Quaternary Science Reviews 151 (2016) 111–126

Contents lists available at ScienceDirect

Quaternary Science Reviews

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

# Joint palaeoclimate reconstruction from pollen data via forward models and climate histories

Andrew C. Parnell [a,*], John Haslett [b], James Sweeney [c], Thinh K. Doan [d], Judy R.M. Allen [e], Brian Huntley [e]

[a] School of Mathematics and Statistics, Insight Centre for Data Analytics, University College Dublin, Ireland
[b] Discipline of Statistics, Trinity College Dublin, Ireland
[c] School of Business, University College Dublin, Ireland
[d] Novartis Ltd, Ireland
[e] School of Biological and Biomedical Sciences, University of Durham, UK

ARTICLE INFO

*Article history:*
Received 11 June 2016
Received in revised form
5 September 2016
Accepted 6 September 2016

*Keywords:*
Palaeoclimate reconstruction
Statistical modelling
Forward models
Climate histories
Joint inference
Palynology
Chronological uncertainty

ABSTRACT

We present a method and software for reconstructing palaeoclimate from pollen data with a focus on accounting for and reducing uncertainty. The tools we use include: forward models, which enable us to account for the data generating process and hence the complex relationship between pollen and climate; joint inference, which reduces uncertainty by borrowing strength between aspects of climate and slices of the core; and dynamic climate histories, which allow for a far richer gamut of inferential possibilities. Through a Monte Carlo approach we generate numerous equally probable joint climate histories, each of which is represented by a sequence of values of three climate dimensions in discrete time, i.e. a multivariate time series. All histories are consistent with the uncertainties in the forward model and the natural temporal variability in climate. Once generated, these histories can provide most probable climate estimates with uncertainty intervals. This is particularly important as attention moves to the dynamics of past climate changes. For example, such methods allow us to identify, with realistic uncertainty, the past century that exhibited the greatest warming. We illustrate our method with two data sets: Laguna de la Roya, with a radiocarbon dated chronology and hence timing uncertainty; and Lago Grande di Monticchio, which contains laminated sediment and extends back to the penultimate glacial stage. The procedure is made available via an open source R package, Bclim, for which we provide code and instructions.



## 1. Introduction

Quantitative methods in palaeoclimate reconstruction, from pollen in slices taken from a sediment core, were first introduced over three decades ago (Bernabo, 1981) and have since replaced their qualitative precursors. However, most methods are still primitive in their modelling of uncertainty. This has greatly inhibited developments on several fronts. In particular, extant reconstruction methods almost never exploit the entire data set available, often ignoring important uncertainties for computational or statistical convenience. As palaeoclimate proxy data are almost always non-standard (non-normal, non-linear, multivariate, etc), this leads to mis-matches between palaeoclimate records across sites, and to a false certainty in the inferences drawn from partial sets of data. Further, real interest often lies in the dynamics of past climate change yet many methods provide a poor basis for inference on such changes. Here we offer a new paradigm, which we refer to as climate histories.

We present software, Bclim, available as an open source R package, as an illustration of our proposed framework. It is the first attempt in palynology at the joint inference of the entire climate history corresponding to a single sediment core. The inputs are:

- A set of fossil proxy data, being pollen counts for each slice.
- The depths of each slice.
- The radiocarbon dates (these may or may not be from a subset of the slices).

* Corresponding author.
*E-mail address:* andrew.parnell@ucd.ie (A.C. Parnell).

http://dx.doi.org/10.1016/j.quascirev.2016.09.007
0277-3791/

In this initial version of Bclim inference is based on a forward model (Salter-Townshend and Haslett, 2012) created from a large set of modern data. The complete model is described in Parnell et al. (Parnell et al., 2015), hereafter referred to as P15. The output of Bclim is a large number of stochastic climate histories, each of which is equally statistically consistent with the available data.

We see a forward model [sometimes known as a proxy systems model, Evans et al., 2013] as the causal chain through which climate is transformed into proxy data stored in an archive. Our definition is broad, ideally encompassing both deterministic and statistical approaches, but with a clear focus on accounting for uncertainty at each stage. This uncertainty may be due to unknown processes which lend themselves to deterministic modelling, such as the means by which pollen is spread through the local area [e.g. Garreta et al., 2009], to stochastic processes such as the probability of detecting a particular variety of pollen through a microscope, given that it is present. In our approach we combine the forward model with a simple stochastic climate model via Bayesian inference, which allows us to produce climate histories with narrower uncertainties than using climate dimensions and pollen slices individually.

In this paper we purposefully avoid mathematical notation unless absolutely necessary. The full technical details can be found in P15. Rather, we aim to explain the concepts behind the statistical model we develop so that users of Bclim (and those who work on alternative proxies) can appreciate why these issues are important, and how they are different from previous approaches. We particularly focus on the output (i.e. climate histories) and the benefit we believe they can offer to those who wish to perform deeper analysis of their proxy data. We emphasise that the use of the software requires no in-depth statistical knowledge of the methods, merely the ability to use some basic R commands and the enthusiasm to find interesting new ways to explore climate histories.

In our first case study [Laguna de la Roya, hereafter Roya; Allen et al., 1996], discussed in Section 5.1, the data comprise pollen counts for 28 taxa from 72 slices, and 6 radiocarbon dates. Our target is three specific dimensions of climate over the period 0 to 16 ka BP at centennial intervals. Our three dimensions represent the length of the growing season, the harshness of the winter, and the moisture available to plants. These are the three climate dimensions to which we refer throughout the paper. The resulting output is thus a collection of climate histories, each of dimension $3 \times 160 = 480$. The key points to note are:

- Our forward model contains a simplified mathematical description of how the 28 pollen taxa respond to these three aspects of climate. We create estimates of the climate by inverting the forward model. Inversion is required because we wish to estimate climate from pollen, which is the opposite direction to that of the forward model. We describe the forward model in Section 3.1.
- The 480 values defining each climate history are reconstructed jointly using all slices in the core and all climate dimensions simultaneously. We produce many sets of the 480 values. Each such set is a climate history. Three climate histories are plotted in Fig. 1, and discussed in detail in Section 3.2.
- The climate histories are temporally constrained by a model of climate dynamics which also takes account of temporal uncertainty. This part of our model is discussed in Section 4.3.

The paper is structured as follows. In Section 2 we point out in overview how our approach differs to those that are traditionally used. In Section 3 we discuss the three main components at a conceptual level; forward models, joint inference, and the creation of climate histories. This section can be skipped by those who want to avoid any technical detail. In Section 4 we describe the details of our software, Bclim, and the necessary inputs and intermediary steps to the creation of climate histories. In Section 5 we apply our method to two sites as outlined above, and show some of the richer inferential possibilities that are now admissible using climate histories. We discuss further possibilities and extensions in Section 6. Computer code, example tutorials, and the ability to request features or point out bugs, are available at github.com/andrewcparnell/Bclim.

## 2. Overview of differences between Bclim and previous approaches

Bclim exploits recent developments in statistical methodology in several ways. These developments involve the use of Bayesian methods for joint statistical inference across all available data, and Monte Carlo algorithms to measure uncertainty as explained in Section 3.2. In our case studies the target of inference is a three-dimensional climate time series defined on an arbitrary user-specified time grid. We refer to this multivariate time series as a climate history.

Table 1 provides a toy illustration of the output of our software. Five univariate climate histories are presented in the rows of the table. They are provided on a regular time grid at times 1, 2, 3, 4, and 5. Each history is thus of length 5. Were climate to be bivariate, there would be two values for each time point. The five climate histories are summarised by their column mean and standard deviation in the last two rows. They can be further summarised, as in the last column, by comparing individual differences. Here we have calculated the difference in climate between time 2 and time 1 for each climate history, and summarised them by their mean and standard deviation. The climate histories allow us to estimate, for example, the standard deviation of the difference between times 2 and 1 which would not otherwise be available without the histories. We elaborate on the possible choices and uses of such summaries for more realistic scenarios in Section 3.3 and apply these techniques to our case studies in Section 5.

Bclim is based on the Bayesian model discussed in P15. It is one of an increasing number of Bayesian approaches to palaeoclimate reconstruction from proxies [e.g. Haslett et al., 2006; Tolwinski-Ward et al., 2014; Vasko et al., 2000] which involve joint inference and forward modelling. In this paper we sketch the methodology only in outline, referring technical readers to P15. The Bclim approach stands in stark contrast to that provided by very many widely used and cited methods [e.g. ter Braak and. Juggins, 1993; Mann et al., 2008]. The most widely used, Weighted Averaging Partial Least Squares (WA-PLS), performs reconstructions separately (i.e. marginally) for each of the chosen dimensions of climate, and for each of the slices in a sediment core. In each case it is using, for each slice, only the corresponding pollen counts for that slice. The reconstruction thus uses only a fragment of the available information. Often only a single 'best' reconstruction is used for each slice. As we elaborate in Section 4, such an approach makes inefficient use of the data and is not to be recommended unless as a crude first step.

More deeply, in WA-PLS and related methods there is no clear modelling of uncertainty. Whilst an attempt at quantifying the uncertainty in each dimension of climate is made (e.g. via a root mean square error of prediction; RMSEP), this is usually a single number which is then used to quantify the uncertainty in the reconstruction. Many users will have the mistaken impression that an underlying normal distribution can be used to interpret the RMSEP, for example by creating the mean plus or minus twice RMSEP. Further, climate dynamics, the implicit focus of many reconstructions, is beyond the reach of the inference. As a simple