The Value of Multi-proxy Reconstruction of Past Climate

Bo Li[1], Douglas W. Nychka[2] and Caspar M. Ammann[2]

Abstract

Understanding the dynamics of climate change in its full richness requires the knowledge of long temperature time series, which is however not available upon direct observation. Fortunately there are other forms of data, known as climate proxies, that can have a statistical relationship to temperature and have been used to infer temperatures in the past before direct measurements. We propose a Bayesian hierarchical model to reconstruct past temperatures that enables to effectively integrate information from different sources, such as proxies with different temporal resolution and forcings acting as the external drivers of large scale temperature revolution. Additionally, this method allows us to rigorously quantify the uncertainty of the reconstruction. Via synthetic climate data from global climate models, our reconstruction method is thoroughly assessed, the skill of each type of proxies, the role of forcings and the sensitivity of our approach to errors in proxies are systematically investigated, and the strength of combining different sources is well demonstrated. These results can be a useful guideline for multi-proxy-based temperature reconstruction of the past.

Some key words: Bayesian hierarchical model; Forcings; Global climate model; Past temperature reconstruction; Proxies

Short title: Past Temperature Reconstruction

[1]Purdue University, West Lafayette, IN 47906. Email: boli@stat.purdue.edu

[2]National Center for Atmospheric Research, Boulder, CO 80307-3000

Email:{nychka,ammann}@ucar.edu

1 Introduction

Understanding the complex dynamics of the Earth's climate system has always been a grand scientific challenge in the geosciences. Recently this problem has assumed additional importance with the scientific consensus that it is likely that human activities are responsible for the global warming observed over the last decades, and also with the possibility that greenhouse gas emissions will continue to increase to levels in the 21st century that the Earth has not seen for millions of years. Not only will these dramatically change the mean climate, but the largest impacts are expected regionally. Part of the difficulty in making projections of climate 50 or 100 years in the future is that the behavior of the Earth system over such time scales has not been well observed. In our work we wish to consider the climate over the past 1000 years as a baseline to study long term variability and responses to variations in solar radiation and other external factors. Also, time spans of this length or longer are useful for testing whether advanced climate models, also known as atmosphere/ocean general circulation models (AOGCM), are able to simulate longer term features of the climate system accurately. AOGCMs are the main scientific tool for determining future climate in response to different scenarios of greenhouse gas emissions and so their development and validation are important for quantifying climate change over the next century.

There are only direct measurements of temperature for the Northern Hemisphere with sufficient sampling for the past 150 years with a few isolated locations extending back into the 17th century. Moreover extensive measurements of the ocean and upper atmosphere are even more restricted in time being largely confined to the last 50 years. Given the paucity of long term and direct measurements of the atmosphere and ocean, the description of long term behavior of the climate system and the development of climate models for long term simulation must rely on other sources of information. Put simply we need other means to determine surface temperatures before there were thermometers.

Fortunately there are other forms of data, known as climate proxies, that can have a statistical relationship to temperature and have been used to infer temperatures in the past before direct measurements. This work proposes Bayesian hierarchical methods that are able to combine different kinds of climate proxies to create a single reconstruction for Northern Hemisphere (NH) average temperature and also provides rigorous measures of uncertainty in the reconstructed temperatures. We evaluate these methods using a high resolution simulation of the Earth's climate from a numerical experiment and quantify the tradeoffs among different kinds of proxies and statistical/physical models. Combining different proxies in a Bayesian framework represents an advance over current methods in the paleoclimate community.

1.1 Multiple proxies for temperature

Different proxies preserve the climate information in different ways and therefore are sensitive to climate meteorological variables at different time scales. One might be good at short time scales, while another better at larger time scales. This is a key to our approach and the Bayesian hierarchical modeling that we employ takes advantage of such complementary skills/charateristics among different proxies. In this work we consider three widely used but distinct climate proxies: tree-rings, borehole temperatures and pollen abundance. Tree-ring width and density are perhaps the most extensive proxy and their relationship to seasonal temperatures has been extensively studied (Fritts, 1976; Cook and Kairiukstis, 1990; Schweingruber, 1996). Although tree-ring measurements typically have their dating accurate to the year, their ability to encode centennial and slower climate variability is often limited by the technique used to remove non-climatic variations in tree-ring time series (Cook *et al.*, 1995). Borehole depth temperature profiles directly preserve the surface temperature variability as the surface heat diffuses downward into the earth and have been recently used to characterize NH continental temperature for the past 500 years (Huang *et al.*, 2000; Harris and Chapman, 2001; Chapman *et al.*, 2004). As opposed to tree-rings, borehole temperatures are only sensitive to climate signals at multi-decadal or longer time scales due to the attenuating feature of the diffusion process. As a third proxy, pollen records provide climate information that can fill the gap between tree-ring and borehole temperatures and are considered sensitive to multi-decadal variability. Although pollen records are widely distributed and have been used for reconstructing discrete periods in the past (Guiot *et al.*, 1993, Williams *et al.*, 2000), they have only rarely been used in large scale temperature reconstructions for the past millennium. For more details on these proxy characteristics, see Guiot *et al.*, 2005.

1.2 External drivers of climate

Besides proxy measurements of past climate there are also observations of the external drivers of the climate system. These are termed external *forcings* because they are assumed to vary independently of the climate system but can have an influence on it. In this work we will focus on solar irradiance, volcanism and greenhouse gases as primary external forcings for temperature evolution. These three forcings drive the large scale climate variation, because the climate system has to react to any variations of these three forcings in order to restore the energy balance (Crowley, 2000; IPCC 2007). A positive solar irradiance forcing tends to warm the surface, whereas a negative one tends to cool it. Volcanism often causes sudden temperature drops because the large amount of aerosols ejected by an explosive volcanic eruption into the atmosphere reduces the radiation reaching the surface. Unlike those two natural forcings, recent increase in greenhouse gases, with CO2 as a major component, is an anthropogenic forcing due to human activities. An important component of our statistical model is the inclusion of an empirical model for temperature that depends on these forcings. This covariate information is an important addition to that from the proxy observations, and also distinguishes our approach from many conventional paleoclimate reconstructions.

1.3 Statistical estimates of past temperatures

Most studies reconstruct the past temperature by relying on only one proxy records of a particular resolution and do not use external forcings as covariate information. For example, Mann *et al.* (1998), Jones *et al.* 1998, Crowley and Lowery (2000) and Guiot *et al.* (2005) reconstructed the Northern Hemisphere (NH) temperature based on annually recorded resolution, with Briffa and Melvin (2008) exclusively based on tree-rings. Viau *et al.* (2006) carried out the reconstruction by focusing on pollen only and Harris and Chapman (2001) drew inference for the past temperature primarily based on borehole. Only a few reconstruction applications have tried to integrate data from sources with very different temporal resolution (e.g., Beltrami *et al.*, 1995; Huang, 2004; Moberg *et al.*, 2005; Haslett *et al.*, 2006). Although they found that data integration improves the reconstruction, none of them has tried to integrate all three types of proxies as well as the external forcings mentioned above. Furthermore, no work has been conducted to systematically investigate the role of proxies and forcings and thus provide a guide for temperature reconstruction.

We develop a Bayesian Hierarchical Model (BHM) to reconstruct the NH mean temperature that incorporates the information from tree-ring, pollen and borehole records altogether, and also makes use of past forcings. BHMs have been demonstrated to be a powerful method in solving complex problems in climatology, ecology and environmetrics (e.g., Berliner *et al.*, 2003; Wikle *et al.*, 2001) by splitting a complicated model into three basic components: an observation level, a process level and a level of prior information on statistical parameter. In this application the observation level relates each proxy to temperature, the process level relates temperature to the external forcings and the prior level specifies prior distributions of regression and variance parameter that tend to stabilize the problem but are otherwise uninformative. The synthesis of proxies and forcings enabled by BHMs will ideally provide more accurate reconstructions because the strength of one component can compensate for weakness of others.

On the one hand, compared to the widely used regression approach (e.g., Li *et al*, 2007) in the temperature reconstruction, our method can avoid the possible attenuation effects caused by errors in explanatory variables (Ammann *et al.*, 2009), because the BHM explicitly models the measurement error in proxies. On the other hand, to our knowledge none of the previous work combined information as thorough and complete as our approach. Among them Haslett *et al.* (2006) and Lee *et al.* (2008) are perhaps the two that are mostly closely related to our method. Haslett *et al.* first proposed a Bayesian reconstruction based on the fossil pollen data, but they did not consider other type of proxies and external forcings. Lee *et al.* proposed a Kalman filter approach to incorporate the external driving factors, yet they did not consider combining different proxies. It is worth noting that a new development in this area is the concept of spatial reconstruction of temperature fields using a vector autoregressive model of order 1 (VAR(1)) as the underlying temperature process by Tingley and Huybers (2009).

1.4 Quantifying the information from proxies and external forcings

In developing and testing this statistical method we take an unconventional approach in its evaluation of scientific impact. Because there is no adequate "ground truth" available for evaluating the fidelity of reconstruction approaches, we use a high resolution state-of-art climate simulation of the past 1150 years as a means for evaluating the method. Specifically we use the output from the climate model to generate synthetic proxy data sets that represent the characteristics of real-world proxies and include the forcings used in the simulation to determine how well our method can reconstruct the model temperatures. The strategy of using synthetic data from climate model output to evaluate the reconstruction methods is well established in the paleoclimatology (Mann and Rutherford, 2002; Rutherford *et al.*, 2003; Zorita *et al.* 2003; von Storch *et al.*, 2004; Ammannn and Wahl, 2007; Lee *et al.*, 2008) and is a practical solution to provide test beds that are complex but where the true tempertures are known. Finally it should be noted that the climate model has substantial complexity relative to any tractable BHM and so studies using this model provide a reasonable measure of how well a statistical model can account for high dimensional and nonlinear geophysical process with stochastic components.

A scientific contribution of this paper is an estimate of the added value of combining proxies with different climate retention characteristics and including external forcings series. To a statistician this is a practical exploration of different statistical designs for proxy and forcing data sets. Assembling a meaningful multi-proxy data set is by itself time consuming and an important question to raise is whether such an effort is worth the additional accuracy in the reconstruction. A companion issue is the value of including external forcing information into the statistical procedure, especially when the forcing series also have observational errors. The inclusion of forcings can be criticized by the reconstruction community as adding physical information and perhaps biases will contaminate a pure reconstruction considering only climate information. Also comparing a reconstruction using external forcing with a climate model simulation may have some circularity as similar forcing series are used for the model results. However, if the forcings provide substantial improvement in the reconstruction and are included with relevant components of statistical uncertainty they might be very useful. In particular, we are interested in determining the tradeoff between using external forcings jointly with a single proxy type and using multiple proxies. These two alternative can have practical implications for the selection of proxies. For example, borehole temperatures required a more complicated observation model than tree-ring or pollen. If external forcings can stand in for the longer time scale information from borehole proxies then borehole data need not be included in the BHM. To our knowledge this is the first deliberate exploration of these statistical design issues for paleoclimate applications.

1.5 Outline of article

Section 2 gives details on the global climate model output used for evaluating the method and the external forcings series for the past 1150 years. Section 3 describes the salient features of different proxy data and how we generate the synthetic proxies. Section 4 presents the BHM for combining proxies and forcings in temperature reconstruction, and discusses several variations of the hierarchical model. Section 5 shows the results from different hierarchical models under different combinations of proxies and forcings, answers the design questions raised in the Introduction, and analyzes the identifiability of parameters. Section 6 discusses the strength and extensibility of this Bayesian hierarchical framework with respect to temperature reconstruction.

2 Climate model output and radiative forcings

Climate system models are large computer codes that implement the basic physical equations for fluid dynamics and for thermodynamics to describe the motion of the atmosphere, ocean and sea ice and their interaction with the land. The models are highly nonlinear and expressed in a differential form where the state of the climate system is stepped from one time point to the next over a short time interval by solving a large system of coupled partial differential equations. The model simulations are started by initial conditions of the ocean, atmosphere and sea ice and are subsequently driven by internal variability that is modulated or changed by external forcings over time such as solar irradiance, volcano aerosols and greenhouse gas concentrations. In addition topography and an annual cycle of land cover are prescribed as boundary conditions for the atmosphere over land. Averaging the results of the model over a specific time period provides an estimate of the climate of the model. The climate simulation used in this work is a run of the National Center for Atmospheric Research (NCAR) Community Climate System Model (CCSM) Version 1.4 (Boville *et al.*, 2001; Otto-Bliesner and Brady, 2001). The model was configured at a resolution of $3.75^0 \times 3.75^0$ or 400km×400km, a 3^0 ocean with meridional resolution of $< 1^0$ at the equator. Details of this experiment with the data available can be found at http://www.ccsm.ucar.edu. Its reported global annual temperature during 850-2000 is shown in Figure 1. A near hockey stick shape of this temperature series results from the increased forcings due to greenhouse gases. This anomaly, while a realistic representation of how temperature has increased, can cause a bias in the reconstruction and is discussed in the section on numerical results.

The three external forcings series are solar irradiance, volcanism and greenhouse gases. Those are also used in Ammann *et al.* (2007). The solar irradiance series is a reconstruction by Bard *et al.* (2000) and is derived from measurements of fluctuations of ^{10}Be production rates which is modulated by solar magnetic variability. The volcanic series is based on a synthesis of individual ice cores (see http://www.ncdc.noaa.gov/paleo/icecore.html) and in some cases on historical records of large eruptions. It is then transformed to be an estimate of volcanic sulfate aerosol mass whose amplitude reflects the radiation reduction. In 20th century of the model simulations, the cooling effect of volcanic aerosols reproduce clearly the pattern in measured global temperatures. The forcing series of greenhouse gases is derived from analysis of air bubbles in ice cores. We use carbon dioxide as a simple representation because other gases erupted very similarly, albeit with different relative concentration. The shape of this forcing series is dominated by the first slow then rapid increase since the beginning of the 19th century in carbon dioxide . Figure 1 illustrates the three forcing series. Because the BHM will estimate the regression relationship between these variables and temperature, the individual scales of these series will not affect the reconstruction. This is useful as some of the bias or error in these series is attributed to the absolute scale. For example, most discrepancies among different solar irradiance estimates are due to different scaling which has little effect to our approach. The year to year variation in the greenhouse effect is very small for centuries before about 1800 and will have negligible effect because the key impact will more likely occur at the end of the time series. The estimates of volcanism, however, bears uncertainty in individual events that can reach 50% of the magnitude (Rind, 1995 and Zielinski, 2000), and thus needs to be taken into account by a statistical model.

3 Climate proxies observations and generating synthetic proxies

3.1 Tree rings

Variation in the width and density of tree-rings represent the most widely known climate proxy. Moreover, the wide geophysical distribution of trees make them well suited to high-resolution paleoclimate research. As a result of strong replication within and between specific sites and regions, and careful checking for common patterns, dating is effectively absolute, i.e., accurate to the year. However, the potentially limited ability of capturing centennial and slower climate variability might inhibit the tree-ring networks from representing climate uniformly across the frequency spectrum, and hence the tree-ring based reconstruction should be viewed with caution for time scales ranging from multi-decadal to multi-centennial.

Synthetic tree-ring observations

To generate synthetic but realistic tree-rings from the climate model simulation, we first select 15 local temperature series from the CCSM output and then generate 15 pseudo tree-ring records by a high pass filter. Specifically, we subtract the 10-year smoothing average from each of the local temperature series to give a filtered result. This might be an extreme approach for representing tree-ring information that perhaps in reality will not occur so drastically. But it serves the purpose of our study that evaluates the capability of our method in combining different proxies. Figure 2 shows the 15 locations and Figure 3 gives an example of generated tree-rings.

3.2 Borehole temperatures

Borehole temperature profiles provide a means to directly estimate temperatures of the past, as surface temperature histories can be obtained through inversion of the down-core temperature anomaly profile against the natural geothermal gradient (e.g., Beltrami, 2001). However, the borehole profile itself is unable to resolve the annual resolution, because the ground essentially acts like a low pass filter of the surface temperature, and the further back in time, the more severe the filtering becomes and thus smear out the temperature anomalies. For this reason, borehole temperatures are only sensitive to climate signals at multi-decadal or longer time scales. Recently, borehole derived temperature estimates have attracted much attention because several studies have systematically pointed towards a larger temperature change since AD 1500 over the NH continents compared to the other reconstructions (e.g., Huang *et al.*, 2000).

Synthetic borehole observations

Unlike generating synthetic tree-ring proxies using individual temperature series, we generate borehole data based on five regional composite temperature series. These five composite temperature series are the local average of model temperature output over a 20×20 grid square as shown in Figure 2. The distribution of those locations reasonably represents the spread of real borehole data (Huang *et al.* 2000). Due to the complexity of the physical process in forming the borehole profile, the algorithm to generate the borehole depth temperature is not straightforward. We follow the POM-SAT model which was originally derived by Carslaw and Jaegar (1959) and recently discussed by Harris (2007) to simulate the borehole profiles up to 500m. The value of the profile at every 5m depth interval is considered the synthetic depth temperature. The POM-SAT model basically describes the diffusion process of surface temperature given an appropriate initial condition while having the attenuation of a thermal perturbation with respect to depth taken into account. To illustrate the characteristic of the borehole data, we show in Figure 4 artificial examples of how a 100-year long constant time series with a perturbation at different ages is represented in the borehole temperature profile. The older a surface perturbation, the more smeared out in signal due to the diffusion.

3.3 Pollen indices

Pollen assemblages retain a smoothed record of climate variation due to the persistence properties of mature plants (Brown *et al.*, 2005), even where the data itself might be available at higher resolution. Since fossil pollen records possess skills in resolving bi-decadal to semi-centennial time resolution, they provide climate information that can fill the gap between tree-rings and borehole data.

Synthetic pollen observations

Similar to the procedure of generating synthetic borehole proxies, we select 10 regional composite temperature series as the local average of 2×2 grid cells to generate pollen data. The locations are also shown in Figure 2. As discussed above that pollen carries information only at multi-decadal temporal resolution, we therefore mimic a pollen assemblage by sampling a 10-year average temperature series at 30 year intervals. Figure 3 gives an example of such a generated pollen series.

3.4 Noise in proxies

Real world proxies are expected to contain extra noise in addition to the uncertainty between temperature at local scales and the hemispheric average. In order to assess the sensitivity of our approach to the noise in proxies, we additionally synthesize another set of proxies with an error component by adding white noise in the local/regional temperature series before they are processed to generate synthetic tree-rings, pollen and borehole. The variance of the noise is chosen to give a signal to noise ratio of 1:4 that consecutively reflects the expected precision in actual data (Mann *et al.*, 2005). The reason that we add random perturbation to the original temperature rather than directly to the pseudo proxies is to preserve the smooth profile of pollen and borehole temperatures, and it more realistically represents the local climate noise that subsequently carries over into the proxy archives.

4 Bayesian Hierarchical Model

BHMs split a complicated model into three basic components. The data model occupies one level of the hierarchy, while the process model resides below it. Typically, a third hierarchical level contains statistical models, also called priors, for unknown parameters that includes additional physical information. The levels are formally generated by a series of conditioning steps where one level is conditioned on knowledge of the levels below it. The reader is referred to Banerjee *et al.* (2004) for an introduction to BHMs. Let $[X, Y]$ denote the joint probability density function of the random variables X and Y and $[X|Y]$ the conditional density of X *given* Y. Now let P denote proxy observations, T the NH temperature process and $\boldsymbol{\theta}$ a set of statistical parameters that are involved in specifying the joint distribution of P and T. The model specification is precisely the joint distribution $[P, T, \boldsymbol{\theta}]$. This form can be built from the product of conditional distributions:

$$[P, T, \boldsymbol{\theta}] = [P|T, \boldsymbol{\theta}][T|\boldsymbol{\theta}][\boldsymbol{\theta}]$$

Regarding the paleoclimate reconstruction problem an adumbration of the hierarchial levels is (i) Data stage $[P|T, \boldsymbol{\theta}]$, (ii) Process stage $[T|\boldsymbol{\theta}]$, (iii) Priors $[\boldsymbol{\theta}]$. Level (i) focuses on modeling statistical errors of the observed data and presents the likelihood of the observed proxies given the true temperature process, while level (ii) models the temperature process from the physical perspective. Level (iii) gives prior distributions of the unknown parameters and closes the hierarchy.

4.1 Full model

Let $\mathbf{D}_i$, $\mathbf{P}_j$ and $\mathbf{B}_k$ be vectors for synthetic tree-ring (**D**endrochronology), **P**ollen and **B**orehole data indexed by their various locations. Note that these groups of proxy vectors will have different lengths due to their sampling. Moreover, each tree-ring and pollen vectors are indexed with respect to time, and the borehole vectors are indexed by depth. Also let $\mathbf{S}$, $\mathbf{V}_0$, and $\mathbf{C}$ be the time series vectors of solar irradiance, volcanism and greenhouse gases, and let $\mathbf{V}$ denote the volcanic series with error.

Let $\mathbf{M}_D$, $\mathbf{M}_P$ and $\mathbf{M}_B$ be the three transformation matrices to link temperature series to tree-ring, pollen and borehole, respectively. These three matrices are determined by the characteristics of tree-ring, pollen and borehole data described in Section 3. More specifically, they represent the linear filters used to generate the corresponding pseudo proxies from the model temperature series. Finally it is useful to partition the full length temperature process $\mathbf{T}$ into the unknown temperatures $\mathbf{T}_1$ requiring reconstruction over the time span of available proxy data, and the observed instrumental temperatures by $\mathbf{T}_2$ (from 1850-present), i.e., $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)'$. Then we have the below three hierarchies:

(i) Data stage:

$$\mathbf{D}_i|(\mathbf{T}_1,\mathbf{T}_2)' = \mu_{iD} + \beta_{iD}\mathbf{M}_D(\mathbf{T}_1,\mathbf{T}_2)' + \boldsymbol{\epsilon}_{iD}, \quad \boldsymbol{\epsilon}_{iD} \sim \mathrm{AR}(2)(\sigma_D^2, \phi_{1D}, \phi_{2D}), \tag{4.1}$$

$$\mathbf{P}_j|(\mathbf{T}_1,\mathbf{T}_2)' = \mu_{jP} + \beta_{jP}\mathbf{M}_P(\mathbf{T}_1,\mathbf{T}_2)' + \boldsymbol{\epsilon}_{jP}, \quad \boldsymbol{\epsilon}_{jP} \sim \mathrm{AR}(2)(\sigma_P^2, \phi_{1P}, \phi_{2P}), \tag{4.2}$$

$$\mathbf{B}_k|(\mathbf{T}_1,\mathbf{T}_2)' = \mathbf{M}_B\{\mu_{kB} + \boldsymbol{\beta}_{kB}(\mathbf{T}_1,\mathbf{T}_2)' + \boldsymbol{\epsilon}_{kB}\}, \quad \boldsymbol{\epsilon}_{kB} \sim \text{iid } N(0, \sigma_B^2), \tag{4.3}$$

$$\mathbf{V}|\mathbf{V}_0 = (1+\epsilon_V)\mathbf{V}_0, \quad \epsilon_V \sim \text{iid } N(0, 1/64); \tag{4.4}$$

(ii) Process stage:

$$(\mathbf{T}_1,\mathbf{T}_2)'|(\mathbf{S},\mathbf{V}_0,\mathbf{C}) = \beta_0 + \beta_1\mathbf{S} + \beta_2\mathbf{V}_0 + \beta_3\mathbf{C} + \boldsymbol{\epsilon}_T, \quad \boldsymbol{\epsilon}_T \sim \mathrm{AR}(2)(\sigma_T^2, \phi_{1T}, \phi_{2T}); \tag{4.5}$$

(iii) Priors:

$\mu_{iL} \sim N(\widetilde{\mu}_{iL}, \widetilde{\sigma}_{iL}^2), \beta_{iL} \sim N(\widetilde{\mu'}_{iL}, \widetilde{\sigma'}_{iL}^2)$, $L \in \{D, P, B\}$;

$\beta_i \sim N(\widetilde{\mu}_i, \widetilde{\sigma}_i^2)$, $i = 0, 1, 2, 3$;

$\phi_{2L} \sim \mathrm{unif}(-1, 1)$, $\phi_{1L}|\phi_{2L} \sim (1-\phi_{2L}) \times \mathrm{unif}(-1, 1)$, $L \in \{D, P, T\}$;

$\sigma_L^2 \sim IG(\widetilde{q}_L, \widetilde{r}_L)$, $L \in \{D, P, B, T\}$.

The target is to estimate $\mathbf{T}_1$ given $\mathbf{T}_2$, the proxies and the forcings. Forward models (4.1) to (4.3) describe the statistical relationship between proxies and the true temperature process, and (4.4) accounts for the uncertainty in the volcanism which is estimated to be between -1/4 to 1/4 of the magnitude of the individual volcanic pulses themselves. The special form of model (4.3) respects the smooth feature of the borehole profile. Proxies of the same type but from different locations, such as the 15 tree-rings, 10 pollens or 5 boreholes generated in Section 3, are allowed to have different regression coefficients including the intercepts and slopes, but they all share the same parameters in the error process to retain the parsimony of the whole model. This restriction is reasonable if the proxies are first standardized before being used, which is a standard method in paleoclimatology (Bradley and Jones, 1993, Osborn and Briffa, 2006). Model (4.5) brings the physical understanding of temperature evolution based on energy balance theory into the reconstruction. Although our BHM contains only linear models, we could of course replace those linear models by more complicated ones. However, we found that the linear models suffice for our data, and the problem at hand, i.e., reconstructing NH mean temperature (Mann *et al.*, 2008). The reason why we choose AR(2) for error processes is justified in Li *et al.* (2007).

4.2 Sampling from the posterior distribution

The specific choice of priors for time lag coefficients (ϕ_{1L}, ϕ_{2L}) guarantees their corresponding AR process to be stationary and causal (Shumway and Stoffer, 2006, ch. 3), and the conjugate priors for all the other parameters allow for an explicit full conditional posterior distribution for those parameters and $\mathbf{T}_1$. There is no closed form for the posterior distribution of time lag coefficients. Thus the posterior is sampled by alternating the Gibbs sampler, which is used for updating $\mathbf{T}_1$ and parameters with explicit full conditional distribution, and the Metropolis-Hasting (M-H) algorithm, which is used for updating autoregressive parameters. More specifically, we generate posteriors by Gibbs sampling for $\mathbf{T}_1$, $\mathbf{V}_0$, (μ_{iL}, β_{iL}) with $L \in \{D, P, B\}$, β_i with $i = 0, 1, 2, 3$ and σ_L^2 with $L \in \{D, P, B, T\}$, and generate posteriors by M-H for (ϕ_{1L}, ϕ_{2L}) with $L \in \{D, P, T\}$.

Whenever the M-H algorithm is used, the acceptance rate is tuned to be roughly between 25% and 50% to secure adequate mixing of posterior samples (Gelman *et al.*, 1996). We choose the hyperparameters $\widetilde{\mu}_{iL} = 0$ and $\widetilde{\mu'}_{iL} = 1$ for $L \in \{D, P, B\}$, because this represents the ideal case when the local/regional temperatures are not biased against the NH temperature. For a similar reason, $\widetilde{\mu}_i$ is set to be $0, 1, 1, 1$ for $i = 0, 1, 2, 3$. We found the results are robust to different choices of those hyperparameters. In order to let the data determine the final estimates of the regression coefficients, we choose a relatively wide variance $\widetilde{\sigma}_{iL}^2 = \widetilde{\sigma'}_{iL}^2 = 1$ for $L \in \{D, P, B\}$ and $\widetilde{\sigma}_i^2 = 1$ for $i = 0, 1, 2, 3$ to make the priors less informative. $(\widetilde{q}_L, \widetilde{r}_L)$ are set to be $(3, 1)$ which corresponds to relatively vague prior knowledge.

4.3 Simplifications of the full BHM

In order to answer questions raised in the introduction, we consider several simplifications of the full model. Here we list the different factors that will figure into our numerical study.

Temperature process model without external forcings

In order to identify the role of forcings in the reconstruction, we compare reconstructions with forcings and without forcings being incorporated. We form BHMs without forcings by omitting (4.4) and replacing model (4.5) in the process stage by

$$(\mathbf{T}_1, \mathbf{T}_2)' = \beta_0 + \boldsymbol{\epsilon}_T, \quad \boldsymbol{\epsilon}_T \sim \mathrm{AR}(2)(\sigma_T^2, \phi_{1T}, \phi_{2T}).$$

This is equivalent to setting the regression coefficients equal to zero in the full process model. The discrepancy between results of this variation and the full model reveals the value of the external forcings at the process level.

An "oracle" proxy

Since we are dealing with climate model output that serves for our data samples, there is an opportunity to test what skill BHM approach can reach under optimal conditions, i.e., if instead of local proxy data there would have existed long prefect instrumental series at the same locations. Therefore, we consider an *oracle*[3] proxy that contains full knowledge of the true temperatures at the gridbox locations. These idealized proxies can be interpreted as having thermometers back in time at these locations and the reconstruction error then is dominated by the skill of these locations in reproducing the hemispheric average. The oracle reconstruction is a benchmark because it is the best reconstruction from a selected sample of local/regional temperature series. To obtain the

[3]After the legendary oracle in the temple at Delphi that provided divine answers to questions posed by visitors.

"oracle" reconstruction, we replace three data models (4.1) to (4.3) by one single model

$$\mathbf{T}_l|(\mathbf{T}_1, \mathbf{T}_2)' = \mu_l + \beta_l(\mathbf{T}_1, \mathbf{T}_2)' + \boldsymbol{\epsilon}_l, \quad \boldsymbol{\epsilon}_l \sim \mathrm{AR}(2)(\sigma^2, \phi_1, \phi_2),$$

where $\mathbf{T}_l$ are the grid box temperature time series (30 in total) that are used to generate all three types of proxies. The other levels in the hierarchy are kept the same.

Subsets of proxies

Although the oracle experiment provides a baseline reference, our main interest is in separating the contribution of each proxy to the reconstruction, and how deficiencies in one could be compensated for by others. This can be accomplished by reducing the full model into a sequence of submodels that only contain a subset of the three proxies, that is, this can be carried out by omitting different sets of proxies from the data level.

Partial temperature process model

The temperature reconstruction usually assumes a stationary relationship between temperatures and proxies across time, and this is also implied in our hierarchical models. In particular, this assumes that the temperatures at the prediction period $\mathbf{T}_1$ and at the calibration period $\mathbf{T}_2$ follow the same model. Thus it imposes the constraint that the mean function of temperatures, either as a function of forcings or as a constant, is identical for both the calibration and prediction periods. One can relax this constraints by modeling only $\mathbf{T}_1$ in the process stage, and investigate the effects caused by this assumption. In this case, the process model (4.5) becomes

$$\mathbf{T}_1|(\mathbf{S}, \mathbf{V}_0, \mathbf{C}) = \beta_0 + \beta_1\mathbf{S} + \beta_2\mathbf{V}_0 + \beta_3\mathbf{C} + \boldsymbol{\epsilon}_T, \quad \boldsymbol{\epsilon}_T \sim \mathrm{AR}(2)(\sigma_T^2, \phi_{1T}, \phi_{2T}).$$

5 Numerical study

Motivated by the issues raised in the introduction we evaluated five different subsets of proxy or related data.

(T) : Oracle proxies

(D) : Tree-ring (Dendrochronology) only

(DP) : Tree-ring + Pollen,

(DB) : Tree-ring + Borehole,

(DBP) : Tree-ring + Borehole + Pollen.

For each of these cases we explored the different BHM choices with a 2^3 factorial design: with/without forcing covariates, with/without proxy noise and with/without $\mathbf{T}_2$ in the process model. Thus our study consists of a total of 5×2^3 separate reconstructions that can be compared to the actual model temperatures. The bias and the variance of differences between the reconstruction and target are two informative measures to evaluate the reconstruction with their sum being the root mean square error (rmse). Both low bias and low variance are desirable for being a good reconstruction, and we report the bias, variance and rmse for all reconstructions in Figure 5. In general the patterns are what one can expect with the rmse decreasing as proxies are added and the oracle proxy having the smallest rmse.

5.1 The value of forcings and proxies

Influence of external forcings

Forcings play a very important role in obtaining a well calibrated and sharp reconstruction. As seen in Figure 5, in all cases inclusion of forcing covariates reduce the bias, the variance and thus rmse. In particular, when there is no pollen data involved, forcings dramatically improve the performance of the reconstruction based on tree-rings or a combination of tree-rings and borehole. Yet once pollen is being employed, the benefit of incorporating forcings becomes less remarkable. Again, this is geophysically not surprising because pollen data contain the signal of decadal to centennial scale variability. In short, external forcings is helpful if the included proxy data does not well represent the low frequency signal.

Skill of proxies

We focus on the five data models (T, D, DP, DB, DBP) where forcings are absent, proxies are not subject to noise, and constant mean process model is assumed for $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)'$ rather than only $\mathbf{T}_1$, to study the contribution due to each type of proxies. This corresponds to the "C" points in the leftmost panels of Figures 5(a) to (c), and the patterns in those plots roughly imply the role of each type of proxy. To more clearly illustrate the benefit of incorporating a particular type of proxy, we compare the spectrum of the reconstruction residuals from the five different data subsets. The residual spectrum at different frequencies measures the variation component that we missed in the reconstruction at that specific frequency. Figure 6 shows that the spectrum based on data models (D) and (DB) look similar and also the spectrum based on data models (DP) and (DBP) are hardly differentiable, although the latter two have smaller power at low frequency due to the pollen proxies involved in the reconstruction. Since pollen was sampled every 30 years from a 10 year smoothing average of temperatures, it would thus be expected to retain the variability at around 30 year period. This is verified by the observation that the spectrum of (DP) and (DBP) departures from the spectrum of (D) and (DB) at about 30 year period after an agreement at high frequencies. Borehole appears useless because the spectrum only describes the variation whereas the information therein is smeared too much to recover any detail of temperature evolution other than a long term trend. However, as seen in Figure 5(a) the borehole does help to reduce the bias a bit. Overall, compared to the oracle proxies (T), the reconstructions using other proxies perform well at high frequency but less precise at low frequency, as part of the low frequency information is lost in pollen and borehole.

5.2 Other inferences

Cause of bias

It can be seen from Figure 5(a) that the reconstruction when $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)'$ is modeled in the process stage carries some systematic positive bias. This positive bias is pronounced in cases where forcings or information at longer time scales is not available, i.e., $\mathbf{T}$ is modeled as an AR(2) with constant mean which corresponds to the "C" points in the figure, and no pollen is used in the reconstruction. The reason for the positive bias is because the observed $\mathbf{T}_2$ which serves the primary source to estimate the mean of $\mathbf{T}$ has higher mean temperature than $\mathbf{T}_1$, but was nonetheless assumed to have the same mean function as $\mathbf{T}_1$ in the model. Therefore, after we remove this a priori assumption and only leave $\mathbf{T}_1$ in the process stage as described in the partial temperature process model in Section 4.3, the positive bias has been largely reduced. Another way to reduce the bias is to incorporate external forcings since this enables the temperature process to be estimated by its dependency on forcings. In this way the difference in mean functions is accounted for by the varying external forcings.

If bias is the primary concern, one should consider the reconstruction that is only based on the model assumption for $\mathbf{T}_1$, though one would have to accept that it will carry more variance. Note that the oracle reconstruction is not necessary to be the optimal reconstruction just in terms of bias. We can see that some reconstructions have even lower bias than their corresponding oracle experiment. Yet due to the bias-variance trade off, the oracle reconstruction is the best optimal in terms of rmse. Not surprisingly, this is particularly visible when realistic noise is applied into proxies.

Sensitivity to noise in proxies

Clearly the noise in proxies introduces both bias and more variability in the reconstructions. We can see from Figure 5 that with the substantial amount of noise added to the proxies, the performance of the reconstruction deteriorates some but not terribly. It is worth noting that our results are only based on one set of contaminated proxies. Given different noise, the performance is slightly different, but does not appreciably change the basic conclusion.

5.3 Posterior samples of the reconstructed temperature

We select model (DP) with forcings to show examples of the reconstruction at three scenarios. Figure 7a is for modeling $\mathbf{T}$ and no noise, Figure 7b for modeling $\mathbf{T}_1$ and no noise, and Figure 7c for modeling $\mathbf{T}$ and with noise. To make this comparison clearer we report posterior reconstructions for decadal average temperature. In all those figures, the 95% uncertainty band calculated from the posterior samples is also displayed together with the reconstruction.

In Figure 7, all the reconstructions follow the trend of the target very well, although they seem to miss some details. Comparison between panel a and b in this figure shows that by modeling only the unknown $\mathbf{T}_1$ in the process stage can effectively reduce the bias caused by assuming $\mathbf{T}_1$ and $\mathbf{T}_2$ to have the same mean function. The difference between panels a and b and panel c illustrates the larger bias and wider uncertainty band introduced by noise. In panels a and b, the uncertainty band covers the target temperatures fairly well, while in panel c the coverage deteriorates due to the bias.

5.4 Identifiability of parameters

We examine the posterior distribution of parameter estimates and compare them to their corresponding priors to make sure that priors only have little influence on the parameter estimates. We focus on the model DBP with forcings included and with proxy errors, because this case has the most complex setting and provides the greatest challenge in determining parameters. We found that in general, the posteriors are not sensitive to the priors and this is suggested by the relatively small variance of the posteriors compared to their priors. However, compared to the stable estimates for regression coefficients for tree-rings, the estimates for borehole data contain more uncertainties. Moreover, the BHM is unable to resolve the variance parameter in the borehole model. We conjecture that this is because the transformation matrix $\mathbf{M}_B$ is an ill-posed matrix that has a small effective rank, on the order of 6 degrees of freedom, hence there is no way to fully recover the temperature information from the borehole profile. This essentially has been illustrated in Figure 4. Despite the difficulty in estimating some of the parameters related to the borehole proxy data the resulting reconstructions are about the same as (DP) combinations or slightly better.

6 Discussion and Conclusions

This paper has proposed a new application of BHMs to reconstruct NH temperatures by jointly using proxy data with different temporal resolutions, and using forcings as external covariates of temperature evolution. With this method we investigated the benefits by combining different proxies and by inclusion of the forcings. Our results showed that a process model that includes the external forcings in the form of an energy balance can dramatically improve the reconstruction, particulary if the applied proxy data is deficit of decadal or centennial scale variability. In our numerical study this improvement can be by a factor of 2 in rmse. However, its role can be partially replaced by our hypothetical pollen proxy that fills this range. Tree rings play a significant role in retaining the high frequency variability, while pollen improved the reconstruction remarkably by capturing the variation at lower frequency band. These results make a case for attempting multi-proxy reconstructions with tree-rings and pollen assemblages and also including external forcing covariates. Although we base these conclusions on a synthetic Monte Carlo experiment, the climate model simulation used as truth is a complex and extensive representation of the actual climate system and so provides confidence that these results will extend well to real world conditions.

Pollen proxies are usually collected from sediment layers and thus possibly subject to dating errors, that is, the date during which pollen was formed or the age of a layer in lake sediment might not be exactly identified due to various reasons, such as the time lag between initial plant introduction to its abundance and different sediment accumulation rates (see Bradley 1999). We account for the dating error to some extent by considering additive proxy errors. However, one area of future work is to consider a more realistic model for dating errors (Haslett and Parnell, 2008). One surprise in this work is that the hypothetical borehole proxies do not improve the reconstruction in a significant way. Nevertheless, in reality pollen records might not perform as well as the synthetic ones, and in such a case borehole information may be used more effectively. Also the use of borehole information might be improved by greater attention to the ill-posed aspects of the data model.

One advantage of the BHM framework is that it readily extends to more complex data level or process models. Because of this flexibility we believe the methods will adapt to more complex statistical features of real data and the experience in this study will transfer to more complicated cases. An important extension will be reconstructing the space-time temperature process instead of the NH temperature, and further reconstructing the multivariate space-time climate process. Climate variables of interest include temperature, precipitation, geopotential heights. The Bayesian framework in this article naturally lends itself to univariate and multivariate space-time random field reconstruction. Achieving this long term goal will provide a valuable analysis to evaluate the next generation of climate system models and improve our understanding of past climate.

References

Ammann, C. M., Flecharda, C. R., Leifelda, J., Neftela, A. and Fuhrera, J. (2007) The carbon budget of newly established temperate grassland depends on management intensity. *Agriculture, Ecosystems & Environment*, 121, 5-20.

Ammann, C. M., Genton, M. G. and Li, B. (2009) Correcting for signal attenuation from noise: Sharpening the focus on past climate. In review.

Ammann, C. M. and Wahl, E. (2007) The Importance of the Geophysical Context in Statistical Evaluations of Climate Reconstruction Procedures. *Climatic Change*, 85, 71-88.

Banerjee S., Carlin, B., Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton.

Bard, E., Raisbeck, G., Yiou, F. and Jouzel, J. (2000) Solar irradiance during the last 1200 years based on cosmogenic nuclides. *Tellus Ser. B*, 52, 985-992.

Beltrami, H. (2001) Surface heat flux histories from inversion of geothermal data: Energy balance at the Earth's surface. *J. Geophys. Res.*, 106, 979-21993.

Beltrami, H., Chapman., D. S., Archambault, S. and Bergeron, Y. (1995) Reconstruction of high resolution ground temperature histories combining dendrochronological and geothermal data. *Earth Plan. Sci. Lett.*, 136, 437- 445.

Berliner, L. M., Milliff, R. F. and Wikle, C. K. (2003) Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research*, 108, doi:10.10292002JC001413.

Boville, B. A., Kiehl, J. T., Rasch, P. J. and Bryan, F. O. (2001) Improvements to the NCAR CSM-1 for transient climate simulations. *Journal of Climate*, 14, 164-179.

Bradley, R. S. (1999) *Paleoclimatology: Reconstructing climates of the Quaternary*, 2nd Ed., Academic Press.

Bradley, R. S. and Jones, P. D. (1993) "Little Ice Age" summer temperature variations:

their nature and relevance to recent global warming trends. *The Holocene*, 3, 367-376.

Briffa, K. R. and Melvin, T. M. (2008) A closer look at Regional Chronology Standardisation of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application. In Hughes, M. K., Diaz, H. F. and Swetnam, T. W., editors, *Dendroclimatology: progress and prospects. Developments in Paleoenvironmental Research*. Springer, in press.

Brown, K. J., Clark, J. S., Grimm, E. C., Donovan, J. J., Mueller, P. G., Hansen, B. C. S. and Stefanova, I. (2005) Fire cycles in North American interior grasslands and their relationship to prairie drought. *Proc. Nat. Acad. Sci.*, 102(25), 8865-8870.

Carslaw, H. S. and Jaeger, J. C. (1959) *Conduction of Heat in Solids*, 2nd Ed., Oxford, UK: Oxford University Press.

Crowley, T. J. (2000) Causes of climate change over the last 1000 years. *Science*, 289, 270-277.

Crowley, T. J. and Lowery, T. S. (2000) How warm was the Medieval warm period? *Ambio*, 29, 51-54.

Chapmann, D. S., Bartlett, M. G. and Harris, R. N. (2004) Comment on "Ground vs surface air temperature trends: Implications for borehole surface temperature reconstructions by M. E. Mann and G. Schmidt, *Geophys. Res. Lett.*, 31, L07205, doi:10.1029/2003GL019054.

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 paleoclimatic studies. *The Holocene*, 5, 229237.

Cook, E. R. and Kairiukstis, L. A. (1990) *Methods of dendrochronology*, International Institute for Applied Systems Analysis, Kluwer Academic Publishers. Yukon Territory, Canada.

Fritts, H. C. (1976) *Tree rings and climate*, Academic Press.

Gelman, A., Roberts, G. O. and Gilks, W. R. (1996) Efficient Metropolis jumping rules. *Baysian Stistics*, 5 eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith. Oxford: Oxford University Press.

Guiot J., Harrison, S. P. and Prentice, C. I. (1993) Reconstruction of Holocene precipitation patterns in Europe using pollen and lake level data. *Quat. Res.*, 40, 139-149.

Guiot, J., Nicault, A., Rathgeber, C., Edouard, J. L., Guibal, F., Pichard, G. and Till, C. (2005) Last-millennium summer-temperature variations in western Europe based on proxy data. *The Holocene*, 15(4), 489-500.

Harris, R. N. (2007) Variations in air and ground temperature and the POM-SAT model: results from the Northern Hemisphere. *Climate of the Past*, 3, 611-621.

Harris R. N. and Chapman, D. S. (2001) Mid-latitude (30-60N) climatic warming inferred by combining borehole temperatures with surface air temperatures. *Geophys. Res. Lett.*, 28, 747-750.

Haslett, J. and Parnell, A. (2008) A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society, Series C*, 57, 399418.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B. and Mitchell, F. J. G. (2006) Bayesian palaeoclimate reconstruction, *J. R. Statist. Soc. A*, 169, 395-438.

Huang, S. (2004) Merging Information from Different Resources for New Insights into Climate Change in the Past and Future. *Geophys. Res. Lett.*, 31, L13205.

Huang, S., Pollack, H. N. and Shen, P. Y. (2000) Temperature trends over the past five centuries reconstructed from borehole temperatures. *Nature*, 403, 756-758.

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change: Solomon, S., D. *et al.* (eds.), Cambridge University Press, Cambridge and

NY.

Jones, P. D., Briffa, K. R., Barnett, T. P. and Tett, S. F. B. (1998) High-resolution Palaeoclimatic Records for the last Millennium: Interpretation, Integration and Comparison with General Circulation Model Control-run Temperatures. *The Holocene*, 8, 455-471.

Lee, T. C. K., Zwiers, F. W. and Tsao, M. (2008) Evaluation of proxy-based millennial reconstruction methods. *Climate Dynamics*, 31(2-3): 263. DOI 10.1007/s00382-007-0351-9.

Li, B., Nychka, W. D. and Ammann, C. M. (2007) The "Hockey Stick" and the 1990s: A statistical perspective on reconstructing hemispheric temperatures, *Tellus*, 59, 591-598.

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

Mann, M. E. and Rutherford, S. (2002) Climate reconstruction using "pseudoproxies". *Geophys. Res. Lett.*, 29, 1501, doi:10.1029/2001GL014554.

Mann, M. E., Rutherford, S., Wahl, E. and Ammann, C. (2005) Testing the fidelity of methods used in proxy-based reconstruction of past climate. *Journal of Climate*, 18, 4097-4107.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K. and Rutherford, S. (2008) Proxy-Based Reconstructions of Hemispheric and Global Surface Temperature Variations over the Past Two Millennia. *Proc. Natl. Acad. Sci.*, 105, 13252-13257.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. and Karlen, W. (2005) Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, 433, 613-617.

Osborn T. J., and Briffa, K. R. (2006) The spatial extent of 20th-Century warmth in the context of the past 1200 years. *Science*, 311, 841-844.

Otto-Bliesner, B. L. and Brady, E. C. (2001) Tropical Pacific variability in the NCAR Climate System Model. *Journal of Climate*, 14, 3587-3607.

Rind, D. (1995) The potential for modeling the effects of different forcing factors on climate during the past 2000 years. In *Global Environmental Change*, NATO ASI Series Vol. 41. Jones, P. D. Ed. Springer-Verlag, 563-581.

Rutherford, S., Mann, M. E., Delworth, T. L. and Stouffer, R. (2003) Climate field reconstruction under stationary and nonstationary forcing. *Journal of Climate*, 16, 462-479.

Schweingruber, F. H. (1996) *Tree Rings and Environment Dendroecology.* Paul Haupt.

Shumway, R. H. and Stoffer, D. S. (2006) *Time Series Analysis and Its Applications: With R Examples*, Springer, New York.

Tingley, M. P. and Huybers, P. (2009) A Bayesian algorithm for reconstructing spatially arrayed temperatures. In Revision.

Viau, A. E., Gajewski, K., Sawada, M. C. and Fines, P. (2006) Millennial-scale temperature variations in North America during the Holocene. *J. Geophys. Res.*, 111, D09012, doi:10.1029/2005JD006031.

von Storch, H., Zorita, E., Jones, J., Dimitriev, Y., Gonzalez-Rouco, F. and Tett, S. (2004) Reconstructing past climate from noisy data. *Science*, 306, 679-682.

Wikle, C. K., Millif, R. F., Nychka, D. and Berliner, L. M. (2001) Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, 96, 382-397.

Williams J. W., Bartlein, P. J. and Webb III, T. (2000) Data-model comparisons for eastern North America Inferred biomes and climate values from pollen data. In: P. Braconnot, Ed.: *Proceedings of the Third Paleoclimatic modeling intercomparison project workshop*, World Climate Research Program, Montreal, Canada, 77-86.

Zielinski, G. A. (2000) Use of paleo-records in determining variability within the volcanism-climate system. *Quaternary Science Reviews*, 19, 417-438.

Zorita, E., Gonzalez-Rouco, F. and Legutke, S. (2003) Testing the Mann *et al.* (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model. *Journal of Climate*, 16, 13781390.

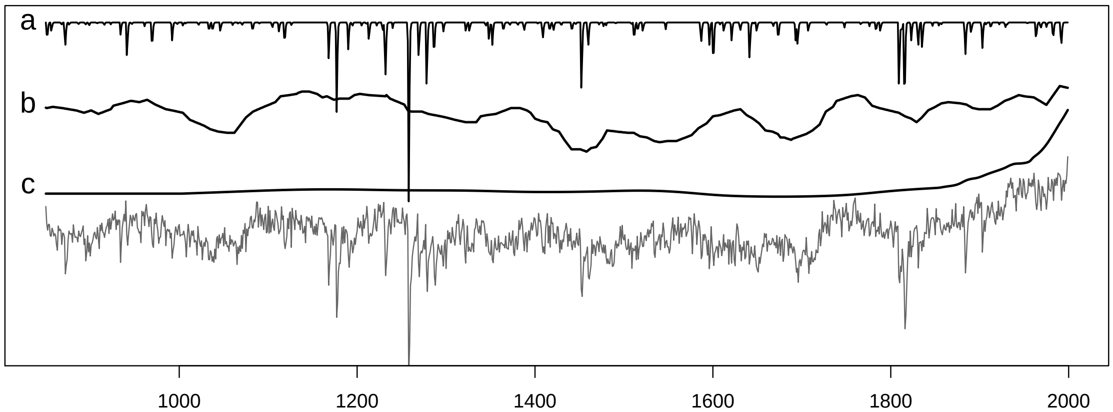


Figure 1: The diagram of three forcings. The grey curve is the NH temperature, and among the black curves **a** is the volcanism, **b** is the solar irradiance and **c** is the greenhouse gases. All the curves are scaled in order to show them clearly in one figure.

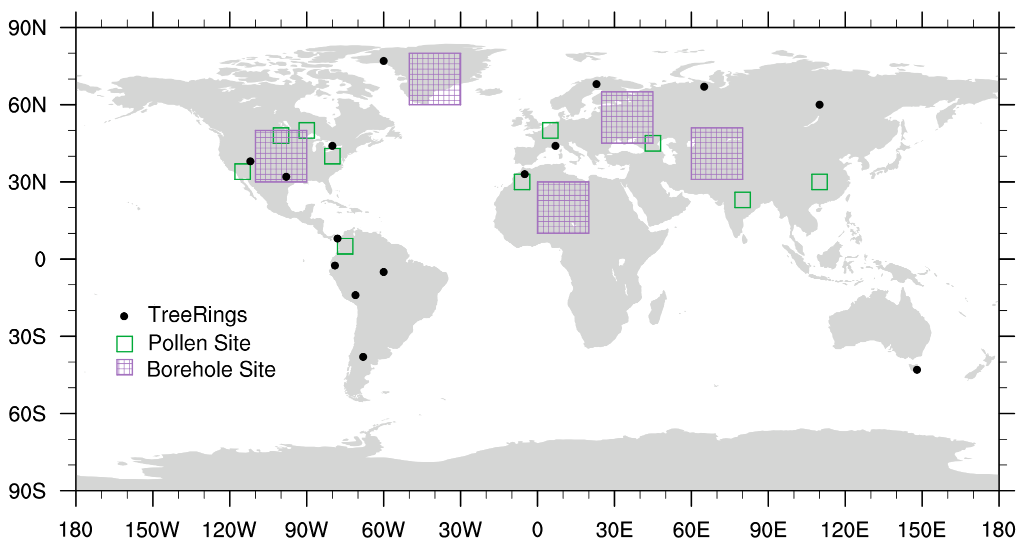


Figure 2: pseudo-proxy sampling locations in CSM

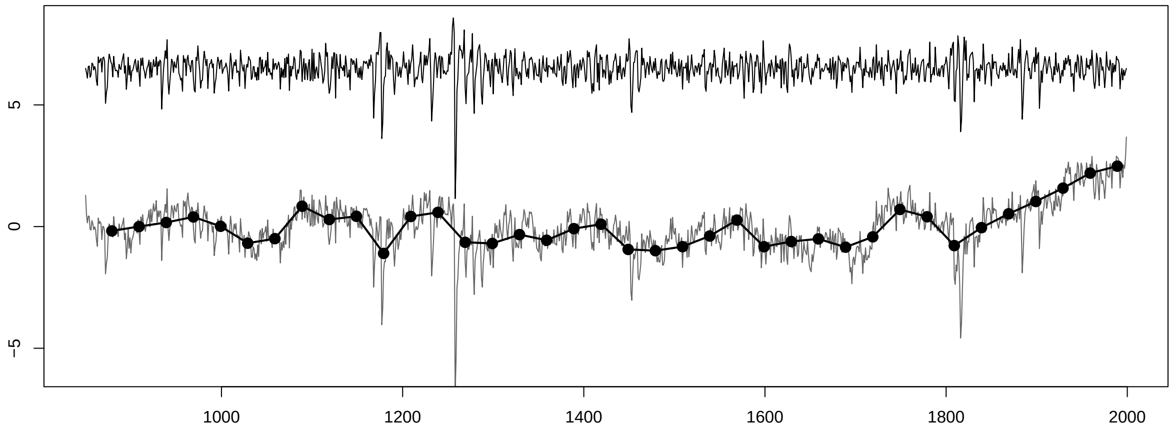


Figure 3: An example of pseudo tree-rings and pseudo pollen together with the temperature (grey curve). The tree-rings are represented by the upper black curve, and the pollen is shown by dots which are observed at every 30 years. The black curve with dots embedded is the 10 year smoothing average of the temperature.

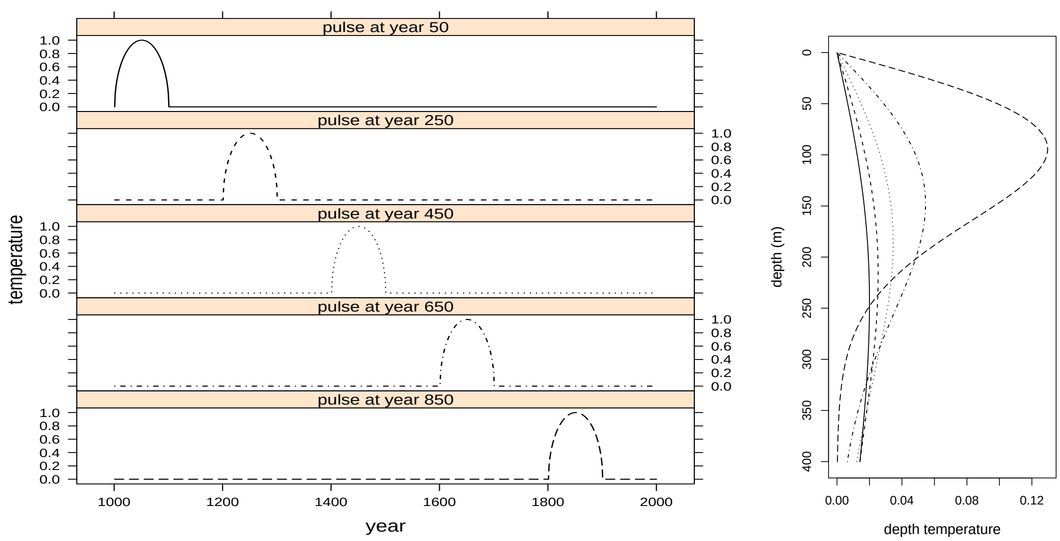
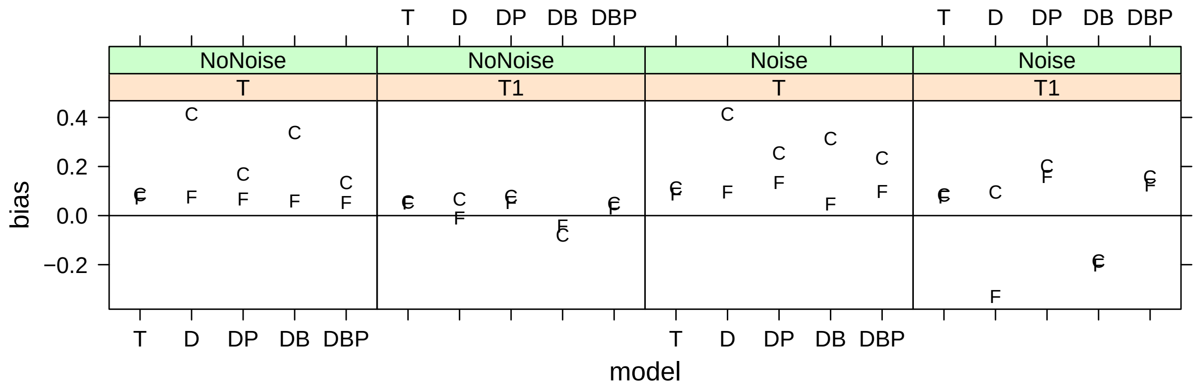
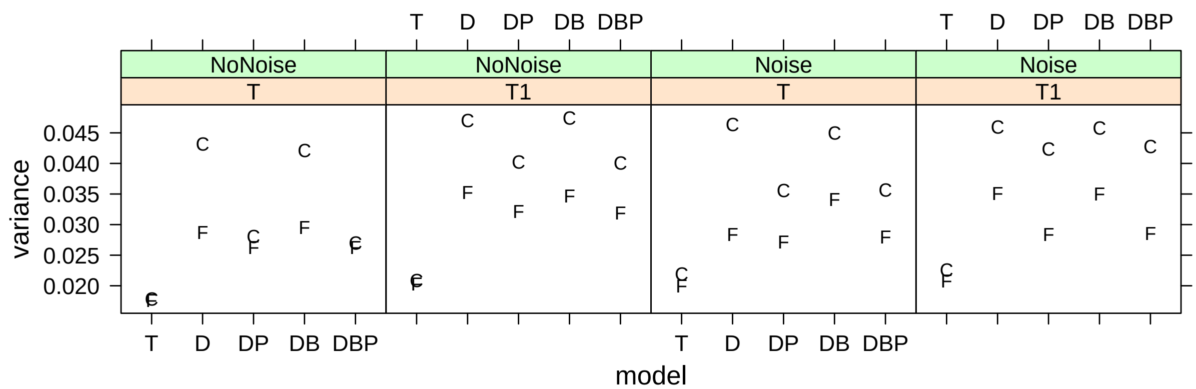


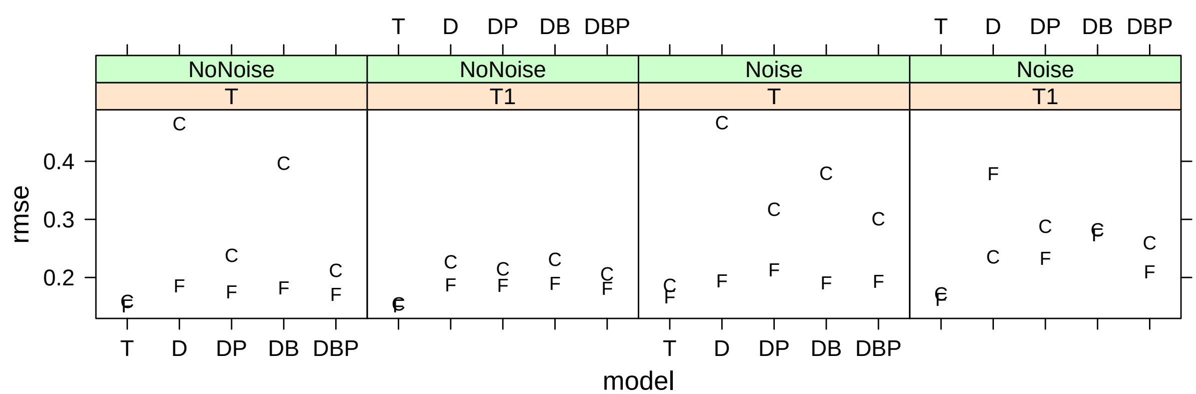
Figure 4: The borehole profile corresponding to the temperature series with a pulse at different time locations



(a) Bias



(b) Variance



(c) Root mean square error

Figure 5: Bias, variance and RMSE of the reconstructions for five data models and 2^3 scenarios that are combinations of with/without forcings, with/without noise and modeling $\mathbf{T}_1/\mathbf{T}$ in the process stage. "C" and "F" are the reconstructions without forcings (with constant mean function) and with forcings incorporated, respectively.

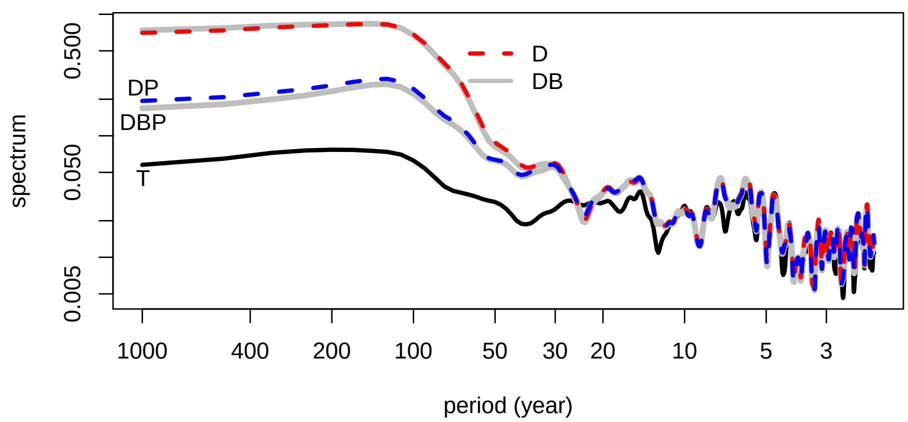


Figure 6: Using smoothed spectrum of reconstruction residuals from the five data models to illustrate the frequency band at which proxies capture the variation of the temperature process.

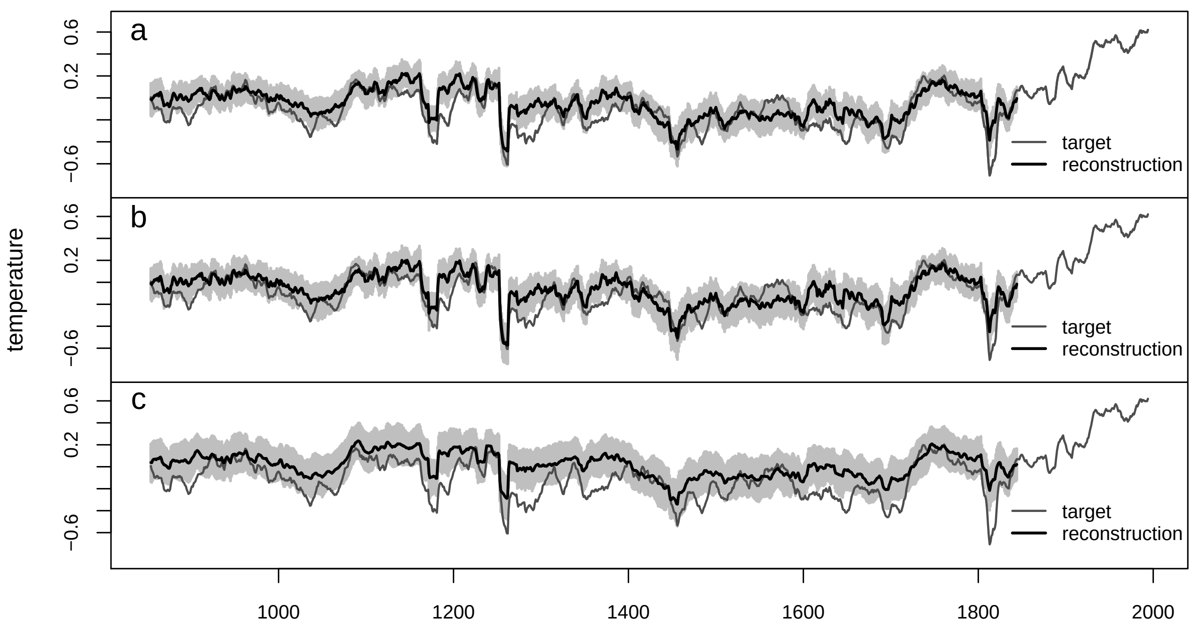


Figure 7: The reconstructions using tree-rings and pollen together with forcings in three scenarios. **a**: modeling $\mathbf{T}$ and without noise; **b**: modeling $\mathbf{T}_1$ and without noise; **c**: modeling $\mathbf{T}$ and with noise.