THE ROYAL SOCIETY

Received 26 May 2003
Accepted 3 February 2004
Published online 19 April 2004

A method for reconstructing climate from fossil beetle assemblages

Amit Huppert and Andrew R. Solow*

Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Fossil beetle remains have been used to reconstruct temperatures. One method by which these reconstructions are made—the Mutual Climatic Range method—is based on the overlap of the observed modern climatic ranges of the beetles present in a fossil sample. A limitation of this method is that it does not exploit variations in the rate of occurrence of a species within its climatic range. We present an alternative method that uses observed variations in this rate in modern data for climate reconstruction. The method is shown to perform well in an experiment using modern data from North America.

Keywords: beetles; calibration; climate reconstruction; logistic regression; Mutual Climatic Range

1. INTRODUCTION

Fossil beetle assemblages have been used for more than 25 years for reconstructing past temperature conditions. Notable contributions include Coope (1977) and Atkinson *et al.* (1987). Elias (1994) provided a good overview. One of the leading methods by which these reconstructions are made is the Mutual Climatic Range (MCR) method. Briefly, under this method, reconstructing the climatic conditions associated with a fossil beetle assemblage involves three steps. In the first step, modern distributional and climatic data are used to determine the climatic ranges or envelopes of the species present in the assemblage. These climate envelopes are typically two-dimensional, one dimension being the mean temperature of the warmest month (TMAX) and the other being either the mean temperature of the coldest month (TMIN) or the difference between TMAX and TMIN (TRANGE). In the second step, the climatic conditions are reconstructed from the overlap of these climatic envelopes. This overlap itself contains a range of climatic conditions. In the third step, this range is converted to single values of TMAX and TMIN (or TRANGE). This conversion is based on linear regression models relating observed modern values of these variables to the corresponding midpoints of the ranges found by applying the first two steps of the MCR method to the modern data.

A limitation of the MCR method is that it does not exploit information contained in the modern data about variations in the rate of occurrence of beetle species within their climatic ranges. The purpose of this paper is to describe and illustrate an alternative method that does exploit this information. The method uses the modern data to fit statistical models of the dependence of the probability of occurrence of each species on TMAX and TMIN. The fitted models are then used to reconstruct these variables from the species present in the fossil assemblage. We present here some results that show that this method works well.

The remainder of this paper is organized in the following way. The method is described in § 2. The results of applying this method to a large North American dataset are presented in § 3. Some concluding remarks are given in § 4.

Before proceeding, the following remark is pertinent. Despite repeated attempts, we have been unable to acquire data to permit a direct comparison of the results of our method with the results of the MCR method. We recognize the value of such a comparison and we regret the unavailability of such data.

2. MATERIAL AND METHODS

Consider a single location j and let the binary random variable $Y_{jk} = 1$ if species k is present, and 0 otherwise. Under our basic model,

$$\text{prob}(Y_{jk} = 1) = p_k(x_j), \tag{2.1}$$

where $p_k(x)$ is an unknown function taking values between 0 and 1 and x_j is the value of a climate variable at location j. Initially, we will focus on the case of a single such variable. We will refer to this as the univariate case and, for concreteness we will refer to this climate variable as temperature. Let s be the set of species present in a fossil assemblage. The problem considered in this paper is to use s to reconstruct the temperature x_0 associated with this assemblage. We describe a two-step approach to this problem. In the first step, the function $p_k(x)$ is estimated from the modern data for each of the species in s. We will refer to this as the regression step. In the second step, the fitted functions are used to reconstruct x_0 from s. We will refer to this as the calibration step.

For species k, the function $p_k(x)$ can be estimated by quadratic logistic regression. Under this model,

$$\begin{aligned}\log\,(p_k(x)/(1 - p_k(x)) &= \beta_{0k} + \beta_{1k}\,x + \beta_{2k}\,x^2 \\ &= f(x)\end{aligned} \tag{2.2}$$

with inverse transformation

$$p_k(x) = \frac{\exp(f(x))}{1 + \exp(f(x))}. \tag{2.3}$$

The unknown parameters β_{0k}, β_{1k} and β_{2k} can be estimated by the method of maximum likelihood. Logistic regression is the standard statistical approach to regression analysis for binary data. Details are provided in Hosmer & Lemeshow (2000). We

* Author for correspondence (asolow@whoi.edu).

use the quadratic model because it allows the estimate of $p_k(x)$ to reach a well-defined maximum within the range of observed temperatures. As an alternative to logistic regression, $p_k(x)$ can be estimated non-parametrically (e.g. Hastie & Tibshirani 1990). In the application described in § 3, the logistic model performed well and there was no need to resort to non-parametric methods.

Turning to the calibration step, let $\hat{p}_k(x)$ be the fitted probability that species k is present at temperature x. The reconstructed temperature $\hat{x}_0$ is the value of x that minimizes

$$\chi^2(x) = \sum \frac{(1 - \hat{p}_k(x))^2}{\hat{p}_k(x)}, \tag{2.4}$$

where the summation is over the species contained in s. The quantity $\chi^2(x)$ is the ordinary χ^2-statistic for testing the hypothesis that $p_k(x_0) = \hat{p}_k(x)$ for the species in s. The value of x that minimizes $\chi^2(x)$ must be found numerically.

It is straightforward to extend this approach to include two (or more) climatic variables. For example, the quadratic logistic model for two variables x and z is

$$\begin{aligned}\log\,(p_k(x, z)/(1 - p_k(x, z)) &= \beta_{0k} + \beta_{1k}x + \beta_{2k}x^2 \\ &\quad + \beta_{3k}z + \beta_{4k}z^2 + \beta_{5k}xz \\ &= f(x, z),\end{aligned} \tag{2.5}$$

with inverse transformation

$$p_k(x, z) = \frac{\exp(f(x, z))}{1 + \exp(f(x, z))}. \tag{2.6}$$

The reconstructed values $\hat{x}_0$ and $\hat{z}_0$ of these variables is the pair (x, z) that minimizes

$$\chi^2(x, z) = \sum \frac{(1 - \hat{p}(x, z))^2}{\hat{p}(x, z)}, \tag{2.7}$$

where the summation is again over the species in s and $\hat{p}_k(x, z)$ is the fitted value of $p_k(x, z)$. We will refer to this as the bivariate method.

3. RESULTS

The data for this study consisted of modern observations of the presence of beetle taxa at sites in North America and the corresponding climate data. The data used in this study include a total of 264 different species at $n = 602$ sites. These data and the associated climate information were extracted from a larger dataset available at http://www.ngdc.noaa.gov/paleo/insect.html. The larger dataset included 269 taxa and 3125 sites. We eliminated from this larger set all species that occurred at fewer than six sites and all sites that contained fewer than six species. The study involved reconstructing mean July temperature (TMAX) and mean January temperature (TMIN). Specifically, for each of these variables, we applied both the univariate and bivariate versions of the method. Performance was assessed via cross-validation. This involved reconstruction at each of the 602 sites, with the site to be reconstructed omitted in the regression step. Cross-validation was used to eliminate any bias that arises when the conditions at a site are reconstructed from a dataset that includes that site.

As an illustration of the results of the regression step, figure 1 shows the fitted univariate quadratic-logistic-regression models for the species *Micropeplus laticollis* (Maklin) and *Holoboreaphilus nordenskioeldi* (Maklin) and figure 2 shows the results for the bivariate quadratic-logistic-regression models. These models were fit using all 602 sites. It is clear from figure 2 that *M. laticollis* prefers relatively warm temperatures in both July and January. By contrast, while *H. nordenskioeldi* favours relatively cold July temperatures, it tolerates a relatively wide range of January temperatures. This pattern, which presumably reflects the ability of *H. nordenskioeldi* to overwinter successfully under a wide temperature range, is exhibited by many species and reduces the value of TMIN in climate reconstruction.

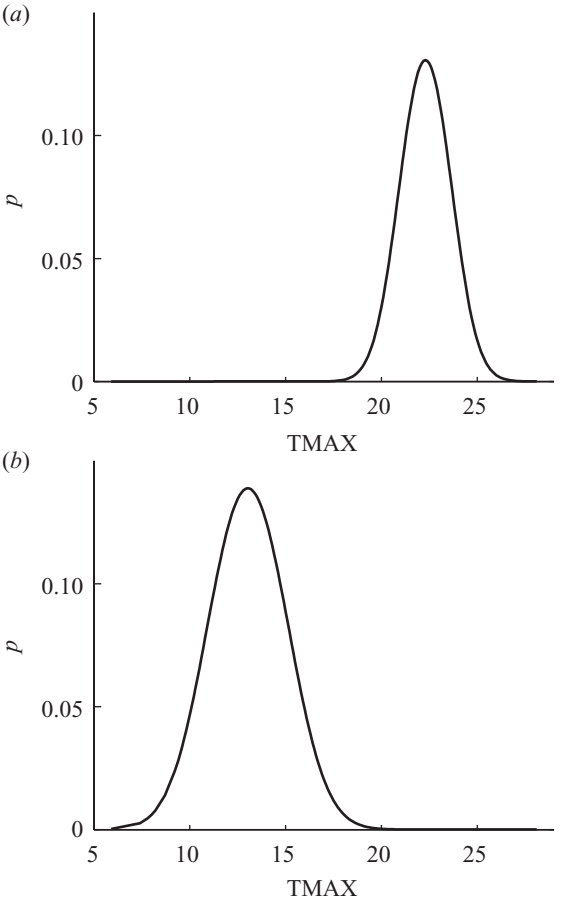

Figure 1. Fitted probability of occurrence based on univariate quadratic logistic regression *versus* TMAX for (*a*) *Micropeplus laticollis* and (*b*) *Holoboreaphilus nordenskioeldi*.

To illustrate the performance of the method, in figure 3, the cross-validated reconstructed values of TMAX and TMIN are plotted against the true values for each of the 602 sites. In this case, reconstruction was based on the bivariate method. In table 1, we report the mean (signed) reconstruction errors and the root-mean-squared reconstruction errors for TMAX and TMIN for both the univariate and bivariate methods. The mean error is a measure of bias, while the root-mean-square error incorporates both bias and error variance. Specifically mean-squared error is the sum of squared bias and error variance. For these data, the two methods perform very similarly in reconstructing both TMAX and TMIN. Therefore, since the bivariate method is somewhat more computationally demanding than the univariate method, the univariate method provides a reasonable compromise. Table 1 also shows that the reconstruction of TMAX is more accurate

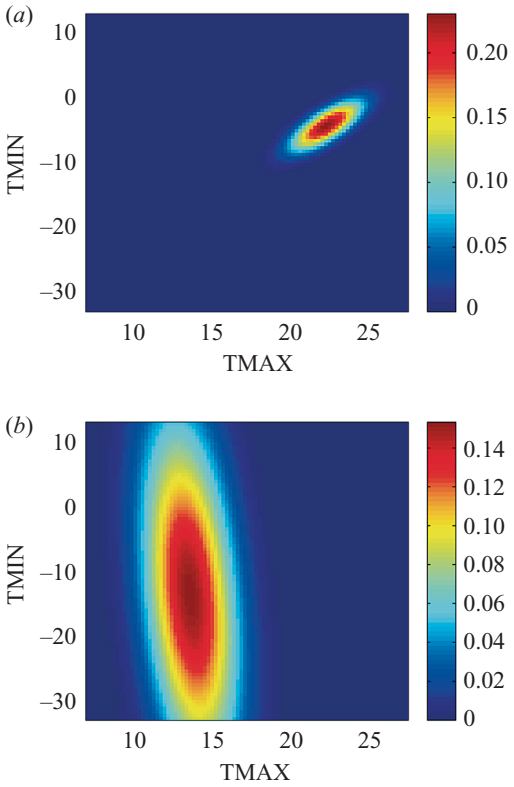

Figure 2. Fitted probability of occurrence based on bivariate quadratic regression versus TMAX and TMIN for (*a*) *Micropeplus laticollis* and (*b*) *Holoboreaphilus nordenskioeldi*.

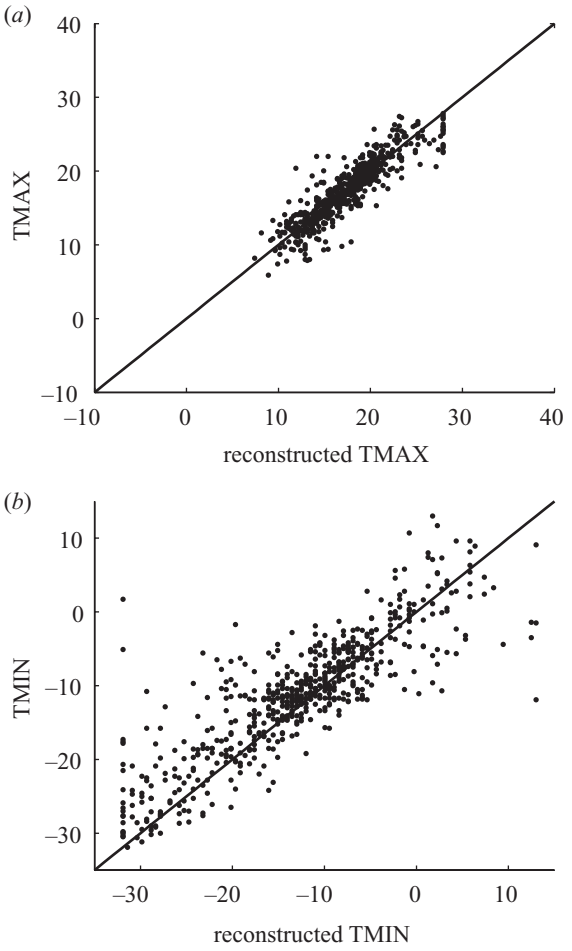

Figure 3. (*a*) Reconstructed TMAX versus true TMAX for all 602 North American sites. Reconstruction is based on the cross-validated bivariate method. (*b*) As in (*a*) for TMIN.

than the reconstruction of TMIN. This result, which is also clear from figure 3, is consistent with those for the MCR method (e.g. Elias 2001) and presumably reflects the ability of these species to overwinter successfully under a wide range of conditions.

4. DISCUSSION

The purpose of this paper has been to describe and illustrate a new method for reconstructing climatic conditions based on fossil beetle assemblages. The method is straightforward and MATLAB code is available from the authors upon request. This method can be viewed as an extension of the MCR method. The main difference is that the MCR method does not exploit variations in the rate of occurrence of a species within its climatic range, whereas the method proposed here is explicitly based on such variation. For reasons beyond our control, we are not able to present a comparison to the MCR method. However, as the modern data certainly exhibit variations in the rate of occurrence of beetle species within their climatic ranges, exploiting this information would seem to be an advantage.

A wide variety of biological proxies have been used to reconstruct past environmental conditions. For example, the relative abundances of fossil foraminifera have been used to reconstruct sea surface temperature (CLIMAP 1976) and similar data for fossil diatoms have been used to reconstruct pH in lakes (Birks *et al.* 1990). A variety of methods are used in these reconstructions. Although the details will vary, the basic two-step method involving regression modelling followed by calibration seems to provide a unified approach.

Table 1. Mean cross-validated error and the root-mean-squared cross-validation error (RMSE) for reconstructing TMAX and TMIN by the univariate and bivariate methods.

	TMAX		TMIN	
method	mean error (°C)	RMSE (°C)	mean error (°C)	RMSE (°C)
univariate	−0.05	1.82	−1.72	5.29
bivariate	0.07	1.96	−1.64	5.11

The helpful comments of John Birks and four anonymous referees are acknowledged with gratitude.

REFERENCES

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

Birks, H. J. B., Juggins, S., Stevenson, A. C. & terBraak, C. J. F. 1990 Diatoms and pH reconstruction. *Phil. Trans. R. Soc. Lond.* B **327**, 263–278.

CLIMAP Project 1976 The surface of the ice-age Earth. *Science* **191**, 1131–1136.

Coope, G. R. 1977 Fossil Coleopteran assemblages as sensitive indicators of climate changes during the Devensian (Last) cold stage. *Phil. Trans. R. Soc. Lond.* B **280**, 313–340.

Elias, S. A. 1994 *Quaternary insects and their environments*. Washington, DC: Smithsonian Institution Press.

Elias, S. A. 2001 Mutual climatic range reconstructions of seasonal temperatures based on Late Pleistocene fossil beetle assemblages in eastern Beringia. *Q. Sci. Rev.* **20**, 77–91.

Hastie, T. J. & Tibshirani, R. J. 1990 *Generalized additive models*. London: Chapman & Hall.

Hosmer, D. W. & Lemeshow, S. 2000 *Applied logistic regression*. New York: Wiley.