Climate Reconstruction Using Tree-Ring Data

Matthew R. Schofield[1]*, Richard J. Barker[2], Andrew Gelman[3] and Edward R. Cook[4]

[1]Department of Statistics, University of Kentucky, Lexington, KY, USA.

[2]Department of Mathematics and Statistics, University of Otago,

PO Box 56, Dunedin, New Zealand.

[3]Department of Statistics, Columbia University, NY, USA.

[4]Lamont–Doherty Earth Observatory, Palisades, NY, USA.

Abstract

blah, blah, blah...

1 Introduction

Instrument records from the past 200 years support a scientific consensus that climate is changing Trenberth et al. (2007). What is less clear is how to interpret these changes in an historical context spanning many hundreds, or thousands, or tens of thousands of years. The statistical question is this: how do we quantify and model historical climate without having direct observations beyond the most recent one or two centuries? The most common solution is to build an index of historical climate change using long-scale proxy observations that are constructed from data such as tree ring measurements, lake sediments, bore-hole observations, coral records and ice cores (see Jones and Mann (2004) for a review as well as Jansen et al. (2007)). Reconstructions based on these sources feature prominently in the 2007 Intergovernmental Panel on Climate Change report (Jansen et al. 2007).

There are two steps necessary to reconstruct climate from proxy data. Step one involves filtering out any non-climate features so that only climatic features remain in the proxy. Step two involves using the filtered proxy data, either alone, or combined with other proxy measurements, to reconstruct climate at either the local or global scale. While some studies have not paid enough attention to the first

*E-mail: matthew.schofield@uky.edu

step, there is nonetheless a rich literature devoted to preserving the climate signal during this filtering process (for example, Briffa et al. 1992, Cook et al. 1995, Briffa et al. 2001, Esper et al. 2002, Osborn and Briffa 2006, Esper et al. 2009, Melvin and Briffa 2008, Briffa and Melvin 2009). Even so, the major focus has been on the second step. This has resulted in several methodological developments in how to reconstruct the climate from the filtered proxy data. In particular, the regularized expectation-maximization (RegEM) algorithm (Schneider 2001, Rutherford et al. 2003) has become increasingly popular. Recently, hierarchical models fitted using Bayesian inferential methods, have been proposed as a possible solution. Haslett et al. (2006) used a Bayesian approach for reconstruction using pollen as a proxy, and Li et al. (2010) and McShane and Wyner (2011) also explore Bayesian approaches for multi-proxy reconstructions.

We view the two-step approach to analysis as unnatural and dangerous. Unnatural because models should be fitted using raw data rather than synthetic data. Dangerous because conditioning on data from the first-stage means that a significant source of uncertainty may be omitted or misrepresented in modelling at the second stage. Problems from conditioning on the output of a statistical model as if it were observed data are fundamental. Any incorrectly modeled variation from the filtering of the proxy measurements will be carried forward to the reconstruction of the climate. At best, this can lead to inefficiencies in the predictions. At worst, it can produce substantial error in the predictions that are not accounted for in the estimated uncertainties.

Another problem is that in the reconstruction step, the underlying statistical model is usually ignored (a default model is used), with all effort focused on the considerable methodological challenges that are faced fitting the default model. An example of this can be seen with the regularized expectation-maximization (RegEM) algorithm (Schneider 2001): the expectation-maximization (EM) algorithm is a statistical method for estimating a model with missing data, not a statistical model itself. Several examples include Rutherford et al. (2003), Mann et al. (2005), Rutherford et al. (2005), where the underlying model is never stated or discussed. Instead, the focus is often on improving the regularization procedure, which while important, is only one element of the model. This results in a default model choice for reconstructions of this kind, irrespective of the data used. This default choice may not be applicable for every dataset and many of the assumptions contained in the default model may have a large effect on climate predictions.

To overcome the shortcomings described above we propose a model-based approach that merges the two steps through specification of a joint model for the raw data and climate. Our approach requires specification of a full likelihood, highlighting the need to make modeling assumptions explicit. This allows us to consider alternative assumptions, and examine sensitivities in both the filtering and reconstruction

components of the model. Furthermore, as the approach jointly standardizes and reconstructs, it does not condition on model output.

While these features of our approach are appealing, they also reveal the limitations of the data and may expose the sensitivity of the estimates to the modeling assumptions. There may even be scenarios where certain assumptions that have appeared to work using the traditional conditional approaches lead to models that are non-identifiable. There is no simple solution to these problems but it is important to acknowledge that they exist and to understand the implications of these limitations for inference about the historical climate record.

2 Data

The dataset that we will explore throughout this manuscript comes from Tornatrask, Sweden (Grudd et al. 2002, Briffa et al. 2008). It consists of tree-ring growth measurements from cores of both living and fossilized Scots Pine, *Pinus sylvestris*. There were 587 cores sampled from individual trees, with a mean segment length, defined to be the length (in years) of each tree-ring series, of 189 years, with a maximum segment length of 609 years. Other than the climate variables discussed below, the only other covariate is tree age, assumed to be 0 when each core is first observed and incremented by one year for each ring.

A feature of these data is the presence of the fossilized trees. Only 121 of the trees had observations in the time period for which we have observed climate variables. If we were to combine all cores together, the 587 cores would give a continuous record from 38bc to 1995.

Throughout the remainder of the manuscript we use a subset of the full dataset. In particular, we use $n = 500$ years of data, meaning any tree observed for at least 20 years after 1496 are used. This leaves $k = 247$ series, with a mean segment length of 179 years with a maximum segment length of 485 years. The number of cores per year varies throughout the last 500 years, with the most data in recent times (figure 1).

These trees are considered to be 'temperature stressed' which means that the limiting factor in their growth is the temperature during the growing season from June to August. Therefore, the climate data we use and predict back in time in this manuscript is the mean summer June - August temperatures (in degrees Celsius) from Abisko, Sweden, the nearest weather station to the trees. The temperature data extends from 1913 to 1995, comprising 83 years of climate observation. We treat June-August temperature from year 1496 to 1912 as missing data, the prediction of which is the focus of this manuscript.

Here we standardize the climate variables (and the log of the climate variables were appropriate) to have mean 0 and standard deviation of 0.5. We also approximately standardize the age variable.

The focus here is on the reconstruction of local climate. However, the ideas we discuss and the models

we fit naturally extend to global and spatial field based reconstructions as we discuss throughout the manuscript.

The following notation will be used throughout the manuscript. The raw tree-ring series are denoted by the $k \times n$ matrix $\boldsymbol{y}$, where y_{it} is the observed tree-ring growth measurement for tree i in year t. Note that tree i has observations from $t = f_i, \ldots, l_i$, where f_i and l_i are the first and last years tree i was observed. We split $\boldsymbol{y}$ into a $k \times m$ matrix $\boldsymbol{y}^{mis}$ and a $k \times (n-m)$ matrix $\boldsymbol{y}^{obs}$. The former denotes values of $\boldsymbol{y}$ that correspond to missing climate values $\boldsymbol{x}^{mis} = (x_1, \ldots, x_m)$, with the later denoting values of $\boldsymbol{y}$ that correspond to observed climate values $\boldsymbol{x}^{obs} = (x_{m+1}, \ldots, x_n)$. The age variable is denoted by the $k \times n$ matrix $\boldsymbol{a}$, where a_{it} is the approximate standardized age of tree i in year t.

3 Standard Methodology

The traditional approach to making predictions about historical climate from tree-ring data requires the raw data to be "standardized" before being statistically calibrated with the observed climate values (Fritts 1976, Cook and Kairiukstis 1990). The standardization procedure serves to:

1. Remove non-climatic growth influences, in particular, age effects from raw data.
2. Compress the information contained in k noisy tree-ring series into 1 synthetic series (from a common location) that expresses the common effect of climate on the trees.

We denote chronologies obtained from $\boldsymbol{y}$ using the standardization approaches by the vector $\boldsymbol{z} = (z_1, \ldots, z_n)$, where the value z_t gives the chronology value for year t. As with $\boldsymbol{y}$, we split the vector $\boldsymbol{z}$ into $\boldsymbol{z}^{mis} = (z_1, \ldots, z_m)$ and $\boldsymbol{z}^{obs} = (z_{m+1}, \ldots, z_n)$.

We outline two standardization procedures, which we refer to as typical standardization (TS) and regional curve standardization (RCS).

3.1 Typical Standardization

TS involves selecting and fitting a statistical model to the annual growth increments of each tree in order to remove the non-climatic growth influences, such as from tree age. After selecting and fitting a model, raw measurements are divided by the expected value under the fitted model to give an index for each tree which is averaged across trees (in a common location) to give a mean index (or chronology) for that location. The variation in this index is presumed to reflect primarily climatic influences with growth-related effects removed by scaling. These indices are subsequently used to reconstruct the climate. Sometimes the deviations for each tree are further modeled to remove correlation before being combined

(Cook and Kairiukstis 1990). Almost all datasets in the International Tree-Ring Data Bank have been standardized in this way.

TS has two major problems. Firstly, the uncertainty about parameters in the standardization is ignored in all further modeling. Treating as certain, quantities that are uncertain, yields estimates and predictions that appear more precise than they really are. Secondly, incorrectly modeled variation in the standardization output can distort modeling of the climate signal inducing bias in the reconstructed climate values. The problem is that standardization effectively detrends the tree-ring series in an attempt to remove the non-climatic growth. This can easily lead to the unintentional removal of climatic influences on growth from the tree-ring chronology – some of the 'baby' can get thrown out with the 'bathwater'.

This second problem has been named the 'segment length curse' (Cook et al. 1995), so-named because using TS we can only expect to recover climate signals that have a higher frequency (in relation to climate cycles) than the segment length of individual tree-ring measurements. As low-frequency climate signal (and possibly some high-frequency signal) is removed during standardization, the examination of changes in climate over hundreds, or thousands of years, becomes problematic.

3.2 Regional Curve Standardization

In an attempt to overcome the segment length curse and preserve the low-frequency climate signal, alternative standardization methods have been proposed (Esper et al. 2002, Briffa et al. 2001). Here we focus on the regional curve standardization (RCS) procedure (Esper et al. 2002).

The idea behind RCS is that we align trees by biological age, irrespective of when the trees were alive. The average for each age (or a group of ages, say in 10-20 year increments) is used to estimate a mean biological growth trend for that group. This is then used to determine a deviation for each measurement, defined as the observed measurement divided by the average for the appropriate age class. The deviations, all detrended by the identical mean biological growth curve, are then back-transformed onto the temporal calendar-year scale and averaged at each time period to get a mean index for that location. The hope with RCS is that non-climatic effects are eliminated by averaging across many trees sampled from across a range of time periods. The variability in the index is assumed to reflect changes in climatic conditions, that can exceed the lengths of the individual tree ring series being detrended, see Briffa et al. (1992) for details. This process requires no assumptions about the shape of the non-climatic growth, although often such constraints may be included.

As noted by Briffa and Melvin (2009), RCS has several potential problems. One problem is that, just as with TS, incorrectly modeled variation can distort modeling of the climate signal and induce bias in the resulting reconstruction. In particular, tree-specific differences in the growth can be misinterpreted

as spurious changes in the climate signal. This is particularly important when the sample size is small to moderate.

A potentially larger problem with the RCS method is that it is inappropriate when samples of trees are not evenly spread through time. This is because certain time periods are dominated by trees from a small number of age classes resulting in bias in the chronology. Unfortunately, for most tree-ring series, this will be the case with early periods having almost exclusively young trees. The Scots Pine data is one of the few datasets where there is a relatively even spread of age classes through time. This makes the RCS standardization technique an attractive option, provided growth can be assumed to be approximately constant between trees.

As with TS, the RCS standardization approach ignores uncertainty about the standardization procedure in all further modeling.

4 Statistical Calibration

The next step in the standard approach is to use location-specific chronologies to reconstruct the climate. Here we focus on the use of linear calibration as a means of reconstructing climate values from tree-ring (and potentially other) data. As we will describe, linear calibration is used both for reconstructing local climate values from one tree-ring series, as well as using multiple proxy series to either reconstruct a global climate variable, or a spatial field of climate variables. Our description of linear calibration below, in particular the comparison between classical and inverse approaches, is largely a summary of Osborne (1991) in the context of climate reconstruction.

Univariate linear calibration assumes a single chronology $\boldsymbol{z}$ has a linear relationship with climate observations $\boldsymbol{x}$. The observed data $\boldsymbol{z}^{obs}$ and $\boldsymbol{x}^{obs}$ is used to estimate the parameters describing the relationship, which are then used to predict $\boldsymbol{x}^{mis}$ using $\boldsymbol{z}^{mis}$. We follow Osborne (1991) and describe two available estimators for univariate calibration:

1. Classical calibration. The calibration model is specified as

$$z_t = \beta_0 + \beta_1 x_t + \epsilon_t, t = 1, \ldots, n.$$

This is the "natural" specification of the model according to the underlying science, as it describes the tree-ring growth chronology in terms of the climate variable. The missing climate values are estimated using

$$\hat{x}_t = \frac{z_t - \hat{\beta}_0}{\hat{\beta}_1}, \quad t = 1, \ldots, m, \tag{1}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated from the linear regression of $\boldsymbol{z}^{obs}$ on $\boldsymbol{x}^{obs}$.

2. Inverse calibration. The calibration model is specified as

$$x_t = \gamma_0 + \gamma_1 z_t + \eta_t, \quad t = 1, \ldots, n.$$

The missing climate predictions are then estimated using

$$\hat{x}_t = \hat{\gamma}_0 + \hat{\gamma}_1 z_t, \quad t = 1, \ldots, m \tag{2}$$

where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are estimated from the linear regression of $\boldsymbol{x}^{obs}$ on $\boldsymbol{z}^{obs}$.

An attraction of the classical estimator is that if we assume normal errors then it is the maximum likelihood estimator. However, if $\hat{\beta}_1$ is not significant, then the confidence interval for $\hat{x}_t$ can be the whole real line or two disjoint semi-finite intervals (Osborne 1991). Despite the mean square error (MSE) of $\hat{x}_t$ being infinite (Williams 1969), conditioning on $|\beta_1| > 0$ in the expectation leads to a consistent estimator with finite MSE (Berkson 1969).

The attraction of the inverse estimator is that for small sample sizes (i.e. small values of $n - m$), the inverse estimator has a lower MSE than the classical estimator (Krutchkoff 1967, Berkson 1969). However, this appears to be the case only when the true value of x_t is close to $\bar{x} = \sum_{i=m+1}^{n} x_i$, with (true) values of x_t outside of the calibration range (i.e. extrapolation) resulting in the classical estimator having superior MSE (Krutchkoff 1969, Berkson 1969). The "advantage" of a lower MSE within the calibration range was called into question by Williams (1969) who suggested that comparisons of methods based on MSE were not sensible since the Rao-Blackwell theorem can be used to show that no unbiased estimator has finite variance. Furthermore, the inverse estimator is also inconsistent, and is only asymptotically unbiased if the true value of $x_t = \bar{x}$ (Berkson 1969). The best justification for the inverse estimator came from Hoadley (1970), who showed that the inverse estimator corresponds to a Bayes estimator with particular informative prior distribution.

In the spirit of Hoadley (1970) we explore the two approaches from a Bayesian perspective to better understand the difference between the classical and inverse regression estimators described above. If we start with the model

$$[z_t|x_t] = \mathcal{N}(\beta_0 + \beta_1 x_t, \sigma^2_{Z|X}), \quad t = 1, \ldots, n,$$

where $[a|b]$ denotes a probability density (or mass) function of a conditional on b. We also assume that the marginal distribution for x_t is also normal:

$$[x_t] = \mathcal{N}(\mu_X, \sigma^2_X), \quad t = 1, \ldots, n.$$

Then by Bayes theorem we can show that

$$[x_t|z_t] = \mathcal{N}(\gamma_0 + \gamma_1 z_t, \sigma^2_{X|Z}),$$

where

$$\begin{aligned}
\gamma_0 &= \mu_X(1 - \gamma_1\beta_1) - \gamma_1\beta_0 \\
\gamma_1 &= \frac{\beta_1\sigma^2_X}{\sigma^2_{Z|X} + \beta_1^2\sigma^2_X} \\
\sigma^2_{X|Z} &= \frac{\sigma^2_X\sigma^2_{Z|X}}{\sigma^2_{Z|X} + \beta_1^2\sigma^2_X}.
\end{aligned}$$

Thus, the model for the inverse estimator (assuming normal errors) can be found by starting with the model for the classical estimator and assuming a common normal distribution for all elements of $\boldsymbol{x}$ with fixed mean and variance μ_X and σ^2_X. This explains why the performance of the inverse estimator deteriorates as it moves away from $\bar{x}$. An intuitive explanation can be seen by considering the case where $m = 1$. If the true value of $x_1 = \bar{x}$ then x_t and $\boldsymbol{x}^{obs}$ are in agreement with respect to the marginal distribution of $\boldsymbol{x}$. As the true value of x_1 moves away from $\bar{x}$ then x_1 and $\boldsymbol{x}^{obs}$ become increasingly in conflict with respect to the marginal distribution. For the case where $m > 1$, we can see that performance of the inverse estimator will depend on how far x_t is from μ_X, a parameter that should be close to $\bar{x}$ under the modeling assumptions. In contrast, in the "Bayesian version" of the classical estimator we have a "vague" prior distribution for x_t [1] in which case the likelihood dominates the prior. It is therefore no surprise that the classical estimator is preferred when x_t is far from $\bar{x}$.

Almost all climate reconstructions use inverse regression approaches (or methods related to this). For example, the description of calibration in Cook and Kairiukstis (1990) only considers regressing $\boldsymbol{x}$ on $\boldsymbol{z}$. As we describe in more detail below, the model underlying the so-called RegEM algorithm (Schneider 2001) follows the inverse estimator in assuming that the marginal distribution for the climate variables $\boldsymbol{x}$ is normal with constant mean and (co)variance. We find the default use of inverse-like estimators curious, especially since the theoretical results above suggest that the classical calibration is preferred over inverse regression in the case of extrapolation. This is especially relevant as many reconstructions are undertaken in the belief the climate in the past is different from the present.

To explore the various standardization and calibration assumptions for the Scots Pine dataset using traditional methods, we first find TS and RCS chronologies. To find the TS chronology we use the process described in the previous section using a negative exponential growth function (we do not adjust for autocorrelation). To find the RCS chronology, we assume that the growth rate is the same for every

[1] Note that vague in this context does not necessarily mean flat. See Ghosh et al. (1995) and Yin (2000) for details.

tree in a given decade of its life (i.e. there is one growth rate from age $1-10$, another from $11-20$, etc).

The Scots Pine dataset is an example where the correlation is strong enough to calculate finite confidence interval using classical calibration, as the regression for both the TS and RCS chronologies is highly significant with the p-value for $\beta_1 < 0.00001$. The temperature estimates for both RCS and TS exhibit some of the differences we expected both in terms of (i) the different standardization approaches, and (ii) differences between inverse and classical calibration (figure 2). When viewing the inverse estimates from a Bayesian perspective, the 'shrinking' effect of the marginal distribution for climate is clearly seen relative to the classical estimates. The comparison of RCS and TS also clearly shows the 'segment length curse' featuring in the TS output. Due to the short segment lengths, the TS approach effectively removes any low-variability climate signals over 200 years of length. This results in reconstructions that are centered on the mean from the observed period, as opposed to showing a general period of lower temperatures from 1600–1800 as is evident in the RCS based reconstruction.

The problem of extrapolation is one of the fundamental themes running through this manuscript. The most striking differences in figure 2 occur in the periods with the most extrapolation, and suggest such extrapolation is sensitive to the assumptions in the various statistical models. Instead of choosing between estimators, we believe that it is better to explore different assumptions as specified through alternative models. Our goal is to use understanding of the model to determine how predictions vary according to different assumptions.

In sections 5 and 6 we focus on two related assumptions that we believe are critical to the standardization and reconstruction processes. The first is the choice we make for the marginal distribution of x_t. The differences evident in figure 2 suggest that the choice of marginal model for climate (as implied by the different analyses) may have a large effect on predictions.

We do not believe that assuming each x_t variable is unrelated, as in the classical estimator, is a sensible model choice. The entire theory of 'forcing' effects on climate suggest structure in the pattern of changes in x_t over time, and it should be our goal to describe these as part of the analysis. Some sort of smoothing would seem desirable.

However, the assumption in the inverse approach that the x_t are independent (given their parameters) and identically distributed seems overly restrictive. It is doubtful that anyone interested in modeling long-term climate believes that this is a reasonable hypothesis. Thus, the two common approaches to calibration represent two extremes in model assumptions with the choice driven by operational convenience rather than scientific sense. Our view is that it is more natural to model $\boldsymbol{x}$ hierarchically using a flexible family of distributions that include these two as special cases, and that admits a wider choice of scientific hypotheses about how $\boldsymbol{x}$ has changed over time. In the next section, we outline one such

approach in which the mean of the marginal distribution for x_t changes smoothly through time.

The second assumption is that of the model describing how the tree-rings depend on climate. Together with the marginal model for x_t, the assumption of linearity can have a significant influence on the predicted values. In particular, we suggest the period with most extrapolation (from 1600–1650 and the period around 1750 from figure 2) will exhibit the most sensitivity to model choice. Using RCS standardization as an example, we can see that various transfer functions (describing how z relates to x) will result in substantial differences when extrapolating, despite the different transfer functions describing the observed data equally well (figure 3).

4.1 Multivariate Calibration

Multivariate linear calibration (Brown 1982) can be used to reconstruct a (potentially multivariate) climate response using multiple proxy observations. These proxies could be from multiple locations, or from different proxy types, or often a combination of the two (for example see Mann et al. 1998).

A heavily used approach in climate reconstruction is the regularized EM (RegEM) algorithm. As described by Schneider (2001), the model underlying the RegEM algorithm assumes that the L_x length vector of climate values in year t, denoted $\boldsymbol{x}_t$ and the L_z length vector of proxies in year t, denoted $\boldsymbol{z}_t$ are jointly multivariate normal

$$\begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right), t = 1, \ldots, n, \tag{3}$$

where $\boldsymbol{\mu}$ is a mean vector of length $L_z + L_x$ and Σ is a $(L_z + L_x) \times (L_z + L_x)$ covariate matrix. This model is then fitted using the EM algorithm. Note that (i) if any value of z_t (or x_t) is missing in year t, then data augmentation (Tanner and Wong 1987) is used as part of the EM algorithm, and (ii) often $L_z > (n - m)$, making the dimensionality of $\boldsymbol{z}$ too large to allow for identifiable estimation of Σ. This leads to the regularization of the RegEM algorithm.

As with univariate calibration, we can express the marginal distribution in terms of a conditional distribution of proxies given climate and a marginal climate distribution:

$$\begin{aligned} \boldsymbol{z}_t &\sim \mathcal{N}(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_t, \Sigma_{Z|X}), \quad t = 1, \ldots, n, \\ \boldsymbol{x}_t &\sim \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_X), \quad t = 1, \ldots, n, \end{aligned}$$

where $\boldsymbol{\beta}_0$ is a vector of length L_z, $\boldsymbol{\beta}_1$ is a $L_z \times L_x$ matrix, $\Sigma_{Z|X}$ is a $L_z \times L_z$ matrix, $\boldsymbol{\mu}_X$ is a vector of length L_x and Σ_X is a $L_x \times L_x$ matrix. These parameters are deterministic functions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from

the marginal model in (3), see Gelman et al. (2004) pp317-324 for details. An important point is that the model underlying the RegEM approach is an extension to the inverse estimator in that it assumes a marginal normal distribution for the climate variables that has constant mean and (co)variance through time.

5 Model-Based Approach

Specifying the problem in terms of a joint model can remove the segment length curse, as well as the other problems caused by conditioning on model output inherent in TS and RCS. In order to isolate the influence of different model assumptions on the reconstructions we specify the model incrementally.

In this section, we start by outlining a simple hierarchical model that provides a model-based analogue of the TS approach. While a stepping stone to a more realistic model, we believe this model is of interest in its own right, as it provides a model-based alternative to the distinct standardization and reconstruction processes described in the previous sections.

The model is specified as

$$\begin{aligned}
\log(y_{it}) &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{it}, \sigma_i^2), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \\
\mu_{it} &= \beta_{0i} + \beta_{1i} a_{it} + \eta_t, \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \\
\eta_t &\overset{\text{iid}}{\sim} \mathcal{N}(\gamma x_t, \sigma_\eta^2), \quad t = 1, \ldots, n, \\
x_t &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2), \quad t = 1, \ldots, n,
\end{aligned} \tag{4}$$

where we assume β_{0i} and β_{1i} are drawn from common distributions, and by 'independent' in iid we mean independent given parameters.

$$\begin{aligned}
\beta_{0i} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_0}, \sigma_{\beta_0}^2), \\
\beta_{1i} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_1}, \sigma_{\beta_1}^2).
\end{aligned}$$

As with TS we model individual tree-ring growth measurements with an effect due to age. In this case we assume a simple linear model on the log-scale (equivalent to a negative exponential model of the original scale as used in the previous section). The major difference from the standard calibration approach is the additional term, common between trees, denoted η_t. This term reflects the fact that in TS we take the average residual (on the log-scale) at each time point in order to find the chronology. The difference here is that we do not simply estimate a value for η_t that we use in a subsequent analysis. Rather, we consider a hierarchical model for η_t in terms of the climate variables x_t. It is this step that links the

standardization and reconstruction processes and allows us to overcome the segment length curse. The final term we specify is the marginal model for x_t. As with the underlying model for inverse regression, the first model we consider treats x_t as normally distributed with a constant mean through time.

The model in (4) is the most simple model-based specification we can consider that provides an analogue to TS. Throughout the rest of this and the next section we consider various extensions we can consider in the model, e.g. autocorrelation. We fit on the log-scale due to the multiplicative nature of the errors evident from plotting the raw data. Taking the log of the raw measurements appears to provide approximately homoscedastic errors.

To fit the model in (4) we use Bayesian inference and Markov chain Monte Carlo (MCMC) implemented in JAGS (Plummer 2003). Prior distributions for all parameters are specified in the supplementary materials. We use three parallel chains and adapt for 5000 iterations per chain before taking a posterior sample of 10,000 iterations per chain.

As expected, results from the model-based approach look like a combination of the TS and RCS estimates in figure 2 (figures 4(a), 4(b) and 4(c)). The most apparent result is that the three approaches all lead to predictions that are similar, with only slight differences in the shape and magnitude of the reconstructions. In particular, the model-based approaches appear to predict the extremes with a similar magnitude to the RCS approach (i.e. compare 1600–1650 and 1750), although they do not have the pronounced period of cool temperatures from 1600 through 1750 as predicted using RCS. Instead, the estimates from the model-based approach suggest that 1600–1650 is a short period of cooling before returning to "average" temperatures from 1650–1750. Note that all three of these models have the same marginal model for x_t.

5.1 Extensions to the Calibration Model

Here we explore how the estimates of linear calibration depend on (i) the response function between the observation $\boldsymbol{y}$ and the measurements $\boldsymbol{x}$, and (ii) the hierarchical/marginal model we place on $\boldsymbol{x}$. We consider extensions and alternative formulations for each of these components in turn.

It is unclear what the appropriate response function should be between tree-ring growth and climate response. Should the log of the tree-ring measurement respond linearly to climate as we specify in (4)? While there exists much theory as to the role of various climatic variables in the process of photosynthesis, the theory for how tree-ring growth responds to summaries of climatic information is much less certain.

Tolwinski-Ward et al. (2011) suggest that scaled temperature-based tree-ring growth g_T be expressed as

$$g_T = \begin{cases} 0 & \text{if } T < T_1 \\ \frac{T-T_1}{T_2-T_1} & \text{if } T_1 < T < T_2 \\ 1 & \text{if } T_2 < T \end{cases}$$

where T is the mean temperature across some time period (they used months), T_1 is a threshold temperature below which growth cannot occur, and T_2 is a threshold temperature above which has optimal growing conditions. In other words, they assume no growth below some temperature T_1, linear growth between T_1 and T_2 up to the optimal growth at and above T_2.

We consider an alternative approach, where instead of specifying a step function, we approximate this with the smooth natural logarithm function. We take the logarithm on the temperature in degrees celsius under the assumption that an appropriate lower bound is zero degrees celsius (recalling that our temperature variable is the mean June-August temperature). We assume that the log of tree-ring growth responds to the log of the temperature, so that (4) is changed, with

$$\begin{aligned} \eta_t &\overset{\text{iid}}{\sim} \mathcal{N}(\gamma \log(x_t), \sigma_\eta^2), \quad t = 1, \ldots, n, \\ \log(x_t) &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2), \quad t = 1, \ldots, n. \end{aligned} \tag{5}$$

Making the change to the response function as specified in (5) does not dramatically change the reconstructions (figure 4(d)), with only minor changes noticeable at the most extreme values. One possibility is that the reconstructions are robust to model misspecification, at least with regard to the choice of transfer function. Another possibility, is that the marginal model for the climate variables is hiding any sensitivity to the choice of transfer function.

We therefore consider an alternative prior distribution for x_t. As we discussed earlier, we do not wish to assume a vague prior for x_t that is independent between years. Instead, we suggest modeling x_t as a normal distribution, but allow the mean to vary according to a smooth function. Here we choose to use a cubic B-spline. That is, (4) is changed, with

$$\begin{aligned} x_t^* &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{xt}, \sigma_x^2), \quad t = 1, \ldots, n, \\ \mu_{xt} &= \sum_{h=1}^{H} \gamma_h B_h(x_t^*), \quad t = 1, \ldots, n, \end{aligned} \tag{6}$$

where x_t^* is used to denote the climate response, including both $x_t^* = x_t$ and $x_t^* = \log(x_t)$, and $B_h(x_t^*)$ is the hth B-spline basis function; see Seber and Lee (2003), Hastie et al. (2009) for an introduction to

B-splines, including a recursive relationship with which to obtain $B_h(x_t^*)$. Our intention is to use the spline to allow for a smoothly evolving mean function through time. To achieve what we consider to be an appropriate level of smoothness, we place a knot every 25 years. However, to be conservative, we also include a hierarchical model on γ_h to "penalize" the spline to avoid overfitting (Eilers and Marx 1996)

$$\gamma_h \overset{\text{iid}}{\sim} \mathcal{N}(\mu_\gamma, \sigma_\gamma^2), \quad h = 1, \ldots, H.$$

Note that if $\sigma_\gamma^2 = 0$ then the marginal mean for x^* is constant through time. We consider the hierarchical spline model for the model with response function both in terms of x_t and that in terms of $\log(x_t)$.

As before we use Bayesian methods to fit the model with the inclusion of the new marginal model for x_t in (6) using Markov chain Monte Carlo (MCMC) in JAGS. The inclusion of the more complex model for x_t makes the model more difficult to fit, with increased autocorrelation in the Markov chains. For this model we use three parallel chains and adapt for 10,000 iterations per chain before taking a posterior sample of 50,000 per chain (Increase this to 100,000 and check σ_X).

The inclusion of the alternative marginal distribution for the climate values has a dramatic effect on the predictions (figure 5). There are several differences that stand out: (i) the amplitudes vary between the models, both between the different marginal models for the temperature, as well as between the different transfer functions, (ii) the inclusion of a flexible model for the mean has led to a difference in the overall mean of the temperature series, and (iii) the smoothness of the estimates differs between the models.

Given the results comparing classical calibration and inverse estimation in figure 2 it is hardly surprising that the amplitudes vary between the marginal models for temperature. Here we have relaxed the assumption of a constant marginal mean for the climate variable which has lead to increased magnitude in the predictions. As well as the increase in magnitude, the more flexible model has also changed the overall mean of the predictions. In particular, with the constant mean through time, the reconstructions have largely been shrunk to the prior distribution (see figures 5(a) and 5(b)). Other than short $\sim$ 50 year periods, the temperature series appears to be centered around a constant mean. However, when we relax the prior distribution for x_t we see that the predictions suggest that, except for a brief period around 1750, the historical temperatures in the area were cooler (see figures 5(c) and 5(d)). This prediction, at least partially agrees with that obtained from the RCS reconstruction (see figure 4(b)). Furthermore, the reconstructions in figures 5(c) and 5(d) clearly gives evidence for the joint modeling approach reconstructing low-frequency variability in the temperature series.

Unlike the model with a constant marginal mean, including the spline leads to a substantial difference in the estimates from the two transfer functions. As suggested in figure 3, this difference is most evident

when we are extrapolating, for example, in the period from 1600–1650.

When comparing figures 5(a) and 5(c) the most striking feature is the extent to which the cubic B-spline model has smoothed out annual temperature variation. Similarly comparing 5(b) and 5(d). We believe the effectiveness of the cubic B-spline model in smoothing out the annual variation is due to the assumption that all trees in the sample respond to climate in the same way. That is, the model is capturing the information about climate in a given year provided by the growth increments from $40-120$ trees and expressing it through the single term η_t. The focus in the next section is describing a model that allows us to relax the assumption that trees have a common response to climate.

6 Tree-Specific Response

The models described in the previous section illustrate how we can use joint model-based inference that provides an equivalent analysis to that provided by commonly used methods. Importantly, it highlights the role of various modeling assumptions in the analysis. However, with a fully-specified model we can go a step further and consider relaxing some of these assumptions.

To relax the assumption that trees all respond to climate in the same way, we can replace the common increment η_t in (4) with an individual response,

$$\mu_{it} = \beta_{0i} + \beta_{1i} a_{it} + \beta_{2i} x_t^*, \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i. \tag{7}$$

As in the previous section we use x_t^* to denote the climate response. As for β_{0i} and β_{1i} we assume a hierarchical model for β_{2i} of the form

$$\beta_{2i} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_2}, \sigma^2_{\beta_2}).$$

By specifying such a response function we have moved from univariate calibration, to multivariate calibration. In our experience, incorporating this change leads to increased precision in the estimates and leads to reconstructed climate signals that have with reduced correlation. However, these benefits come at a cost, one being a substantial increase in the difficulty of computational methods required to efficiently fit the model, as discussed later in this section, and the other increased sensitivity of the resulting predictions to modeling assumptions.

Our intuitive explanation for the latter, is that allowing each tree to respond separately to climate, combined with a flexible marginal models for climate, allows the x_t variables to reconstruct a range of various patterns in the data. The resulting reconstruction then helps to explain important sources of variability in the data that are not included in the model, at the cost of having a reconstruction that no longer reflects the climate variable of interest. Two examples where this is important for the

Scots Pine data both relate to correlation: (i) between trees in a given year, and (ii) autocorrelation. With individual tree-specific response to climate, the difference in estimates of the missing climate values between a model that assumes autocorrelation in the tree-ring series and one that does not, is substantial. In particular, if we ignore autocorrelation in the growth increments, the reconstruction appears to include the autocorrelation component through the climate process. This is manifest through the values of β_{2i} being small in magnitude, with corresponding values of x_t very large in magnitude. The flexibility of the model means that this combination does a reasonable job of modeling the autocorrelation for $\boldsymbol{y}^{mis}$ without overly affecting the model fit of $\boldsymbol{y}^{obs}$ since $\beta_{2i} \approx 0$.

There will be few observational datasets where it is reasonable to assume that errors are independent. Much more likely is the presence of unobserved variables, that confound the effect of climate on growth. Examples might include changes in soil fertility over time or factors that affect the amount of light that trees receive. We assume that such confounders do not interact with the climate process of interest, and account for them through assuming the errors in any year are correlated between trees. Including these changes leads to further changes of (4) so that

$$\log(\boldsymbol{y}_t) \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_{yt}, \Sigma_{yt}), t = 1, \ldots, n, \tag{8}$$

where $\boldsymbol{y}_t$ is a vector whose elements consist all values of y_{it} observed in time t, $\boldsymbol{\mu}_{yt}$ is a vector, with elements μ_{it} as given in (7) for those individuals observed in time t and Σ_{yt} is a submatrix of the full variance-covariance matrix Σ_y for those individuals observed at time t. Note that we could have also included missing y_{it} values into the model using data augmentation to make these vectors/matrices of the same dimension in every time period. However, we found that this specification was much less efficient, particularly when reconstructing the full 2000 year series.

The inclusion of correlated errors in (8) provides a direct (statistical) link to the underlying statistical models of Schneider (2001), Mann et al. (2005, 2007) and others who use the RegEM algorithm. Instead of having multiple trees, the multi-proxy series of Mann et al. (2005, 2007) have multiple tree-ring chronologies (see section 3) as well as other proxies. That is, if instead of modeling $\log(\boldsymbol{y}_t)$ in (8), we model a vector of proxy observations $\boldsymbol{z}_t$, set $\beta_{1i} = 0$ and use

$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$$

for a univariate climate variable, or

$$\boldsymbol{x}_t \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_x, \Sigma_x)$$

for a multivariate climate variable, we obtain an identical statistical model to that underlying the RegEM

algorithm as described by Schneider (2001). The only distinction is that we have expressed the model conditionally instead of jointly. This suggests that the results we find regarding the sensitivity of estimates to the specification of (i) response function, and (ii) marginal model for the climate process will also apply to their approach, although the extent of such sensitivity will be data dependent.

As we describe in section 4.1, this model can be difficult to fit due to having more variables than we have data points, resulting in non-identifiability of Σ_y. In our case, it occurs when we have more trees than we have years of observed climate responses, i.e. $k > n - m$, as is the case here. The approach of Schneider (2001), Mann et al. (2007) and others is to consider different regularization/penalized methods for reducing the dimensionality. In a Bayesian context such approaches can be viewed as a certain hierarchical distribution on the parameters in the model. The approach we take here is to use Bayesian factor analysis, e.g. see West (2003), Carvalho et al. (2008), Ghosh and Dunson (2009). In particular, we allow for n_f latent factors and again specify a hierarchical model for the factor loading variables to prevent over-fitting. Including the Bayesian factor analysis means that the model from (8) becomes

$$\begin{aligned} \log(y_{it}) &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{it}, \sigma_y^2), i = 1, \ldots, k, \quad t = f_i, \ldots, l_i \\ \mu_{it} &= \beta_{0i} + \beta_{1i} a_{it} + \beta_{2i} x_t^* + \sum_{h=1}^{n_f} \phi_{ih} \delta_{ht}, \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i \end{aligned} \tag{9}$$

where ϕ_{ih} are the factor loadings for tree i to factor h and δ_{ht} is the value of the hth factor in the tth year. The marginal variance for $\boldsymbol{y}_t$ is then $\boldsymbol{\phi}_t \boldsymbol{\phi}_t' + \sigma_y^2$, where $\boldsymbol{\phi}_t$ is the matrix of ϕ_{ih} values for trees observed in year t. As is standard we set

$$\delta_{ht} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad h = 1, \ldots, n_f, \quad t = 1, \ldots, n.$$

We specify a hierarchical model for the factor loadings

$$\begin{aligned} \phi_{ih} &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{h\phi}^2), \quad i = 1, \ldots, k, \; h = 1, \ldots, n_f, \\ \sigma_{h\phi}^2 &\sim \mathcal{N}(0, \sigma_{\phi\phi}^2) T(0, \infty), \quad h = 1, \ldots, n_f, \end{aligned}$$

where $T(a, b)$ denotes truncation of a random variable between a and b. In practice, we specify n_f to be larger than we expect is necessary, and use the hierarchical specification to shrink any factors that are not required in the model in order to avoid over-fitting. We note that this representation is strictly non-identifiable, since (i) the signs of ϕ and δ can reverse and leave the products unaffected, and (ii) label-switching can occurs between the factors. This does not concern us here as the factors

themselves are of little interest in this application and we use the model as specified above as a means to specify a reduced dimension variance-covariance matrix. For applications where the factors are of interest, additional constraints can be placed on some of the ϕ_{ih} values to ensure identifiability(see, for example Ghosh and Dunson 2009).

As mentioned earlier, it is also important to allow for autocorrelation in the tree-ring growth observations. To include autocorrelation for p lags we include an $w+p$ additional "missing" growth observations per tree before the first observation using data augmentation. As specifying the exact marginal distribution of $\boldsymbol{y}_t$ for models with $p > 1$, these additional variables are included as a "buffer", allowing the marginal variance of the $\boldsymbol{y}_t$ values to reach an equilibrium for all observations. The complete model we consider is:

$$\begin{aligned} y_{it} &\overset{\text{iid}}{\sim} \mathcal{N}\left(\mu_{it}, \kappa\sigma_y^2\right), \quad i = 1, \ldots, k, \quad t = f_i - w - p, \ldots, f_i - w - 1 \\ y_{it} &\overset{\text{iid}}{\sim} \mathcal{N}\left(\mu_{it} + \sum_{j=1}^{p} \rho_j (y_{it-j} - \mu_{it-j}), \sigma_y^2\right), \quad i = 1, \ldots, k, \quad t = f_i - w, \ldots, l_i, \end{aligned} \tag{10}$$

where the value κ can be chosen so that the marginal variance of $\boldsymbol{y}_t$ reaches equilibrium quickly (here we use $\kappa = 2$). We set the value of $\mu_{it} = 0$ for all augmented "observations": $i = 1, \ldots, k$, $t = f_i - m - p, \ldots, f_i - 1$, while μ_{it} for all observed values is specified in (9): $i = 1, \ldots, k$, $t = f_i, \ldots, l_i$. To ensure that the model is identifiable, we include the constraint that all roots of the complex polynomial in B,

$$1 - \rho_1 B - \rho_2 B^2 - \ldots - \rho_p B^p = 0$$

have modulus greater than the unit circle.

The complete model also includes many of the hierarchical specifications previously mentioned. In particular, we have

$$\begin{aligned} \beta_{0i} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0}), \quad i = 1, \ldots, k \\ \beta_{1i} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_1}, \sigma^2_{\beta_1}), \quad i = 1, \ldots, k \\ \beta_{2i} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_2}, \sigma^2_{\beta_2}), \quad i = 1, \ldots, k \\ \delta_{ht} &\overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad h = 1, \ldots, n_f, \quad t = 1, \ldots, n \\ \phi_{ih} &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{h\phi}), \quad i = 1, \ldots, k, \; h = 1, \ldots, n_f, \\ \sigma^2_{h\phi} &\sim \mathcal{N}(0, \sigma^2_{\phi\phi}) T(0, \infty), \quad h = 1, \ldots, n_f, \end{aligned} \tag{11}$$

We deliberately do not specify the hierarchical model for $\boldsymbol{x}^{mis}$ as we consider models where the mean is either constant as in (4), or specified using a cubic B-spline as in (6).

As mentioned above, fitting this model is challenging[2]. We were unable to have adequate mixing/convergence of the Markov chain when using numerous self-written MCMC samplers, including univariate Gibbs samplers, blocked Gibbs samplers and ASIS samplers (Yu and Meng 2011). The only MCMC algorithm we found that could generate a Markov chain with adequate mixing was Hamiltonian Monte Carlo (Duane et al. 1987, Neal 1996). We use Hamiltonian Monte Carlo to update all parameters in one-step.

Full details of Hamiltonian Monte Carlo approach are in Neal (2011); here we provide a brief summary that we try to keep non-technical. The key idea is that it is theoretically possible to use Hamiltonian dynamics to move efficiently through, and sample from the posterior distribution. This requires an augmenting "momentum" variable is included for each parameter in the model. If you consider the current parameter values as a particle in space, we move this particle across the surface of the posterior distribution with the given momentum (which can be negative) in the direction of the gradient, a process that continuously changes the parameter values, the momentum variables and the gradient.

A problem is that we cannot do this in practice as it is a continuous process. We can however, discretize the process by using an appropriate iterative equation (here we use the leapfrog algorithm) that introduces error into the process. The introduction of error means that instead of using Hamiltonian dynamics to sample from the posterior distribution, we use it to generate an efficient proposal within a Metropolis-Hastings updater.

The Hamiltonian Monte Carlo algorithm can be used to update some, or all of the continuous parameters in the statistical model under study (Neal 2011). Here we update all parameters, including all hyperparameters, using the Hamiltonian Monte Carlo algorithm. For a model over 500 years with 200 trees and 20 factors, this amounts to updating over $15,000$ parameters simultaneously.

The Hamiltonian Monte Carlo has not been extensively used as it is difficult to implement. One reason for this is that partial derivatives of the log-posterior distribution are required for every parameter being updated. For all models we consider here we found explicit derivatives for all parameters. Another reason implementation has been limited is that optimization of the algorithm is difficult. There are at least two variables in the leapfrog algorithm that can affect the performance of the algorithm. These are the step-size and number of leapfrog steps.

The step-size effectively describes "how much" discretization is used. If the step-size is too small, many leapfrog steps will be required to move through the parameter space. If the step-size is too large, we will move through the parameter space quickly, but at the cost of increased discretization error that will result in decreased Metropolis-Hastings acceptance rate. Furthermore, the step-sizes can be scaled

[2]Much of the computational difficulty appears to come from the inclusion of the fossilized trees. When all trees have at least some observations in $\boldsymbol{y}^{obs}$ we found computation was less difficult.

for each parameter separately to further improve the performance of the algorithm.

The number of leapfrog steps, for a given step-size, also helps determine how well we move through the parameter space. If there are too many steps, then we can actually move through the parameter space and back again, leading to a proposal very close to the current point. If there are too few steps, we do not move very far through the sample space and the proposal is little better than an entirely random update (i.e. not taking the gradient into account).

Here we chose the number of leapfrog steps using trial and error, settling on 250 steps per iteration. We then use a predetermined and fixed adaptive phase to find appropriate values for the step-sizes. During this phase we find a global step-size across all parameters that increases with each Metropolis-Hastings acceptance and is reduced with each Metropolis-Hastings rejection in order to approximately give the optimal acceptance rate of 65% (Neal 2011). Each parameter has their step-size scaled according to a (far from optimal) estimate of the variance of their marginal posterior distribution found by taking the inverse of the diagonal of the Hessian matrix. At the end of the adaptive phase, we can either take the last value for the step-size variables, or we can take some sort of average over the last h values. We then use this estimate of step-size for all subsequent iterations in the Markov chain from which we draw our posterior sample. Even though we take the non-diagonal elements of the Hessian matrix to be zero, we see a considerable improvement over assuming constant step-sizes for each parameter. Note that this approach will not work for variables that are log-concave. See Neal (2011) and Girolami and Calderhead (2011) for alternative optimization strategies.

The MCMC consisted of using three chains with over-dispersed starting values. With 250 leapfrog steps the model mixed quickly, appeared to converge well and the Markov chain had low autocorrelation. We run the model for an adaptive phase of 1000 iterations followed by 1000 iterations as a draw from the posterior distribution. On the lead authors desktop machine, using Scots Pine data from only the last 500 years the model run in approximately XX hours. The full data (2000 years) runs in approx X days. Full details of the models fitted using Hamiltonian Monte Carlo, including the derivatives are in the supplementary materials.

The analysis needs to be rerun as these results are from a dataset with last 30 years of climate data removed.

Here we fit four different models to the full model specified in (10) and (11). In particular, we set $p = 1$ and $n_f = 25$ and explore the two different transfer functions with the two different marginal models for temperature. The estimates for the mean summer temperature for the four different versions of the full model are shown in figure 6. The three features of interest are: (i) the uncertainty in the predicted climate values is much smaller than in figure 5, (ii) the sensitivity of the model to the marginal model

for climate is reduced, and (iii) the sensitivity to the transfer function is still apparent.

As expected, allowing the climate response to vary according to tree has lead to a large increase in the precision of the climate estimate, reducing the effect of the assumed marginal distribution. However, even with this increase in information about $\boldsymbol{x}$ we can still see substantial variability in climate estimates between different models for how $\boldsymbol{y}$ depends on $\boldsymbol{x}$ (figures 6(c) and 6(d)), with predictions differing by as much as seven degrees celsius in the extreme years from $1600 - 1650$. We argue that the data is unable to conclusively distinguish between these large differences as these differences occur in the extremes of the reconstructions where extrapolation is the most evident. In years where the tree-ring measurements look more like what we observed in the observed period, there is a much smaller difference between the models, for example consider $1500 - 1550$ and $1850 - 1900$.

7 Model Checking

As discussed above, there is substantial sensitivity in the climate estimates from exploring alternative specifications of both the response function and the marginal model for the climate observations. To determine which predictions are better supported by the data it is useful to consider model fit; models that do not reconstruct the observed data well should be ruled out.

Model checking has the potential to add considerable value to climate reconstruction. Model checking a reconstruction (or a calibration model in general) differs in one respect from model checking normally carried out for statistical models: our interest is in the prediction of the latent covariate $\boldsymbol{x}$ instead of the observations $\boldsymbol{y}$. While $\boldsymbol{x}$ is of primary concern, exploring the model fit of both $\boldsymbol{x}$ and $\boldsymbol{y}$ can help the overall assessment of model fit.

We follow McShane and Wyner (2011) and use cross-validation (see Hastie et al. 2009, for example) to explore the predictive performance of the models during the period for which we have climate observations. The standard approach for ensuring validity of any climate reconstruction is to leave out a portion of climate observations and check to ensure that the fitted model improves the prediction of the held out values over a constant mean model (see Cook and Kairiukstis 1990, for additional information on traditional model verification). As discussed by McShane and Wyner (2011) this approach suffers due to the default choice of climate values to hold out. Here we use five-fold cross-validation, where the climate series is split into five approximately evenly sized blocks, one of which is held in each of the five resulting cross-validation datasets. This allows for a more robust check of predictive performance that is less dependent on any one default choice. However, as discussed by McShane and Wyner (2011) the cross-validation is limited in that we are unable to ascertain the performance of the different models for extreme values, as none of our hold-out blocks contain values that require extrapolation.

RESULTS.

As well as exploring the performance of the model predicting the values of held out climate values, we also believe the tree-ring measurement observations $\boldsymbol{y}$ also contain some information as to model fit of the $\boldsymbol{x}$ values. In particular, we compare how $\hat{y}_{it} \approx f(\hat{x}_t)$, the predicted value of observation y_{it} (a value that depends of the climate prediction $\hat{x}_t$) varies between the different models. Do some of the models do a better job at describing the observed variable y_{it}? Unfortunately, this comparison does not help in determining which climate predictions are "correct", but it can be useful for ruling out poorly performing models. For example, consider a case where we have two (or more) models where the predicted values of x_t are different, as we have above. If the posterior predictive checks are unable to distinguish between these models then this suggests that the data (or at least our model check) is unable to separate between the two sets of modeling assumptions based on the predicted climate series.

The exact model fitting check we use here is simple. In each iteration of the Markov chain, we determine the residuals $\delta_{it} = \hat{y}_{it} - y_{it}$ and average these over all iterations to give $\tilde{\delta}_{it}$. We then take the median of $\tilde{\delta}_{it}$ across trees to find $\tilde{\delta}_t$, an average predicted deviation in year t, which we use to compare between the different models.

RESULTS

As a complement to model checking it would be useful to compute Bayes factors (for example, see Kass and Raftery 1995) to determine the relative support the data give each model. While an attractive option theoretically, the complexity of the models leads to considerable practical difficulty in estimating these quantities. See Sisson (2005) for some of the possible approaches to estimating these quantities.

Simulation and pseudoproxies (Mann et al. 2005, 2007) have been used to assess the performance of different methods in reconstructing known values. Although simulation is a useful tool, it can lead to over-confidence if alternative data generating models are not considered. Most, if not all, pseudoproxy experiments simulate their proxies under a certain transfer function, and then test their model/method of reconstruction assuming the same transfer function. For example, Mann et al. (2005) constructed proxies by adding white noise to the climate series (a linear transfer function). This same linear response was then assumed in the model underlying the RegEM procedure they use to fit the data. To better understand the robustness of climate estimates from a reconstruction, any pseudoproxy experiment should consider departures from the statistical model used to fit the model. These departures include the transfer function, as well as other areas of the model that we did not focus on here, but discuss in the next section.

Where the current implementation of pseudoproxy experiments do a good job, is in exploring the effect of the marginal model for the climate variable. This is because most use a climate series from a

(physical) climate model, which is not the same as the statistical model used to fit the data. We suggest the results we have presented here help to explain why the RegEM approach often underestimates low-frequency variability (see, for example von Storch et al. 2004). That is, the underestimation may be due to the marginal model assumed for the prior distribution shrinking the estimates toward the marginal mean as can be seen with the modest shrinkage toward the observed period seen in figure 6(b) compared with figure 6(d). Considering alternative marginal models for the climate variables could potentially overcome such underestimation.

Another potential approach for assessing model fit comes from the statistical calibration literature. Brown and Sundberg (1987) and Brown and Sundberg (1989) discuss situations where the model used to calibrate the relationship between $\boldsymbol{x}^{obs}$ and $\boldsymbol{y}^{obs}$ no longer holds for the corresponding $\boldsymbol{y}^{mis}$ and $\boldsymbol{x}^{mis}$ data. One way they suggest determining whether the model is appropriate is comparing the estimates from a partial likelihood with those from the full likelihood. That is, we can compare our estimates of $\boldsymbol{x}^{mis}$ from the full likelihood approach we specify here, with estimates obtained using a partial likelihood approach, where the parameters determining the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ are estimated only from $\boldsymbol{x}^{obs}$ and $\boldsymbol{y}^{obs}$, with these estimates used to predict $\boldsymbol{x}^{mis}$. While this is an appealing diagnostic check, many of the observations in this dataset are from fossilized trees and have no overlap with the observed climate data. This means that any comparison would be comparing two different datasets. In such a case, there would be no surprise if the predictions differ.

The most important model check available is that of scientific validity. For example, most will consider the reconstruction (and underlying model) in figure 6(c) to be scientifically invalid based on the reconstruction. This is because it is unlikely that these trees can survive a summer with a mean temperature below freezing. While this may be a valid criticism, it somewhat misses the point that the data appears to be unable to distinguish between various models that lead to very different reconstructions. Moreover, other seemingly valid reconstructions may also have scientifically implausible implications that are not immediately recognizable.

This does not mean that we think that scientific validity is not important, but it means that we see the value in separating prior information from data driven estimation. In other words, we believe that being able to quantify the influence of prior information (in whatever form) on the estimates is important.

8 Discussion

The prediction of historical climate from tree-ring data has many sources of uncertainty. Instead of trying to remove (or ignore) these in a stepwise manner, conditioning on model output each time, we have attempted to build a model-based framework for including these sources in a coherent way. In

saying this, we acknowledge there are numerous additional sources of uncertainty that we either have not, or have only briefly mentioned. Two of these are uncertainty in the specification of a tree-ring growth model, and uncertainty in the dating of tree-rings.

Despite changing many of the modeling assumptions, including the transfer function and the marginal model for climate through time, we assumed that tree-ring growth responds to age according to a negative-exponential function (linear on the log scale) in all of the models fitted. The reason for this was because we wanted to focus on what we consider important issues related to extrapolation and the model sensitivity corresponding to that. Having the same growth function in every model may help explain why the shape of the climate predictions (i.e. the dates where climate is assumed higher/lower than average) is as "robust" as it is, despite the sometimes dramatic effect on the magnitude. Other aging functions can be used here, although care needs to be taken to ensure the overall model continues to be identifiable. Assuming a flexible aging function, such as a polynomial regression or cubic spline will lead to identifiability problems if the climate variable also has a flexible mean function, such as the cubic spline used here. Effectively the inherent sparseness of the (climate) data leads to assumptions about the aging process being necessary. For example, we can consider a model-based RCS analogue by assuming that each tree has the same (possibly flexible) aging function.

Despite assuming here that the data are dated correctly this is very unlikely to be the case. The fossilized trees included in the dataset are dated according to a process called crossdating. Fritts (1976) considers this procedure to be the most important principle in dendrochronology. The process aligns the trees according to the growth year and uses correlation in the tree-ring growth measurements to (i) determine locally absent, or missing rings, (ii) false rings, and (iii) date series with unknown dates. Fritts (1976) describes how in a year of extreme climate a tree "may not form a ring on all portions of the stem", resulting in a missing ring. A false ring can occur when there is "a change in cell structure ... within an annual growth increment ... [that] resembles the boundary of a true annual ring." Finally, crossdating can be used to help align and assign the lifetime of fossilized series. The problem, as it relates to our discussion here, is that the crossdating process ignores the uncertainty in making all of these assignments. Furthermore, this could lead to potential bias propagating back through time. The ideal solution, is to include the cross-dating process within probabilistic statistical models. In saying this, we acknowledge this task is difficult, not only in specifying the cross-dating process, but also in the related computational challenges from fitting such a model.

Some may view the sensitivity in the predictions of climate we observe across various models as a failure of using a model-based joint standardization and reconstruction approach. We take the opposite view. We openly acknowledge the sensitivity of the resulting predictions to the choice of model and

believe this shows the limitations of the data and the role modeling assumptions have on the predictions. We believe that it also shows the sensitivity of the current standardization and reconstruction approaches to conditioning on parameter estimates from statistical models. While a reconstruction obtained using the traditional approaches may be appropriate for (i) the parameter estimate conditioned on in the standardization phase, and (ii) the underlying assumption about the distribution for the climate variable implicitly assumed in the calibration procedure, these analyses show that the parameter values are far from certain and the assumptions about the climate distribution are far from innocuous. In particular, any dataset that involves extrapolation beyond the range of the observed data, as is the case here, will be subject to sensitivity in the resulting predictions with respect to modeling assumptions.

Many of the ideas that we explore for tree-ring measurements are also applicable if using multiple proxies to reconstruct climate. Unfortunately, other proxy variables include other omplexities that must be accounted for in the modeling. These complexities can be placed into four broad categories:

1. There are non-climatic as well as climatic influences on the proxy measure. The ideas we discuss here directly relate to this problem.
2. There is uncertainty in the dating of the proxy. We are unsure what point in time (and hence what climate values) should be related against the proxy measurement.
3. Variation in the coarseness of the data. For example, in recent history the value of the proxy may reflect climate over 1 year, while in more distant history the value of the proxy may relate to climate of 150 years (Li et al. 2010). To make matters worse, often these periods of time are themselves uncertain.
4. It is difficult to formalize how the proxy responds to the climate.

These differences are incorporated either through the model for how the proxy measurements depend on the climate, or through hierarchical extensions. Despite these difficulties, including multiple proxies in a reconstruction can be of value, particularly when different proxy types reflect different aspects of the climate process (for example, see Li et al. 2010).

Multiple proxies may also help in the central problems mentioned in this manuscript, that of extrapolation and model sensitivity. In particular, the multiple proxies may provide information about the calibration curve across a range of climate values. However, this advantage is not automatic, in that, including observations from multiple proxies will not automatically lead to robust predictions. For example, consider a dataset where we have multiple proxies from multiple locations. The goal of the model is to estimate a spatial field of climate values, that we will also assume align spatially with our proxy measurements. One possible model for reconstructing such data is to consider a local model for each proxy dataset, with a hierarchical spatial model on the underlying local climate values that describes how

the climate changes both across space and time. Such a spatial hierarchical model could look something like the models considered in Tingley and Huybers (2010). Considering such a model makes it clear that the inclusion of multi-proxy data does not necessarily overcome the problem of extrapolation. If we have specified an incorrect model describing how the proxies respond to climate, it is likely that our models will yield a local approximation to the response function in each location.

The value of including data from multiple proxies is greater flexibility in model checking. Consider again, the multi-proxy model described above. Under the assumption that a proxy in two (or more) locations should respond in a similar manner to climate, then we can check whether there are any systematic differences in the estimated response functions. For example, if the parameters estimates for a linear response to climate depend on the location (i.e. they are all high in one location and low in another), this could be indicative of a misspecified model with a local approximation fitted in each location. However, such statements depend critically on the assumptions we make, as it may just be that the proxy in one location has a much stronger relationship with climate than the proxy in the other location.

There are many hierarchical extensions that we could consider. These include spatial components, such as the models considered by Tingley and Huybers (2010) that would allow us to describe not only how the climate changes through time, but how these changes vary across space. Other possibilities include models for climate that include covariates (for example Li et al. 2010) or that incorporate physical models. The inclusion of such extensions provides two substantial future challenges. The first is computational: even with only one location, 500 years and 247 trees, there was a substantial computational burden for our analysis. Computation time further increased for the full 2000 year reconstruction, to the point where it was impractical to fit numerous models and to perform cross-validation. With numerous locations the computational burden appears daunting and we need an approach that can scale models of this sort. The second challenge is one of complexity and applicability: with additional hierarchical components included with additional data, we need to be able to specify and check any necessary assumptions. Related to this is the role of data collection: strategies are needed that justify the simplifying assumptions needed to fit the models. This is especially the case as climate models based on calibration necessarily involve extrapolation.

References

Berkson, J. (1969), "Estimation of a linear function for a calibration line; consideration of a recent proposal," *Technometrics*, 11, 649–660.

Briffa, K. R., Jones, P. D., Bartholin, T. S., Eckstein, D., Schweingruber, F. H., Karlen, W., Zetterberg, P., and Eronen, M. (1992), "Fennoscandian summers from AD 500: temperature changes on short and long timescales," *Climate Dynamics*, 7, 111–119.

Briffa, K. R. and Melvin, T. M. (2009), "A closer look at Regional Curve Standardisation of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application." in *Dendroclimatology: Progress and Prospects*, eds. Hughes, M. K., Diaz, H. F., and Swetnam, T. W., Springer Verlag, In Press.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A. (2001), "Low-frequency temperature variations from a northern tree ring density network," *Journal of Geophysical Research*, 106, 2929–2942.

Briffa, K. R., Shishov, V. V., Melvin, T. M., Vaganov, E. A., Grudd, H., Hantemirov, R. M., Eronen, M., and 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, 2269–2282.

Brown, P. and Sundberg, R. (1989), "Prediction diagnostics and updating in multivariate calibration," *Biometrika*, 76, 349–361.

Brown, P. J. (1982), "Multivariate Calibration," *Journal of the Royal Statistical Society B*, 44, 287–321.

Brown, P. J. and Sundberg, R. (1987), "Confidence and Conflict in Multivariate Calibration," *Journal of the Royal Statistical Society B .*, 49, 46–57.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008), "High-dimensional sparse factor modeling: Applications in gene expression genomics," *Journal of the American Statistical Association*, 103, 1438–1456.

Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A., and Funkhouser, G. (1995), "The 'segment length curse' in long tree-ring chronology development for palaeclimatic studies," *The Holocene*, 5, 229–237.

Cook, E. R. and Kairiukstis, L. A. (eds.) (1990), *Methods of Dendrochronology*, Kluwer Academic Publishers.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987), "Hybrid Monte Carlo Algorithm," *Physics Letters B*, 195, 216–222.

Eilers, P. H. C. and Marx, B. D. (1996), "Flexible Smoothing with B-splines and Penalties," *Statistical Science*, 11, 89–121.

Esper, J., Cook, E. R., and Schweingruber, F. H. (2002), "Low-Frequency Signals in Long Tree-Ring Chronologies for Reconstructing Past Temperature Variability," *Science*, 295, 2250–2253.

Esper, J., Krusic, P. J., Peters, K., and Frank, D. (2009), "Exploration of long-term growth changes using the tree-ring detrending program "Spotty"," *Dendrochronologia*, 27, 75–82.

Fritts, H. C. (1976), *Tree Rings and Climate*, Academic Press: London.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Chapman & Hall/CRC, 2nd ed.

Ghosh, J. and Dunson, D. (2009), "Default prior distributions and efficient posterior computation in Bayesian factor analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320.

Ghosh, M., Carlin, B. P., and Srivastava, M. S. (1995), "Probability matching priors for linear calibration," *Test*, 4, 333–357.

Girolami, M. and Calderhead, B. (2011), "Riemann manifold langevin and hamiltonian monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 123–214.

Grudd, H., Briffa, K. R., Karlen, W., Bartholin, T. S., Jones, P. D., and Kromer, B. (2002), "A 7400-year tree-ring chronology in northern Swedish Lapland: natural climatic variability expressed on annual to millennial timescales," *The Holocene*, 12, 657–665.

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

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning. Second Edition.*, Springer.

Hoadley, B. (1970), "A Bayesian Look at Inverse Linear Regression," *Journal of the American Statistical Association*, 65, 356–369.

Jansen, E., Overpeck, J., Briffa, K. R., Duplessy, J. C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D. (2007), "Contribution of Working Group I to the Fourth Assessment Report of the

Intergovernmental Panel on Climate Change," in *Climate Change 2007: The Physical Science Basis*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jones, P. D. and Mann, M. E. (2004), "Climate Over Past Millennia," *Reviews of Geophysics*, 42, RG2002.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Krutchkoff, R. G. (1967), "Classical and Inverse Methods for Calibration," *Technometrics*, 9, 425–439.

— (1969), "Classical and inverse regression methods of calibration in extrapolation," *Technometrics*, 11, 605–608.

Li, B., Nychka, D. W., and Ammann, C. M. (2010), "The value of multiproxy reconstruction of past climate," *Journal of the American Statistical Association*, 105, 883–895.

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

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C. (2005), "Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate," *Journal of Climate*, 18, 4097–4107.

— (2007), "Robustness of proxy-based climate field reconstruction methods," *Journal of Geophysical Research*, 112, D12109.

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

Melvin, T. M. and Briffa, K. R. (2008), "A "Signal-Free" approach to Dendroclimatic Standardisation," *Dendrochronologia*, 26, 71–86.

Neal, R. M. (1996), *Bayesian Learning for Neural Networks*, Springer.

— (2011), *MCMC using Hamiltonian dynamics*, Chapman & Hall/CRC, chap. 5, pp. 113 – 162.

Osborn, T. J. and Briffa, K. R. (2006), "The Spatial Extent of 20th-Century Warmth in the Context of the Past 1200 Years," *Science*, 311, 841–844.

Osborne, C. (1991), "Statistical Calibration: A Review," *International Statistical Review*, 59, 309–336.

Plummer, M. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria*.

Rutherford, S., Mann, M. E., Delworth, T. L., and Stouffer, R. J. (2003), "Climate Field Reconstruction under Stationary and Nonstationary Forcing," *Journal of Climate*, 16, 462–479.

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., and Jones, P. D. (2005), "Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain," *Journal of Climate*, 18, 2308–2329.

Schneider, T. (2001), "Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values," *Journal of Climate*, 14, 853–887.

Seber, G. A. F. and Lee, A. J. (2003), *Linear Regression Analysis*, Wiley.

Sisson, S. A. (2005), "Transdimensional Markov Chains," *Journal of the American Statistical Association*, 100, 1077–1089.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation (with discussion)," *Journal of the American Statistical Association*, 82, 529–550.

Tingley, M. and Huybers, P. (2010), "A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 1: Development and applications to paleoclimate reconstruction problems." *Journal of Climate*, 23, 2759–2781.

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J. (2011), "An efficient forward model of the climate controls on interannual variation in tree-ring width," *Climate Dynamics*, 36, 2419–2439.

Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J. A., Rusticucci, M., Soden, B., and Zhai, P. (2007), "Observations: surface and atmospheric climate change," in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press.

von Storch, H., Zorita, E., Jones, J., Dimitriev, Y., Gonzalez-Rouco, F., and Tett, S. (2004), "Reconstructing Past Climate from Noisy Data," *Science*, 306, 679–682.

West, M. (2003), "Bayesian factor regression models in the ?large p, small n? paradigm," *Bayesian statistics*, 7, 723–732.

Williams, E. J. (1969), "A note on regression methods in calibration," *Technometrics*, 11, 189–192.

Yin, M. (2000), "Noninformative Priors for Multivariate Linear Calibration," *Jounral of Multivariate Analysis*, 73, 221–240.

Yu, Y. and Meng, X.-L. (2011), "To Center or Not to Center: That Is Not the Question–An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency," *Journal of Computational and Graphical Statistics*, 20, 531–570.

Figures

Figure 1: The number of cores (series) per year in the dataset.

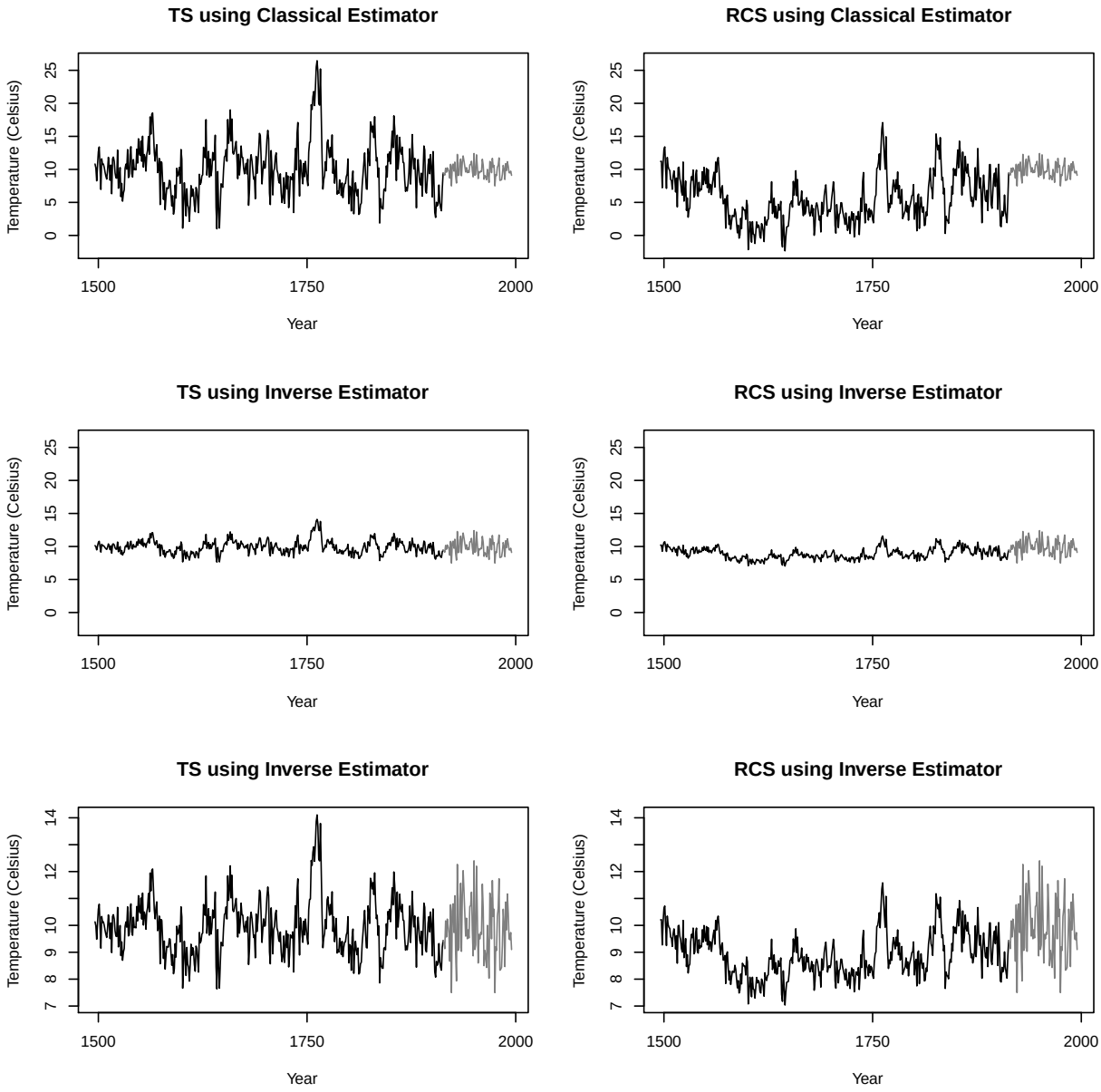


Figure 2: Reconstructions using both classical and inverse estimators for TS and RCS chronologies for the Scots Pine data. The grey values are observed temperature observations and the black values are estimated using linear calibration. The y-axis is the same in plots in the first two rows. The only difference between the last two rows is the scale of the y-axis.

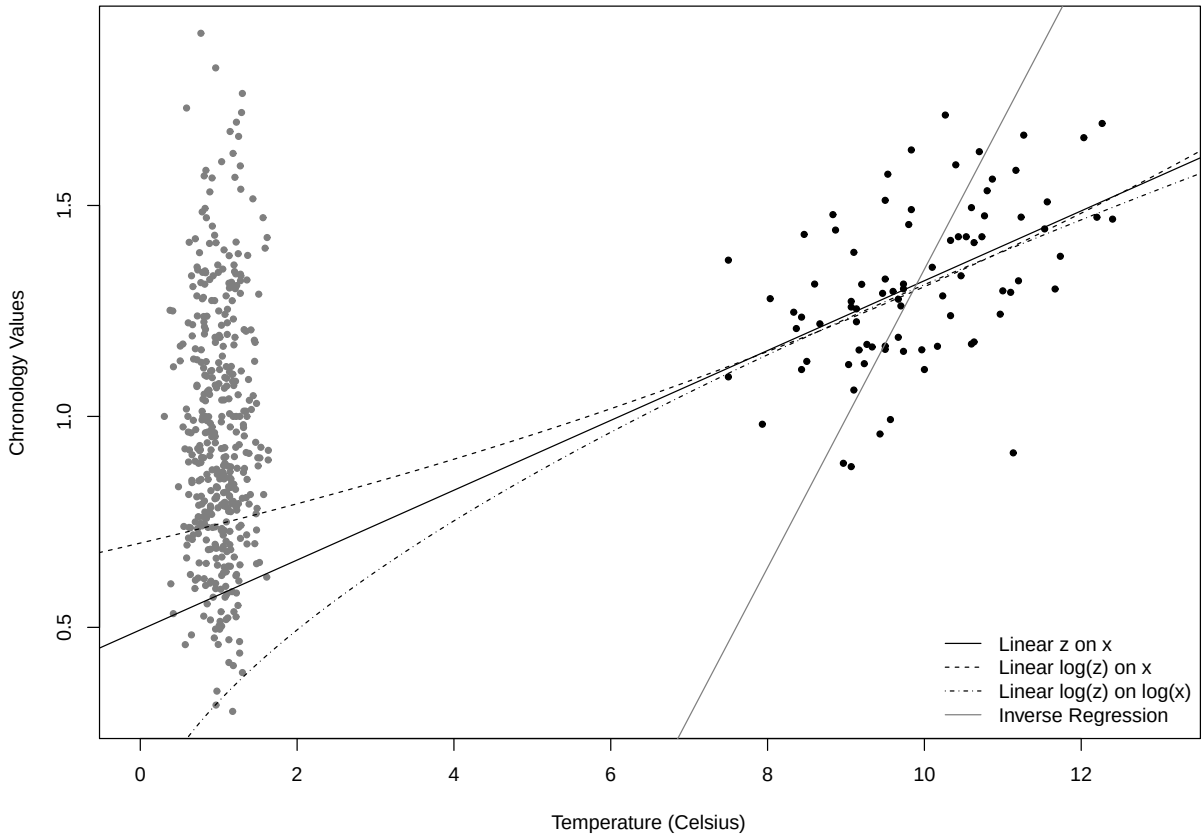
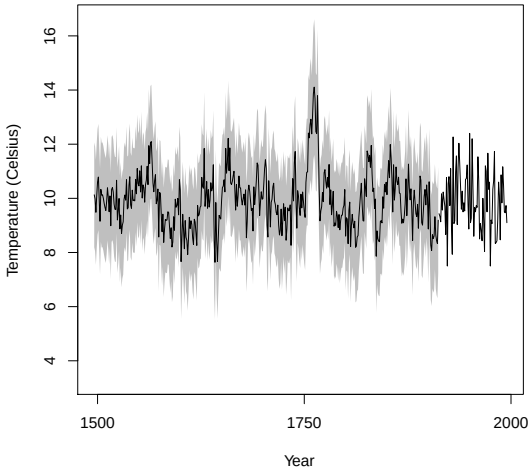
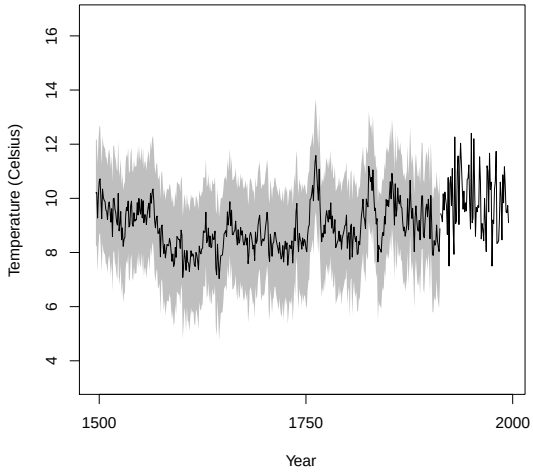


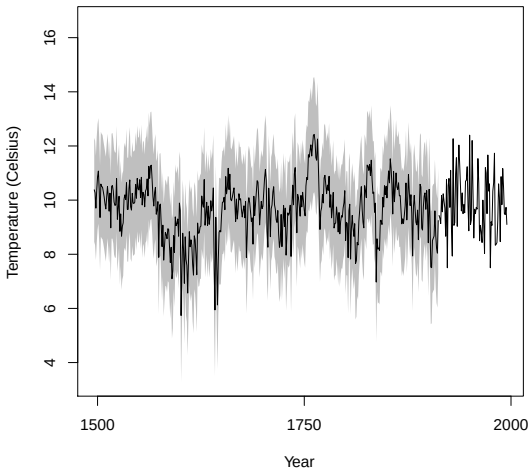
Figure 3: A plot displaying the effect of transfer function on the reconstruction of the temperature variable for the Scots Pine data. The black points are the $(\boldsymbol{z}^{obs}, \boldsymbol{x}^{obs})$ pairs. The grey values are the $\boldsymbol{z}^{mis}$ that we have jittered around the value 1 on the x-axis. The four lines denote four (or many) possible models we could consider for the data, including three different models under the classical approach, as well as inverse regression.



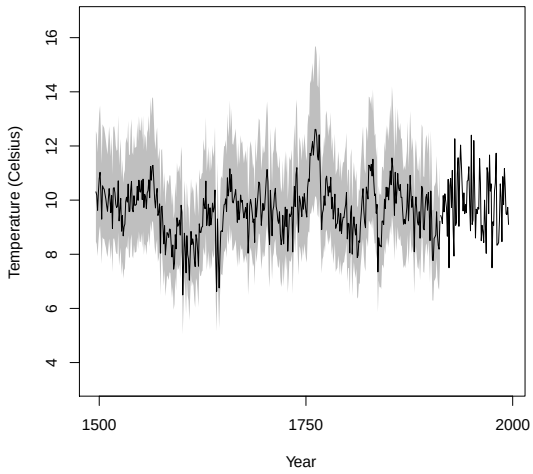
(a) TS using Inverse Estimator



(b) RCS using Inverse Estimator

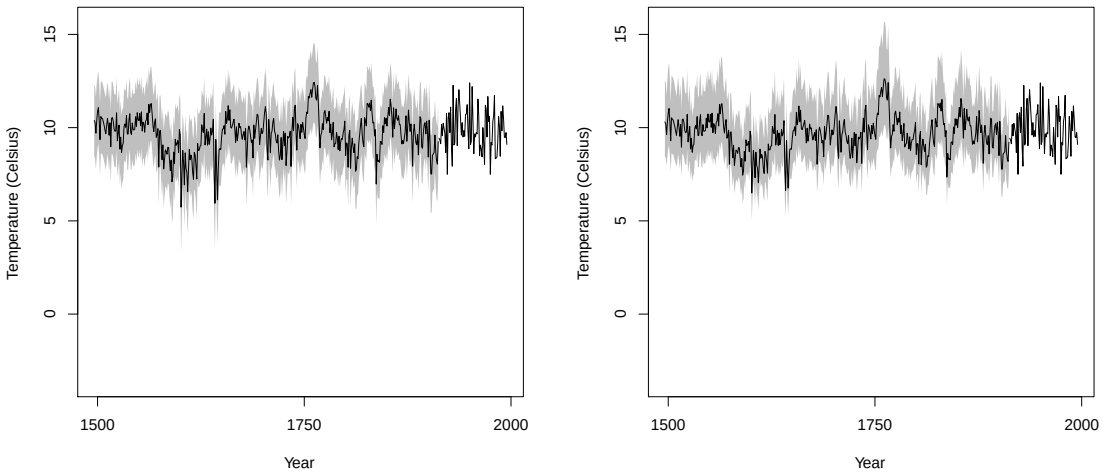


(c) Model-based estimates: $\log(y)$ on x



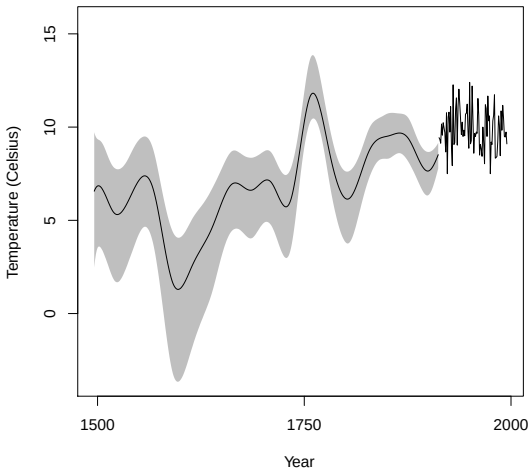
(d) Model-based estimates: $\log(y)$ on $\log(x)$

Figure 4: Reconstructions comparing traditional approaches: inverse estimates for both TS and RCS with estimates from the model-based approach. For the traditional approaches, the uncertainty bands are 95% prediction intervals. For the model-based approach, the uncertainty bands are central 95% credible intervals for the missing temperature values. In all models, the black lines corresponding that have no uncertainty are the observed temperature data. All other black lines are predicted values (traditional approaches) and the estimated median of the posterior distribution for the missing climate values (model-based approaches). All plots have a common y-axis.

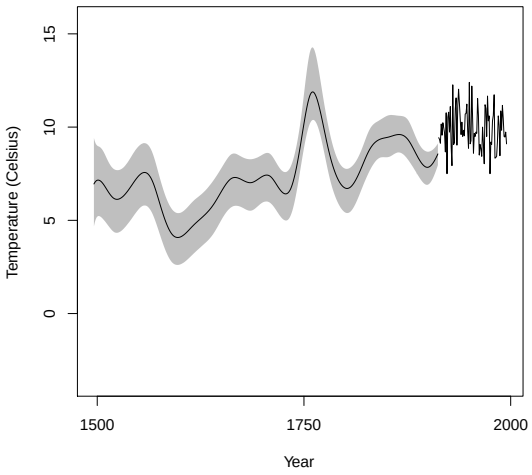


(a) $\log(y)$ on x; Constant Mean

(b) $\log(y)$ on $\log(x)$; Constant Mean

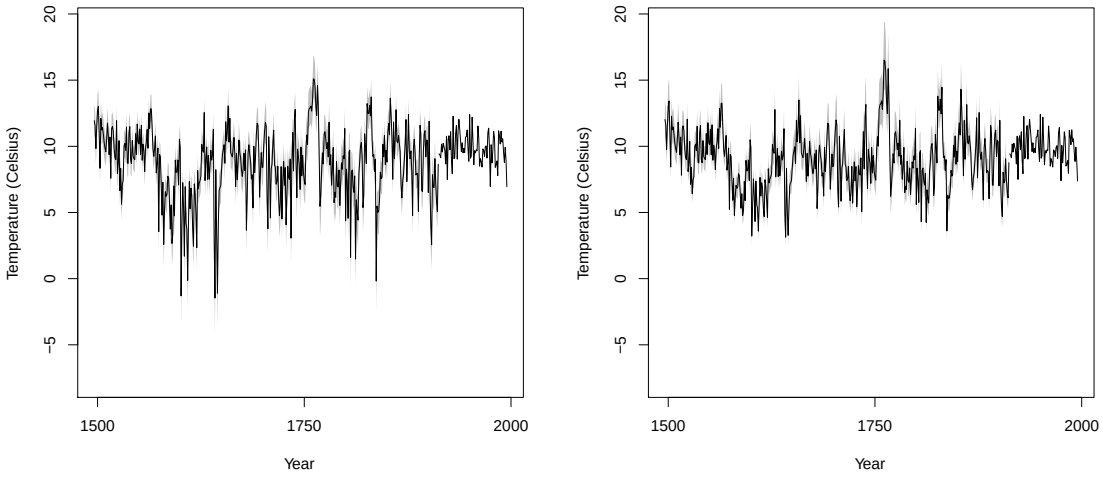


(c) $\log(y)$ on x; Smoothly Evolving Mean



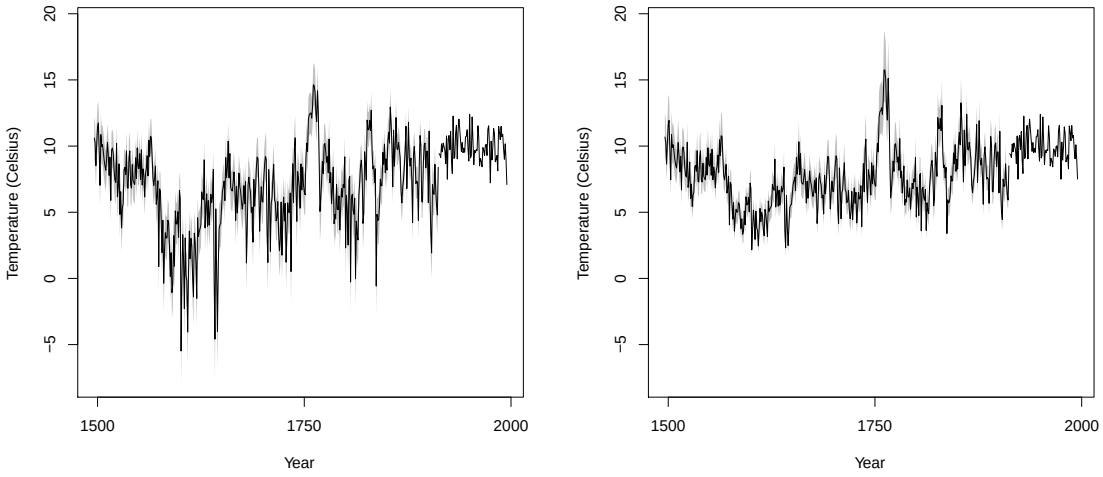
(d) $\log(y)$ on $\log(x)$; Smoothly Evolving Mean

Figure 5: Reconstructions comparing model-based approaches with various (i) transfer functions describing the tree-ring growth model, and (ii) marginal (prior) models for the temperature variable. The uncertainty bands are central 95% credible intervals for the missing temperature values. In all models, the black lines corresponding that have no uncertainty are the observed temperature data. All other black lines are the estimated median of the posterior distribution for the missing climate values. All plots have a common y-axis.



(a) $\log(y)$ on x; Constant Mean

(b) $\log(y)$ on $\log(x)$; Constant Mean



(c) $\log(y)$ on x; Smoothly Evolving Mean

(d) $\log(y)$ on $\log(x)$; Smoothly Evolving Mean

Figure 6: Reconstructions assuming that response to climate varies by tree. The different reconstructions compare model-based approaches with various (i) transfer functions describing the tree-ring growth model, and (ii) marginal (prior) models for the temperature variable. The uncertainty bands are central 95% credible intervals for the missing temperature values. In all models, the black lines corresponding that have no uncertainty are the observed temperature data. All other black lines are the estimated median of the posterior distribution for the missing climate values. All plots have a common y-axis.