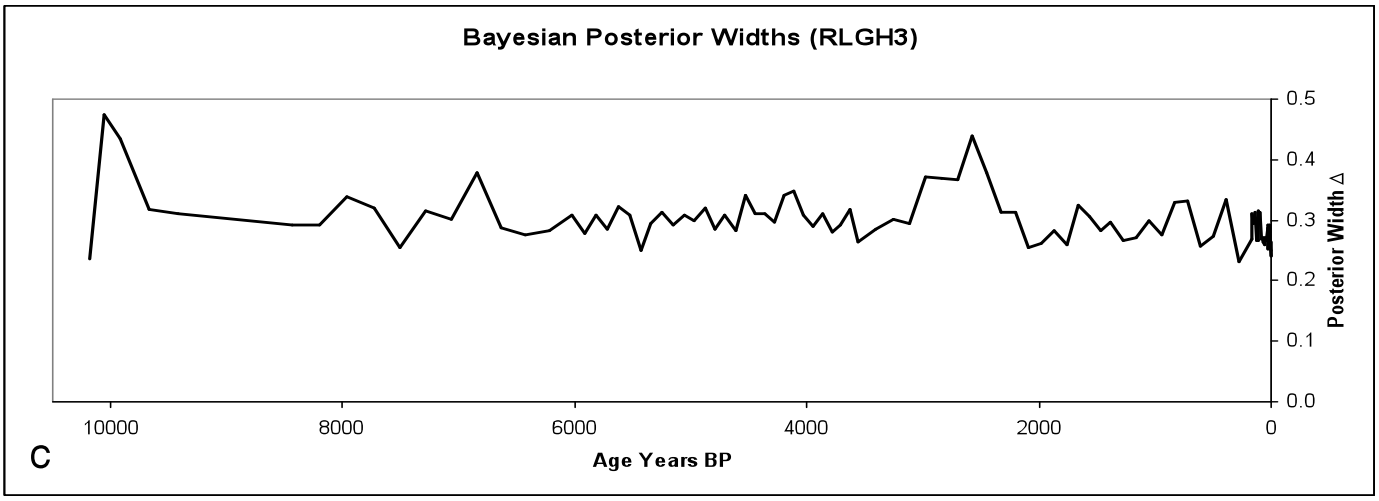

Jack-knifed Residuals
y = 0.0035x - 0.026
Jack-knifed Residual
Measured pH
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
b

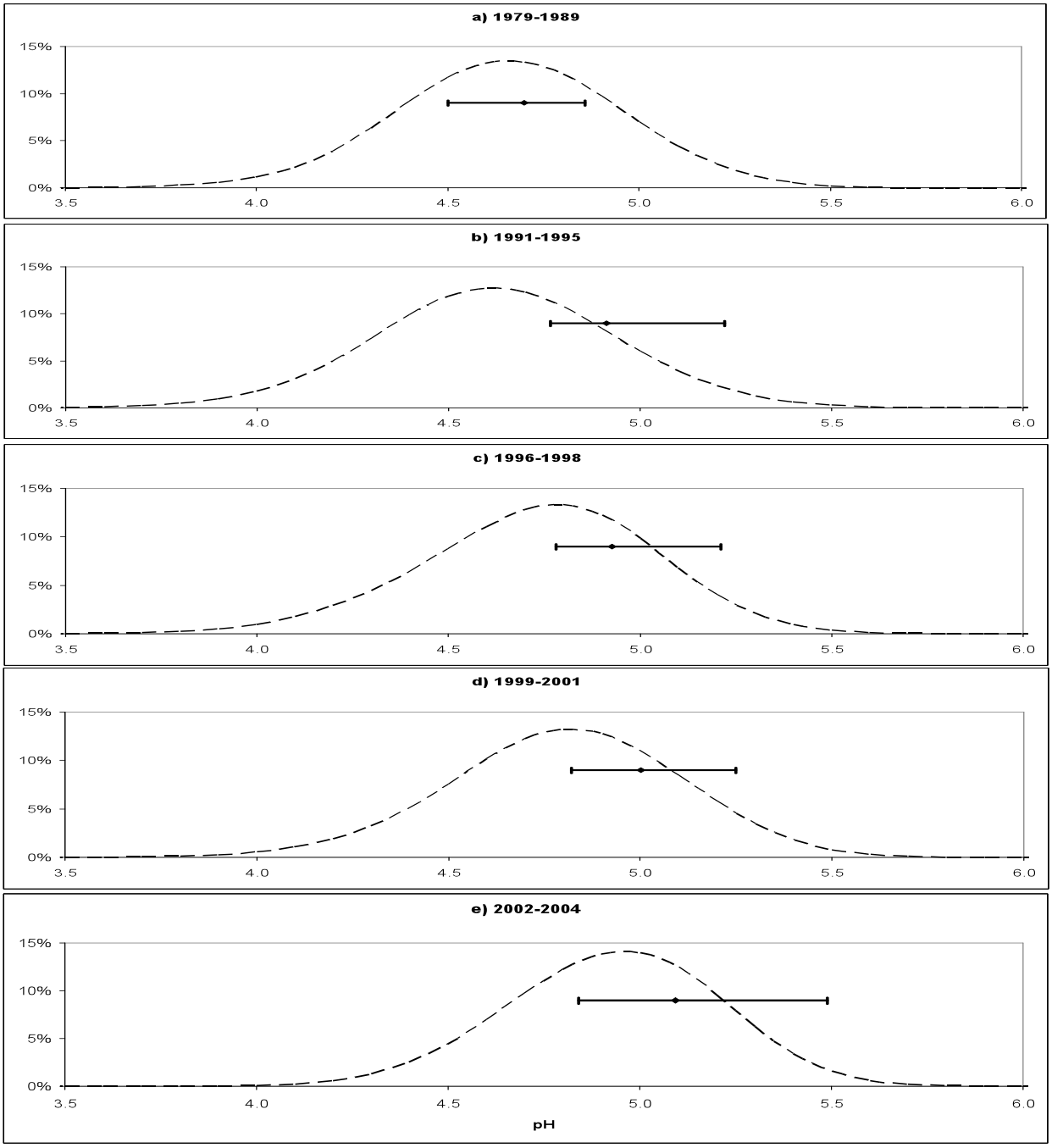

Residual Histogram
Posterior Widths <0.3 pH units
Percentage of Lakes
35%
30%
25%
20%
15%
10%
5%
0%
-1.1 to -0.9
-0.9 to -0.7
-0.7 to -0.5
-0.5 to -0.3
-0.3 to -0.1
-0.1 to 0.1
0.1 to 0.3
0.3 to 0.5
0.5 to 0.7
0.7 to 0.9
0.9 to 1.1
Residual (pH units)
a

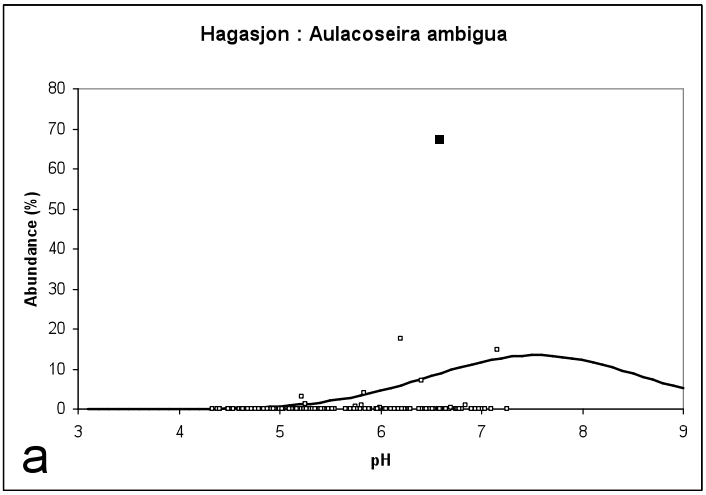

Residual Histogram
Posterior Widths >0.3 pH units
Percentage of Lakes
35%
30%
25%
20%
15%
10%
5%
0%
-1.1 to -0.9
-0.9 to -0.7
-0.7 to -0.5
-0.5 to -0.3
-0.3 to -0.1
-0.1 to 0.1
0.1 to 0.3
0.3 to 0.5
0.5 to 0.7
0.7 to 0.9
0.9 to 1.1
Residual (pH units)
b


WAPLS Reconstruction (RLGH3)
6.5
6.0
5.5
5.0
4.5
4.0
10000
8000
6000
4000
2000
0
Age Years BP
Reconstructed pH
a


Bayesian Reconstruction (RLGH3)
6.5
6.0
5.5
5.0
4.5
4.0
10000
8000
6000
4000
2000
0
Age Years BP
Reconstructed pH
b


Bayesian Posterior Widths (RLGH3)
0.5
0.4
0.3
0.2
0.1
0.0
10000
8000
6000
4000
2000
0
Age Years BP
Posterior Width Δ
c


a) 1979-1989
b) 1991-1995
c) 1996-1998
d) 1999-2001
e) 2002-2004
15%
10%
5%
0%
3.5
4.0
4.5
5.0
5.5
6.0
pH


Hagasjon : Aulacoseira ambigua
Abundance (%)
80
70
60
50
40
30
20
10
0
3
4
5
6
7
8
9
pH
a

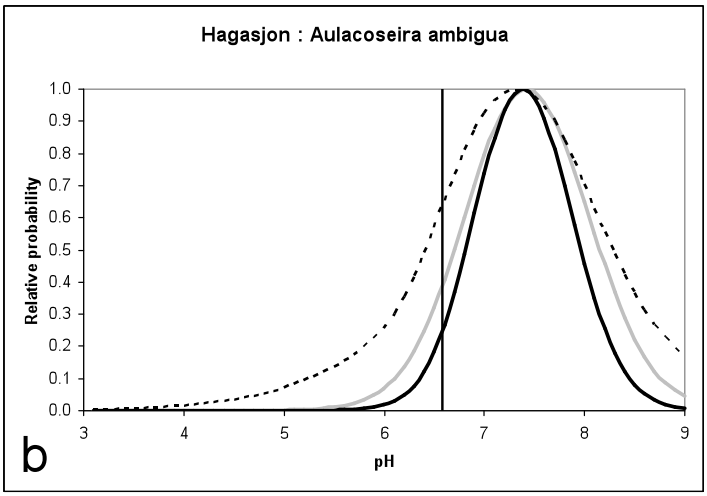

Hagasjon : Aulacoseira ambigua
Relative probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
pH
b

Table 1 SRC variables: minimum and maximum values allowed for each variable and the number of values ("resolution") tested at even intervals across this range. See Equations 1-3 for a description of the SRC variables. x_{min} and x_{max} are the extremes of the training set pH measurements, $N_{k\ max}$ is the maximum percentage abundance of a given species within the training set, $\%occ_k$ is the percentage of training set sites in which a given species is present. These values define 8,000 (20x4x5x5x4) SRCs for each species. These SRCs are assigned an equal *a priori* probability which is refined by Bayes' Equation for each training set count (including zero counts).

SRC Variable	Minimum	Maximum	Resolution
u_k	$x_{min} - 0.5 = 3.83$	$x_{max} + 0.5 = 7.75$	20
N_k	$0.2\ N_{k\ max}$	$N_{k\ max}$	4
t_k	0.4	1.7	5
P_k	0.4	1.0	5
p_k	$1.0 \times \%occ_k$	$2.5 \times \%occ_k$	4

Table 2 Comparison of performance statistics for various models applied to the SWAP training set. WA Inv: weighted averaging with inverse deshrinking. WA Cla: weighted averaging with classical deshrinking. WA-PLS1: weighted averaging partial least squares 1st component. GLR/ML: Gaussian logit regression/maximum likelihood. For description of these models see e.g. Birks (1995). See main text for description of performance characteristics.

	RMSE	RMSEP	R^2	Maximum Bias	LLSESP
WA Inv	0.276	0.307	84.2%	0.278	-0.180
WA Cla	0.295	0.317	84.3%	0.169	-0.058
WA-PLS1	0.276	0.310	83.9%	0.322	-0.187
GLR/ML	0.274	0.334	82.5%	0.235	-0.072
Bayesian	0.295	0.328	84.7%	0.180	0.004