Title	Discussion Paper: A Bayesian multinomial regression model for paleoclimate reconstruction with time uncertainty
Authors(s)	Parnell, Andrew C.
Publication date	2016-11
Publication information	Parnell, Andrew C. “Discussion Paper: A Bayesian Multinomial Regression Model for Paleoclimate Reconstruction with Time Uncertainty.” Wiley, November 2016. https://doi.org/10.1002/env.2401.
Publisher	Wiley
Item record/more information	http://hdl.handle.net/10197/8332
Publisher's statement	This is the author's version of the following article: Andrew Parneel (2016) "Discussion Paper: A Bayesian multinomial regression model for paleoclimate reconstruction with time uncertainty" Environmetrics, 27(7) : 431-433 which has been published in final form at http://dx.doi.org/10.1002/env.2401.
Publisher's version (DOI)	10.1002/env.2401

Discussion of 'A Bayesian multinomial regression model for paleoclimate reconstruction with time uncertainty'

Andrew C. Parnell

June 7, 2016

1 Introduction

It is a pleasure to be a discussant on this paper; one that exhibits the best of both applied statistics and cutting edge methodology. The authors should be congratulated for the clarity of their expression and I hope this paper reaches a wide audience of palaeoclimate scientists, statisticians and environmental scientists more generally. Palaeoclimate is important: it provides useful constraints on the speed at which the climate changes; it helps us learn more about the fragility of our existence (all human history is encompassed in a uniquely warm and stable climate), and provides a method by which to test our physics-based models which allow us to predict future climate change.

In this discussion I hope to highlight some of the real contributions of this paper, to point out some of the important non-statistical considerations (which, as applied statisticians in this area we should be cognisant), and to contrast with the rapidly expanding mostly-Bayesian palaeoclimate statistics literature. In particular, accounting for time uncertainty is, I suspect, almost a unique challenge for time series analysis in palaeoclimate science. For this reason it has been ignored for decades. Now with tools as in this paper, they can start to draw proper inferences on climate over time with suitably quantified uncertainties.

2 Statistical considerations

The model the authors fit provides a set of tools that have been previously beyond the capabilities of most palaeoclimate scientists. The traditional paradigm has the scientist fitting the models pollen slice by pollen slice with either simple or opaque methods used to turn pollen slices into climate estimates. The approach taken by the present authors moves far beyond these restrictions, with joint inference across slices, and a clear likelihood which quantifies the transformation between proxy (here pollen) and climate. Perhaps more usefully, the authors show in their Figure 8 three of what I call *climate histories*, posterior draws of the individual climate time series. The inference to be drawn from these is perhaps the most important, as they can be used to quantify past climate dynamics, as the authors ably demonstrate with the quantified behaviour of the 8.2ka event. As mere statisticians, we can only guess at their potential use, but once explained to palaeoclimate scientists, I believe we will see a large adoption by the user community.

My preference for explaining these techniques and data sets is to use the language of state-space models (e.g. Wikle et al., 1998). Following the notation in the paper, we have initially:

$$\begin{aligned} \boldsymbol{Y}_i^f &= g_\theta(x_i^f) \\ x_i^f &= x_{i-1}^f + \epsilon_i \end{aligned}$$

so the pollen Y are 'calibrated' via some function g (parameterised by θ) against the climate variable x, which is itself given a Weiner Process prior. The fun extensions arise because extra layers can be added to the state space model. One such is a set of extra data in the form of the modern calibration set which takes the same form as the state equation:

$$\boldsymbol{Y}_i^m = g_\theta(\boldsymbol{x}_i^m)$$

This provides priceless extra information about the parameters θ. The authors use the space-for-time substitution where this modern data set comes from other sampled local lakes, though other calibration sets are possible, such as those where there is a period of instrumental temperature measurements which overlap with the fossil record as in Mann et al. (1998) and, more recently, Schofield et al. (2015). Both approaches fit in with the state space approach.

Focussing on the modern calibration, a key modelling choice the authors make is that g_θ above is a realisation from a multinomial, where the link-transformed proportions on each pollen taxa are restricted to be Gaussian in shape. When back-transformed the relationships between the climate variable and the pollen response can be flexibly non-Gaussian shaped, but it remains unclear to me exactly what hidden restrictions are being placed on the relationships through this approach. Further, the posterior predictive checking provided in the appendix seems to suggest that many of the pollen taxa have under- or over-estimated uncertainties. It would be interesting to contrast this with the spline approach of e.g. Cahill et al. (2016) or the non-parametric approach of Salter-Townshend and Haslett (2012).

As with all Bayesian models, the choice of prior structures has a key bearing on the output of the models. Perhaps the most explicit aspect of this choice is the Weiner Process applied in the evolution equation. Whilst such a process appears appropriate for the 1100 years of the climate model simulation, it is less clear to me that such an approach can capture broader dynamics such as at the 8.2k event or the transition to the Holocene known as the Younger Dryas, as seen around 10,000 years ago in Figures 5–7. If the goal is to learn about the speed of climate change, then the choice of these priors is vital.

3 Non-statistical considerations

As applied statisticians it is our job to gain at the very least a basic understanding of the domains in which we work. The authors are to be congratulated in providing a paper which addresses many of the issues of palaeoclimate data whilst retaining a parsimonious statistical approach. One of the most interesting facets of the paper from my point of view is that quite a few of the authors' choices conflict with previous advice I have been given from palynologists with whom I collaborate.

Perhaps the most divergent of these non-statistical choices is the use of a training set containing only modern samples located in the neighbourhood of the fossil pollen core. The authors argue that this provides a simple and relevant set of plausible past climates for reconstruction purposes. Of course, once this modern set is chosen, it is impossible for the past temperature estimates to lie outside the range of modern temperatures; the common no modern analogue problem. Further it restricts the size of potential climate changes, and so may point to smoother climate changes over time. The advantage of having a much larger set of potential modern climates is that the past climate values are much less likely to lie on the edge of modern 'climate space'. The price paid for using the larger data set is the extra noise and struggles in model fitting.

A second important conflict is the choice of climate variable(s) to reconstruct. Of course, we can choose to reconstruct any climate variable(s) we have available, no matter how relevant it is to our knowledge or the proxy. The authors here choose 30-year annual mean temperatures. A poorly chosen climate variable will yield large uncertainties, as it may not be sensitive to changes in the proxy variable. There is work (see e.g. Huntley, 2012, and references therein) discussing whether plants are sensitive to broad measures such as mean annual temperatures. Other authors (Haslett et al., 2006; Parnell et al., 2015) have chosen to focus on only variables which are explicitly relevant to plants, perhaps at the expense of some relevant anthropogenic interpretation. Going further, it may be that pollen here responds to more than just temperature (a common second climate variable is moisture, e.g. Tolwinski-Ward et al., 2014), which would require bivariate reconstructions and a bivariate time series prior. A worry, advanced by Huntley (2012), is that a proxy which responds primarily to moisture might produce a spurious signal in temperature if moisture is not included in the reconstructed climate variables.

4 Summary

I would like to again applaud the authors on an excellent, readable, and methodologically-thorough study. As a final note, it may be worth considering where the next steps lie for statistical palaeoclimate science. As with all such research enterprises, it is useful to think about the data we are *not* currently using which may contribute to our knowledge of the palaeoclimate. The most obvious is data from other proxies from the same site, such as tree rings, isotope signatures, etc. Following my state space notation above, each proxy requires its own state equation and corresponding calibration step. Another data source is the same proxy at different sites. This presents a real challenge, as we are required to analyse a multivariate temporally-misaligned series (a first attempt at such problems is in Doan et al., 2014). Yet another is mechanistic information about the relationship between proxy and climate, and the dynamics of climate itself. The authors have made a first step in this area by using output from a climate model to constrain climate smoothness. I look forward to seeing the authors' future work in this area.

References

Cahill, N., Kemp, A. C., Horton, B. P., and Parnell, A. C. (2016). A Bayesian hierarchical model for reconstructing relative sea level: from raw data to rates of change. *Climate of the Past*, 12(2):525–542.

Doan, T. K., Parnell, A. C., and Haslett, J. (2014). Joint Inference of Misaligned Irregular Time Series with Application to Greenland Ice Core Data. *Advances in Statistical Climatology, Meteorology and Oceanography*, 1(1):15–27.

Haslett, J., Whiley, M., Bhattacharya, S., Mitchell, F. J. G., Allen, J. R. M., Huntley, B., Wilson, S. P., and Salter-Townshend, M. (2006). Bayesian palaeoclimate reconstruction. *Journal of the Royal Statistical Society, Series A*, 169:395–438.

Huntley, B. (2012). Reconstructing palaeoclimates from biological proxies: some often overlooked sources of uncertainty. *Quaternary Science Reviews*, 31:1–16.

Mann, M. E., Bradley, R. S., and Hughes, M. K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392(6678):779–787.

Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R. M., Huntley, B., and Haslett, J. (2015). Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1):115–138.

Salter-Townshend, M. and Haslett, J. (2012). Fast inversion of a flexible regression model for multivariate pollen counts data. *Environmetrics*, 23(7):595–605.

Schofield, M., Barker, R., Gelman, A., Cook, E. R., and Briffa, K. R. (2015). A Model-Based Approach to Climate Reconstruction Using Tree-Ring Data. *Journal of the American Statistical Association*, 111(536):93–106.

Tolwinski-Ward, S. E., Tingley, M. P., Evans, M. N., Hughes, M. K., and Nychka, D. W. (2014). Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response. *Climate Dynamics*, 44(3-4):791–806.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154.