Received: 29 May 2017 | Revised: 22 December 2017 | Accepted: 24 April 2018
DOI: 10.1002/wics.1437

WILEY WIREs COMPUTATIONAL STATISTICS

OVERVIEW

Statistical challenges in estimating past climate changes

James Sweeney[1] | Michael Salter-Townshend[2] | Tamsin Edwards[3] | Caitlin E. Buck[4] | Andrew C. Parnell[5]

[1]School of Business, University College Dublin, Dublin, Ireland
[2]School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
[3]Department of Geography, Kings College London, London, UK
[4]School of Mathematics and Statistics, University of Sheffield, Sheffield, UK
[5]Insight Centre for Data Analytics, Dublin, Ireland

Correspondence
James Sweeney, School of Business, University College Dublin, Dublin, Ireland.
Email: james.sweeney@ucd.ie

Funding information
EPSRC-funded Past Earth Network, Grant/Award Number: EP/M008363/1; Leverhulme International Network, Grant/Award Number: F/00118/BE

We review the statistical methods currently in use to estimate past changes in climate. These methods encompass the full gamut of statistical modeling approaches, ranging from simple regression up to nonparametric spatiotemporal Bayesian models. Often the full inferential challenge is broken down into many submodels each of which may involve multiple stochastic components, and occasionally mechanistic or process-based models too. We argue that many of the traditional approaches are simplistic in their structure, handling, and presentation of uncertainty, and that newer models (which incorporate mechanistic aspects alongside statistical models) provide an exciting research agenda for the next decade. We hope that policy-makers and those charged with predicting future climate change will increasingly use probabilistic paleoclimate reconstructions to calibrate their forecasts, learn about key natural climatological parameters, and make appropriate decisions concerning future climate change. Remarkably few statisticians have involved themselves with paleoclimate reconstruction, and we hope that this article inspires more to take up the challenge.

This article is categorized under:
Applications of Computational Statistics > Computational Climate Change and Numerical Weather Forecasting

KEYWORDS
Bayesian methods and theory, computational Bayesian methods, Paleoclimate reconstruction, statistical modelling of climate

1 | INTRODUCTION

The study of past climate or paleoclimate is an international focus of research effort, as evidenced by the work of the Intergovernmental Panel on Climate Change (Stocker, 2014). This is because paleoclimate provides a useful test-bed for estimating natural climate variability, for judging the size and speed of potential changes, and for calibrating our complex models of the climate system (Haslett et al., 2006; Li, Nychka, & Ammann, 2010). However, the study of paleoclimate is impeded by the fact that, in general, we do not have direct observational measurements of past climate. Instead we rely on proxy (or fossil) climate markers, which take the form of imprecise chemical, geological, and biological records that have been left behind in long environmental archives such as lake, ocean, and ice deposits. There are several statistical challenges of note. First, the individual proxy sources are on different temporal scales and observed at multiple distinct spatial locations. Second, the chronology of the fossil proxy data is largely unknown, associated with perhaps a few samples of the fossil record with age estimates from scientific dating methods, such as radiocarbon dating. Third, as reconstruction approaches typically rely on the uniformitarianism principle, that is, the knowledge of an organism's present-day environmental preferences can be used to make statements about the past environmental conditions of a fossil sample, an additional challenge is to incorporate knowledge of the climate system supplied by mechanistic vegetation and climate models to guide reconstructions when this assumption of uniformity is inappropriate.

In this article, we provide an overview of these and further statistical issues, including computational challenges, in the context of modern paleoclimate reconstruction methods. In the following section, we provide a brief introduction to some of the proxy climate data sources used for the reconstruction of past climates. The remainder of this article is structured as follows. In section 2, we broadly sketch the process of past climate estimation from multiple sources of uncertain information and highlight a number of challenging obstacles. In section 3, we provide a brief overview of the literature for classical climate reconstruction methods. We consider three commonly used classical methods for paleoclimate reconstruction, outlining the limitations and statistical challenges encountered by each approach. In section 4, we present a Bayesian implementation of classical models including a discussion on chronological uncertainty and Bayesian inference. Section 5 presents an extension of the approaches to the spatiotemporal, multiproxy setting where all information and sources of uncertainty are accounted for in a coherent manner. Finally, section 6 contains a summary of the broad statistical challenges that remain.

1.1 | Proxy sources of information for past climate

Climate is a multidimensional space–time process which, for the purposes of statistical modeling, needs to be quantitatively defined. Thus, climate is usually described in terms of the familiar elements of observed weather and is often measured as 30-year averages of these weather-related variables, assuming stationarity of the climate system over this time frame. In the examples we discuss later in the paper, climate might simply be northern hemisphere mean temperature over time (Mann, Bradley, & Hughes, 1998), or a more complex measurement, such as multivariate temperature and moisture variables across a region or continent (Parnell et al., 2015). Climate data may be chemical or biological, involving simple direct measurements of climate or intricately indirect observations. Direct measures of climate may come from the more recent past where available, such as climate measurements from satellites. However, we do not discuss the use of direct temperature measurements in estimating past climate changes because these, though often useful, are only available from the very recent past.

Indirect measurements of climate are broadly described as *proxy data* and here we review some of the most common types that might form part of a paleoclimate reconstruction. Many reconstructions rely on just one or two types of proxy; a major research challenge is the combining of multiple proxies into a suitable model. Many papers that use more than two types of proxy (e.g., Mann et al., 1998) suffer from the uncertainty quantification problems we outline in the remainder of the paper. In Figure 1, we provide a simplified diagram of the sequence of steps involved in obtaining proxy data from which we attempt to make inference on past climate.

Perhaps the most common and widely used proxy data type in paleoclimate reconstruction is that of tree rings (dendroclimatology; e.g., Briffa et al., 2008). These proxies, for some species, exhibit a very high temporal resolution down to a yearly or sometimes even seasonal signal. The traditional approach has been to calibrate the width of the rings with an overlapping instrumental temperature period (e.g., see Jones, Briffa, Barnett, and Tett, 1998; Mann et al., 1998; Mann et al., 2008). More modern approaches (Tolwinski-Ward, Tingley, Evans, Hughes, & Nychka, 2015) use richer versions of this calibration where the relationship between proxy and climate is tempered by some limitations of the growth rate of the rings. The ages for the rings can be estimated via dendrochronology (matching tree-ring widths across trees and sites with known ages) to produce a very high resolution reconstruction. Since good matching requires lots of overlapping records, most dendro-based reconstructions extend only to the previous 1,000 years. The major issue with such reconstructions is the unknown extent of their spatial link with perhaps local or regional climate features. Further complications exist in that younger trees tend to grow rings faster so the growth rate needs to be taken into account. For a more detailed description of dendroclimatology, see Hughes, Swetnam, and Diaz (2011).

For reconstructions going back into the Holocene (approximately 10,000 years before present), pollen is the most common proxy data source, and the proxy we primarily focus on in this article. The attraction of plant pollen as a climate proxy is its ubiquity and diversity; for example, Wilson, Peet, Dengler, and Pärtel (2012) and Mora, Tittensor, Adl, Simpson, and Worm (2011) cite the number of plant species worldwide as being in the hundreds of thousands. Each plant species has a preferred range of climate(s), and thus the presence or absence of an individual species provides a clue, albeit extremely noisy, to the prevailing climate at the time the pollen was produced. Fossil pollen can be found in lake and ocean sediments and, under expert analysis, can be recognized down to the species (i.e., a grouping of similar plant subspecies) level. This higher level grouping is due to the difficulty in distinguishing the pollen of similar subspecies from one another, for example,

FIGURE 1 General overview of the various processes that lead to the proxy paleodata used in climate reconstructions. In the example of pollen, the sensor system is the plant ecosystem. The archive systems are the lakes or mires where pollen is deposited. The observation system includes the field and laboratory measures such as core sampling, pollen counting, and radiocarbon dating among others

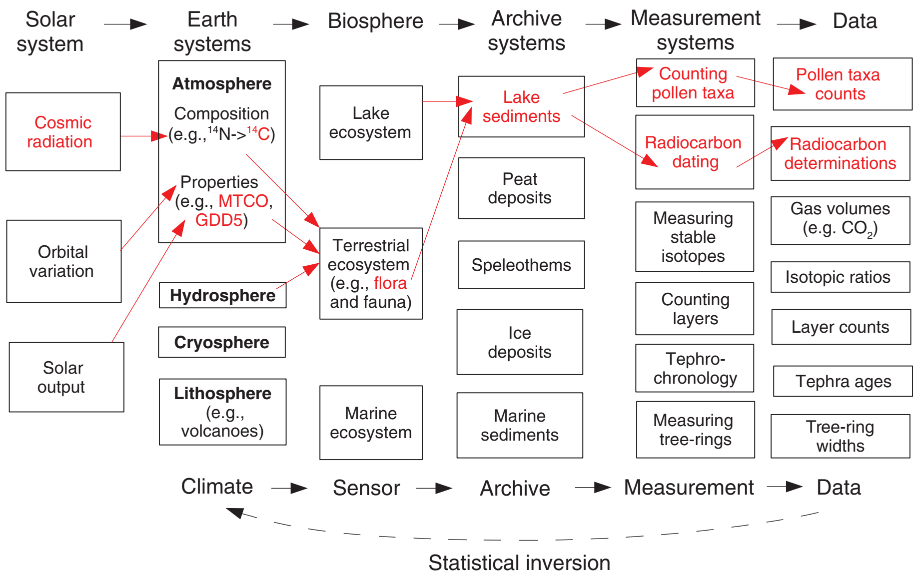

FIGURE 2 Overview of some of the climatological processes, which lead to the proxy paleodata used in climate reconstructions, including an overview of the intermediate stages involved in data acquisition. The arrows represent the flow or causal direction of the steps, which lead to the proxy data. As an example, the processes that lead to fossil pollen data obtained from lake sediment are highlighted in red with two climate variables of interest identified. One is GDD5 (growing degree days above 5 °C), a measure of the length of the growing season (days above 5 °C), and the other is MTCO (mean temperature of the coldest month), a temperature measure which captures the harshness of winter

distinguishing between the pollen of a mountain ash tree versus that of a river ash tree. The pollen counts from these similar subspecies are thus aggregated to a species level, that is, "ash." A slice from a core can contain hundreds of different species, and usually the top 50 or so are counted to produce a compositional vector of, for example, 400 pollen grains. This compositional vector can be compared/calibrated against modern samples to determine the past climate. The age information associated with the proxy data is harder to reconstruct, as usually only imprecise radiocarbon dates can be taken from the core. This adds a considerable blurring of uncertainty to the reconstructions, which makes it more difficult to obtain the underlying climate signal. For a more detailed description of the statistical issues in reconstructing climate from pollen data, see Parnell et al. (2015), or Parnell et al. (2016) for a less technical description.

The main method for reconstructing climate from nonbiological proxies concerns the use of stable isotopes. These are geochemical measurements of the abundance of a particular element compared to a reference standard. Many different elements are often collected and these are variously interpreted to be representative of past climate. For example, the stable isotope of oxygen, measured as δ^{18}O, is often considered to be a proxy for the temperature of summer rainfall, and is measured, over hundreds of thousands of years, in ice cores (Dansgaard et al., 1993). Surprisingly the quantification of uncertainty in ice core reconstructions is still very simplistic, often given only as a percentage value. Perhaps because of the simplified uncertainty structure, such reconstructions disagree at even local spatial scales (see Doan, Haslett, and Parnell, 2015), and counting layers/seasons in ice to provide the age of these reconstructions can also prove problematic (Klauenberg et al., 2011).

Although the above three represent the most-used proxies for paleoclimate in general, hundreds of others exist. These include chironomids (nonbiting midges), speleothems (cave formations, e.g., stalactites), diatoms (microalgae), corals, foraminifera (single cell, shelled marine species), and many others. From a statistical perspective the issues involved in each are similar. The field or laboratory measurements must be transformed into estimates of climate using mechanistic/statistical methods, which may involve modern calibration datasets, and they must each be dated to provide the time scale for reconstructions. We provide a pictorial overview of the process in Figure 2. However, a minutiae of detail remain in how each may represent aspects of climate and their spatial and temporal resolutions. Much of this can be modeled using Bayesian inference with appropriate expert information, and this is our preferred paradigm for reconstructing paleoclimate with uncertainty.

2 | THE GRAND CHALLENGE

The ultimate goal of paleoclimate reconstruction is estimation of the mechanics of past climate given all available data. In order to make inference on the paleoclimate from all such data, a statistical model is required. Once the model has been described we may choose to proceed using classical or Bayesian approaches. In either scenario the focus is on estimating the paleoclimate with suitably quantified uncertainties. For simple methods the uncertainty might just be a single measure such as root mean square error (RMSE), but for the richer more recent Bayesian approaches it is likely be a set of simulations or *climate histories* in multidimensional space and time, which capture the full joint probability distribution of all climate variables. Figure 3 displays a more detailed flow chart of the paleoclimate reconstruction process for pollen, with radiocarbon dating providing the chronological information.

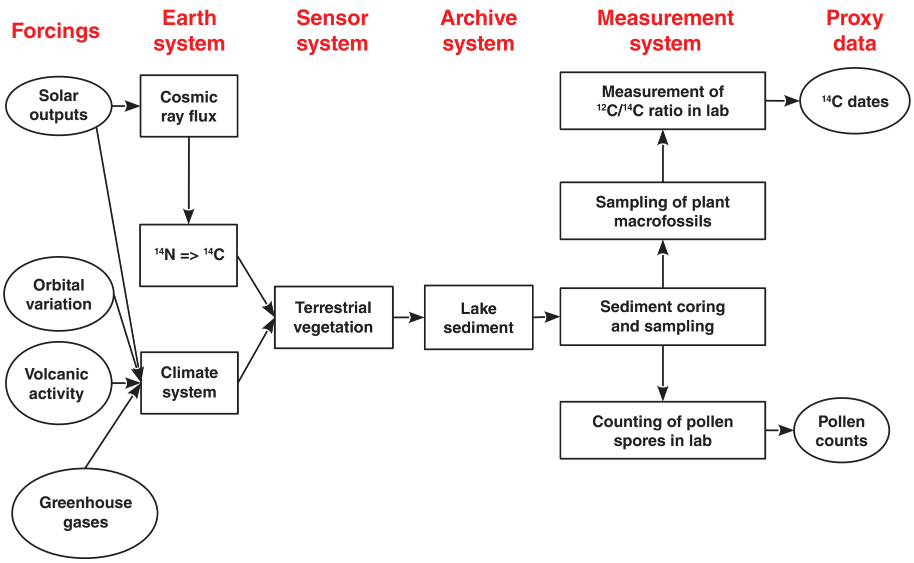

FIGURE 3 Key components of a model for the link between climate forcings and pollen counts from lake sediments. Ellipses represent numerical input and output values. Rectangles represent components of the model for which detailed models need specifying and arrows represent the flow of numerical values from one part of the model to another and illustrate places in which conditional independence assumptions are typically made. For clarity, we do not show any feedbacks, but deciding which feedbacks to model and how to represent them is a key part of implementing such models

To meet this challenge, we have to define the climate variables that we want to reconstruct. Unfortunately, however, the climate variables that are often used by climate scientists (e.g., global sea levels, global mean temperature) are not the climate variables that can accurately be inferred from paleodata—in the case of pollen, it is well known (Huntley, 2012) that many plants and trees do not respond to broad brush measures of climate. For example, where we to set up a model to estimate global mean temperature from pollen counts taken from a core in central Italy, we are likely to obtain a poor reconstruction. If we additionally used a simplistic model to describe the pollen–climate relationship, such as a Gaussian linear model with parameter estimation via classical least squares, such a reconstruction might be naively precise and lead to false inference. This is due to the pollen response to climate being highly nonlinear for most species (ter Braak, 1995). However, it is important to note that even with a richer model and inference approach the reconstruction is likely to be highly uncertain, which is at least honest, if not useful. Ideally we wish to choose a climate measurement, which is both reasonably informed by the proxy data, and yet of interest to those who need to evaluate climate models and make decisions.

It is easy to find large sets of proxy data online (e.g., Pangea: https://www.pangaea.de), and it is relatively simple to produce paleoclimate estimates by treating these proxy data as explanatory variables in a regression type model, such as we will observe in the overview of classical reconstruction methods. However, we would caution against such an approach for three main reasons.

1. Cause and effect: It is inadvisable to model climate as a function of proxy response as this is an inversion of the true causal relationship.
2. Combining all of the various uncertainties involved in the precision of the proxy response, the climate measurements, and that of estimated model parameters is extremely difficult.
3. It is difficult to see how underlying physical processes, which govern the generating of response data, and which vary across proxies, can explicitly be accounted for.

As an example, different proxy variables will respond to different aspects of the (multivariate) climate, possibly over different time ranges, and this response might change across time (Garreta et al., 2009). For example, it might take many years to grow an oak forest, and so pollen counts taken from a fossil core beneath a lake nearby are likely to change slowly. In contrast, oxygen isotope measurements from an ice core can reflect much faster changes in the temperature/precipitation regime and so will provide a richer, higher resolution record, albeit only in places where ice cores exist (Doan et al., 2015). A further important issue to note is that response of vegetation to climate will also depend on atmospheric CO_2 concentrations, which change over time.

Many of the more basic models we discuss focus on creating statistical approximations of the proxy–climate relationship. More advanced approaches use combined physical/statistical models of the proxy–climate relationship with a hope of capturing its changing dynamics. We term any model that provides estimated proxy data from given climate data, rather than the reverse, a *forward* or *proxy systems model.* A key part of the grand challenge is combining many of these models (i.e., for multiple different proxies) together. Figure 3 provides a clear schematic guide for how a forward model could be created for pollen proxy data.

The usual scenario when creating proxy datasets is that a core is extracted from a long environmental archive (e.g., ice, lake sediments, speleothems or tree rings) and partitioned into slices. Each slice is analyzed to produce the proxy data, and represents a time window of past climate. The size of this time window will be highly dependent on the accumulation rate of the cored deposit. If the accumulation is slow, a slice may contain decades or even hundreds of years' worth of proxy information. Thus, a considerable effort associated with paleoclimate reconstruction is the creation of accumulation models (Parnell et al., 2015) to estimate the ages of the proxy slices. The accumulation models are usually created from a smaller set of slices that have been scientifically dated (e.g., radiocarbon dated, which is expensive), though some archives (e.g., ice and trees) allow for more precise relative dating via annual or other layer counting. In either scenario this adds a considerable statistical hurdle to the overall challenge, because the timing of the proxy slices is uncertain. A further challenge for proxies such as pollen is the issue of zero-inflation within the dataset. It is important to recognize that excesses of zeros observed for a given proxy may be due either to sampling error or to environmental factors at individual sites. If unaccounted for, this zero-inflation may result in the underestimation of response models.

The proxy data and the climate variables are usually separated into two parts. First, there is the *modern calibration period* where all the proxy data and all the climate variables are known. For this period the timing of the data is usually known exactly and there is no need to resort to accumulation models. The second part is the *fossil period* when we have only the proxy data, and usually only the accumulation rate as a guide to the age of each slice. There are thus several statistically challenging parts to the grand challenge. One part is to estimate the relationship between the modern proxy and climate data, another is to estimate the relationship between accumulation and age, and yet another is to infer past climate based on the modern relationships and accumulation models. Deeper goals might include estimating the mechanics or underlying parameters governing this climate change, or incorporating mechanistic information in the proxy–climate relationship, again with the goal of inferring underlying parameters. Once estimated, we often would like to create a map or time series of how past climate has changed on a regular location/time grid with properly quantified measures of uncertainty. We will observe in later sections how each of these goals pose challenges of computation, particularly so when a Bayesian approach is chosen.

2.1 | Notation and estimation for the grand challenge

We now describe a statistical framework for the grand challenge by introducing the notation we will use throughout the paper. We define the following:

- $c(s, t)$ is a multidimensional measurement of climate at location s and time t. We assume both s and t are continuous, with the former also being multivariate. In the paleoclimate literature, time is often written in years before present (Years BP,) where present refers to the year 1950 AD.
- $y_k(s, d)$ is a multidimensional proxy measurement taken from a slice at depth d for proxy k at location s. y_k might be a set of multivariate counts of N species (possibly multinomial) for a given proxy, such as pollen counts for several plant species, or be a continuous multidimensional variable (e.g., isotope measurements from an ice core). The depth d is usually treated as a univariate continuous measurement.
- $a_k(s, d)$ is the age of the slice at depth d for proxy k at site s in years before present. In many cases $a_k(d)$ at an individual site is necessarily a monotonic function of d, as older slices must lie deeper in the core. An alternative approach is provided by working with radiocarbon age $r_k(d)$ instead, as we do in later sections, which sidesteps the issues involved in converting radiocarbon age to calendar age; we refer the interested reader to Blackwell and Buck (2008) and Parnell, Buck, and Doan (2011) for a more in-depth introduction to the difficulties involved.

We further superscript the three above objects with m to indicate modern (or calibration) data for which both the proxy, time, and climate variables are all known, and f to indicate fossil measurements where the climate variables are missing. The grand challenge can be elucidated thus:

Estimate $c^f(s, t)$ with quantified uncertainty for a set of chosen s and t values, given $y_k^f(s,d), a_k^f(s,d), r_k^f(s,d), y_k^m(s), c^m(s)$ for a set number of proxies $k = 1, \ldots, M$.

The grand challenge is thus to find $\pi\left(c^f(s,t)|y_{1:M}^f(s,d), a_{1:M}^f(s,d), r_{1:M}^f(s,d), y_{1:M}^m(s), c^m(s)\right)$, where all of the sources of uncertainty involved in the climate reconstruction process are represented via a probability distribution on climate at each time t and location s.

In the following sections, we provide an overview of existing climate reconstruction strategies and, within each section, sketch the main statistical and computational challenges, which must be overcome for the grand challenge to be achieved.

3 | CLASSICAL APPROACHES TO PAST CLIMATE ESTIMATION

Here we provide a review of classical approaches to paleoclimate reconstruction for the *single* proxy setting and defer discussion of the more complex multiproxy approaches to later sections. In the following we refer to as "classical" any method where inference approaches are non-Bayesian in nature. Typically these reconstruction methods do not consider temporal uncertainty in the fossil record and reconstruct climate on a slice-by-slice basis at individual sites with a focus on a single proxy at a time. As a result, these methods are less subject to the problems of computation which plague the Bayesian approaches introduced in subsequent sections. As the focus is on individual sites, we temporarily omit the explicit s notation in the following.

Classical methods for paleoclimate reconstruction can be divided into two contrasting approaches, namely the choice of whether to model the modern proxy data y_k^m as a function of modern climate variables c^m, for example, $y_k^m = f(c^m) + error$, or conversely, to model c^m as a function of y_k^m, for example, $c^m = g\left(y_k^m\right) + error$. This latter case is an inversion of what is understood as the typical cause and effect mechanism in that the environment variable is treated as the response variable and the proxy data the explanatory variable. The former approach, which follows along conventional cause and effect lines, that is, climate → proxy response is referred to by various authors (Haslett et al., 2006; Salonen et al., 2012) as a "forward" modeling approach, and is the foundation of many of the Bayesian approaches to the reconstruction problem. ter Braak (1995) refers to the latter method as an "inverse" modeling approach, a terminology we continue here.

These contrasting choices of approach are inspired in part by the nature of the datasets available for model training, with many proxy datasets (e.g., for pollen, chironomids, or foraminifera) each comprised of up to 300 species (Juggins & Birks, 2012; ter Braak, 1995), and often multivariate climate measurements. If a forward modeling approach is pursued, then the first stage will involve the consideration of models for extremely high-dimensional sum-constrained species counts data, which typically cannot be reasonably explained by simple functional forms of the multivariate climate. The challenges of computation in fitting such models increase with the number of species jointly considered for each proxy and the more nonlinear (or multimodal) the species response is in respect to multivariate climate. In contrast, inverse models avoid these problems by modeling individual climate variables as a function of the multivariate species response, drastically reducing the challenges of computation.

In the interests of brevity we limit our exploration to three of the more commonly used classical approaches for past climate estimation, which include:

1. Modern analogue techniques (MAT)
2. Weighted averaging (WA) and weighted averaging partial least squares (WAPLS)
3. Response surface methods

The first two are so called "inverse" modeling approaches, with the third a "forward" modeling approach.

3.1 | Modern analogue technique

The modern analogue technique (MAT) is the simplest and most intuitive method of estimating the past climate of a fossil proxy sample (Juggins & Birks, 2012), following along the lines of the traditional k-nearest neighbors approach (ter Braak, 1995). Essentially, given a modern training dataset comprised of counts at $i = 1, \ldots, n$ sites for the N species of an individual climate proxy, say pollen, and known climate variables of interest, we find a measure of dissimilarity $\delta_i(d)$ between the fossil sample of an individual proxy slice at depth d, $y_k^f(d)$, and those at each of the $i = 1, \ldots, n$ sites in the modern training dataset. The typical dissimilarity measure for $\delta_i(d)$ is the sum of the squared differences between the fossil pollen of the slice at depth d and the modern pollen at site i. The closest modern analogue for the fossil sample at depth d is the climate of the modern training dataset sample that has the smallest $\delta_i(d)$. A form of smoothing, or robustness, is provided by taking a weighted average (WA) of the climate values of the K most similar modern analogues, ordered by magnitude of $\delta_i(d)$. K is usually chosen as the value that minimizes the RMSE between the observed climates in the training dataset and those predicted for these data by WA of the K most similar (ter Braak, 1995) analogues.

The approach avoids the specification of complex models for climate–proxy interaction, and provides additional benefits: If the magnitudes of the $\delta_i(d)$ for the fossil values of a given slice are large compared to those observed in the training set, then this is an indication that none of the modern analogues are a good match for the fossil sample (Juggins & Birks, 2012). However, Birks, Heiri, Seppä, and Bjune (2010) outline several statistical limitations. First, there is a problem of bias of the estimates at the edges of climate space due to the minority of samples in these regions. Furthermore, extremely large training sets are typically required in order for the method to be effective in providing accurate reconstructions as the method requires a broad coverage of samples in climate space; this becomes increasingly difficult for increasing number of climate variables being considered jointly due to the curse of dimensionality. The method also provides no way to interpolate or extrapolate to climates unobserved in the training set. In terms of challenges of computation, the training datasets considered are not typically large enough to encounter temporal bottlenecks familiar to nearest neighbor methods in the identification of the k-nearest modern analogues.

3.2 | WA and WA-partial least squares

Juggins and Birks (2012) note the popularity of WA approaches to past climate estimation in paleolimnology, citing as a key reason the ecologically appealing conformity of these approaches with *Shelford's law of tolerance* (Shelford, 1931). Shelford's law in principle states that an organism, plant species, or otherwise has a preferred optimal environmental range. On the basis of this, unimodal response models may adequately describe the relationship climate and species response. Toivonen, Manila, Korhola, and Olander (2001) also cite good performance of WA approaches in settings involving noisy, compositional data, that is, data where the counts of individual species are correlated due to the data collection process, which involves counting until a predefined total number of samples is obtained. As each of the species for a given proxy tend to be most abundant at sites with a climate variable close to the species optimum, an estimate of the optimum is thus obtained by a simple WA of the climate values over the n sites at which the species is observed. Model inversion is extremely simple, with the climate estimate for a fossil proxy $y_k^f(d)$ provided as a WA of the $j = 1, \ldots, N$ species optima of that proxy in the sample.

ter Braak (1995) outlines how species tolerances in terms of the breath of the growing range either side of the optima can also be taken into account in a down-weighting fashion, by accounting for the "tolerance" of the species to climate values away from the optimum. This is achieved by giving more weight to the counts of species with more precisely identified (lower-tolerance) optima. Juggins and Birks (2012) note that this can produce moderate improvements over non–down-weighted versions. However, the authors note drawbacks of WA methods including their sensitivity to an uneven distribution of climate values in the training dataset, particularly where the training set is not large. The method also suffers from edge effects, which potentially result in biases in predicted values (ter Braak, 1995). In addition, the method does not account for variability or error in the species record, with zero counts reflecting species unavailability as opposed to sampling error.

These problems motivated the improvement of this simple method by harnessing further information available in the compositional species data, resulting in the WA-partial least squares method (WAPLS) (ter Braak & Juggins, 1993). The approach is simply a combination of WA and partial least squares (PLS), and combines the unimodal response models of WA with the dimension reduction benefits of PLS to address both multicollinearity and residual structure in the species counts (ter Braak, 1995). There are several important limitations to the WAPLS method, however, foremost of which is that the method is designed for the situation where the species–climate relations for a given proxy are unimodal, exhibiting one absolute climate preference, which is not typical for species where subspecies data may be grouped together (such as in the case of pollen). PLS is used to guard against multicollinearity; however, it also implies linearity in the relationships across species which is not necessarily a reasonable assumption. Furthermore, the method may identify structure or patterns in the species observations which are due to other climate variables, as opposed to relationships between species, resulting in biases (Birks et al., 2010).

3.3 | Response surface methods

The response surface approach is a form of modern analogue technique (Birks et al., 2010) and is a forward modeling approach. As opposed to modeling each species response to each climate variable separately, the forward model provides a smoothing of the data over a multidimensional climate domain, which is then used in place of the species compositions to predict the climate associated with a fossil sample. The primary benefits of the approach are both conceptual (modeling proxy response as a function of climate) as well as ecological in that the response surface method allows for more than one climate preference for each species, a problem noted and encountered by several authors (Haslett et al., 2006; Huntley, 1993; ter Braak, 1995).

This multimodal response was first modeled by Bartlein, Prentice, and Webb (1986), who use polynomial regression to estimate the response surfaces for eight pollen species for two climate variables jointly. The global nature of the polynomial bases used for the response surfaces resulted in undesired boundary effects however. Prentice, Bartlein, and Webb (1991) surmount the boundary effects problem by using locally weighted regression to infer nonparametric response surfaces, and thus obtain response surfaces for 13 different pollen species considering three climate variables jointly. Quantitative climate reconstructions are provided from the fitted response surfaces by "inverting" the model as follows:

1. Climate values are inferred for the fossil pollen data by scanning the predicted pollen percentages, discretized to a regular grid, and comparing them to the observed pollen percentages.
2. The 10 climates whose associated pollen compositions are closest to the observed fossil pollen compositions are identified using a squared distance dissimilarity measure. To address multimodality in the output, the final (single) inferred climate value is taken as the centroid of the 10 proposed climate values, each weighted by their inverse squared distances to the fossil sample.

Huntley (1993) cites a number of benefits of the approach over competing methods, noting that they provide a useful explanation of the climate–proxy distribution or abundance patterns and increased resistance to outliers in the pollen record. However, he also notes that, similar to MAT methods, the approach suffers from the "no modern analogue problem", though it does allow for limited interpolation and extrapolation. Further, there is a problem of multiple modern analogues, where individuals among the 10 closest identified can be extremely contrasting in their climate predictions. Taking the centroid, as per Prentice et al. (1991), will result in an aggregated estimate of climate which is potentially far from the 10 nearest identified. Birks et al. (2010) identified further issues including the necessity that modern and fossil information are from the same sedimentary environment in order to minimize the impact of further variation on the process, a result that potentially limits the amount of data available for model training.

3.4 | Further challenges and the uncertainty of estimates

Forward modeling approaches that primarily focus on modeling the observed proxy response as a function of one or more climate variables are hampered by a number of additional challenges, the majority of which are computational in nature:

1. *Likelihood*: When the proxy data are compositional in nature, likelihoods for sum-constrained data such as the multinomial should be specified. However, the complex functional form of the multinomial can result in challenges of inference (Baker, 1994), as the sum constraint requires that parameters of the models for the responses to climate for all species are jointly estimated. As the number of species within a given dataset is potentially large, the number of parameters requiring inference can be much larger than can be feasibly considered in the available computing time. Even Bayesian approaches are not immune to this problem—Haslett et al. (2006) consider a flexible Bayesian nonparametric smoothing model for the multinomial response of 14 pollen species to two climate variables with inference on model parameters taking the order of weeks. Furthermore, their model did not account for additional complications such as zero-inflation within the counts dataset. Addressing this feature would introduce additional modeling complexity and thus further exacerbate the computational burden of inference.
2. *Forward models*: Shelford's law of tolerance (Shelford, 1931) is typically invoked, which states that each species has a preferred optimal climate range, and results in the fitting of simple unimodal models for the climate–proxy relationship. However, these relationships often cannot be described by simple models (Birks et al., 2010; Toivonen et al., 2001; Vasko, Toivonen, & Korhola, 2000), especially in the case of pollen where the counts data for an individual species are formed by aggregating the counts of a number of subspecies, each of which may have a distinct preferred climate range. For example, the pollen of both Pinus (ter Braak, 1995) and Graminaeae (Salonen et al., 2012) exhibit signs of multimodality in their preferred climate ranges. As a result, more flexible (and thus more parameter heavy) models allowing for multiple climate preferences per species are to be preferred, introducing further challenges of computation. If the CO_2 dependence of vegetation response is also accounted for via a mechanistic model, computational challenges worsen further.
3. *Model inversion*: The prediction of past climates using the fitted models is challenging due to the computational complexity of inverting models for prediction, which involves numerical optimization over potentially multimodal response surfaces in several climate variables. This can be difficult due to the multimodal nature of the climate–proxy interaction, and particularly so if several climate variables are jointly considered.

Inverse modeling approaches, which seek to avoid the difficult inversion step required for forward approaches by instead modeling the inverse relationship, also encounter several further difficulties:

1. The climate variables are modeled as a function of highly correlated species counts/proportions. Models based on linear methods, which harness the species counts as predictor variables, common in the paleolimnology literature, will thus suffer from multicollinearity in the species compositions due to the high correlations between species with similar climate preferences.

2. It is not clear how to account for correlation in the relationships between the climate variables as they are unknown or not fully understood, and are difficult to model in this inverse format (ter Braak, 1995). As a result, inverse modeling approaches typically focus on single climate variables at a time which ignores the fact that the proxy response can jointly depend on several climate variables (Huntley, 2012).

In addition to these challenges of whether to adopt forward or inverse modeling approaches, the primary weakness of classical approaches to the climate estimation problem is that there appears to be no consistent way to make statements of uncertainty in the quantitative reconstructions that are produced. None of the introduced approaches adequately quantify and propagate the full range of uncertainties involved in both the modeling and sedimentation processes to the final estimates of climate that are produced. These include issues of temporal and spatial correlation—classical reconstruction methods do not typically consider temporal uncertainty and reconstruct climate on a slice-by-slice basis at individual sites. Paleoclimate reconstructions are typically presented in terms of single climate values that are estimated from multimodal outputs with only cursory measures of uncertainty provided, such as an RMSE or a squared chord distance. The models used are typically simple in nature, and involve the consideration of a limited range of relationships. This deficiency is noted by Holden, Mackay, and Simpson (2008) who state "the major weakness of these [classical] approaches is that they do not explicitly model the uncertainty associated with individual reconstructions," a sentiment also expressed in Haslett et al. (2006) and Birks et al. (2010). Furthermore, Birks et al. (2010) cite "an obsession with models with the lowest RMSE" as being a particular problem with the use of classical approaches and state that the best manner of dealing holistically with the various sources of uncertainty is via the harnessing of the modern Monte Carlo simulation methods of Bayesian statistics.

4 | BAYESIAN JOINT MODELS

The attraction of a Bayesian approach is the potential to allow for the various sources of uncertainty impacting on the reconstruction problem in a holistic and coherent manner (Birks et al., 2010). Whereas classical approaches learn about model parameters from the training datasets and then treat these parameters as fixed constants for prediction (Birks et al., 2010), Bayesian implementations involve the consideration of joint models for the probability of the climate variables of interest, the proxy data, and all other model parameters. The result is a full joint probability distribution on climate which is neatly summarized via climate histories and/or maps. These are individual simulations of climate through time and/or space (Parnell et al., 2016), which carefully reflect each source of climate information and dependence. An example is presented in Figure 4.

However, when performing reconstructions in a Bayesian setting there is a severe computational barrier to be overcome. Typically inference is via Markov Chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1995), a mechanism for simulating from probability distributions with unknown normalizing constants. This is computationally intensive and thus far the challenging task of climate reconstruction in a Bayesian setting has been performed using one of two approaches: (a) simplification of the model to one for which inference is tractable or (b) approximation of the inferential routines. We will

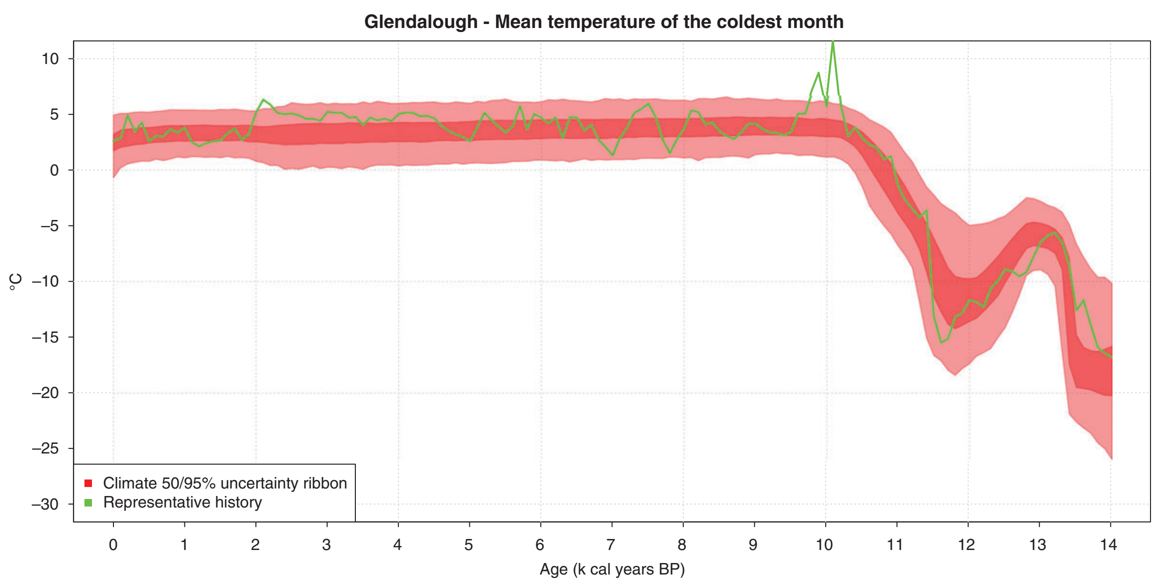

FIGURE 4 Glendalough reconstruction of the mean temperature of the coldest month. The red region represents the 95% probability intervals for climate over time. The darker shading represents the 50% intervals. Overlain in green is a "most representative" climate history across all of the sampled climates

first define the task of Bayesian paleoclimate reconstruction and then introduce approaches under these two categories. We initially constrain the discussion to climate reconstruction at a single site s given a single proxy k, and once more omit explicit s dependence. A further simplification in the following is that the calendar age $a_k(d)$ at each depth is assumed known, and thus no temporal uncertainty in the age of fossil samples is considered.

A key element of Bayesian approaches to climate reconstruction is the forward model which describes the data-generating process (Salonen et al., 2012), that is, the model specified for the response surface which incorporates a priori ecological knowledge to describe the relationship between proxy and climate. As previously, the primary interest is in the predictive distribution for unknown paleoclimate $c^f(t)$ at an individual site s given a sample of fossil proxy information from that site $y_k^f(d)$, that is, $\pi\left(c^f(t)|c^m, y_k^m, y_k^f(d)\right)$, where the end product is a list of plausible climate values with associated posterior probabilities. In the following we denote by θ the unknown parameters of the forward model describing this relationship, which must first be inferred. Following the notation outlined in section 2, again omitting s dependence due to the focus on a single site, the Bayesian formulation is:

$$\pi\left(c^f(t), \theta | c^m, y_k^m, y_k^f(d)\right) \propto \pi\left(y_k^f(d), y_k^m | c^m, c^f(t), \theta\right) \pi\left(c^f(t), \theta\right). \tag{1}$$

where $\pi\left(y_k^m, y_k^f(d)|c^m, c^f(t), \theta\right)$ is the likelihood of observing the proxy data, given the climate measurements, and the model parameters and $\pi(c^f(t), \theta)$ represent any prior climatological beliefs (Salonen et al., 2012). The challenge of inference in the Bayesian setting is that the normalizing constant of the left-hand side is unknown. Brute force estimation of it is intractable due to the high dimensionality of $(c^f(t), \theta)$.

MCMC (Brooks, Gelman, Jones, & Meng, 2011) methods proceed by iterative sampling from a distribution without requiring the normalizing constant. These samples may then be summarized or otherwise interrogated to provide information about $\pi(c^f(t), \theta)$. In Metropolis–Hasting MCMC each successive sample is generated by proposing a stochastic perturbation of the current sample and then either rejecting it or accepting it, thus producing a chain of samples. Whether to accept or reject each proposed sample is based on examination of the product of the ratio of the unnormalized posteriors (left-hand side of Equation 1) and proposal probability densities $\pi\left(c^{f^*}(t), \theta^* | c_i^f(t), \theta_i\right)$, which denotes the probability of proposing a move from sample i in the chain to a new sample indexed with an *.

Detailed theory shows that this scheme does in fact sample from the target, but that it is only guaranteed to do so after an infinite number of iterations. Examination of this routine shows that samples are not independent and that a suitable proposal density must be specified that will allow the chain to move around the posterior target density (mixing). A proposal density that generates large changes in θ will be inefficient as it will rarely leave areas of high posterior probability. Conversely, a proposal that generates conservative moves in $(c^f(t), \theta)$ will generate highly correlated samples and move slowly around the target. Therefore to create an efficient sampler, a sensible choice of proposal is required.

Finally, by integrating this over θ, the posterior distributions for climate will fully reflect the uncertainty in model parameters. In the following we expand on the simplified setting presented here to identify the main statistical challenges hindering Bayesian approaches to past climate estimation, including addressing the uncertainty in the chronology of the fossil record.

4.1 | Unimodal response surfaces based on Shelford's law

We now discuss Bayesian approaches to paleoclimate reconstruction by building up from simple models for which MCMC-based inference is practical to more complex models that necessitate approximate inference, with reference to the relevant literature.

A Bayesian framework for the problem of paleoclimate reconstruction was first described in a series of important papers by authors in the University of Helsinki. First, Toivonen et al. (2001) (released as a working paper in 2000, referenced in Vasko et al., 2000) proposed a Bayesian unimodal response model BUM, invoking Shelford's law of tolerance in order to achieve tractable inference.

Furthermore, in BUM the compositional nature of the data was ignored so that chironomid species could be modeled as responding independently to univariate climate variables (summer surface-water temperature or mean July air temperature). Comparison with WA, WAPLS, and other classical calibration techniques was favorable under cross validation. Vasko et al. (2000) then extend the approach to a Bayesian hierarchical multinomial regression model to address the compositional constraint. They demonstrate that this approach, named BUMMER, outperforms BUM and classical WA-based methods in terms of cross validation to surface-sediment chironomid data; Korhola, Vasko, Toivonen, and Olander (2002) then presented extensive results of the BUMMER model applied to long-term summer temperatures to reconstruct Holocene climate patterns in Finnish Lapland.

More recent Bayesian work by Holden et al. (2008) also invokes Shelford's law and avoids MCMC inference entirely by discretizing the low-dimensional posterior of their simple model. As the data are zero-inflated, presence and abundance-when-present are modeled as functions of a single underlying process, which reduces the number of model parameters. Reconstructions of mean annual temperature based on chironomids or pollen, and pH based on diatoms, using the Holden et al. (2008) model, are available in the the BUMPER (Bayesian user-friendly model for paleo-environmental reconstruction; Holden et al., 2017) software package. They find good performance for chironomids and diatoms but poorer performance for the pollen-based reconstructions, which they attribute to some pollen types comprising multiple species and thus having multimodal responses, violating Shelford's law. This result is also experienced by Salonen et al. (2012), who note the poor performance and bimodal response of several plant species in a pollen application. These examples illustrate the challenging problem of response surface modeling—the computational conveniences of harnessing parametric unimodal response surfaces for the climate–proxy relationship are offset by their unsuitability in applications where species are potentially comprised of several subspecies, such as in the pollen setting.

Holmström, Ilvonen, Seppä, and Veski (2015) attribute the potential multimodal pollen problems to the use of European-wide pollen vegetation datasets. The authors circumvent the issue in a pollen application by limiting the training set to locations in Scandinavia and the Baltic region coincident to the fossil proxy sites; however, this is an undesirable solution as the full amount of model training data potentially available is not utilized, resulting in uncertainty estimates for reconstructions that are potentially naively precise due to the exclusion of subspecies data.

4.2 | Multimodal response surfaces

We now turn our attention to more sophisticated models that require more approximating assumptions in the inferential algorithms or other computational efficiencies to be made. The first serious attempt to address the complexity required by fully Bayesian models for climate reconstruction came in by Haslett et al. (2006). They address the issues of univariate climate variable modeling and multimodality in the pollen response surfaces. Unlike the BUMMER model, which fits to a single climate variable at a time, responses are jointly modeled on two climate dimensions. This is performed nonparametrically to allow for the multimodal responses observed in pollen species. The term "climate-space" is used to refer to this two-dimensional (2D) climate and the variables chosen were aspects of climate that plants (and thus pollen) are sensitive to, namely growing degree days above 5 °C (GDD5, a measure of the length of growing season) and mean temperature of the coldest month (MTCO, a measure of harshness of winter). Fourteen pollen species were selected, with each having a distinct preferred climate in terms of these two variables.

However, computational overhead was the primary obstacle, with MCMC-based inference of the high-dimensional posterior having run times being of the order of weeks, despite measures taken to improve efficiency of the algorithms and running on high-performance computers. Thus, cross validation to compare models and to assess accuracy of reconstruction was impossible. In order to model the response surfaces in a nonparametric fashion each species response in 2D climate space is modeled as smooth, but with no constraint on shape of response. This model was well suited to pollen data where species that respond quite differently to climate may have indistinguishable pollen spores. This required a response parameter in θ for each of the approximately 8,000 modern sampling sites and a large-scale hierarchical Bayesian model was formed with inference via MCMC. The article was also the first to attempt to coherently account for temporal correlation in the fossil record by sampling paleoclimates conditional on the fitted models and the fossil pollen data—a t_8 distribution for the smoothness of the paleoclimate was imposed as a prior, informed by Greenland ice core data, and the paleoclimates were thus modeled jointly in a temporal sense. However, despite of the number of advances, the paper also identified several remaining challenges:

- The nonparametric modeling of responses requires high numbers of unknown parameters. This leads to a long running time for the MCMC-based methodology, and poor mixing and convergence, that is, the typical model fitting issues when using MCMC methods.
- Zero-inflation of the pollen counts where sampling sites may not have had particular species present, despite a suitable climate, is not addressed. This results in many additional zeros in the data over and above that explained by simple counts models and potential underestimation of species responses.
- Dependency among species over and above that caused by the constraint of sum-to-one nature of compositional counts. Vasko et al. (2000) showed that accounting for the compositional nature of data collection methods improved model fit to chironomid assemblage data; however, there is dependency beyond this simple model that is due to competition among pollen species/species.
- The laminar nature of the Greenland ice core that inspired a t_8 model for climate change is unsuitable for the uneven time sampling of the fossil proxy data such as occurs with pollen.

- The dates of the fossil pollen are assumed known rather than uncertain. Radiocarbon dating of a subset of the slices of the sediment core and linear interpolation are crude approximations to the true processes involved.

In light of these shortcomings, several attempts to improve the model have been attempted while simultaneously addressing the computational complexity issue. In particular, Salter-Townshend and Haslett (2006) introduce a parsimonious model for the overabundance of zero counts. The probability of absence from a sampling site is assumed to be functionally related to the abundance when present so that the response surfaces now account for both the abundance when present and the probability of presence. This model introduces a single additional parameter for each pollen species modeled and model fit is shown to be superior in terms of cross-validation prediction accuracy. Salter-Townshend and Haslett (2012) then used this model along with a nesting structure on the species to account for additional dependency (richer covariance structure) and demonstrate superior performance in terms of cross validation of the modern data.

In order to accommodate these modeling extensions, MCMC-based inference is replaced by an integrated nested Laplace approximations approach (Rue, Martino, & Chopin, 2009), which speeds up the inference tasks by several orders of magnitude. However, this comes at the cost of enforcing compromises in the likelihood structure. The continuous 2D climate space is approximated by a regularly spaced 50×50 lattice and the climate measurements of each observation adjusted to their nearest grid point. Flexibility in the response surfaces, and efficiencies in computation, are achieved by imposing a Gaussian Markov random field (GMRF; Rue & Held, 2005) on this regular lattice, making a discrete approximation to the continuous nonparametric multivariate Gaussian response surface model. A GMRF approximation of the gridded θ response surfaces posterior is then found, without recourse to MCMC or other sampling methodology. This approximation is demonstrated to be highly accurate; however, the GMRF-based approach does not currently extend to climate dimensions greater than three due to the substantial computational cost imposed by the discretization of multidimensional climate space.

4.3 | Accounting for temporal uncertainty in the chronology

In order to fully account for temporal uncertainty in the climate reconstructions, the prediction for climate at time t must take into account the uncertainty in the relationship between the unknown calendar ages $a_k(d)$ of the fossil slices at depth d, which are estimated from the associated radiocarbon age $r_k(d)$ obtained from a laboratory. Addressing some further limitations in Haslett et al. (2006), Haslett and Parnell (2006) introduce a model for radiocarbon-dated depth-chronologies to address varying sedimentation rates, where depth d and age a are not linearly related, and model the uncertainty in the fossil dates jointly. An accompanying R package Bchron (Parnell, 2016) performs age-depth modeling and date calibration with uncertainty. Parnell et al. (2015) use this model, and a normal inverse gamma process prior, to model the stochastic volatility of paleoclimate for a number of pollen cores. By making two small and conservative simplifying assumptions to the model (firstly that unobserved paleoclimate and fossil pollen contribute negligible information to learning response surface model parameters and secondly that the expected impact of a changing climate on the sedimentation process is zero), the reconstruction task can be broken down into three discrete stages:

1. Response surface module: $\pi\left(\theta | y_k^m, c^m\right)$
2. Chronology module: $\pi(a, \psi | r, d)$
3. Reconstruction module:

$$\pi\left(c^f(t), a_k, \theta, \psi, \nu | y_k^f(d), r, d, y_k^m, c^m\right) \propto,$$

$$\prod_{i=1}^{N^f} \pi\left(y_k^f(d_i) | c^f(a_k(d_i)), \theta\right) \pi\left(c^f(t) | a_k(d_i), \nu\right) \pi(a_k, \psi | r, d) \pi\left(\theta | y_k^m, c^m\right) \pi(\nu).$$

where $a_k(d_i)$ is the unobserved calendar age at depth d_i and N^f is the number of fossil pollen slices, each of which contain the counts of the N species (i.e., $N^f \times N$ pollen counts). Furthermore, ν are parameters for the climate process, ψ are a set of parameters governing the sedimentation process (i.e., linking age and depth), r is the radiocarbon age of a sample, and d is the depth, as previously. Each of the stages is still computationally intensive in their own right. Computational savings are made by *pre-processing* the response surface posteriors of the forward model stage, resulting in efficient low-dimensional MCMC inference of the jointly inferred posterior for paleoclimate in stage 3. Specifically, marginal data posteriors (MDPs) are first calculated—these are independent posteriors for climate given pollen only (i.e., no other fossil slices) for slice i. Assuming known y_k^m, c^m,

$$\pi\left(c^f(a_k(d_i))|y_k^f(d_i),y_k^m,c^m\right) \propto \pi\left(y_k^f(d_i)|c^f(a_k(d_i)),y_k^m,c^m\right)\pi\left(c^f(t_i)\right), \tag{2}$$

$$\propto \int \pi(y^f(d_i)\,|\,c^f(a_k(d_i),\theta)\pi\left(\theta|y_k^m c^m\right)d\theta. \tag{3}$$

$c^f(t = a_k(d_i))$ is assigned a flat prior; changes in climate are modeled without making a priori statements about marginal values at a slice. The MDPs are approximated with a mixture of Gaussians to simplify integration steps, with the mixture approximation performed once per slice, before integration with the depth-chronology part of the model.

Most importantly, this modular form is computationally attractive as new sites for paleoclimate reconstruction can be analyzed without re-doing the computationally expensive response surface module. In Figure 4 we present a temperature reconstruction (mean temperature of the coldest month) for a pollen core at Glendalough, Ireland, over the past 14k years which coherently brings together uncertainty in the sedimentation history age, in addition to model uncertainty, using the approach outlined in Parnell et al. (2015) and implemented in the R package Bclim (Parnell & Sweeney, 2016).

Other authors have since utilized the benefits of this modular form for the reconstruction problem—Ilvonen, Holmström, Seppä, and Veski (2016) reconstruct Finnish mean annual temperature, this time using the BUMMER model for the response surface module, but in conjunction with the Bchron model for depth-chronologies. Unimodal responses are justified as only a single climate variable is modeled, modern training data are carefully chosen to be very focused, and the unimodal assumption is appropriate.

5 | EXTENSIONS TO SPATIAL, MULTIPROXY, AND MECHANISTIC MODELS

In this section, we review some recent approaches that build on the Bayesian approaches outlined in previous sections. These fall into the broad categories of spatial models, where multiple datasets are combined across sites with a view to a spatiotemporal climate reconstruction; mechanistic models, which aim to incorporate physical processes into the model; and multiproxy models, where multiple climate proxies are combined in a consistent manner to utilize more information, and so reduce uncertainty. In Figure 5, we seek to provide a summary of the structure of four of the approaches we review, linking them back to

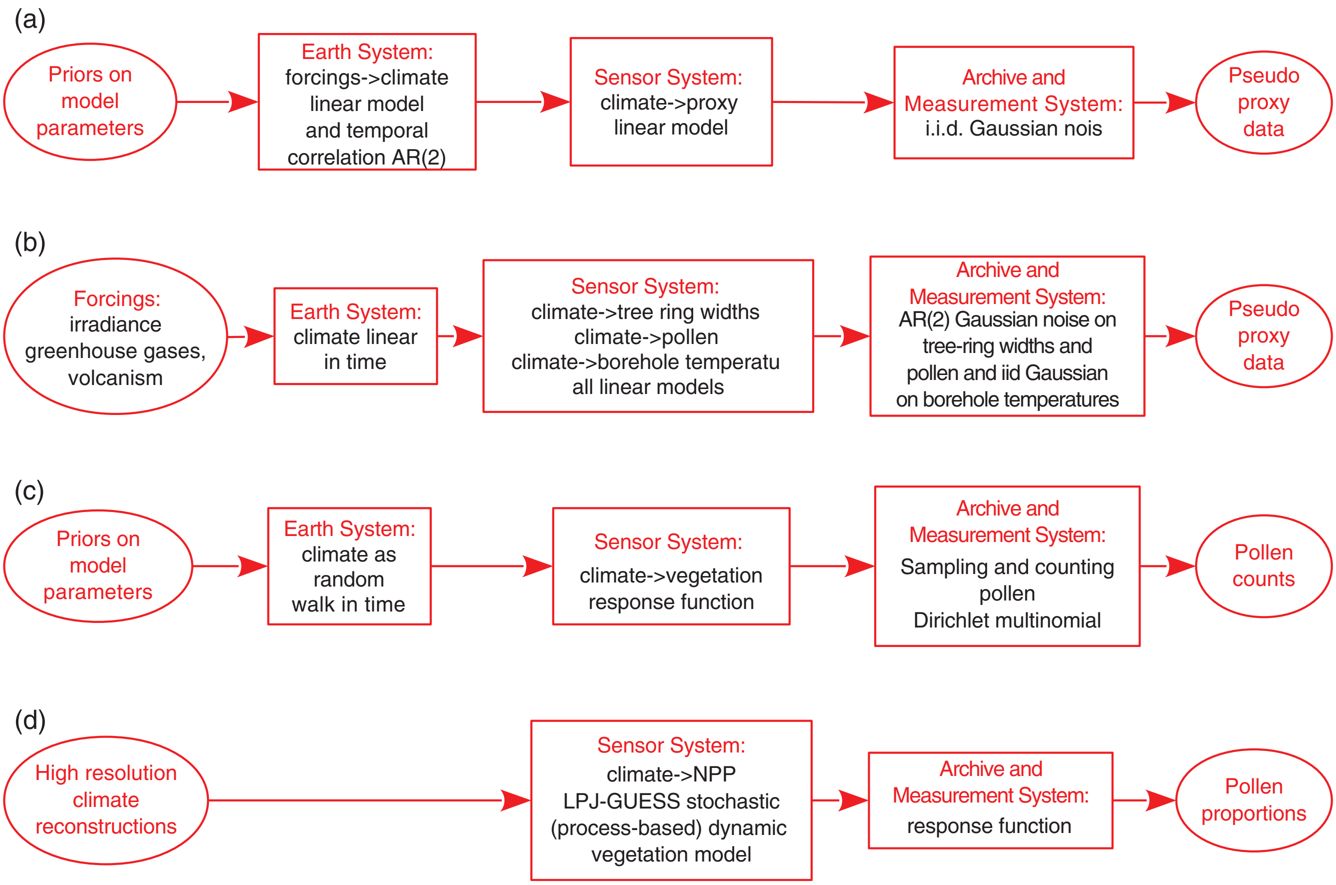

FIGURE 5 Schematic representations of the link between climate and proxies for four of the journal articles summarized in this paper. (a) Relates to Tingley and Huybers (2010), (b) to Li et al. (2010), (c) to Haslett et al. (2006) and (d) to Garreta et al. (2009)

the structure for pollen-based climate reconstruction in Figure 3. The idea with all these extensions is simultaneously to reduce uncertainties and allow for more detailed causal analysis of the parameters governing climate change. When fitted into the Bayesian framework, all of these approaches are only in their infancy. Many of the papers referenced are proof-of-concept attempts toward the goal of combining data in joint probabilistic models. There has been even slower progress made on the meta-combination of spatial, multiproxy, *and* mechanistic models, and we hope that this is a key goal of future research.

5.1 | Spatial and spatiotemporal approaches

In spatial approaches to paleoclimate reconstruction, the target of inference is $c^f(s)$, where s denotes a location in space. This may be two- or three-dimensional if altitude is further included with latitude and longitude. When time is included as well the target is $c^f(s, t)$. Much progress has been made in developing advanced spatial statistical models for uncertain data (Lindgren, Rue, & Lindström, 2011), and some of these approaches have been applied in paleoclimate research (Salter-Townshend & Haslett, 2012). However, in the main, the approaches taken in the paleoclimate literature use traditional Gaussian process approaches (Banerjee, Gelfand, Finlay, & Sang, 2008). This Gaussian process approach proceeds by defining a correlation function by which neighboring sites will have similar climate values. The degree to which sites are deemed "neighboring" is determined by the correlation function chosen and the distance between sites, and is controlled by one or more unknown parameters.

We cover two of the most widely read and cited papers, which also follow the Bayesian forward approach outlined previously, and so are compatible with many of the other ideas in this section. A key challenge is that the data are usually irregularly observed time series at each site. The challenge is to temporally align the series so as to produce a spatiotemporal grid of climate values. This is most effectively achieved by using a statistical model that works in continuous time, for example, the continuous time stochastic volatility model of Parnell et al. (2015).

The first approach we discuss is that of Tingley and Huybers (2010) (known as BARCAST), which aims to produce a spatial reconstruction of temperature based on the Climatic Research Unit (CRU) dataset (Brohan, Kennedy, Harris, Tett, & Jones, 2006) using pseudo-proxies (simulated proxy data) to validate the approach. Adjusting their notation slightly, they work with discretized time, and write:

$$\begin{bmatrix} c_t^f(s_1) \\ c_t^f(s_2) \\ \vdots \\ c_t^f(s_n) \end{bmatrix} = \alpha \begin{bmatrix} c_{t-1}^f(s_1) \\ c_{t-1}^f(s_2) \\ \vdots \\ c_{t-1}^f(s_n) \end{bmatrix} + \epsilon_t,$$

where $c_t^f(s_i)$ indicates the mean temperature at time t and location s_i. The discrete time approach is acceptable here because the CRU dataset they use is gridded and so is amenable to standard auto-regressive models. The spatial aspect is captured in ϵ_t which is given a multivariate Gaussian process with covariance matrix Σ, such that:

$$\Sigma_{ij} = \sigma^2 \exp\left(-\varphi|s_i - s_j|\right).$$

Thus the model is space–time separable with the spatial field unchanging over time. This is a severe simplification of reality but, given the complexities of the datasets involved, remains computationally tractable. The above equations form the spatiotemporal process part of the model, with further parts being added to take account of the proxy data.

A more advanced approach is that of Holmström et al. (2015), which allows for more realistic data with differing chronologies (i.e., differing time scales) for different sites, and still produces gridded spatiotemporal climate reconstructions. Again, adjusting their notation to match ours, they have:

$$c^f\left(s_i, t_j\right) = c^f\left(s_i, t_{j-1}\right) + \frac{1}{\sqrt{\kappa\left(t_j - t_{j-1}\right)}} \epsilon_{t_j},$$

where t_j now represents continuous time point j at location s. κ here represents a time smoothing parameter, and ϵ_{t_j} captures the spatial covariance, again given the exponential form as in Tingley and Huybers (2010) above. The time difference $t_j - t_{j-1}$ accounts for nonregular temporal differences between sites. Holmström et al. (2015) fit their model to a set of four lakes using MCMC techniques. They use informative prior distributions on many of the parameters of interest but ignore the time uncertainty in each of the four lake chronologies. The spatial smoothness parameter is informed by climate model simulations. The space–time separability of the covariance, despite still suffering from many of the drawbacks of the BARCAST model, enables some computational shortcuts.

5.2 | Mechanistic approaches

In mechanistic approaches, physical processes are included in the model. These physical processes can range from the inclusion of simple differential equations governing climate change over time, to advanced models involving multidimensional stochastic partial differential equations. Although the goal is, as always, to reduce uncertainties and increase explanation, these types of models involve a considerable computational overhead which is exacerbated when incorporated into a Bayesian model due to the repeated simulations/iterations that are often required to capture uncertainty. There is a long literature of mechanistic models used in paleoclimate (Crucifix, 2012). Here we focus solely on papers that discussed the embedding of mechanistic models in a statistical framework, and can be included in the general Bayesian solution as posed at the start of the paper. An excellent discussion of the issues involved in using statistical methods with mechanistic models can be found in Crucifix (2012).

There are two primary places where mechanistic models can be incorporated. The first is in the climate component of the model, for example, replacing the statistical time series model (Ilvonen et al., 2016; Parnell et al., 2015) with a set of differential equations. Although the time series approaches have parameters that represent, say, smoothness or volatility of climate over time, the differential equations can allow for parameters that capture mechanistic climate feedbacks or the complex effects of other variables or forcings. In some cases, for example, Li et al. (2010), the time series model may incorporate both statistical and mechanistic ideas. The second place where mechanistic models can be incorporated is in the transfer from proxy to climate: the forward model. As described above, a statistical model may capture the main features of the proxy–climate relationship, but may not allow for known mechanistic actions of the proxy, such as being able to incorporate the CO_2 dependence of the response. Cases where proxies may compete, or when differential lags occur between proxy and climate, may also be particularly suited to mechanistic involvement.

Tolwinski-Ward et al. (2015) present a forward model for tree rings, which contain some mechanistic elements. They reconstruct two climate variables: temperature and moisture, and define ramp functions for each, which represents the tree-ring growth response. The parameters of these growth response functions represent the limits at which the trees will grow. The remainder of the model is fitted using the Bayesian approach.

A far more sophisticated forward model is used by Garreta et al. (2009). They build a statistical framework that incorporates the stochastic LPJ-GUESS vegetation model (Smith et al., 2014), which simulates pollen counts from climatic inputs. The vegetation model accounts not only for the production of pollen based on climate, but also includes pollen dispersal and spatial accumulation. This involves estimating a far richer set of parameters governing such relationships, which causes considerable computational challenges.

The state of the art in mechanistic modeling of paleoclimate over time is that of Carson, Crucifix, Preston, and Wilkinson (2017). They evaluated a competing set of stochastic differential equation models over the glacial–interglacial cycle. Using some of the more recent statistical methods, for example, particle MCMC (Andrieu, Doucet, & Holenstein, 2010), they were able to efficiently estimate the parameters of the competing models, and subsequently the marginal likelihoods and Bayes factors. The results and methods used in the paper seem highly promising for future research directions when including mechanistic models.

5.3 | Multiproxy

The idea of combining multiple proxies together dates back to the seminal papers of Mann, Bradley, and Hughes (1999) and before, where classical (non-Bayesian approaches) are adopted. These methods, such as the composite plus scaling method (Jones et al., 2009), regress standardized and weighted multiple modern proxies (e.g., tree-ring, marine sediment, speleothem, lacustrine, ice core, and coral data) against the modern instrumental record to combine into an average representation of the temperature histories originally constructed only on the basis of the individual records (Holmström et al., 2015). This represents a multiproxy extension of the inverse methods introduced in earlier sections. We caution against this approach to multiproxy analysis as, similar to the classical inverse approaches, sources of uncertainty within and across proxies are not fully and coherently accounted for. In this regard, McShane and Wyner (2011) carried out a careful Bayesian analysis of the Mann et al. (1999) dataset, and concluded that the reconstructions provided in the article are perhaps unreliable; although their mean reconstructions do replicate the "hockey stick" shape found by Mann et al., they found very large uncertainties and speculated that the long "handle" shape is due to regression to the mean of the model rather than a climate signal.

The overriding challenge in multiproxy reconstructions is to take account of the differing relationships between the proxies and the climates, and to account for uncertainty in both. A list of the potential problems in these relationships can be found in Huntley (2012). The Bayesian solution to this problem is to stitch together forward models in a Bayesian likelihood assuming some conditional independencies:

$$\pi\left(y_1^m, y_2^m, \ldots, y_M^m \mid c_1^m, c_2^m, \ldots\right) = \prod_{k=1}^{M} \pi\left(y_k^m \mid c_1^m, c_2^m, \ldots\right)$$

where y_k represents the proxy data for proxy $k = 1, \ldots, M$, and $c_1, c_2, \ldots$ represents the different climate variables. The assumption here is that, when all the important climate variables are known, the proxies are conditionally independent and forward models can be built for them separately. In this sense, even multiple variables of the same proxy type (e.g., pollen counts from different species) can be treated as separate proxies. This framework is in direct contrast to the approach of Mann et al. (1999), which assumes that proxies are observed without uncertainty and marginally independent, that is, independent sources of information.

Surprisingly, given that the above framework allows for simple combinations of proxies, there is relatively little literature on the combination of substantially different proxy types. This may be in part because, although the mathematics is relatively simple, in practice different proxy types respond to different but related aspects of climate, so tying them together can be a challenge. For example, some plant pollen counts may respond to the harshness of the winter, while certain trees may respond to the length of the growing season. Both these climate variables are correlated, and so any climate model (either stochastic or mechanistic) must estimate these jointly. Another challenge is that the proxies may respond to climate variables on different timescales, but this problem is already present in many multispecies single proxy reconstructions (Haslett et al., 2006).

The approach outlined above was first described in detail by Li et al. (2010) using a simulation (pseudo-proxy) dataset combined with a simple climate model to reconstruct a univariate temperature variable. They reconstruct northern hemisphere mean temperature using tree-ring, pollen, and borehole data, with different regression type models on each. This is a clear improvement on the multiproxy methods of Mann et al. (1999) but lacks the richness of the forward models proposed by, for example, Ohlwein and Wahl (2012).

A more focused approach using real data from multiple proxies in the Bayesian framework is that of Cahill, Kemp, Horton, and Parnell (2016). In their example, the variable to be reconstructed is sea level at a specific site. The proxy data are foraminifera which live within the tidal range, and a stable isotope measurement (δ^{13}C) which provides an additional constraint. The forward model is a Bayesian nonparametric spline, with a Gaussian process to model the changing rates of sea level. The multiproxy model works well here because both proxies provide information on a single climate variable of interest. We hope such models will find more widespread use.

6 | DISCUSSION

The ultimate goal of paleoclimate reconstruction is to estimate the mechanics of past climate given all available data. These reconstructions provide an understanding of past climate and of environmental changes, provide a method for the evaluation of climate models and the uncertainty in their estimates, and help to improve our predictions of the future. In this article, we have provided an overview of the methods currently used to achieve this goal and identified that the challenges involved are multidisciplinary, comprising problems of an ecological, computational, and statistical nature. We conclude the article by touching briefly on a number of these issues, and proposing further areas for development.

From an ecological point of view, the challenges include a proper addressing of the quality and consistency of the data used for model training (Salonen et al., 2012), which are subject to errors in the identification of the proxy data, as well as errors of omission such as the expression of proxy data in proportion rather than count form. Another challenge is the addition of further sources of proxy information to the modern training record (Birks et al., 2010), with the hope that this will result in improvements in the precision of climate inferences. Furthermore, there is a requirement to develop a broader understanding of the climate variables, which drive the response of individual proxies—the absence of important explanatory climatological variables results in confounding correlations between proxies being identified, and potentially erroneous inferences being made.

From a statistical point of view, the challenges are numerous. In this article, we have presented an attractive modular form for paleoclimate estimation, which breaks the paleoclimate reconstruction challenge into separable modules of forward model building, the addressing of spatial and temporal uncertainty, and the harnessing of mechanistic models. This enables the embedding of the reconstruction process in a Bayesian statistical framework, which allows for coherent and holistic accounting of all the sources of uncertainty that impact at each stage of the process. This modular form allows the isolation of a number of key statistical challenges, each of which offers the scope for substantial methodological contributions. One challenge is the requirement to move from simplistic one-dimensional unimodal forward models, to flexible modeling approaches which allow for individual species to express multiple climate preferences in multiple climate dimensions. A further progression will hopefully involve the simultaneous harnessing of several forward models for climate estimation, as opposed to the present use of individual models, by weighting models with a model averaging approach (Raftery, Madigan, & Hoeting, 1997). Surprisingly little research has been carried out in this regard and we see it as an area of substantial research potential, in addition to the development of more refined forward models.

A more fundamental challenge is to move away from the uniformitarianism principle of current methods via the incorporation of mechanistic models into the estimation process. This offers the scope to address the issue of a lack of modern analogues for fossil samples; however, these models require understanding of complex processes, and testing and evaluation with data, and may present substantial computational challenges. Indeed, perhaps the most pressing and useful contribution is via the development of software for the dissemination of Bayesian approaches and methods for the reconstruction problem, and the speeding up of inference through computational advances. Unfortunately, existing Bayesian approaches are often regarded as slow and computationally intensive and these problems are perhaps the most substantial impediment to their adoption by researchers (Birks et al., 2010) who currently favor classical approaches.

ACKNOWLEDGMENTS

The research of Professor Parnell, Professor Buck, and Dr Edwards was supported by the Leverhulme International Network Grant F/00118/BE: Studying Uncertainty in Palaeoclimate Reconstructions: A Network (SUPRA-net). Further research funding was also provided by the EPSRC-funded Past Earth Network, grant reference: EP/M008363/1.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLES

On nearest-neighbor Gaussian process models for massive spatial data
Spatial modeling with system of stochastic partial differential equations
Computational methods for climate data

REFERENCES

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *72*(3), 269–342. https://doi.org/10.1111/j.1467-9868.2009.00736.x

Baker, S. G. (1994). The multinomial-Poisson transformation. *Journal of the Royal Statistical Society, Series D: The Statistician*, *43*(4), 495–504.

Banerjee, S., Gelfand, A. E., Finlay, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B*, *70*(4), 825–848.

Bartlein, P. J., Prentice, I. C., & Webb, T. (1986). Climate response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, *13*, 35–57.

Birks, H. J. B., Heiri, O., Seppä, H., & Bjune, A. E. (2010). Strengths and weaknesses of quantitative climate reconstructions based on late-quaternary biological proxies. *The Open Ecology Journal*, *3*, 68–110.

Blackwell, P. G., & Buck, C. E. (2008). Estimating radiocarbon calibration curves. *Bayesian Analysis*, *3*(2), 225–248. https://doi.org/10.1214/08-BA309

Briffa, K. R., Shishov, V. V., Melvin, T. M., Vaganov, E. A., Grudd, H., Hantemirov, R. M., … Naurzbaev, M. M. (2008). Trends in recent temperature and radial tree growth spanning 2000 years across northwest Eurasia. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *363*(1501), 2269–2282. https://doi.org/10.1098/rstb.2007.2199

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research*, *111*(D12), D12106. https://doi.org/10.1029/2005JD006548

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Cahill, N., Kemp, A. C., Horton, B. P., & 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. https://doi.org/10.5194/cp-12-525-2016

Carson, J., Crucifix, M., Preston, S., & Wilkinson, R. D. (2017). Bayesian model selection for the glacial-interglacial cycle. *Journal of the Royal Statistical Society: Series C: Applied Statistics*. https://doi.org/10.1111/rssc.12222

Crucifix, M. (2012). Traditional and novel approaches to palaeoclimate modelling. *Quaternary Science Reviews*, *57*, 1–16.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., … Jouzel, J. (1993). Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, *364*, 218–220. https://doi.org/10.1038/364218a0

Doan, T. K., Haslett, J., & Parnell, A. C. (2015). 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. https://doi.org/10.5194/ascmo-1-15-2015

Garreta, V., Miller, P. A., Guiot, J., Hély, C., Brewer, S., Sykes, M. T., & Litt, T. (2009). A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model. *Climate Dynamics*, *35*(2–3), 371–389. https://doi.org/10.1007/s00382-009-0629-1

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. CRC Press.

Haslett, J., & Parnell, A. (2006). A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society, Series C*, *57*(4), 399–418.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., … Mitchell, F. J. G. (2006). Bayesian palaeoclimate reconstruction. *Journal of the Royal Statistical Society, Series A*, *169*(3), 1–36.

Holden, P. B., Mackay, A. W., & Simpson, G. L. (2008). A Bayesian palaeoenvironmental transfer function model for acidified lakes. *Journal of Paleolimnology*, *39*(4), 551–566.

Holden, P. B., John B Birks, H., Brooks, S. J., Bush, M. B., Hwang, G. M., Matthews-Bird, F., … van Woesik, R. (2017). BUMPER v1. 0: A Bayesian user-friendly model for palaeo-environmental reconstruction. *Geoscientific Model Development*, *10*(1), 483–498.

Holmström, L., Ilvonen, L., Seppä, H., & Veski, S. (2015). A Bayesian spatiotemporal model for reconstructing climate from multiple pollen records. *Annals of Applied Statistics*, *9*(3), 1194–1225.

Hughes, M. K., Swetnam, T. W., & Diaz, H. F. (2011). *Dendrochronology: Progress and prospects Developments in Paleoenvironmental Research*. Springer Netherlands (Vol. 11).

Huntley, B. (1993). The use of climate response surfaces to reconstruct paleoclimate from quaternary pollen and plant macrofossil data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *341*, 215–224.

Huntley, B. (2012). Reconstructing palaeoclimates from biological proxies: Some often overlooked sources of uncertainty. *Quaternary Science Reviews*, *31*, 1–16. https://doi.org/10.1016/j.quascirev.2011.11.006

Ilvonen, L., Holmström, L., Seppä, H., & Veski, S. (2016). A Bayesian multinomial regression model for palaeoclimate reconstruction with time uncertainty. *Environmetrics*, *27*(7), 409–422. https://doi.org/10.1002/env.2393

Jones, P. D., Briffa, K. R., Barnett, T. P., & Tett, S. F. B. (1998). High-resolution palaeoclimatic records for the last millennium: Interpretation, integration and comparison with general circulation model control-run temperatures. *The Holocene*, *8*(4), 455–471. https://doi.org/10.1191/095968398667194956

Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., Van Ommen, T. D., Vinther, B. M., … Xoplaki, E. (2009). High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *The Holocene*, *19*(1), 3–49.

Juggins, S., & Birks, H. J. B. (2012). Quantitative environmental reconstructions from biological data. In J. B. H. Birks, A. F. Lotter, S. Juggins, and J. P. Smol (Eds.), *Tracking environmental change using lake sediments: Data handling and numerical techniques* (pp. 431–494). Springer Netherlands.

Klauenberg, K., Blackwell, P. G., Buck, C. E., Mulvaney, R., Röthlisberger, R., & Wolff, E. W. (2011). Bayesian glaciological modelling to quantify uncertainties in ice core chronologies. *Quaternary Science Reviews*, *30*(21), 2961–2975. https://doi.org/10.1016/j.quascirev.2011.03.008

Korhola, A., Vasko, K., Toivonen, H. T. T., & Olander, H. (2002). Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling. *Quaternary Science Reviews*, *21*, 1841–1860.

Li, B., Nychka, D. W., & Ammann, C. M. (2010). The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association*, *105*(491), 883–911.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, *73*(4), 423–498.

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

Mann, M. E., Bradley, R. S., & Hughes, M. K. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, *26*(6), 759–762.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., & Ni, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, *105*(36), 13252–13257. https://doi.org/10.1073/pnas.0805721105

McShane, B. B., & Wyner, A. J. (2011). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, *5*, 5–44.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biology*, *9*(8), 1–8. https://doi.org/10.1371/journal.pbio.1001127

Ohlwein, C., & Wahl, E. R. (2012). Review of probabilistic pollen-climate transfer methods. *Quaternary Science Reviews*, *31*, 17–29.

Andrew Parnell (2016). *Package Bchron*. Retrieved from https://github.com/andrewcparnell/Bchron

Parnell, A. C., Buck, C. E., & Doan, T. K. (2011). Invited review. *Quaternary Science Reviews*, *30*(21–22), 2948–2960. https://doi.org/10.1016/j.quascirev.2011.07.024

Parnell, A. C., Haslett, J., Sweeney, J., Doan, T. K., Allen, J. R. M., & Huntley, B. (2016). Joint palaeoclimate reconstruction from pollen data via forward models and climate histories. *Quaternary Science Reviews*, *151*, 111–126.

Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R. M., Huntley, B., & 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.

Andrew Parnell and James Sweeney. (2016). *Package Bclim*. Retrieved from https://github.com/andrewcparnell/Bclim

Prentice, I. C., Bartlein, P. J., & Webb, T. (1991). Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology*, *72*(6), 2038–2056.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications Monographs on Statistics and Applied Probability* (Vol. 104). London, England: Chapman & Hall.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, *71*, 319–392.

Salonen, J. S., Ilvonen, L., Seppä, H., Holmström, L., Telford, R. J., Gaidamavičius, A., … Subetto, D. (2012). Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative climate reconstruction. *The Holocene*, *22*(4), 413–424.

Salter-Townshend, M., & Haslett, J. (2006). Zero-inflation of compositional data. *Proceedings of the 21st International Workshop on Statistical Modelling*, *21*, 448–456.

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

Shelford, V. E. (1931). Some concepts of bioecology. *Ecology*, *12*(3), 455–467.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., & Zaehle, S. (2014). Implications of incorporating *n* cycling and *n* limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, *11*(7), 2027–2054. https://doi.org/10.5194/bg-11-2027-2014

Stocker, T. (2014). *Climate change 2013: The physical science basis*. Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

ter Braak, C. J. F. (1995). Non-linear methods for multivariate statistical calibration and their use in palaoecology: A comparison if inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. *Chemometrics and Intelligent Laboratory Systems*, *28*, 165–180.

ter Braak, C. J. F., & Juggins, S. (1993). Weighted averaging partial least squares regression (WA-PLS): An improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, *269*(1), 485–502. https://doi.org/10.1007/BF00028046

Tingley, M. P., & Huybers, P. (2010). A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *Journal of Climate*, *23*(10), 2759–2781.

Toivonen, H. T. T., Manila, H., Korhola, A., & Olander, H. (2001). Applying Bayesian statistics to organism-based environmental reconstruction. *Ecological Applications*, *11*(2), 618–630.

Tolwinski-Ward, S. E., Tingley, M. P., Evans, M. N., Hughes, M. K., & Nychka, D. W. (2015). 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.

Vasko, K., Toivonen, H., & Korhola, A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *Journal of Paleoclimnology*, *24*, 243–250.

Wilson, J. B., Peet, R. K., Dengler, J., & Pärtel, M. (2012). Plant species richness: The world records. *Journal of Vegetation Science*, *23*(4), 796–802. https://doi.org/10.1111/j.1654-1103.2012.01400.x

How to cite this article: Sweeney J, Salter-Townshend M, Edwards T, Buck CE, Parnell AC. Statistical challenges in estimating past climate changes. *WIREs Comput Stat*. 2018;10:e1437. https://doi.org/10.1002/wics.1437