A Model-Based Approach to Climate Reconstruction Using Tree-Ring Data

Matthew R. Schofield[12]*, Richard J. Barker[2], Andrew Gelman[3], Edward R. Cook[4] and Keith R. Briffa[5]

[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.

[5]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

October 16, 2015

Abstract

Quantifying long-term historical climate is fundamental to understanding recent climate change. Most instrumentally recorded climate data are only available for the past 200 years, so proxy observations from natural archives are often considered. We describe a model-based approach to reconstructing climate defined in terms of raw tree-ring measurement data that simultaneously accounts for non-climatic and climatic variability.

In this approach we specify a joint model for the tree-ring data and climate variable that we fit using Bayesian inference. We consider a range of prior densities and compare the modeling approach to current methodology using an example case of Scots pine from Torneträsk, Sweden to reconstruct growing season temperature. We describe how current approaches translate into particular model assumptions. We explore how changes to various components in the model-based approach affect the resulting reconstruction.

We show that minor changes in model specification can have little effect on model fit but lead to large changes in the predictions. In particular, the periods of relatively warmer and cooler temperatures are robust between models, but the magnitude of the resulting temperatures are highly model dependent.

*E-mail: mschofield@maths.otago.ac.nz

Such sensitivity may not be apparent with traditional approaches because the underlying statistical model is often hidden or poorly described.

1 Introduction

Instrumental meteorological 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 of years. The statistical problem is to infer historical climate without having direct observations beyond the most recent two centuries. A common solution is to index historical climate values using proxy observations from natural archives such as tree rings, lake sediments, bore holes, corals and ice cores (see Jones and Mann (2004) and Jansen et al. (2007) for examples). Reconstructions based on these sources feature prominently in the Fifth Assessment Report of Working Group 1 of the Intergovernmental Panel on Climate Change (Masson-Delmotte et al. 2013). This report and other work, including that of the Past Global Changes (PAGES) 2K group (PAGES 2k Consortium 2013), have demonstrated robustness in the general pattern of variability in reconstructions representing large regions at decadal to centennial timescales.

The standard approach to reconstructing climate from tree-ring data involves (i) filtering out any signal from non-climatic sources so that only climate signal (and random noise) remain, (ii) combining multiple filtered tree series from the same macro-environmental setting into one mean chronology, and (iii) using the resulting chronology, alone or combined with other proxy measurements, to reconstruct local, hemispheric or global climate. The first two steps are referred to as "standardization" in tree-ring studies (e.g. Fritts 1976). There is a rich literature describing dendroclimatic methods devoted to preserving the climate signal during standardization (e.g. Briffa et al. 1992, Cook et al. 1995, Melvin and Briffa 2008, Esper et al. 2009, Briffa and Melvin 2011). However, the major focus of recent research has been on the third step (for example, see Jones et al. 2009, Tingley et al. 2012, Wahl and Smerdon 2012). This has resulted in several methodological developments for reconstructing climate values from filtered proxy data. More recently, Bayesian inference has been proposed. Haslett et al. (2006) used a Bayesian approach for reconstruction using a pollen proxy. Li et al. (2010), Tingley and Huybers (2010a,b) and McShane and Wyner (2011) explore Bayesian approaches for multi-proxy reconstructions.

The standardization process results in a chronology that is easily shared among the community for use in climate reconstructions. However, in subsequent use the chronologies are treated as data and the standardization process does not feature in the climate reconstruction model. One consequence is that uncertainties from steps (1) and (2) above do not propagate through to the final predictions. Any incorrectly modeled variation from the standardization of the proxy measurements will be carried forward to the climate reconstruction. The multi-step procedures also make it different to check overall model fit. Assumptions made in the second or

third stages cannot be checked against the raw data. To overcome these problems we merge the three steps above, specifying a joint model for the raw data and climate measure. Simultaneously describing climatic and non-climatic influences on the raw data makes it possible to overcome difficulties encountered in multi-step procedures, such as reconstructing long-scale climate variability, often referred to as the segment-length curse (Cook et al. 1995).

Use of a joint model ensures that assumptions are made explicit. As a first step we need to find model-based descriptions for the underlying assumptions in both the standardization and reconstruction processes described above. Statistical calibration is often used to reconstruct the climate in step (3) above, with different approaches typically presented as a choice between various statistical procedures. In our framework, the different approaches can be viewed as differences in modeling assumptions and comparison between procedures becomes a choice between models.

Once a modeling framework has been specified, we can consider alternative assumptions and examine how the predictions of historical climate depend on changes in both the standardization and reconstruction components of the model. If the predictions change substantially, this shows sensitivity of the modeling assumptions and can help reveal limitations of the data.

We first introduce the data in section 2 and summarize currently used methodology for standardizing tree-ring data in section 3. We then comment on the problems with the current methodology in section 4 before defining and exploring the model-based approach in section 5 and considering model adequacy in section 6. We then discuss the implications of the modeling approach in section 7, that includes a summary for the paleoclimate community.

2 Data

Our tree-ring data are records of annual radial growth increments (i.e. ring widths) of living and well-preserved remnant (subfossil) Scots pine (*Pinus sylvestris*) growing near the latitudinal tree-line in Torneträsk, northern Sweden (Grudd et al. 2002, Briffa et al. 2008). Ring widths are measured along one or more radii, from the earliest ring at or near the center of the tree bole outward to include the most recently formed growth ring. They represent the annual radial expansion of woody tissue in the bole of a tree. They are typically measured from samples taken at breast height. While ring widths vary from tree to tree and also at different heights in the same tree, the overall pattern of year to year relative ring width changes is similar. The absolute dimensions of these ring widths vary according to what part of the tree they represent and to the overall growth rate of the tree in question. Many trees of the same species are sampled from a site, defined as an area within the same macro-environment, to ensure within site replication.

It is assumed that ring width variation reflects, among other things, climatic variation (Fritts 1976). For the Torneträsk trees it is believed that the important

climatic influence is the growing season temperature (Grudd et al. 2002). However, even in an unchanging environment ring widths change over time due to the changing allocation of growth products around the outside of an expanding bole as the tree grows. Trees exhibit various and sometimes complex growth forms but a pattern of early ring width increase followed by a slower reduction in ring widths in older age is generally observed. Because such changes in measured ring width time series do not represent climate variability it is important to account for them in the analysis. Traditionally, such variations have been filtered out via the standardization process referred to in section 1.

It is important that each ring width is assigned to its exact calendar year of formation. This is done by a process called "crossdating" (Fritts 1976) that matches high-frequency patterns of wide and narrow rings between the sampled trees. This helps identify and enable the correction of dating errors that might arise by simple ring counting (Fritts and Swetnam 1989). More details about crossdating are in the supplementary materials, section 1 (herein S1). The data we use here are obtained after crossdating. We discuss potential uncertainty resulting from the crossdating procedure in section 7.

The full dataset consists of ring widths from 587 trees. We analyze a subset comprising $k = 247$ trees with at least 20 ring width observations after the year 1496, leading to a 500 year series. The number of ring width observations for each tree (defined as the segment length) varied from 26 to 485 years, with an average segment length of 179 years. Additional details about the data, including plots, are available in S1.

As the climatic factor limiting growth of these trees is believed to be growing season temperature, the climate variable we reconstruct is the mean temperature during June-August for the Abisko weather station, the closest weather station to Torneträsk. Temperature has been directly measured from 1913 to 1995. In all models that follow, we assume that these temperature values are observed without error. Our goal is to predict the mean June-August temperature values from 1496 to 1912. There are 121 trees that have measurements during the period of observed temperatures. To account for the non-climatic ring growth, the only other covariate we consider is the age of the tree in each year. By convention, the age of a tree is defined as the number of years since the innermost ring of the core sample; this corresponds to the year in which the tree reached the height at which the sample was taken for living trees or to the innermost year for the remnant wood samples.

We use the Torneträsk data for two reasons. Firstly, the correlation between the growth increments and climate is strong enough to make viable local reconstructions (see section 4 for details). Secondly, the spread of age classes of the trees is relatively even through time. This makes the use of regional curve standardization (discussed in section 3.2) viable for these data as well as traditional standardization (discussed in section 3.1).

Although our focus here is on reconstructing growing season temperature, the

models we fit and the ideas we discuss can be generalized to other climate variables and to global and spatial field based reconstructions.

2.1 Notation

The raw data are denoted by the $k \times n$ matrix $\boldsymbol{y}$, where y_{it} is the observed ring width for tree i in year t. We use f_i and l_i to denote the first and last year tree i was measured. 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 $\boldsymbol{y}$ for which the climate values $\boldsymbol{x}^{mis} = (x_1, \ldots, x_m)$ are missing; the latter denotes values of $\boldsymbol{y}$ for which climate values $\boldsymbol{x}^{obs} = (x_{m+1}, \ldots, x_n)$ have been observed. We denote tree age by the $k \times n$ matrix $\boldsymbol{a}$, where a_{it} is the age of tree i in year t.

Throughout we distinguish between climate and temperature. When describing general methodology, we refer to $\boldsymbol{x}$ as a climate variable. When discussing the Torneträsk data, we refer to $\boldsymbol{x}$ as temperature.

3 Standard methodology

In this section, we outline the traditional approach to the multi-step procedure for reconstructing climate variables. The first step involves standardizing the ring widths, $\boldsymbol{y}$ (Fritts 1976, Cook and Kairiukstis 1990). Standardization serves to (1) filter out non-climatic growth influences, in particular, age effects and tree to tree differences in growth, and (2) compress the information contained in k noisy tree-ring series at a particular site into one synthetic chronology that contains a common signal, assumed to be climatic in nature. We denote the standardized chronologies obtained from $\boldsymbol{y}$ by the vector $\boldsymbol{z} = (z_1, \ldots, z_n)$, where z_t is the value for year t. As for $\boldsymbol{y}$, we split $\boldsymbol{z}$ into $\boldsymbol{z}^{mis} = (z_1, \ldots, z_m)$ and $\boldsymbol{z}^{obs} = (z_{m+1}, \ldots, z_n)$.

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

3.1 Traditional standardization

TS involves selecting and fitting a statistical model to the annual growth increments of each tree separately in order to remove the non-climatic growth influences such as those associated with tree age (Fritts 1976). Raw measurements, y_{it} are then scaled to give an index for each tree, $w_{it} = y_{it}/\hat{y}_{it}$, where $\hat{y}_{it}$ is the fitted value under the assumed model. This index is then averaged across trees at a site to give a mean index (or chronology) for that site, $z_t = m(w_{1t}, \ldots, w_{kt})$. The function $m(\cdot)$ is a measure of central location and may be chosen to minimize the effect of outliers. Variation in z_t is presumed to reflect primarily climatic influences with tree-specific growth-related effects removed by scaling. Deviations for each tree may be further modeled to remove autocorrelation before being combined (Cook and Kairiukstis

1990). Almost all datasets in the International Tree-Ring Data Bank (`http://www.ncdc.noaa.gov/paleo/treering.html`) have been standardized in this way.

TS has two major problems. First, the uncertainty about parameters in the standardization is ignored in all further modeling. Second, 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 TS effectively detrends each 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 (Cook et al. 1995); i.e. some of the low-frequency climate "baby" can get thrown out with the non-climate "bathwater."

This second problem has been named the "segment length curse" (Cook et al. 1995): one can only expect to recover climate signals that have a higher frequency (in relation to climate cycles) than the inverse of segment length of individual tree-ring measurements. As low-frequency climate signal (and possibly some medium-frequency signal) is removed during standardization, the examination of changes in climate over hundreds, or thousands of years, becomes problematic. We use simulated temperature and tree-ring series to demonstrate the problem in S3.

3.2 Regional curve standardization

RCS represents an attempt to overcome the segment length curse and preserve the low-frequency climate signal (Briffa et al. 1992). RCS assumes that there is a common growth rate for all trees of the same age (or age class, e.g. we could assume a common growth rate applies to binned ages of 10-20 year increments). We define $\hat{G}_j$ to be the estimated growth rate for age j and obtain an index $w_{it} = y_{it}/\hat{G}_{a_{it}}$. As before, we then average the index to obtain the chronology $z_t = m(w_{1t}, \ldots, w_{kt})$.

The idea behind RCS is that climatic effects on the standardization curve can be largely eliminated by averaging measurements across many trees sampled from across a range of overlapping time periods; in effect, removing the climate influence on $\hat{G}_j$. 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. RCS makes minimal assumptions about the shape of the non-climatic growth trend, although such constraints may be included.

RCS has several potential problems (Briffa and Melvin 2011). As with TS, incorrectly modeled variation can distort modeling of the climate signal and induce bias in the resulting reconstruction, e.g. tree-specific differences in the growth can be misinterpreted as changes in the climate signal, particularly when the sample size is small. Furthermore, the RCS approach also ignores uncertainty about the standardization procedure in all further modeling.

RCS is inappropriate when samples are taken from trees that are all about the same age since they form a common age class and have been exposed to the same climate conditions. Fortunately, many chronologies have a range of ages at each time

period, especially when they consist of living and sub-fossil trees. This is the case for the Torneträsk Scots pine data set which has a relatively even distribution of tree age classes spread over a relatively long time. This makes simple implementation of the RCS technique a viable option, provided the mean biological age trend can be assumed to be approximately constant between trees. More complex implementations of RCS have been developed, such as processing subsets of the full data set to allow trees to have different growth rates (Melvin et al. 2013). The analysis we describe here uses a single standardization curve RCS implementation.

3.3 Statistical calibration

Following standardization, statistical calibration (Osborne 1991) is used to reconstruct climate from the site-specific chronologies. We initially focus on univariate linear calibration, where we assume that a single chronology $\boldsymbol{z}$ has a linear relationship with climate observations $\boldsymbol{x}$.

Observed data $\boldsymbol{z}^{obs}$ and $\boldsymbol{x}^{obs}$ are used to estimate parameters describing the relationship, which are then used to predict $\boldsymbol{x}^{mis}$ using $\boldsymbol{z}^{mis}$. This can be done using either: (i) classical calibration, where we regress $\boldsymbol{z}^{obs}$ on $\boldsymbol{x}^{obs}$, or (ii) inverse calibration, where $\boldsymbol{x}^{obs}$ is regressed on $\boldsymbol{z}^{obs}$. The two corresponding procedures, along with their properties are described in S2. In S2 we show how the underlying model for inverse calibration can be obtained if (i) we start with the model for classical calibration, (ii) consider $x_1, \ldots, x_n$ to be iid realizations of n normal random variables, and (iii) use Bayes rule to find the distribution of x_t conditional on z_t. The implied normal model for $\boldsymbol{x}$ leads to predictions[1] from inverse calibration being shrunk toward the mean value of $\boldsymbol{x}^{obs}$ compared with the predictions from classical calibration. This helps explain why inverse calibration is preferred in terms of MSE only if x_t^{mis} is close to the mean $\boldsymbol{x}^{obs}$ value.

Extensions to the multivariate case are also discussed in S2. In particular, we show how the model underlying the RegEM algorithm of Schneider (2001), commonly used for multiple proxy climate reconstructions, is equivalent to a multivariate extension of inverse calibration. See Christiansen (2011), Christiansen and Ljungqvist (2011) and Christiansen (2014) for discussion of calibration assumptions with respect to various climate reconstructions.

4 Problems with existing approaches

We begin by examining the local reconstruction for the Scots pine dataset using traditional methods. We construct TS and RCS chronologies using the approaches outlined in sections 3.1 and 3.2. We use a negative exponential growth curve in TS to

[1]Throughout we use prediction to refer to making inference regarding $x_t,\ t = 1, \ldots, m$.

account for tree age and we assume the same growth rate for all trees in binned (10-year) age groups in RCS , with more details provided in S4. Of interest is how resulting predictions compare with the assumptions and properties of both the standardization approaches, described in section 3, as well as the calibration approaches, described in section 3.3 and S2. We refer to the reconstructions found using the traditional multi-step approach as TS_INV, TS_CLASS, RCS_INV and RCS_CLASS. Table 1 describes the labels for all reconstructions we consider.

The observed temperatures are a highly significant predictor of the chronologies found using both TS and RCS standardizations (both have p-value < 0.00001). The resulting temperature predictions exhibit differences we would expect given the properties of the procedures outlined in sections 3.1, 3.2, 3.3 and S2 (Figure 1). The predictions from inverse calibration are shrunk toward the distribution of observed temperatures relative to the predictions from classical calibration (Figure 1(a) vs 1(c) or Figure 1(b) vs 1(d)). The predictions from TS are centered around the mean of the observed temperature (Figure 1(a) and 1(c)) and do not show the lower temperatures from 1600–1800 evident in the predictions from RCS (Figure 1(b) and Figure 1(d)). This is evidence of the segment length curse. Due to the short segment lengths, TS effectively removes any low-frequency temperature signals with period over 200 years of length.

[Figure 1 about here.]

Almost all climate reconstructions use inverse calibration or related approaches. For example, the description of calibration in Cook and Kairiukstis (1990) only considers regressing climate ($\boldsymbol{x}$) on the tree-ring chronologies ($\boldsymbol{z}$). In the example above, reconstructions using classical calibration result in scientifically implausible predictions, with some below 0 °C (Figure 1). Judging the reconstructions and choice of procedure based on whether the predictions are plausible seems appealing. However, the statistical properties of the calibration procedures (section 3.3 and S2) provide important context to a discussion on choice of method. One such property is that classical calibration is preferred in terms of MSE over inverse calibration when the true value x_t^{mis} is far from the mean $\boldsymbol{x}^{obs}$ value. The implication is that classical calibration should be preferred when inferring unusual (in a historical context) climate values.

The conflict between statistical and non-statistical arguments is concerning. Many reconstructions are undertaken in the belief that climate conditions are likely to change with time, in spite of the assumed model. Thus, a likely explanation for the predictions from inverse calibration appearing more plausible that those from classical calibration is the insensitivity to misspecification of the model for $\boldsymbol{y}|\boldsymbol{x}$ arising from the strong model for $\boldsymbol{x}$ that shrinks the predictions toward the observed climate values. This shrinkage may suggest plausibility, but at the expense of masking the amplitude of temporal variation in climate. Likewise, scientific implausibility evident

in the predictions from classical calibration suggests that the underlying assumptions are inappropriate.

Instead of choosing between procedures, we prefer to consider a modeling framework that includes these two procedures as special cases. This provides flexibility to allow other assumptions that may be more appropriate for the observed data. Such a model should include the joint effects of age and climate to allow us to propagate uncertainty from the standardization through to the reconstruction. A benefit of this approach is that it allows us to overcome the segment length curse (more details are provided in S5).

We will focus on two related assumptions that we believe are critical to the standardization and reconstruction processes. The first is the choice of the model for climate ($\boldsymbol{x}$). The differences evident in Figure 1 suggest that the probability model specified for the growing season temperatures (as implied by the different analyses) may have a large effect on predictions. The second focus is on the description of how the ring widths depend on growing season temperature.

5 Model-based description

We initially present two models that are model-based analogues of the current TS and RCS standardization approaches. That is, we specify probability models that include assumptions that are consistent with those used in the multi-step TS and RCS procedures.

5.1 TS model-based analogue

We model individual tree-ring growth measurements with an effect due to age. The model is linear in age on the log scale (equivalent to a negative exponential model on the original scale as used in the previous section) to account for the assumption of multiplicative errors implied by the standardization processes. The model is

$$\log(y_{it}) \overset{\text{ind.}}{\sim} \mathcal{N}(\beta_{0i} + \beta_{1i} a_{it} + \eta_t, \sigma_y^2), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \tag{1}$$

$$\eta_t \overset{\text{ind.}}{\sim} \mathcal{N}(\beta_2 x_t, \sigma_\eta^2), \quad t = 1, \ldots, n, \tag{2}$$

$$x_t \overset{\text{iid}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2), \quad t = 1, \ldots, n, \tag{3}$$

where ind. denotes an independent sequence of random variables. We further assume that β_{0i} and β_{1i} are drawn from common distributions,

$$\beta_{hi} \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_{\beta_h}, \sigma_{\beta_h}^2), \quad h = 0, 1, \quad i = 1, \ldots, k. \tag{4}$$

An important inclusion is the term η_t, assumed to be common between trees. This term reflects that TS uses an average residual on the log scale at each time 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. That is, the ring widths are modeled simulatenously in terms of both age and climate. As described in S5, it is this step that links standardization and reconstruction and allows us to overcome the segment length curse. The final term we specify is the model for x_t. As with the implied model for inverse calibration, the first model we consider treats x_t as normally distributed with a constant mean through time. We refer to the model given by (1) - (4) as MB_TS_CON.

To aid in our understanding of the model (and how it compares to models we consider later) we can express MB_TS_CON as

$$\log(y_{it}) \overset{\text{ind.}}{\sim} \mathcal{N}(\beta'_{0i} + \beta_{1i}a_{it} + \beta_2 x'_t + \eta'_t, \sigma_y^2), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \tag{5}$$

where the parameters are

$$\begin{aligned}
\beta'_{0i} &\equiv \beta_{0i} + \beta_2\mu_x \text{ such that } \beta'_{0i} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{\beta_0} + \beta_2\mu_x, \sigma^2_{\beta_0}),\\
x'_t &\equiv x_t - \mu_x \quad \text{ such that } x'_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_x^2),\\
\eta'_t &\equiv \eta_t - \beta_2 x_t \quad \text{such that } \eta'_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2).
\end{aligned}$$

That is, an alternative way of considering MB_TS_CON is that the model for $\log(y_{it})$ has three features: (i) a linear aging term, (ii) a term that describes the effect of the climate relative to its mean, and (iii) a zero-mean yearly error term that is common between the trees. The year to year variability that is common between trees and is not accounted for by the linear aging term is then shared between x'_t and η'_t. The corresponding variances σ_x^2 and σ_η^2 are identifiable due to $x^{obs}_{m+1}, \ldots, x^{obs}_n$ being observed.

5.2 RCS model-based analogue

To include assumptions consistent with RCS we need only change the term (1). An assumption of RCS was that all trees of the same age have the same growth rate. We include this assumption by replacing (1) with

$$\log(y_{it}) \overset{\text{ind.}}{\sim} \mathcal{N}(\zeta_{a_{it}} + \eta_t, \sigma_y^2), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i,$$

where ζ_j is the growth rate for all trees aged j years. Each ζ_j is relatively uninformed by the data, as it is informed by at most k observations (and usually far fewer). One possible solution is to specify a hierarchical model for $\boldsymbol{\zeta}$ that reduces this variability and allows the values to change smoothly over time. Another possibility, that we implement here, is to bin the ages (we use 10 yearly bins) so that ζ_1 is the growth

rate for all trees aged 1 to 10 years, ζ_2 is the growth rate for all trees aged 11 to 20 years, etc,

$$\log(y_{it}) \overset{\text{ind.}}{\sim} \mathcal{N}(\zeta_{\lceil a_{it}/b \rceil} + \eta_t, \sigma_y^2), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \tag{6}$$

where $b = 10$ and $\lceil x \rceil$ is the smallest integer value greater than or equal to x. We refer to the model given by (6), (2) and (3) as MB_RCS_CON.

5.3 Model fitting

To fit MB_TS_CON and MB_RCS_CON to the Torneträsk data we use Bayesian inference and Markov chain Monte Carlo (MCMC) implemented in JAGS (Plummer 2003). Prior distributions for all parameters and details of the algorithm are specified in S6. This approach differs from that taken by Haslett et al. (2006) who did not use the full likelihood for inference. Instead they updated the parameters in the model for $\boldsymbol{y}|\boldsymbol{x}$ using only the observed data ($\boldsymbol{y}^{obs}$ and $\boldsymbol{x}^{obs}$) and then used these values to update $\boldsymbol{x}^{mis}$. Other approaches to inference are also possible for fitting the models we have described, e.g. maximum likelihood estimation could be implemented via the EM algorithm (Dempster et al. 1977).

The reconstructions using multi-step procedures (TS_INV and RCS_INV) and their corresponding model-based analogues (MB_TS_CON and MB_RCS_CON) lead to similar overall patterns of variability in the predictions (Figure 2). The main difference between the predictions appears to be the magnitude of warming/cooling. The actual time periods of relative warming and cooling are largely in common between all approaches.

The difference in reconstructions between assuming TS and RCS assumptions appears remarkably similar within a given modeling approach; i.e. compare Figure 2(b)-(a) to 2(d)-(c). The RCS assumptions lead to a reconstruction with a pronounced period of cool temperatures from 1600 through 1750, whereas the TS assumptions favor a period of cooling from 1600–1650 followed by a return to stable temperature from 1650–1750.

There are also differences that depend on the modeling approach. In general, the model-based approach has resulted in predictions where the coolest periods (e.g. 1600 – 1650) are colder than with the corresponding multi-step approach; i.e. 2(d)-(b) for RCS and 2(c)-(a) for TS.

[Figure 2 about here.]

Examination of various quantities in both the TS and RCS models can help us determine the adequacy of the underlying assumptions in the alternate model. If the ring growth parameters β_{0i} and β_{1i} in MB_TS_CON are similar for every tree in the sample then this supports the assumption made in RCS that growth is constant between individuals of the same age. A simple approach is to examine the posterior

densities for σ_{β_0} and σ_{β_1} to determine whether there is substantial mass near 0. For both parameters there is no evidence of mass near 0 (Figures 3(a) and 3(b)), suggesting that the constant growth assumption in RCS may not be appropriate. Likewise, if the growth parameters ζ_j in MB_RCS_CON show a close to linear trend with age this supports the assumption in TS about how ring width growth varies with age. While far from definitive, the ζ_j values suggest that assuming linear growth on the log scale may be a reasonable assumption (Figures 3(c) and 3(d)), especially for the earlier age classes. While the values for older age classes do depart from the general trend, the number of trees in each age class is relatively small. The age classes with the most trees are $61-70$ and $71-80$ with 226 trees. In contrast, age class $301-310$ has 29 trees, $401-410$ has seven trees, $501-510$ has three trees and all age classes $541-550$ and older have only one tree.

[Figure 3 about here.]

5.4 Extensions to the calibration model

The comparison above examined the predictions from the joint model-based approach and those from a multi-step procedure. A default choice was assumed for the model for temperature in all cases. The choice was made so that the model-based approaches and multi-step procedures incorporated the same assumptions and were comparable. We now investigate the effect this default choice has on the resulting reconstruction. For this we only consider the ring width model in (1) that uses TS and do not consider RCS any further.

The specific choice of model in (3) seems overly restrictive and it is doubtful that anyone interested in modeling long-term temperature believes that this is a reasonable hypothesis. One possible alternative is to treat $x_1, \ldots, x_n$ as if they were distinct parameters. Such an assumption is implied by classical calibration as described in S2. However, it seems reasonable to assume that there will be some relationship between the year to year values of growing season temperature (or many climate variables). This suggests any possible model should include structure in the pattern of changes in $\boldsymbol{x}$ over time. A model for $\boldsymbol{x}$ that includes some sort of smoothing thus seems desirable. One such example of this is given by Haslett et al. (2006) who looked at variance structures on random-walk models for the climate variable.

Our approach here is to model $\boldsymbol{x}$ hierarchically using a flexible family of distributions that admits a wide choice of scientific hypotheses about temporal changes in $\boldsymbol{x}$. We do this by modeling x_t as normally distributed, but allow the mean to vary according to a smooth function. We use a cubic B-spline. Now (3) is expressed as

$$\begin{aligned} x_t &\overset{\text{ind.}}{\sim} \mathcal{N}(\alpha_t, \sigma_x^2), \quad t = 1, \ldots, n, \\ \alpha_t &= \textstyle\sum_{h=1}^{H} \gamma_h B_h(x_t), \quad t = 1, \ldots, n, \end{aligned} \tag{7}$$

where α_t is the year-specific mean of the climate, and $B_h(x_t)$ is the hth B-spline basis function (see e.g. Hastie et al. 2009). We place a knot every 25 years and include a hierarchical model on γ_h to penalize the spline (Eilers and Marx 1996),

$$\gamma_h \overset{\text{iid}}{\sim} \mathcal{N}(\mu_\gamma, \sigma^2_\gamma), \quad h = 1, \ldots, H. \tag{8}$$

If $\sigma^2_\gamma = 0$ then the model for $\boldsymbol{x}$ in (7) collapses to the original model for $\boldsymbol{x}$ in (3) with a constant mean through time. We refer to the model given by (1), (2), (7), (8) and (4) as MB_TS_SPL.

As before, we can express MB_TS_SPL as

$$\log(y_{it}) \overset{\text{ind.}}{\sim} \mathcal{N}(\beta_{0i} + \beta_{1i} a_{it} + \beta_2 \alpha_t + \beta_2 x'_t + \eta'_t, \sigma^2_y), \quad i = 1, \ldots, k, \quad t = f_i, \ldots, l_i, \tag{9}$$

where the parameters are

$$x'_t \equiv x_t - \alpha_t \quad \text{such that } x'_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_x),$$
$$\eta'_t \equiv \eta_t - \beta_2 x_t \text{ such that } \eta'_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\eta).$$

The models (5) and (9) are similar: both include a term that describes the effect of the climate relative to its mean and a zero-mean yearly error term that is common between trees. The difference is the inclusion of the term $\beta_2 \alpha_t$ in (9). That is, we can think of MB_TS_SPL as extending the standardization model in MB_TS_CON to include additional variation that is (i) common between trees, (ii) smooth through time, and (iii) directly related to the mean of the climate process.

We fit MB_TS_SPL using MCMC, implemented in JAGS with details in S7. A feature of this model is the presence of two modes in the likelihood. We discuss the implications in S8.

Changing the model for $\boldsymbol{x}$ has a substantial impact on the reconstructed temperatures (Figures 4(a) and 4(b)). The amplitude of the predictions varies considerably between MB_TS_CON and MB_TS_SPL. There is a difference of more than 4 °C between the most extreme prediction made in the two models (in the period $1600 - 1650$). Given the sensitivity to the method of calibration shown in Figure 1, it is unsurprising that the amplitudes change upon a generalization of the model for temperature. Here we have relaxed the assumption of a constant mean for temperature which has increased the magnitude of the predictions.

The more flexible model for $\boldsymbol{x}$ has also led to a change in the overall mean of the predictions. In particular, Figure 4(b) suggests an increasing trend in temperatures throughout the entire series, apparent when comparing the two reconstructions in Figure 4(b)-(a). Such a result supports the notion that the simultaneous modeling of ring-width observations in terms of both climatic and non-climatic variables enables us to reconstruct low-frequency variability in the climate series. It is only once we relax the default model for $\boldsymbol{x}$ that such a trend is evident. The information about the trend appears insufficient to overwhelm a model that assumes a constant mean for $\boldsymbol{x}$ through time, i.e. Figure 4(a).

[Figure 4 about here.]

5.5 Incorporating additional information

The change in model for $\boldsymbol{x}$ has lead to predictions of temperature that are not scientifically reasonable. There is a prolonged period where the temperature series is too cold for the trees to survive based on ecophysiological research (Körner and Paulsen 2004, Körner 2008). A plausible lower bound for growing season temperature is around 4 °C or 5 °C.

There are many ways that additional information such as this can be incorporated into an analysis. We discuss two such possibilities. These analyses are intended to illustrate how such information can be included and the effect this has on the resulting predictions. We have chosen the models and priors for the problem at hand, but we recognize that (i) many other possibilities exist, and (ii) some may disagree with the choices/approaches we have taken. Even in these cases, we believe that exploring how the assumptions influence the predictions is instructive and aids in understanding of the data and model.

The first approach is to consider a model with informative or weakly-informative priors for all parameters. It is difficult to assign informative priors for the parameters that describe ring width. Factors such as historical site condition play an important role and cannot be assessed until the tree-rings have been inspected during the cross-dating procedure. Therefore, we place heavy-tailed weakly-informative independent prior distributions on the parameters μ_{β_0}, σ_{β_0}, μ_{β_1}, σ_{β_1}, σ_y, σ_η and β_2.

We specify informative priors on the parameters related to the model for $\boldsymbol{x}$. We place independent prior distributions on σ_x, μ_γ and σ_γ that lead to implied prior probabilities $\Pr(8\,°\text{C} < x_t < 12\,°\text{C}) = 0.81$, $\Pr(x > 6\,°\text{C}) = 0.97$ and $\Pr(x > 4\,°\text{C}) = 0.99$. Additional details are given in S9. We refer to the model given by (1), (2), (4), (7) and (8) with the priors specified in S9 as MB_TS_SPL_IP.

A second approach is to change the model for how ring widths are influenced by temperature, similar to that given in Tolwinski-Ward et al. (2011). We assume that there is a linear effect of temperature on log ring width growth between two temperature bounds that we refer to as $x_{\min}$ and $x_{\max}$. Below $x_{\min}$ the tree does not grow (as we are on the log-scale this corresponds to a value of $-\infty$). At $x_{\max}$ the growth rate is assumed optimal and all temperatures above $x_{\max}$ attain this optimal growth rate. We include this in the model by changing (2) to

$$\eta_t \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_t, \sigma_\eta^2), \quad t = 1, \ldots, n, \text{ where,} \tag{10}$$
$$\mu_t = \begin{cases} -\infty & \text{if } x_t < x_{\min} \\ \beta_2 x_t & \text{if } x_{\min} \leq x_t \leq x_{\max} \\ \beta_2 x_{\max} & \text{if } x_t > x_{\max} \end{cases}$$

and ensuring the prior density for β_2 has positive support. The values $x_{\min}$ and $x_{\max}$ are set by the user (we use 4 °C and 20 °C respectively). As this model is made up of

piecewise linear growth increments, we refer to the model given by (1), (4), (7), (8) and (10) as MB_TS_SPL_PL.

For the Torneträsk data, the observed ring widths $\boldsymbol{y}^{obs}$ are all strictly positive. Thus values $x_t < x_{\min}$ are inconsistent with the observed data for model MB_TS_SPL_PL. This is a strong constraint, but can be relaxed if desired (e.g. we could assume that when the temperature is below $x_{\min}$ the expected growth due to temperature is small, but non-zero).

We fit MB_TS_SPL_PL and MB_TS_SPL_IP using MCMC, implemented in JAGS. The MCMC details are practically equivalent to those given in S7 except for the changes to prior distributions.

The informative prior has little effect on the predictions (Figure 4(c)-(b)). Despite a prior probability of 0.01 that $x_t < 4\,^{\circ}\mathrm{C}$, many of the observations between 1600 – 1650 are below this value. The data have overwhelmed the prior. The prior densities we have specified have not been strong enough to ensure reasonable predictions from an ecophysiological perspective.

The model MB_TS_SPL_PL has had the expected effect on the prediction of temperature (Figure 4(d)) with no predictions below the value $x_{\min} = 4$. Comparing the predictions from MB_TS_SPL and MB_TS_SPL_PL reveals interesting differences. Not only does the amplitude change (to within the ecophysiological limits defined by the modified model) but the overall temperature trend changes. There is still an increasing trend through time (Figure 4(d)) but the trend is smaller than in model MB_TS_SPL (Figure 4(d)-(b)).

6 Model checking

To explore predictive performance of the models, we alternatively hold-out the first and second halves of the observed temperature data. This approach is commonly used in dendrochronology (e.g. Cook and Kairiukstis 1990) as the primary goal of such analyses is prediction outside of the observed sample. However, we note that any approach to assess the predictive performance is limited to the range of observed temperature values. Furthermore, since the hold-out blocks remove half of the observed temperatures used for model fitting, the resulting predictions will be correspondingly less certain.

There appears to be little difference between the various approaches in terms of predictive performance (Figure 5). All approaches are likely to overestimate when the true temperature value is below the observed mean and vice versa. There appears to be little difference between the models with the smooth mean for temperature and those with constant mean. This is somewhat surprising given the difference in predictions based on these models (the differences observed in Figure 4 also occur when modeling with the hold-out samples removed). Based on the values from the hold-out samples, all approaches (including those with scientifically implausible extremes) appear to be conservative at predicting values that are further from the observed mean.

RMSE values based on the hold-out are presented in Table 1 in the supplementary materials. We have not presented model checks for MB_TS_SPL_IP as its predictions were nearly identical to those from model MB_TS_SPL.

[Figure 5 about here.]

A key advantage of the model-based approach compared to the multi-step procedures is the ability to explore other measures of model fit. One such approach is looking at the fit of the conditional model $\boldsymbol{y}|\boldsymbol{x}$. We do this using averaged residuals. That is, we find the residual $\delta_{it} = y_{it} - \hat{y}_{it}$, where $\hat{y}_{it} = \bar{\beta}_{0i} + \bar{\beta}_{1i}a_{it} + \bar{\eta}_t$ for TS and $\hat{y}_{it} = \bar{\zeta}_{\lceil a_{it}/10 \rceil} + \bar{\eta}_t$ for RCS, with a bar denoting the posterior mean of the unknown quantity. We then find the average residual, $\tilde{\delta}_t$ by taking the mean of these residuals across all trees with observations in year t. For both models with constant mean for x_t, there is a pattern in the average residuals over time (Figures 6(a) and 6(b)). No such pattern is obvious in the models that include the smooth mean for climate (Figures 6(c) and 6(d)). However, these residual plots suggest that there may be non-constant variance through time. This appears to be an issue for both the period of observed temperatures, as well as the periods where temperature was not measured.

The RCS residuals appear more variable that the TS residuals. Few averaged residuals in the TS models exceed 0.02 units (ring width on the log-scale), whereas the RCS model has many averaged residuals beyond 0.02 units and some beyond 0.04. This is consistent with the posterior distribution for σ_y under the different models. The central 95% credible interval for σ_y is $(0.545, 0.552)$ in MB_RCS_CON whereas the interval is $(0.362, 0.366)$ in MB_TS_CON. We believe this is due to unmodeled variability; the RCS model assumes a common age-based growth rate.

[Figure 6 about here.]

Model MB_TS_SPL allows mean temperature to evolve smoothly over time, but results in scientifically implausible predictions. Despite the obvious shortcomings with the predictions from this model, we have presented the results and explored model checks. These checks suggest that the fit and predictions of the model (for the hold-out samples) are as good or better than the other models we have presented that do have scientifically plausible predictions. This demonstrates that the data alone are insufficient to discriminate between the various models we have explored. External information is needed to resolve the differences. What is less clear is how that should be achieved. We considered two approaches to include that information in the modeling process. The first considered soft constraints given by informative priors. The data overwhelmed the prior and the predictions remained implausible. In the second approach, we included a hard constraint associated with a change of response function (describing how tree-rings respond to temperature). The predictions changed considerably with the hard constraint ensuring plausible values (at least according to a minimum temperature). Other changes to the response function that we have

investigated (and not presented) suggest that the model for describing how climate (in our case temperature) relates to ring widths has a large impact on the resulting predictions while often leaving the fit of the observed data unchanged.

We believe that the statements of uncertainty we present for the temperature predictions understate the true uncertainty as all of our predictions condition on an assumed model. For these data we appear unable to distinguish between various assumptions, even when there are substantial differences in predictions. As there is no model clearly favored by the data and many modeling assumptions that could be considered (some additional assumptions are mentioned in section 7), the results suggest: (i) modeling decisions matter, and (ii) the uncertainty we should have in the predictions is greater than that expressed by any one model.

One way to account for model uncertainty is to model-average uncertainty intervals for the unknown temperature values (e.g. Hoeting et al. 1999). We have not done this for two reasons. First, it is difficult to calculate the marginal likelihoods required to find the Bayes factors or, equivalently, specify a trans-dimensional MCMC algorithm, such as reversible jump algorithm (Green 1995) to explore joint model and parameter space. Second, Bayes factors are known to be sensitive to priors placed on parameters even if the data overwhelm the priors within specific models. For these reasons we have instead examined how the model fit and predictions of missing temperature observations change with the model assumed.

7 Discussion

Inferring historical climate from ring-width 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 built a model-based framework for including these sources. Care is needed when specifying alternative assumptions to ensure that the overall model continues to be identifiable. From the TS model statement in (9) we can see that including a flexible aging function such as a polynomial regression or spline would lead to identifiability problems if the climate variable also has a flexible mean function. However, for the RCS representation the constraint that all trees of the same age have the same growth response makes it is possible to have a flexible aging function and allow the climate variable to have a flexible mean.

We have explored changes to many of the model assumptions, including the description of ring-width growth (TS vs RCS), how the ring widths depend on climate (including using or not using additional information), and the model for climate. We have shown that changes to these assumptions can lead to important differences in resulting predictions. Moreover, the validity of many of these assumptions cannot be determined based solely on the data. Reconstructions found using chronologies formed by multi-step procedures and obtained using default calibration approaches may hide such sensitivity to underlying modeling assumptions.

We have restricted ourselves to one transformation on $\boldsymbol{y}$ that was chosen to provide

a link between the traditional multi-step approaches and the model-based approach. While breaking the link to the traditional multi-step approaches, other transformations of $\boldsymbol{y}$ could be used in place of the log transform. For example, similar results and conclusions are obtained for this data when using a square root transform instead of the log transform. Other transformations, such as a Box-Cox transformation, could also be used.

The crossdating process (described in S1) can be thought of as obtaining an estimate of the overall chronology. There may be uncertainty in the estimates of ring-width dates (Wigley et al. 1987). Any such unmodeled uncertainty will not propagate through to the reconstruction and will result in predictions of historical climate that appear more precise than they should be. Potential remedies include (i) specifying the crossdating process within a probabilistic statistical model, or (ii) having the dendrochronologist estimate uncertainty in the cross-dating process, i.e. giving various plausible dates for the series with corresponding estimates of certainty in that being the correct date (that sum to one). Inclusion of the crossdating process would be a difficult task. Setting up such a model (option (i) above) would require research effort, because of the challenge of including all the information that is currently used in the human judgment involved in the visual aspect of crossdating through a microscope. Moreover, there would be considerable computational difficulties involved in including any dating uncertainty.

A benefit of including the crossdating process within the statistical model is that it would avoid using the data twice. Many crossdating procedures use the data, and the assumption that trees respond in a related way, to estimate the correct alignment of the cores. Once these estimates are obtained, the data are then used again to predict historical climate, assuming that the relatedness in the tree response depends on the climatic variable of interest.

Another assumption we have not varied is that the growth increments have a common response to temperature. Despite the trees existing in the same macro-environment, there may be fine scale local differences in the temperature, leading to a tree-specific response. Extending the models in this way is ongoing research and allows the inclusion of more realistic assumptions at the expense of several practical difficulties. Model fitting becomes difficult, with care required to ensure the model remains identifiable. Our experience is that there is considerable sensitivity to the specification of the error structure in the model. Furthermore, some extensions necessitated the use of Hamiltonian Monte Carlo algorithms that are difficult to implement; Gibbs/Metropolis approaches appear impractically slow.

The idea of a common modeling framework can be extended to multiple proxies (Evans et al. 2013). Unfortunately, other proxy variables also include complexities that must be accounted for in the modeling. 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). In particular, the multiple proxies may provide information about the calibration curve

across a range of climate values. However, including observations from multiple proxies will not automatically lead to robust predictions. An advantage of including data from multiple proxies is that it allows for more expansive model checking. That is, we may be able to use information from different proxies to check the appropriateness of different aspects of our assumed model.

There are many hierarchical extensions that we could consider. These include data from different geographic locations with appropriate spatial models, such as the those considered by Tingley and Huybers (2010a). 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: developing scalable methodology that allows us to model the raw data across numerous locations without resorting to multi-step procedures. The second challenge is one of complexity and applicability: developing appropriate model checks for the assumptions that, as shown here, are critical to historical prediction of climate.

7.1 Message for the paleoclimate community

We have demonstrated model-based approaches for tree-ring based reconstructions that are able to incorporate the assumptions of traditional approaches as special cases. The modeling framework allows us to relax assumptions long used out of necessity, giving flexibility to our model choices. Using the Scots pine data from Torneträsk we show how modeling choices matter. Alternative models fitting the data equally well can lead to substantially different predictions. These results do not necessarily mean that existing reconstructions are incorrect. If the assumptions underlying the reconstruction is a close approximation of reality, the resulting prediction and associated uncertainty will likely be appropriate (up to the problems associated with the two-step procedures used). However, if we are unsure whether the assumptions are correct and there are other assumptions equally plausible *a-priori*, we will have unrecognized uncertainty in the predictions. We believe that such uncertainty should be acknowledged when using standardized data and default models.

As an example consider the predictions from model MB_TS_CON for Abisko, Sweden. If we believe the assumptions underlying model MB_TS_CON then there is a 95% probability that summer mean temperature in 1599 was between 8.1 °C and 12.0 °C as suggested by the central credible interval (Figure 4(a)). However, if we adopt the assumptions underlying model MB_TS_SPL_PL we would believe that the summer mean temperature in 1599 may have been much colder than 8.1 °C with a 95% credible interval between 4.1 °C and 7.8 °C. In practice, unless the data are able to discriminate between these assumptions (which they were not able to do here as shown in Section 6), there is more uncertainty about the summer mean temperature in 1599 than that found in any one model considered. We believe that such model uncertainty needs to be recognized by the community as an important source of uncertainty associated

with predictions of historical climate. The use of default methods makes evaluation of such uncertainty difficult.

Acknowledgements

This research was supported by NSF #0934516. We thank three annonymous referees and the Associate Editor for valuable comments that improved this manuscript. Lamont-Doherty Contribution Number 0000.

References

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. (2011), "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, pp. 113 – 145.

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.

Christiansen, B. (2011), "Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided?" *Journal of Climate*, 24, 674–692.

— (2014), "Straight line fitting and predictions: on a marginal likelihood approach to linear regression and errors-in-variables models," *Journal of Climate*, 27, 2014–2031.

Christiansen, B. and Ljungqvist, F. C. (2011), "Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability," *Journal of Climate*, 24, 6013–6034.

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.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, 39, 1–38.

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

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.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J. (2013), "Applications of proxy system modeling in high resolution paleoclimatology," *Quaternary Science Reviews*, 76, 16–28.

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

Fritts, H. C. and Swetnam, T. W. (1989), "Dendroecology: a tool for evaluating," *Advances in Ecological Research*, 19, 111.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711 – 732.

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.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian model averaging: a tutorial," *Statistical Science*, 14, 382–417.

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., Briffa, K. R., Osborn, T. J., Lough, J. M., Van Ommen, T. D., Vinther, B. M., Luterbacher, J. W. E. R. Z. F. W., Wahl, E. R., Zwiers, F. W., Mann, M. E., Schmidt, G. A., Ammann, C. M., Buckley, B. M., Cobb, K. M., Esper, J., Goosse, H., Graham, N., Jansen, E., Kiefer, T., Kull, C., Küttel, M., Mosley-Thompson, E., Overpeck, J. T., Riedwyl, N., Schulz, M., Tudhope, A. W., Villalba, R., Wanner, H., Wolff, E., and Xoplaki, E. (2009), "High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects," *The Holocene*, 19, 3–49.

Jones, P. D. and Mann, M. E. (2004), "Climate over past millennia," *Reviews of Geophysics*, 42, RG2002.

Körner, C. (2008), "Winter crop growth at low temperature may hold the answer for alpine treeline formation," *Plant Ecology & Diversity*, 1, 3–11.

Körner, C. and Paulsen, J. (2004), "A world-wide study of high altitude treeline temperatures," *Journal of Biogeography*, 31, 713–732.

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.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez-Rouco, J. F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A. (2013), "Information from Paleoclimate Archives," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

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.

Melvin, T. M., Grudd, H., and Briffa, K. R. (2013), "Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data," *The Holocene*, 23, 364 – 373.

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

PAGES 2k Consortium (2013), "Continental-scale temperature variability during the past two millennia," *Nature Geoscience*, 6, 339–346.

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*.

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.

Tingley, M. and Huybers, P. (2010a), "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.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B. (2012), "Piecing together the past: Statistical insights into paleoclimatic reconstructions," *Quaternary Science Reviews*, 35, 1–22.

Tingley, M. P. and Huybers, P. (2010b), "A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: Comparison with the regularized expectation-maximization algorithm," *Journal of Climate*, 23, 2782–2800.

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.

Wahl, E. R. and Smerdon, J. E. (2012), "Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors," *Geophysical Research Letters*, 39, L06703.

Wigley, T. M. L., Jones, P. D., and Briffa, K. R. (1987), "Cross-dating methods in dendrochronology," *Journal of Archaeological Science*, 14, 51–64.

Tables

TS_INV	Multi-stage approach using TS and inverse calibration
TS_CLASS	Multi-stage approach using TS and classical calibration
RCS_INV	Multi-stage approach using RCS and inverse calibration
RCS_CLASS	Multi-stage approach using RCS and classical calibration
MB_TS_CON	Model-based approach using (1) – (4)
MB_TS_SPL	Model-based approach using (1), (2), (4), (7) and (8)
MB_TS_SPL_PL	Model-based approach using (1), (4), (7), (8) and (10)
MB_TS_SPL_IP	Model MB_TS_SPL with informative priors described in S9.
MB_RCS_CON	Model-based approach using (2), (3) and (6)

Table 1: A description of all reconstructions considered.

Figures

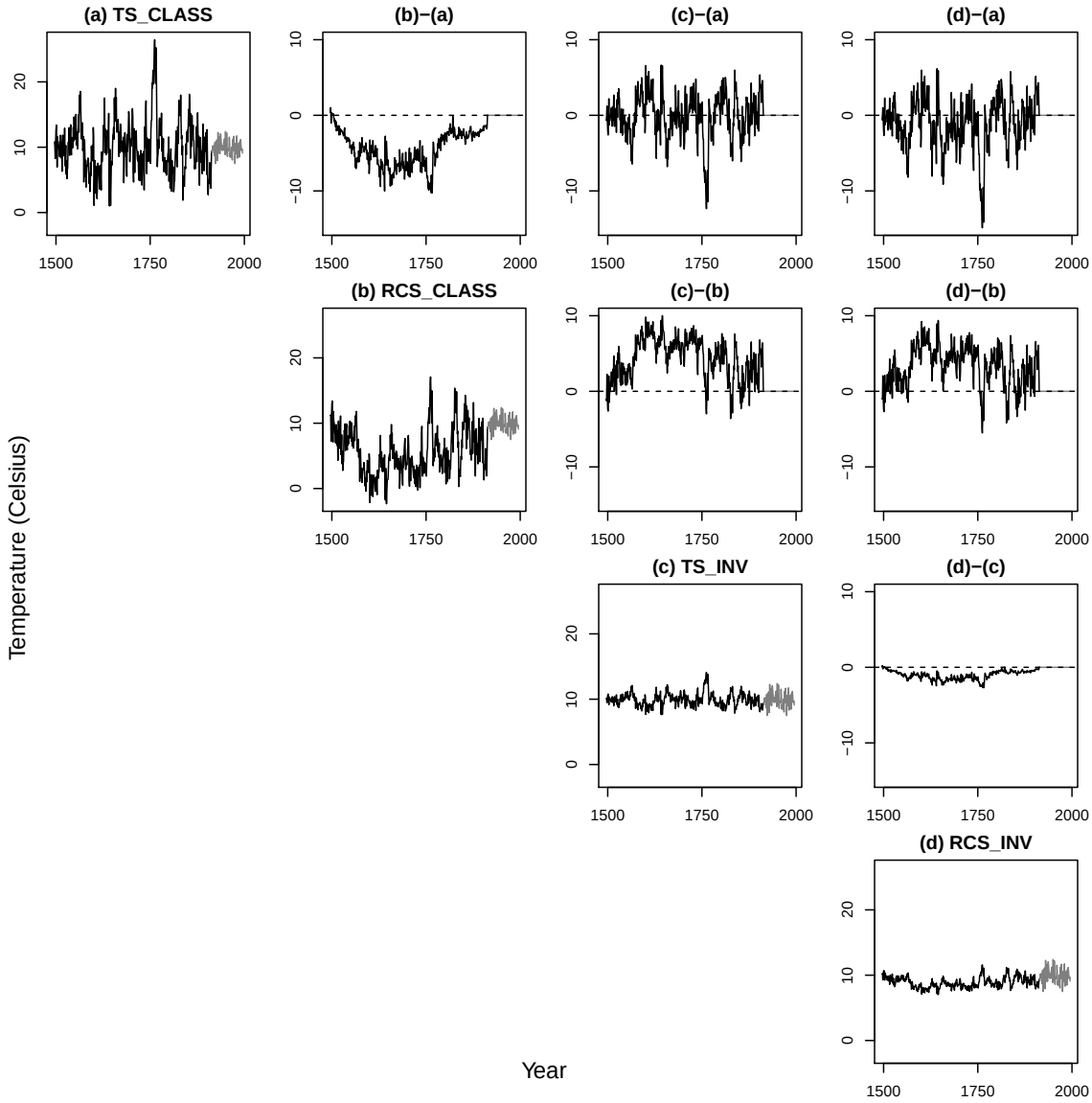

Figure 1: Predictions for different (i) standardization approaches and (ii) calibration procedures. In the diagonal plots, the black lines are the predicted reconstruction and the gray lines are the observed data. The off diagonal plots give the difference between the predictions in the corresponding diagonal plots.

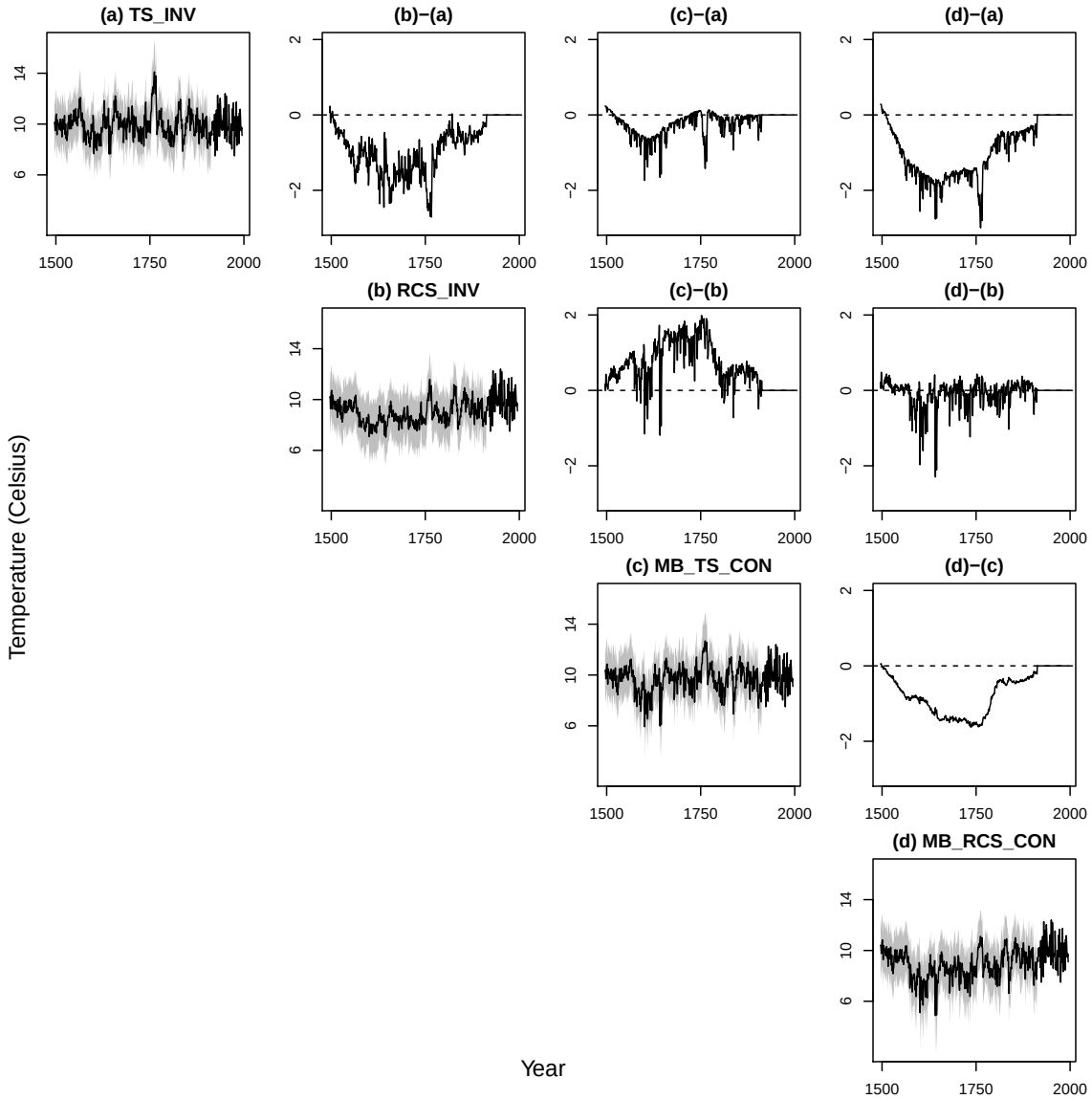

Figure 2: Predictions from traditional approaches and the model-based analogues. In the diagonal plots, the black lines are either the predicted $\hat{x}_t^{mis}$ or the median of the posterior distribution for x_t^{mis} and the gray areas are 95% prediction/credible intervals. The off diagonal plots give the difference between the predictions in the corresponding diagonal plots.

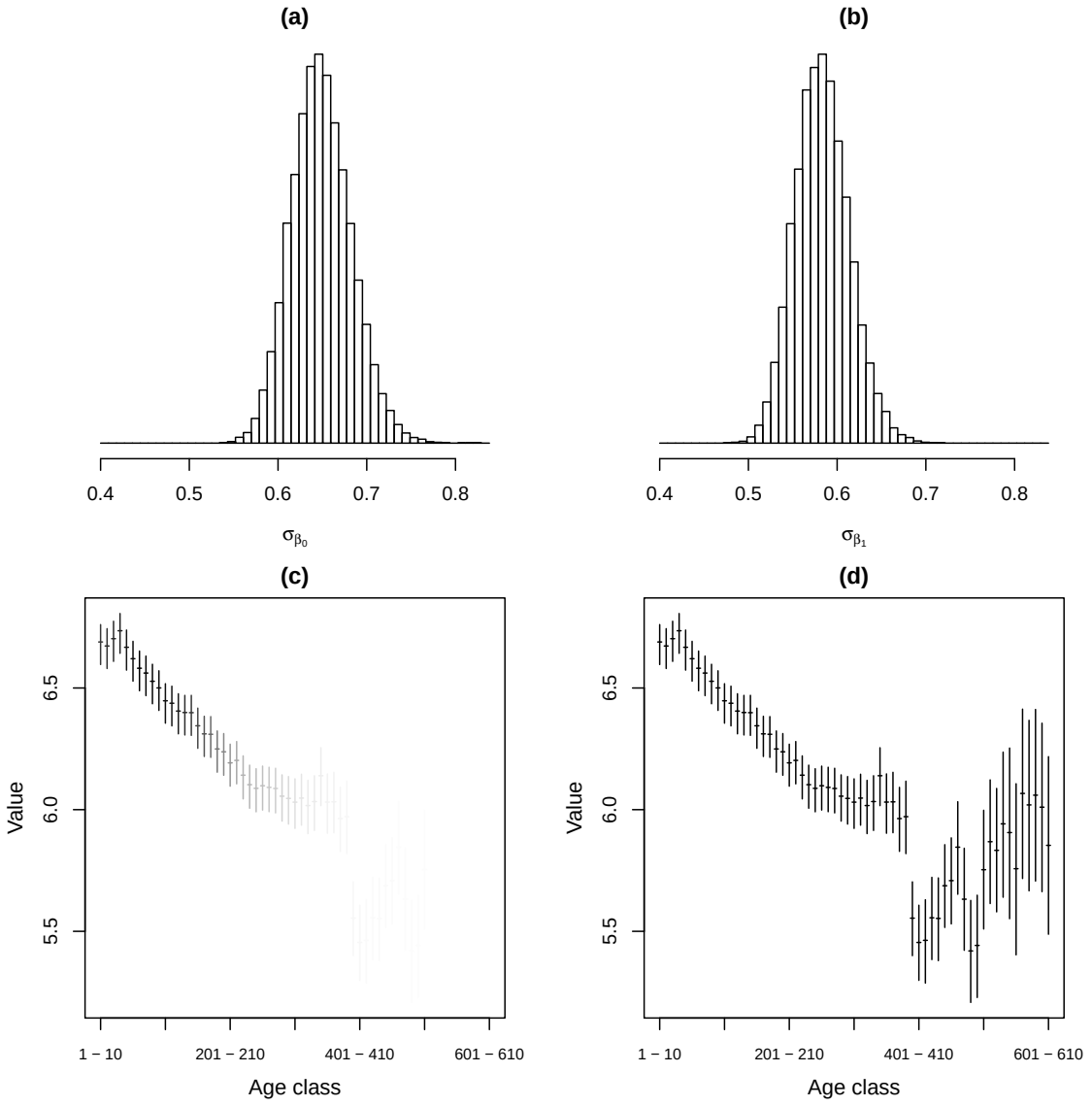

Figure 3: Various summaries from model-based TS and RCS implementations. Estimates of the posterior density of σ_{β_0} (a) and σ_{β_1} (b) in MB_TS_CON. 95% central credible intervals for ζ_j from MB_RCS_CON are shown in (c) and (d), with the horizontal line denoting the posterior median. In (c) the lines are shaded relative to the number of trees with observations in that age class.

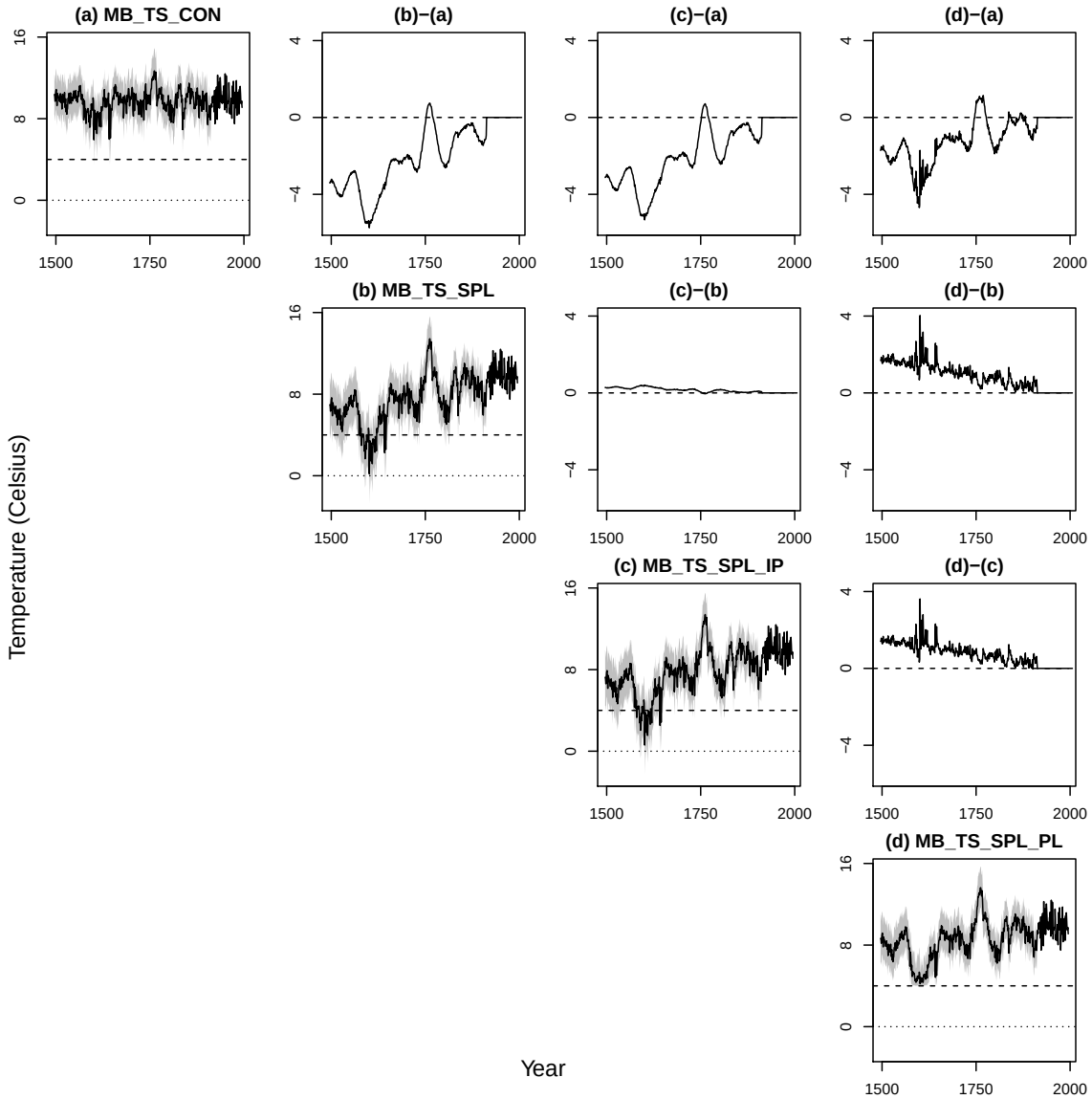

Figure 4: Predictions when making different assumptions in the model-based approaches. In the diagonal plots, the black lines are the median of the posterior distribution for x_t^{mis} and the gray areas are 95% credible intervals. The off diagonal plots give the difference between the predictions in the corresponding diagonal plots. In all plots the horizontal dashed line is at 4 °C and the horizontal dotted line is at 0 °C.

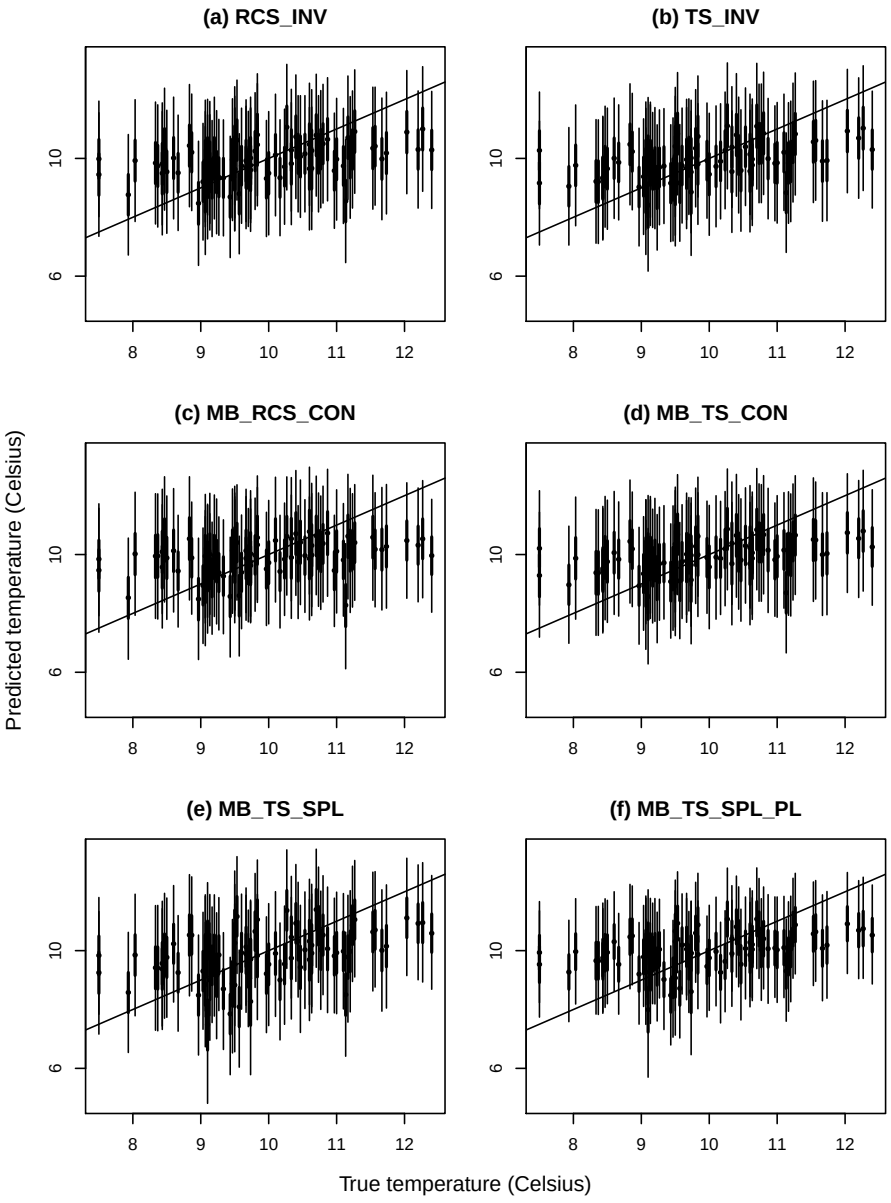

Figure 5: Predictions (and corresponding uncertainty intervals) plotted against the true values for the held-out observations. In plots (a) and (b), the thin lines give a 95% prediction interval, the thick lines are the 50% prediction interval and the point is the prediction for the held-out values. For (c) – (f) the thin lines are 95% posterior predictive credible intervals, the thicker lines are 50% credible intervals and the point is the median of the posterior predictive distribution for the held-out values.

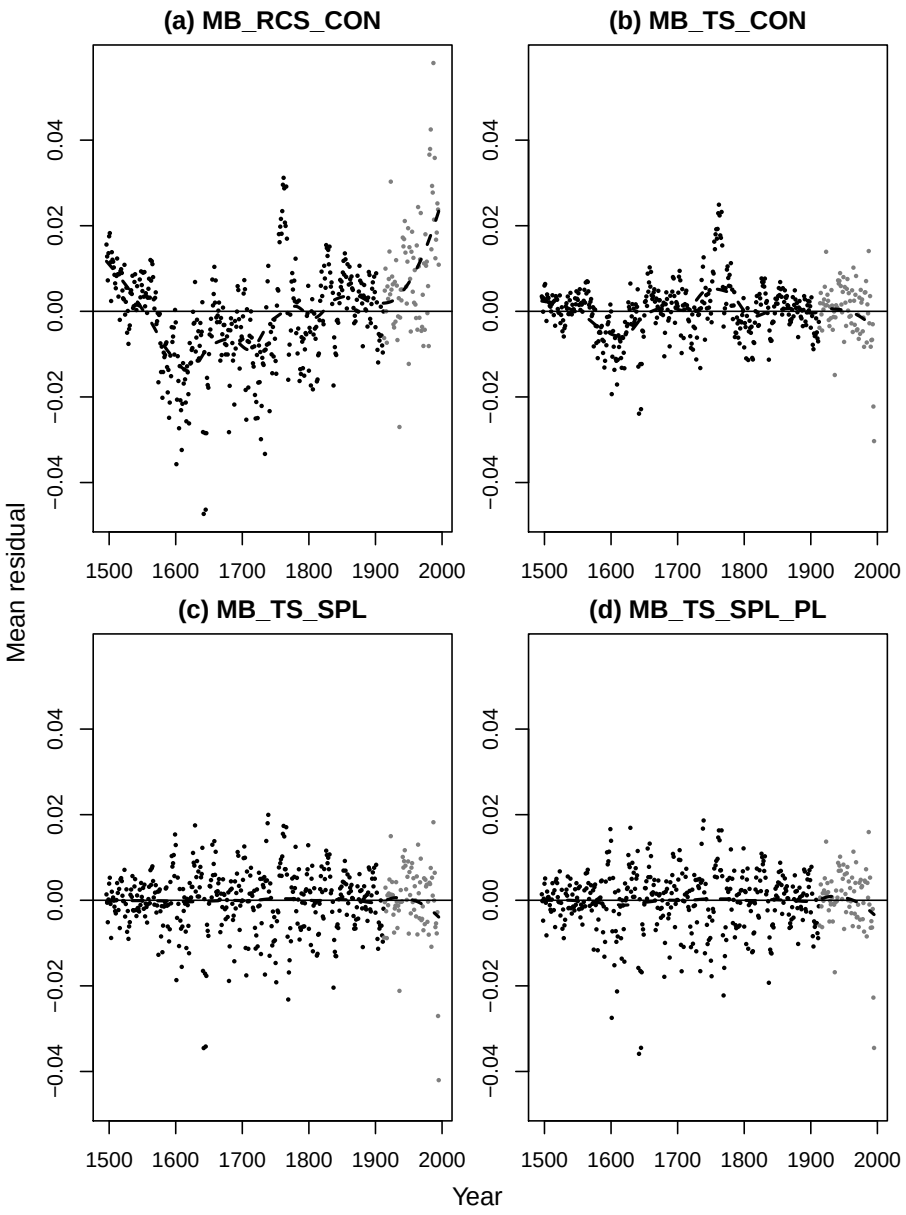

Figure 6: The average residual plotted against year. The average residual in year t is found by averaging the residuals (found at the posterior mean of the parameters) across all trees observed in year t. The gray dots refer to years in which the temperature values were observed and black dots to years in which the climate values were not observed. The dashed line is a loess fit.

A Model-Based Approach to Climate Reconstruction Using Tree-Ring Data: Supplementary Materials

Matthew R. Schofield[12]*, Richard J. Barker[2], Andrew Gelman[3],
Edward R. Cook[4] and Keith Briffa[5]

[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.

[5]Climatic Research Unit, School of Environmental Sciences, University of East Anglia,
Norwich NR4 7TJ, UK.

1 Data

1.1 Crossdating

Once the ring widths of each sampled tree are measured to a high precision, typically ± 0.01 mm or better, the ring-width measurements are evaluated for strength of crossdating and quality of measurement using a quality control program such as COFECHA (Holmes 1983). The most common dating errors are caused by (i) missing rings (i.e., the incomplete formation of radial growth around the circumference of the tree) at the point of sampling, and (ii) false rings (i.e., the occurrence of intra-annual growth bands that might be wrongly identified as true rings). See Fritts (1976) and Cook and Kairiukstis (1990) for details. Missing

*E-mail: mschofield@maths.otago.ac.nz

rings are rare for trees growing in northern Sweden (St. George et al. 2013) and this is true for the data we use here. When crossdating is done with proper quality control to a sufficiently replicated data set, as is the case here, it is possible to have a high degree of confidence in the resulting chronology. That is, if the crossdating was performed independently by two experienced dendrochronologists, it is highly likely that they would obtain the same chronology. Similar confidence is also associated with the dating of the sub-fossil wood samples.

1.2 Torneträsk data

The full Common Era data comprise ring-width measurements collected from 587 trees. The mean segment length is 189 years, with a maximum of 609 years. Combining all cores together gives a continuous record from 38 B.C. to 1995.

As described in the manuscript, we analyze a subset of the full data comprising of $k = 247$ series with at least 20 measurements after the year 1496, leading to a 500 year series. The raw data for 9 of the trees is in Figure 1. The mean number of measurements per tree is 179 years, with a maximum of 485 years (Figure 2). The number of cores per year varies throughout the period, with the most data in recent times (Figure 3). This is due to the transition from remnant wood to living trees.

The original owner of the Torneträsk data is Hakan Grudd (Grudd et al. 2002). The data used here are available from `www.maths.otago.ac.nz/home/resources/matthew_schofield/`.

2 Statistical calibration

The two main approaches used to predict $\boldsymbol{x}^{mis}$ using $\boldsymbol{x}^{obs}$ and $\boldsymbol{z}$ are:

1. **Classical calibration.** The natural specification of the model according to the underlying science describes the tree-ring growth chronology in terms of the climate variable:

$$z_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad t = 1, \ldots, n. \tag{1}$$

The missing climate values are then predicted using

$$\hat{x}_t = \frac{z_t - \hat{\beta}_0}{\hat{\beta}_1}, \quad t = 1, \ldots, m, \tag{2}$$

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. Alternatively, the calibration model can be specified as

$$x_t = \gamma_0 + \gamma_1 z_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \quad t = 1, \ldots, n,$$

with the missing climate values predicted using

$$\tilde{x}_t = \hat{\gamma}_0 + \hat{\gamma}_1 z_t, \quad t = 1, \ldots, m, \tag{3}$$

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

The predictions from classical calibration (1) correspond to the maximum likelihood estimator. However, if $\hat{\beta}_1$ is not statistically significant, then the confidence interval for x_t can be the whole real line, or two disjoint semi-finite intervals (Osborne 1991). The mean square error of $\hat{x}_t$ is infinite (Williams 1969), but conditioning on $|\beta_1| > 0$ in the expectation leads to a consistent estimator with finite mean square error (Berkson 1969). The relative MSE of $\hat{x}_t$ vs $\tilde{x}_t$ depend on how far the true value of x_t is from $\bar{x} = \text{mean}(\boldsymbol{x}^{obs})$. When x_t and $\bar{x}$ are close, predictions from inverse calibration can have a lower mean squared error, with the predictions from classical calibration preferred otherwise (Krutchkoff 1967, 1969, Berkson 1969). The predictions from inverse calibration can also be derived as a Bayes estimator with a specific informative prior distribution (Hoadley 1970).

Here we explore the two approaches from a Bayesian perspective. We start by modeling $z_t|x_t$ as exchangeable observations with probability model

$$p(z_t|x_t) = \mathcal{N}(\beta_0 + \beta_1 x_t, \sigma_{Z|X}^2), \quad t = 1, \ldots, n.$$

We also assume that the marginal distribution for x_t is normal:

$$p(x_t) = \mathcal{N}(\mu_X, \sigma_X^2), \quad t = 1, \ldots, n.$$

Then Bayes theorem yields

$$p(x_t|z_t) = \mathcal{N}(\gamma_0 + \gamma_1 z_t, \sigma_{X|Z}^2), \tag{4}$$

where

$$\gamma_0 = \mu_X(1 - \gamma_1\beta_1) - \gamma_1\beta_0$$
$$\gamma_1 = \frac{\beta_1\sigma_X^2}{\sigma_{Z|X}^2 + \beta_1^2\sigma_X^2}$$
$$\sigma_{X|Z}^2 = \frac{\sigma_X^2\sigma_{Z|X}^2}{\sigma_{Z|X}^2 + \beta_1^2\sigma_X^2}.$$

The model for inverse calibration, assuming normal errors, can be found by starting with the model for classical calibration and assuming a common normal distribution for all elements of $\boldsymbol{x}$ with fixed mean and variance μ_X and σ_X^2. This explains why the performance of the inverse estimator improves as the true value of x_t approaches $\bar{x}$. First consider the case $m = 1$. If the true value of x_1 is $\bar{x}$, then x_t and $\boldsymbol{x}^{obs}$ are in close agreement with respect to the marginal distribution of $\boldsymbol{x}$. As the true value of x_1 moves away from $\bar{x}$, there is increasing discordance between x_1 and $\boldsymbol{x}^{obs}$ with respect to the marginal distribution. When $m > 1$, the performance of the inverse estimator depends on how far x_t is from μ_X, a parameter that should be close to $\bar{x}$ under the modeling assumptions. In contrast, the Bayesian equivalent of classical calibration is obtained assuming a vague [1] prior distribution for the unknowns, including x_t. It is therefore no surprise that classical calibration is preferred when x_t is far from $\bar{x}$.

2.1 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, different proxy types, or a combination of the two (for example see Mann et al. 1998).

To give a multivariate extension of inverse calibration, we (i) model $\boldsymbol{z}_t$ conditional on $\boldsymbol{x}_t$ and (ii) express the marginal distribution of $\boldsymbol{x}_t$ as normally distributed:

$$\begin{aligned}
\boldsymbol{z}_t &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, n, \\
\boldsymbol{\epsilon}_t &\overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{Z|X}), \quad t = 1, \ldots, n, \\
\boldsymbol{x}_t &\overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_X), \quad t = 1, \ldots, n,
\end{aligned} \tag{5}$$

where $\boldsymbol{\beta}_0$ is a vector of length L_z, $\boldsymbol{\beta}_1$ is a $L_z \times L_x$ matrix, $\boldsymbol{\epsilon}_t$ is a vector of length L_z, $\Sigma_{Z|X}$ is a $L_z \times L_z$

[1] "Vague" in this context does not necessarily mean constant. See Ghosh et al. (1995) and Yin (2000) for details.

matrix, $\boldsymbol{\mu}_X$ is a vector of length L_x and Σ_X is a $L_x \times L_x$ matrix.

As described by Schneider (2001), the regularized expectation-maximization (RegEM) algorithm is based on a model of $\boldsymbol{x}_t$ (the L_x length vector of climate values in year t) and $\boldsymbol{z}_t$ (the L_z length vector of proxies in year t) as jointly normally distributed:

$$\begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{x}_t \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma), t = 1, \ldots, n, \tag{6}$$

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 as there are missing values in $\boldsymbol{x}$ and possibly $\boldsymbol{z}$. Often $L_z > (n - m)$, making the dimensionality of $\boldsymbol{z}$ too large to allow for identifiable estimation of Σ, hence the need for a regularization step.

As shown in Gelman et al. (2004) pp. 317–324, the parameters in (5) are deterministic functions of the parameters in (6). That is, we can think of the model underlying the RegEM algorithm as being an extension of inverse calibration, where the marginal model for the climate variables is a normal distribution with constant mean and covariance through time.

3 Segment length curse

Here we show how traditional standardization (TS) can lead to the unintentional removal of climatic influences on growth from the tree-ring chronology, referred to as the segment length curse (Cook et al. 1995). For ease of description and presentation we simulate and standardize ring width series on the log-scale. This means that instead of finding $I_{it} = y_{it}/\hat{y}_{it}$ we fit the growth model to the log of the ring widths and find the residuals. We note that we simulate only one tree per site for pedagogical reasons and similar results hold in the case where many trees per site were sampled.

We consider a simulated temperature series $x_{1t},\ t = 1, \ldots, n$. This is a stationary series with zero mean (it is an AR 1 process with correlation 0.95). From this series we construct a second temperature series x_{2t}, that differs only in terms of a linear trend:

$$x_{2t} = x_{1t} - 4 + \frac{4t}{n}, \quad t = 1, \ldots, n.$$

One such realization for $n = 500$ that is used in the remainder of this section is shown in Figure 4.

A commonly used growth function that we use throughout the manuscript and supplement is a linear effect on the log scale (negative exponential on the original scale). For each temperature series we simulate a ring width series,

$$\log(y_{jt}) = 2 - 0.4a_t + x_{jt} + \epsilon_t, \quad j = 1, 2, \quad t = 1, \ldots, n$$

where a_t is the age of the tree (centered and scaled so that $a_1 = -2$ and $a_n = 2$) and $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.5^2)$. Note that we use common parameters and error terms, so that the only difference between the two ring width series is the underlying temperature series. Realized simulations of $\log(y_{1t})$ and $\log(y_{2t})$ series are shown in Figure 4.

We have only one ring series (for each temperature series), and the assumed growth model is linear on the log scale. This means the log of the standardized chronologies, $\log(z_{jt})$, are found by regressing $\log(y_{jt})$ on a_t and finding the residuals. The resulting chronologies $\log(z_{1t})$ and $\log(z_{2t})$ are identical (Figure 5). Standardization has led to the trend in $\log(y_{2t})$ being removed as an age effect, despite truly being the result of a trend in temperature.

4 Details of standardizations implemented

To find the chronology using TS we use the process described in manuscript section 3.1 using a negative exponential growth function (not adjusting for autocorrelation). To find the chronology using regional curve standardization (RCS), we assume that the growth rate is the same for every tree in a given decade of its life (with one growth rate from age 1–10, another from 11–20, and so forth). We do not use pith offset estimates, making the simplifying assumption that each sample series begins with the innermost ring at that sampling height. Furthermore, this simple form of RCS that assumes all tree measurement series can be standardized with the same expectation of growth at each specific ring age (i) may not be optimal where some trees exhibit notably different growth rates under the same climate conditions, and (ii) cannot account for differences in sample make up through time. Where relatively large trees are preferentially selected for modern sampling, the recent section of a chronology may exhibit a positive bias ("modern sample bias", Briffa and Melvin 2011). This issue is addressed for these trees in Melvin et al. (2013), though not in the approach adopted here.

5 Modeling ring widths in terms of age and temperature

Here we explore the data simulated in section 3 for the purpose of demonstrating how modeling ring widths in terms of both age and temperature together (we call this a simultaneous model) can overcome the segment length curse. We compare reconstructions using the chronologies evaluated in section 3 to those found from a simultaneous model.

We purposefully simulated the data in section 3 so that the correlation between the ring widths and temperature was so strong that both classical and inverse calibration give nearly identical predictions[2]. Here we use the classical estimator.

Recall that there were two temperature series, x_{1t} and x_{2t} that differ according to a linear trend. We treat the temperature values associated with $t = 1, \ldots, m = 400$ as unknown for both series. The goal is to then use the observed temperature values from $t = 401, \ldots, n = 500$, along with the ring widths, to predict the temperature values from $t = 1, \ldots, m$.

The reconstructions using the chronologies from TS found in 3 are shown in the first column of Figure 6. This approach has done well at reconstructing x_{1t}, but as expected has failed to reconstruct the trend in temperature in x_{2t}.

A simultaneous model uses the original ring width series y_{1t} and y_{2t}. For the period for which we have observed temperature data ($t = m + 1, \ldots, n$) we consider a model that describes the ring widths in terms of both their age effect and a temperature effect:

$$\log(y_{jt}) = \beta_{j0} + \beta_{j1} a_t + \beta_{j2} x_{jt} + \epsilon_{jt}, \quad j = 1, 2, \quad t = m + 1, \ldots, n,$$

where $\epsilon_{1t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_1^2)$ and $\epsilon_{2t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_2^2)$. We then fit this model using least squares and use estimated parameters from this regression, along with the known age covariates to predict the missing climate values. We use an estimator that allows for the known age terms,

$$\hat{x}_{jt} = \frac{\log(y_{jt}) - \hat{\beta}_{j0} - \hat{\beta}_{j1} a_t}{\hat{\beta}_{j2}}, \quad j = 1, 2, \quad t = 1, \ldots, m.$$

The resulting reconstructions are shown in the second column of Figure 6. The simultaneous model has overcome the segment length curse, with the trend in temperature evident. The two reconstruction approaches are directly compared in Figure 7.

[2]This also allowed for only one ring width series to be considered, making the example easier to follow.

The simultaneous model we use here is very simple (it is used for pedagogical reasons). Criticisms include (i) conditioning on parameters estimates to reconstruct the temperature, (ii) not using the information about tree age for $t = 1, \ldots, m$ to improve the fit of the model and subsequent accuracy of the predictions of past temperature, (iii) it being defined for only one ring width series with limited error structure, and (iv) it being defined for classical calibration only. The model described in section 5 of the manuscript is a joint model for ring widths and temperature that has been developed to overcome these criticisms. It uses the full likelihood, so does not condition on parameter estimates and uses all information (including age information for $t = 1, \ldots, m$) for parameter estimation. It allows for multiple tree-ring series and includes a hierarchical model for the climate variable (in this case temperature) that can be used to express different calibration assumptions.

6 Modeling details (MB_TS_CON and MB_RCS_CON)

Prior to fitting, we standardized the climate variables x_t (by subtracting the sample mean and dividing by twice the sample standard deviation) to help specification of the prior distribution. For all models listed below JAGS (Plummer 2003) was used to generate a posterior sample. We use three parallel chains and check convergence with the Brooks-Gelman-Rubin diagnostic $\hat{R}$ (Brooks and Gelman 1998). We combine the chains and compute the effective sample size (ess) using the `coda` package in R (Plummer et al. 2006). Further details for each model is given below.

6.1 TS model (MB_TS_CON)

6.1.1 Prior distributions

All prior distributions are independent and specified as:

$$
\begin{aligned}
\mu_{\beta_h} &\overset{iid}{\sim} \mathcal{N}(0, 100000), \quad h = 0, 1, \\
\sigma_{\beta_h} &\overset{iid}{\sim} t_3(0, 5^2)T(0, \infty), \quad h = 0, 1, \\
\beta_2 &\sim \mathcal{N}(0, 100000), \\
\sigma_\eta &\sim t_3(0, 5^2)T(0, \infty), \\
\mu_x &\sim \mathcal{N}(0, 100000), \\
\sigma_x &\sim t_3(0, 5^2)T(0, \infty),
\end{aligned}
$$

$$\sigma_y \sim t_3(0, 5^2)T(0, \infty),$$

where $T(a, b)$ denotes that the random variable is truncated between the values a and b.

6.1.2 MCMC information

We adapt the algorithm for $5,000$ iterations per chain before running the algorithm for an additional $10,000$ iterations per chain. The $\hat{R}$ values for all unknown quantities was < 1.05. The ess for $x_1, \ldots, x_m$ was at least $1,100$ with a median ess $= 10,900$. For the other unknown quantities the ess was at least 200 with a median ess $= 1300$.

6.2 RCS model (MB_RCS_CON)

6.2.1 Prior distributions

All prior distributions are independent and specified as:

$$\begin{aligned}
\zeta_h &\overset{iid}{\sim} \mathcal{N}(0, 100000), \quad h = 1, \ldots, 61, \\
\beta_2 &\sim \mathcal{N}(0, 100000), \\
\sigma_\eta &\sim t_3(0, 5^2)T(0, \infty), \\
\mu_x &\sim \mathcal{N}(0, 100000), \\
\sigma_x &\sim t_3(0, 5^2)T(0, \infty), \\
\sigma_y &\sim t_3(0, 5^2)T(0, \infty),
\end{aligned}$$

where $T(a, b)$ denotes that the random variable is truncated between the values a and b.

6.2.2 MCMC information

We adapt the algorithm for $5,000$ iterations per chain before running the algorithm for an additional $10,000$ iterations per chain. The $\hat{R}$ values for all unknown quantities was < 1.05. The ess for $x_1, \ldots, x_m$ was at least 350 with a median ess $= 3,600$. For the other unknown quantities the ess was at least 80 with a median ess $= 175$.

7 Modeling details (MB_TS_SPL)

Prior to fitting, we standardized the climate variables x_t to help specification of the prior distribution.

7.1 Prior distributions

All prior distributions are independent and specified as:

$$
\begin{aligned}
\mu_{\beta_h} &\overset{iid}{\sim} \mathcal{N}(0, 100000), \quad h = 0, 1, \\
\sigma_{\beta_h} &\overset{iid}{\sim} t_3(0, 5^2)T(0, \infty), \quad h = 0, 1, \\
\beta_2 &\sim \mathcal{N}(0, 100000), \\
\sigma_\eta &\sim t_3(0, 5^2)T(0, \infty), \\
\mu_\gamma &\sim \mathcal{N}(0, 100000), \\
\sigma_\gamma &\sim t_3(0, 5^2)T(0, \infty), \\
\sigma_x &\sim t_3(0, 5^2)T(0, \infty), \\
\sigma_y &\sim t_3(0, 5^2)T(0, \infty).
\end{aligned}
$$

7.2 MCMC information

We adapt the algorithm for $10,000$ iterations per chain before running the algorithm for an additional $50,000$ iterations per chain. The $\hat{R}$ values for all unknown quantities was < 1.05. The ess for $x_1, \ldots, x_m$ was at least 700 with a median ess $= 2,500$. For the other unknown quantities the ess was at least 300 with a median ess $= 3000$.

8 Modeling Difficulties

There appear to be two modes in the likelihood. If we use "good" starting values (e.g. based on an RCS reconstruction) only one mode was ever encountered (we call this mode 1). However, when exploring the effect of starting value on the performance of the algorithm we discovered the presence of a second mode (we call this mode 2). This mode is encountered when both the missing climate values $\boldsymbol{x}^{mis}$ and the corresponding response variable β_2 have their signs reversed, with $\mu_t^{mis} = x_t^{mis}\beta_2$ approximately the same in the two modes. Note that $\mu_t^{obs} = x_t^{obs}\beta_2$ differ between the two modes.

We wish to know the comparative volumes of the two modes to gauge their relative importance. Using default MCMC algorithms (JAGS uses the Gibbs sampler when possible and one-at-a-time slice sampling otherwise), we have never once observed switching between modes. When exploring this behavior using simulated data, we found that as m, the number of missing climate observations increases, the probability of the chain switching between modes decreases.

To ascertain the relative volumes of the two modes, we considered an MCMC algorithm that included a proposal designed to switch between modes. We used a "default" algorithm (we use Gibbs sampling when possible with one-at-a-time Metropolis-Hastings sampling otherwise) but also include an additional Metropolis-Hastings step (with deterministic proposal) with the intention of enabling mode-switching. The description below assumes the marginal model for x_t is a cubic B-spline, i.e. MB_TS_SPL. The ith MCMC iteration of θ is denoted $\theta^{(i)}$ with proposal denoted θ^*.

- The proposed value for $x_t^{(i)}$ is $x_t^* = -x_t^{(i)},\ t = 1, \ldots, m$, with $x_t^* = x_t^{(i)},\ t = m+1, \ldots, n$ (as these are observed quantities).
- The proposed value for β_2 is $\beta_2^* = -\beta_2^{(i)}$.
- The proposed values for $\boldsymbol{\gamma}$ (B-spline parameters) is

$$\begin{aligned}\boldsymbol{\gamma}^* &= (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}^* + (\boldsymbol{\gamma}^{(i)} - (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}^{(i)}) \\ &= (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'(\boldsymbol{x}^* - \boldsymbol{x}^{(i)}) + \boldsymbol{\gamma}^{(i)}\end{aligned}$$

where $\boldsymbol{B}$ is the matrix of B-spline basis functions. Applying these proposals twice returns us to the original values – thus denoting the proposal function $J(\cdot|\cdot)$ satisfies $J(\boldsymbol{\gamma}^*, \boldsymbol{x}^*, \beta_2^*|\boldsymbol{\gamma}^{(i)}, \boldsymbol{x}^{(i)}, \beta^{(i)}) = 1 = J(\boldsymbol{\gamma}^{(i)}, \boldsymbol{x}^{(i)}, \beta_2^{(i)}|\boldsymbol{\gamma}^*, \boldsymbol{x}^*, \beta_2^*)$. We then accept or reject the jointly proposed $\boldsymbol{x}^*, \beta_2^*$ and $\boldsymbol{\gamma}^*$ values in a single Metropolis-Hastings step.

To implement this algorithm, we chose starting values to ensure at least one chain in each mode. We ran the default Gibbs/Metropolis component of our algorithm for 2000 iterations (to allow for burn-in) before including the "mode-switching" proposal once every 100 iterations. Every time we have run this algorithm the chain has jumped from mode 2 to mode 1 on the first attempt. In contrast, we are yet to observe a jump from mode 1 to mode 2. This suggests that mode 2 is far enough in the tail of the posterior distribution that it can be ignored. In practice, we followed a three-step procedure to model-fitting. We first fit the models using disperse starting values. As this suggested that two modes were present we explored

the "mode-switching" algorithm above. As we found only one mode to be of practical significance we then re-fitted the models using default algorithms (i.e. using JAGS), choosing starting values that did not lead to the chain finding mode 2. As in almost every applied problem there is the possibly that more than two modes exist and we failed to find them. However, in numerous runs of these models from varied starting values we have only ever encountered two such modes.

The multiple modes are only a problem in the models that do not include additional information. Restricting the support of β_2 to be positive overcomes the problem observed. However, we note that there may be situations where we may be unwilling to restrict the support as we did here.

9 Modeling details (MB_TS_SPL_IP)

Prior to fitting, we standardized the climate variables x_t to help specification of the prior distribution.

9.1 Prior distributions

The weakly-informative prior distributions for the tree-ring model are:

$$\begin{aligned}
\mu_{\beta_h} &\overset{iid}{\sim} t_3(0, 10^2), \quad h = 0, 1, \\
\sigma_{\beta_h} &\overset{iid}{\sim} t_3(0, 1^2)T(0, \infty), \quad h = 0, 1, \\
\beta_2 &\sim t_3(0, 10^2)T(0, \infty), \\
\sigma_\eta &\sim t_3(0, 5^2)T(0, \infty), \\
\sigma_y &\sim t_3(0, 5^2)T(0, \infty).
\end{aligned}$$

The informative prior distributions for the climate model are given below.

- The prior for μ_γ is

$$\mu_\gamma \sim t(0, 0.5^2, 3).$$

  This implies a prior 50% interval for μ_γ of $(-0.38, 0.38)$. That is, we have a prior probability of 0.5 that the overall 500 year mean is within ± 0.38 sample standard deviations of the observed sample mean (from the past 100 years). A prior 90% interval is $(-1.17, 1.17)$.

- The prior for σ_γ is

$$\sigma_\gamma \sim t(0, 0.5^2, 3)T(0,).$$

This implies a prior 50% interval for μ_γ of $(0, 0.38)$. That is, we have a prior probability of 0.5 that the standard deviation of the mean temperature (how much the mean of the temperature variable changes through time) is between 0 to 0.38 times the observed sample standard deviation of the temperature observations. A prior 90% interval is $(0, 1.17)$.

Based on the informative priors specified we simulate from the implied prior distribution for α_t (the underlying mean of the temperature observation x_t). We simulate 10000 draws from the prior distribution and find that 93.7% of values fall between 8 °C and 12 °C. Moreover, 99.8% of values are greater than 4 °C and 99.5% are greater than 6 °C.

The prior for σ_x is

$$\sigma_x \sim t(0, 1, 3)T(0,).$$

This implies that a prior 50% interval for the year to year standard deviation is $(0, 0.76)$. That is, we have a prior probability of 0.5 that the year to year standard deviation of temperature values from the mean is between 0 and 0.76 times the observed sample standard deviation. Likewise, a prior 90% interval is $(0, 2.35)$.

Simulating 10000 draws from the implied prior distribution for x_t we find that 81.1% of values are between 8 °C and 12 °C. Moreover, 99.0% are above 4 °C and 97.4% are above 6 °C.

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. and Melvin, T. M. (2011), "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, pp. 113 – 145.

Brooks, S. P. and Gelman, A. (1998), "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

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

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.

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

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.

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

Holmes, R. L. (1983), "Computer-assisted quality control in tree-ring dating and measurement," *Tree-Ring Bulletin*, 43, 69–78.

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.

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.

Melvin, T. M., Grudd, H., and Briffa, K. R. (2013), "Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data," *The Holocene*, 23, 364 – 373.

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.*

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: convergence diagnosis and output analysis for MCMC," *R News*, 6, 7–11.

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

302 St. George, S., Ault, T. R., and Torbenson, M. C. A. (2013), "Absent growth rings are rare in Northern
303 Hemisphere forests outside the American Southwest," *Geophysical Research Letters*.

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

305 Yin, M. (2000), "Noninformative priors for multivariate linear calibration," *Journal of Multivariate Analysis*,
306 73, 221–240.

Tables

Model	MSE for hold-out sample 1	2	Overall RMSE	RE
RCS_INV	1.02	0.93	0.99	0.26
TS_INV	0.95	0.87	0.95	0.31
MB_RCS_CON	1.11	1.00	1.03	0.20
MB_TS_CON	0.98	0.86	0.96	0.30
MB_TS_SPL	1.13	1.02	1.04	0.19
MB_TS_SPL_PL	1.01	1.04	1.01	0.22
No tree-ring data; constant mean	1.42	1.23	1.15	—

Table 1: The MSE for each hold-out sample and overall (i) RMSE for mean summer temperature across five different models and (ii) reduction of error statistic (RE). For more details on the RE statistic, see Cook and Kairiukstis (1990). The final model assumes that there is no tree-ring data, with temperatures assumed normally distributed with constant mean.

Figures

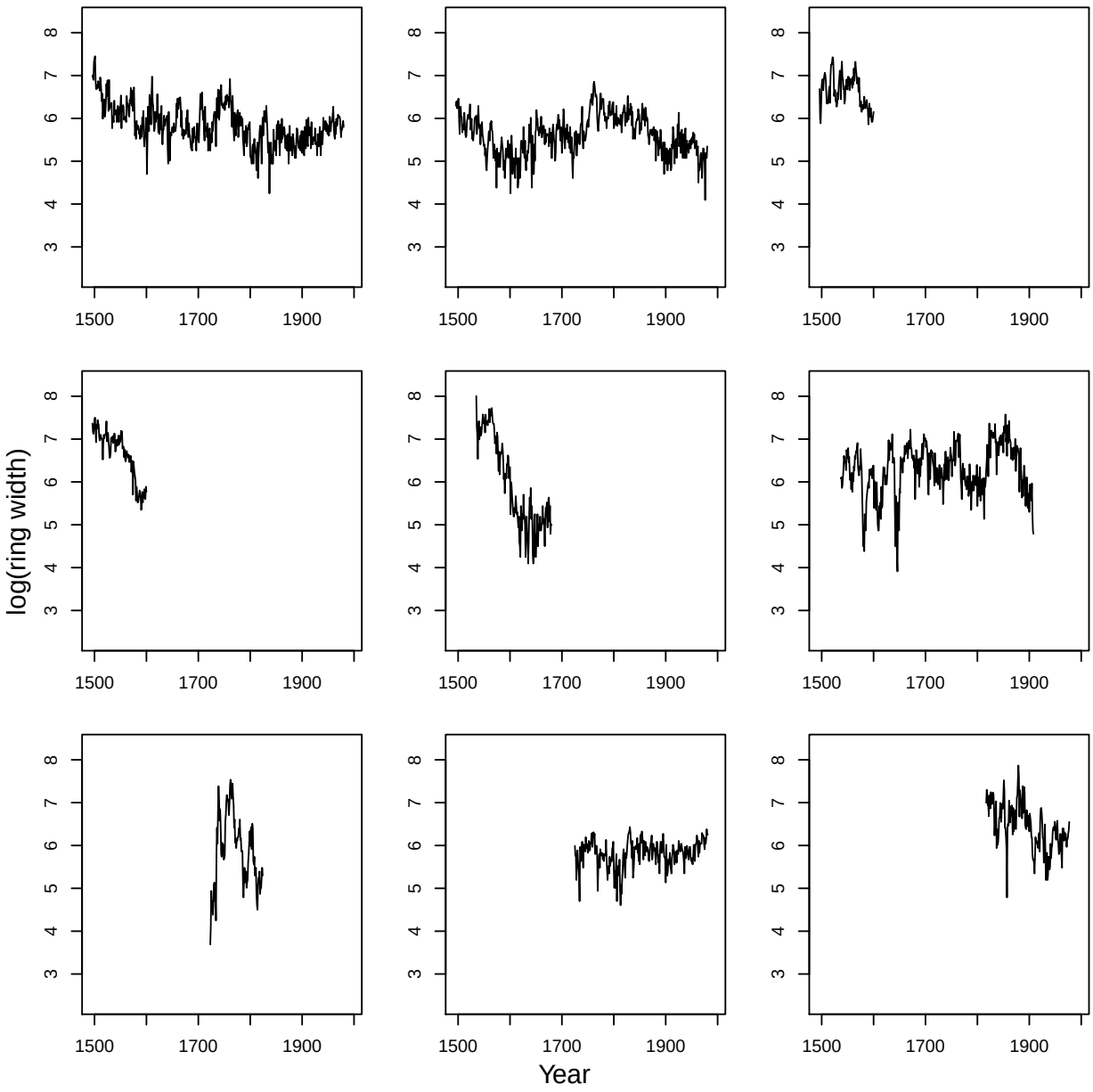

Figure 1: The raw data plotted for 9 (out of 247) trees. The trees were chosen so that there was a range of both segment lengths and start times.

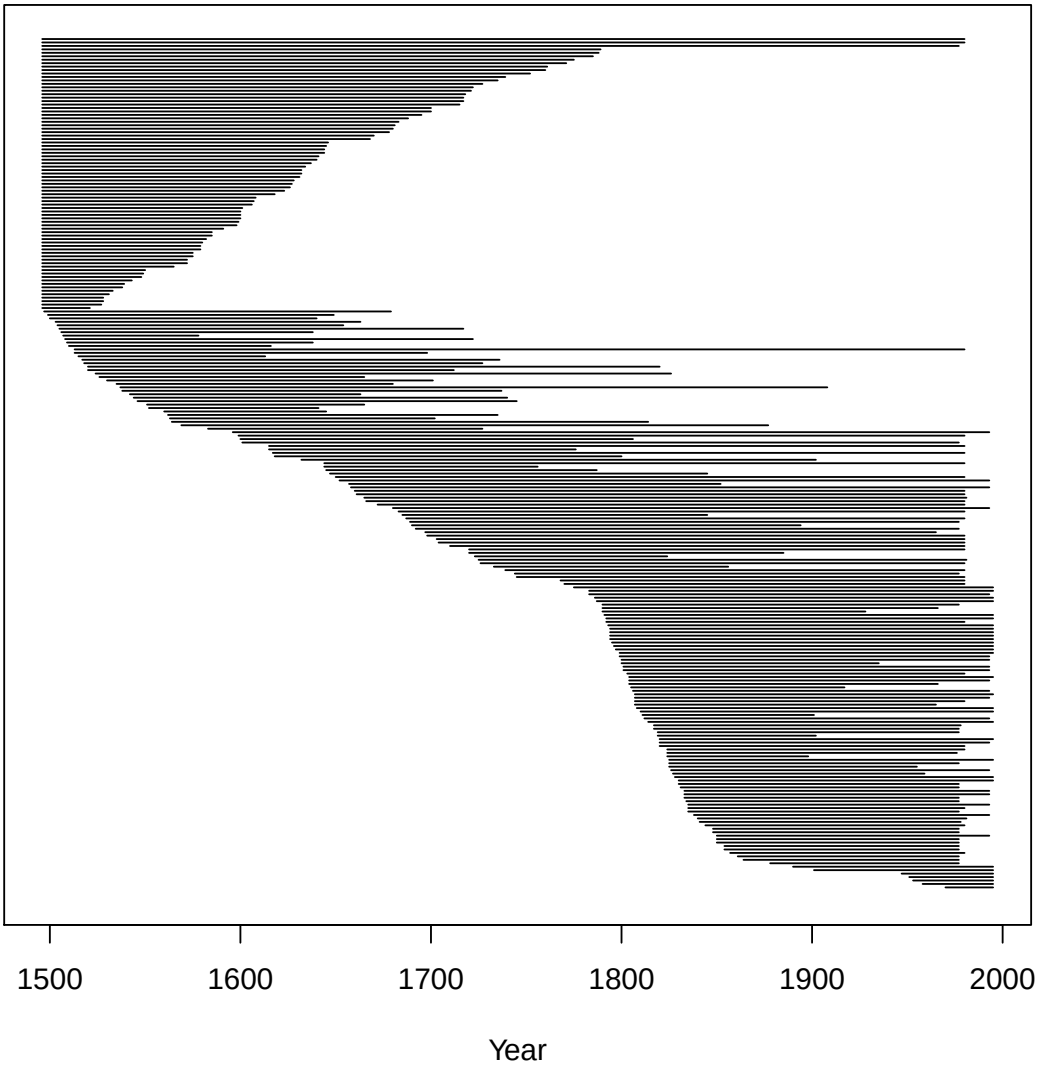

Figure 2: For each of the 247 trees used in the analysis a line is drawn to denote the segment length.

Figure 3: The number of cores (series) per year in the dataset.

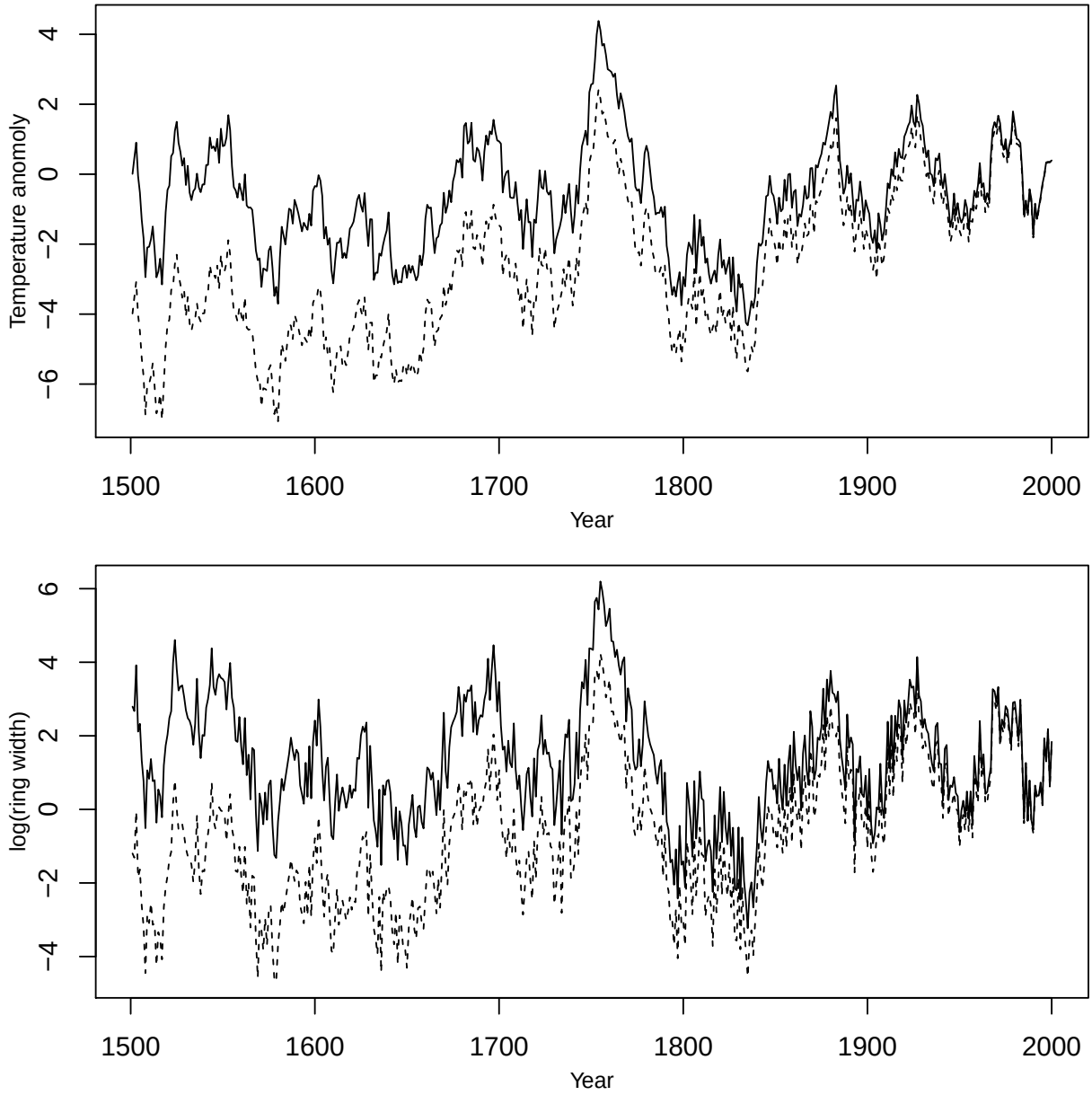

Figure 4: Upper panel: simulated temperature series x_{1t} (solid line) and x_{2t} (dashed line). Lower panel: simulated ring width series $\log(y_{1t})$ (solid line) and $\log(y_{2t})$ (dashed line)

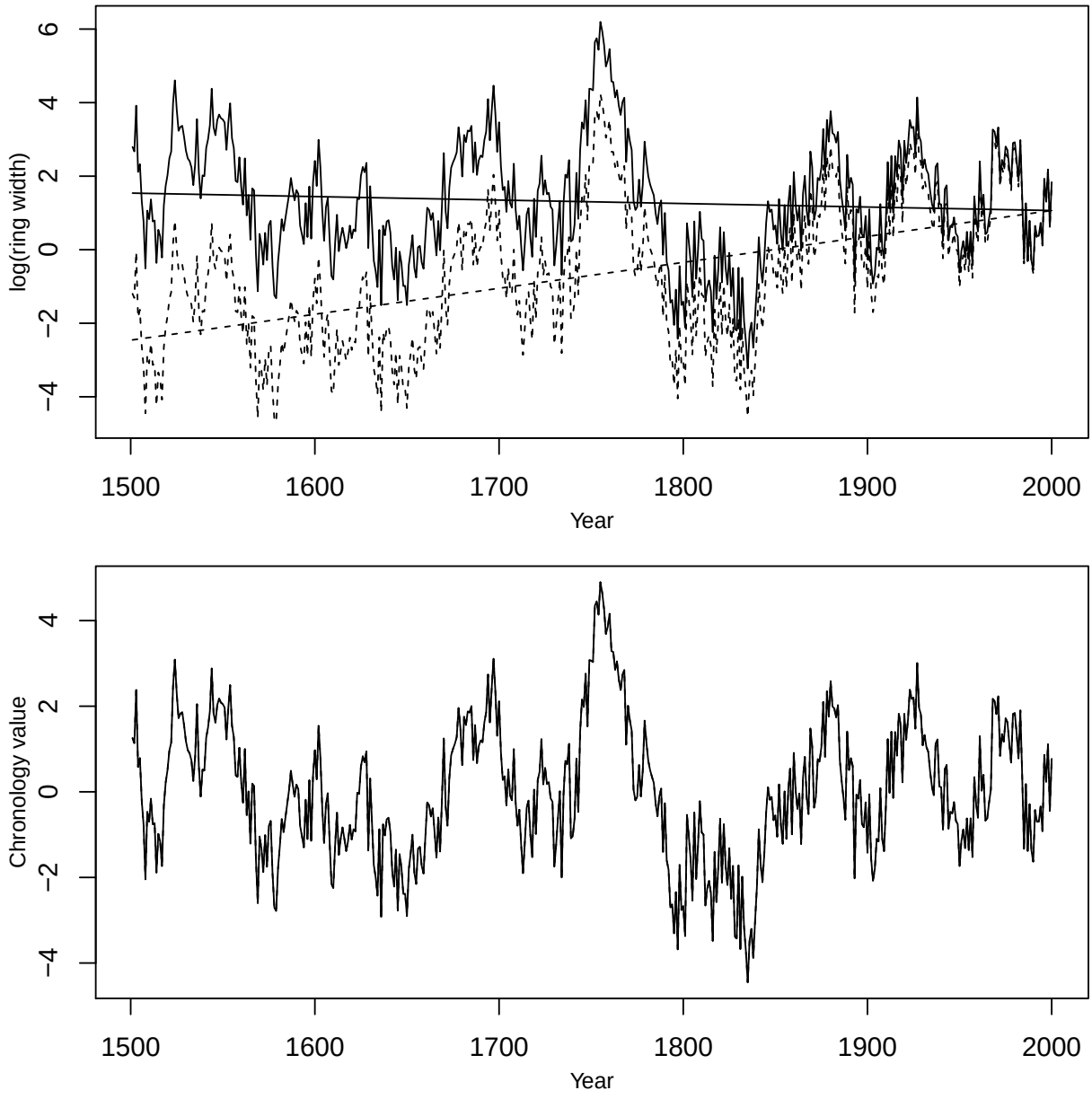

Figure 5: Upper panel: simulated ring width series $\log(y_{1t})$ (solid line) and $\log(y_{2t})$ (dashed line) with fitted aging function superimposed. Lower panel: chronology $\log(z_{1t})$ (solid line) and $\log(z_{2t})$ (dashed line). The two chronologies are indistinguishable.

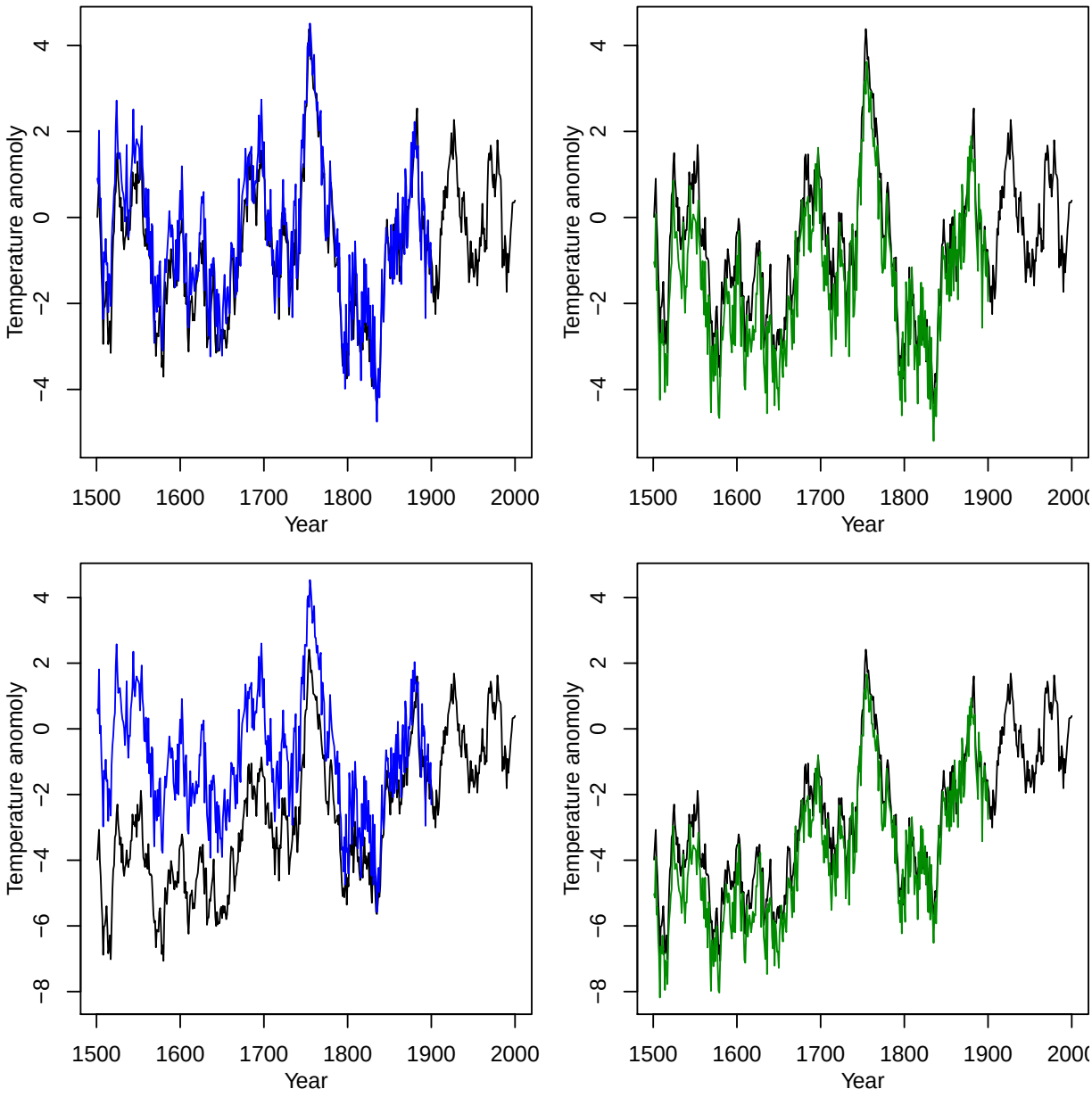

Figure 6: Upper row: reconstructions of temperature series x_{1t}. Lower row: reconstructions of temperature series x_{2t}. Left column: reconstructions found via chronology from TS. Right column: reconstructions found via simultaneous model. In all plots the black line is the true temperature series; the blue line is the reconstruction via TS; the green line is the reconstruction via the simultaneous model.

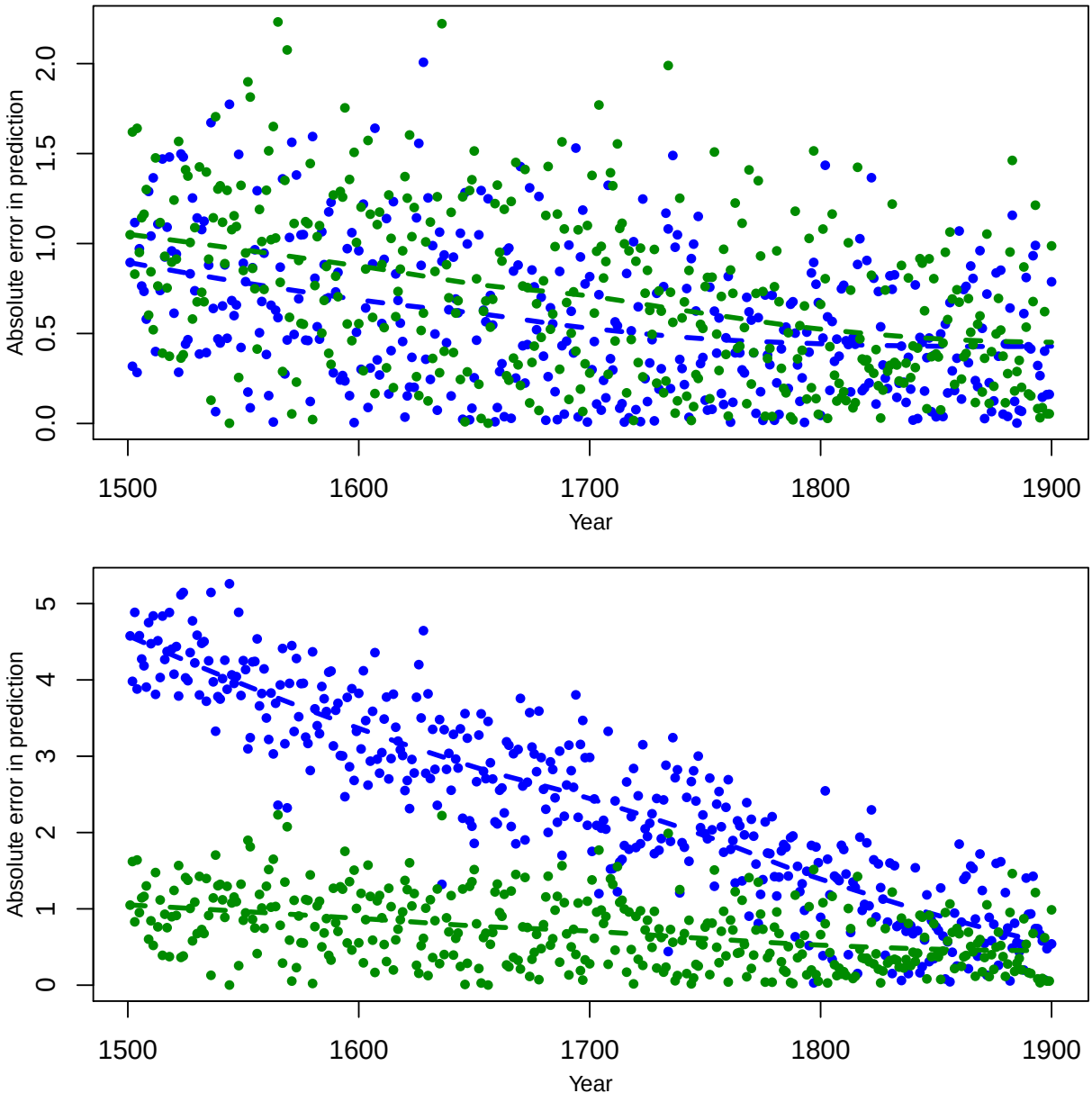

Figure 7: Plots of absolute error for each of the two temperature series, x_{1t} (upper plot) and x_{2t} (lower plot). In both plots, the blue points correspond to the reconstruction via TS; the green points correspond to the reconstruction via the simultaneous model. The dashed lines are the fitted values from a loess smoother.