Quaternary Science Reviews 30 (2011) 1272–1278

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages

R.J. Telford [a,b,*], H.J.B. Birks [a,b,c,d]

[a] Department of Biology, University of Bergen, Thormøhlensgate 53 A, N-5006 Bergen, Norway
[b] Bjerknes Centre for Climate Research, Allégaten 55, N-5007 Bergen, Norway
[c] School of Geography and the Environment, University of Oxford, UK
[d] Environmental Change Research Centre, University College London, UK

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form
28 February 2011
Accepted 1 March 2011
Available online 23 March 2011

Keywords:
Palaeoecology
Palaeoclimate
Reconstruction
Significance test
Transfer functions

ABSTRACT

We present a method to test the statistical significance of a quantitative palaeoenvironmental reconstruction inferred from biotic assemblages with a transfer function. A reconstruction is considered statistically significant if it explains more of the variance in the fossil data than most reconstructions derived from transfer functions trained on random environmental data. Given reconstructions of several environmental variables from the same fossil proxy, the method can determine which is the best reconstruction, and if there is sufficient information in the proxy data to support multiple independent reconstructions. Reconstructions that fail this test have limited credibility and should be interpreted with considerable caution.

1. Introduction

Since Imbrie and Kipp (1971) introduced the methodology, quantitative reconstructions have been made for scores of environmental variables, using dozens of proxies, at thousands of sites. Many of these reconstructions are valid and useful. Some are not. The difficulty is to determine which. This is a two-part problem. The first part is to decide if a variable could *potentially* be reconstructed. The second is to determine if a variable can be reconstructed at a *specific* site.

The first problem is normally addressed by using a constrained ordination to test if a variable explains a significant proportion of the variance in the biotic data in the modern training-set, and then choosing an appropriate transfer function method and estimating its performance with cross-validation (Birks, 1995). Recent research has highlighted the potential pitfalls caused by spatial autocorrelation (Telford and Birks, 2005, 2009), model selection bias (Telford et al., 2004), and unevenly sampled gradients (Telford and Birks, submitted for publication), that can potentially make ecologically unimportant variables appear possible to reconstruct, and make model performance statistics over-optimistic.

Passing the first test is necessary but not sufficient. With a judicious choice of enough training-set sites, maximising variation in the environmental variable of interest, and minimising nuisance gradients (Birks, 1995), any ecologically relevant variable will be a significant predictor of the biotic data. Statistical significance of an environmental variable as a predictor of the training-set biota is a property of the modern training-set. It reveals which variables can potentially be reconstructed, but it does not provide a guarantee that a variable can be meaningfully reconstructed at a specific site. Any fossil data can be run through the transfer function and, provided there are species in common, numeric results will be obtained. But these results are not necessarily useful. On rare occasions, it is possible to validate the reconstruction against instrumental records (e.g., Lotter, 1998). While reassuring when successful, validation is valid only for the site, and time-window, for which it was performed. For most sites, no validation is possible, and instead a range of metrics have been used to assess reconstructions. These include the presence of good analogues for the fossil samples in the training-set, and the goodness-of-fit of the fossil observations passively placed in an ordination of the modern observations and the environmental variable being reconstructed (Birks, 1995). While these can highlight reconstructions that are

---

* Corresponding author. Department of Biology, University of Bergen, Thormøhlensgate 53 A, N-5006 Bergen, Norway. Tel.: +47 55583422.
E-mail address: richard.telford@bio.uib.no (R.J. Telford).

0277-3791/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.quascirev.2011.03.002

unlikely to be reliable because the fossil assemblages have aberrant compositions, they cannot identify sites where changes in the environmental variables being reconstructed were not important drivers of assemblage composition.

Many authors reconstruct multiple variables from the same fossil data. There are two ways to interpret these multiple reconstructions. The first is when a group of variables is reconstructed that are expected to be highly correlated. For example, Telford et al. (1999) reconstructed conductivity, pH, anion ratio and cation ratio from diatoms in a soda lake. These should be viewed as multiple alternative reconstructions, little more than a rescaling and change of units. The second interpretation is that the each reconstruction contains unique information about the palaeoenvironment. Unfortunately, it is seldom clear which of these two interpretations the author believes, and the assumption of the second interpretation, that each reconstruction contains unique information, is rarely, if ever, tested. Indeed, except where validation against instrumental data is possible, the utility of a single reconstruction is not assessed either.

Since it is always possible to obtain numeric results for a reconstruction of any variable, regardless of its ecological relevance, some test of utility is required. We propose the following test, that a single reconstruction should explain more of the variation in the fossil data in a constrained ordination than a transfer function trained on random environmental data applied to the same fossil data. Multiple independent reconstructions should each explain more of the fossil data than a random variable after the other reconstructions have been partialled out. The remainder of this paper explores the properties of these tests using transfer functions from several training-sets on a variety of fossil data. In particular, we wish to test: are reconstructions that most palaeoecologists would think robust statistically significant; are reconstructions that we think questionable not significant; and how robust is the method to various possible causes of Type I and Type II errors?

2. Data and methods

Published reconstructions from a range of organisms and transfer function methods, and using training-sets with different properties were analysed (Table 1).

The proportion of variance in the fossil data explained by a single reconstruction is estimated using a constrained ordination. We use redundancy analysis (RDA) as the species turnover is relatively low in most of our examples (Table 1). Then, using the biotic data from the same training-set, reconstructions are inferred from transfer functions trained on random environmental variables drawn from a uniform distribution. The proportion of the variance explained by these random reconstructions is estimated. If the observed environmental variable explains more of the variance than 95% of the random reconstructions, the reconstruction is deemed statistically significant. The proportion of the variance explained by the first axis of a principal components analysis (PCA) is also recorded, as this represents the maximum proportion of the variance in the fossil data that the reconstruction could possibly explain.

If there are multiple reconstructions, a forward selection procedure is adopted. First, the reconstruction that explains the most variance is accepted. Then other reconstructions are tested to determine if they explain significantly more variance than random reconstructions when the first reconstruction is partialled out. This procedure can be repeated until there are no more significant reconstructions.

In all cases, 999 random environmental variables were generated to produce the null distribution. Significance values are given as the fraction of random variables that explain as much as or more of the fossil data than the observed variable.

In this paper, we chose to develop the null distribution by using uniformly distributed random variables rather than by permuting the environmental variable, the procedure used by CANOCO *(*ter Braak and Šmilauer, 2002*)*. This choice allows us to compare multiple reconstructions against the same null distribution, and, more importantly, to use spatially structured random variables when the training-set is autocorrelated. Quantile–quantile plots (not shown) of the proportion of variance explained by reconstructions derived from permuted or uniformly distributed variables do not deviate from the 1:1 line. This result is not method dependent.

We use weighted averaging (WA) with inverse deshrinking on square-root transformed data and the modern analogue technique (MAT) with squared chord distance and five analogues. The

Table 1

Reconstructions analysed. Effective number of species is estimated using N2 (Hill, 1973). Turnover is estimated as the length of the first detrended correspondence analysis axis (standard deviation units). Significance is based on 999 trials. WA – weighted averaging, MAT – modern analogue technique.

Site Name	Proxy	Effective No. Species	No. observations	Turnover SD	Time span	Variable (s)	Transfer Function	Significance	Training-set	Reconstruction Citation
Round Loch of Glenhead (RLGH)	Diatoms	15.6	20	0.7	Industrial	pH	WA	0.006	SWAP (Birks et al., 1990)	Allott et al. (1992)
Kråkenes	Chironomids	9.9	44	3.2	Lateglacial	Temperature	WA	0.004	Brooks and Birks (2001)	Brooks and Birks (2000b)
Bjørnfjelltjørn	Chironomids	12.8	61	1.4	Holocene	Temperature	WA	0.183	Brooks and Birks (2001)	Brooks (2006)
Whitrig Bog	Chironomids	14.8	76	2.4	Lateglacial	Temperature	WA	0.001	Brooks and Birks (2001)	Brooks and Birks (2000a)
Holebudalen	Chironomids	12.5	53	1.8	Holocene	Temperature	WA	0.009	Brooks and Birks (2001)	Velle et al. (2005)
Vøring Plateau MD95-2011	Planktonic foraminifera	3.3	380	1.4	Lateglacial – modern	Annual SST	MAT	0.001	Pflaumann et al. (2003)	Risebrobakken et al. (2003)
Chukchi Sea	Dinoflagellates	4.0	86	1.2	Holocene	Summer SST, sea-ice duration, summer salinity	MAT	0.003, 0.007, 0.146	Radi and de Vernal (2008)	McKay et al. (2008)
Bjørnfjelltjørn	Pollen	4.8	63	1.2	Holocene	July temperature, January temperature, precipitation	MAT	0.029, 0.160, 0.743	(Bjune et al., 2010)	(Birks and Peglar, unpublished).
Holebudalen	Pollen	8.0	94	1.2	Holocene	July temperature, January temperature, precipitation	MAT	0.001, 0.173, 0.153	(Bjune et al., 2010)	Eide et al. (2006)