*The Annals of Applied Statistics*
2015, Vol. 9, No. 3, 1194–1225
DOI: 10.1214/15-AOAS832


# A BAYESIAN SPATIOTEMPORAL MODEL FOR RECONSTRUCTING CLIMATE FROM MULTIPLE POLLEN RECORDS

BY LASSE HOLMSTRÖM*, LIISA ILVONEN*, HEIKKI SEPPÄ[†,1] AND SIIM VESKI[‡]

*University of Oulu*, University of Helsinki*† *and Tallinn University of Technology*‡

Holocene (the last 12,000 years) temperature variation, including the transition out of the last Ice Age to a warmer climate, is reconstructed at multiple locations in southern Finland, Sweden and Estonia based on pollen fossil data from lake sediment cores. A novel Bayesian statistical approach is proposed that allows the reconstructed temperature histories to interact through shared environmental response parameters and spatial dependence. The prior distribution for past temperatures is partially based on numerical climate simulation. The features in the reconstructions are consistent with the quantitative climate reconstructions based on more commonly used reconstruction techniques. The results suggest that the novel spatio-temporal approach can provide quantitative reconstructions that are smoother, less uncertain and generally more realistic than the site-specific individual reconstructions.

**1. Introduction.** Instrumental temperature records rarely cover more than the past 100–200 years. On the other hand, temperature proxy data, such as fossil pollen, tree rings or ice cores, provide a continuous and long record of climatic changes where instrumental data do not exist [Jansen et al. (2007), Masson-Delmotte et al. (2013)]. The present article proposes Bayesian statistical methodology for pollen-based paleotemperature reconstruction at multiple locations that takes into account spatial and temporal dependencies between the sites and along the cores. The method is then applied to reconstruct Holocene, that is, post Ice Age mean annual temperature variation at four locations in southern Finland, Sweden and Estonia based on fossil pollen data extracted from lake sediment cores.

The standard approach to temperature reconstruction from multiple proxy records is the so-called Composite Plus Scaling (CPS) method that uses the modern instrumental record and a suitable regression technique to combine into an average representation the temperature histories originally constructed only on the basis of

Received May 2014; revised May 2015.

[1]Supported by the Academy of Finland project Ecological History and Long-Term Dynamics of the Boreal Forest Ecosystem (EBOR) and the Nordic top-level research initiative Cryosphere-Atmosphere Interactions in a Changing Arctic Climate (CRAICC).

*Key words and phrases.* Bayesian modeling, paleoclimate, regression, space–time modeling, temperature proxy.

the individual records [e.g., Jones et al. (2009) and the references therein; see also NRC (2006)]. We propose a novel method that effectively combines the data from all the original individual proxy records, in our case the pollen taxon abundances, and reconstructs their temperature histories in a joint estimation process that allows the histories to interact through shared environmental response parameters and spatial dependence. Our approach therefore represents a deeper integration of the information in the proxy records than the standard methodology.

The usefulness of pollen and other organisms as temperature proxies is based on the fact that different organisms tend to have different optimal temperatures, that is, temperatures in which they fare particularly well. Therefore, the relative abundances of different pollen types in a sediment core layer reflect the temperature at the time when the sediment layer was formed. Pollen data is widely used in quantitative climate reconstructions because pollen is abundant and widely dispersed and because the importance of climate for the distribution and abundance of plants is well studied and documented [Dahl (1998), Woodward (1987)]. For recent reviews on climate reconstruction methodology, see Birks et al. (2010), Jones et al. (2009), Juggins and Birks (2012) and, for pollen-based methods viewed from a Bayesian perspective, see Ohlwein and Wahl (2012).

The Bayesian BARCAST model discussed in Tingley and Huybers (2010a) and Tingley and Huybers (2010b) aims to reconstruct a spatially and temporally complete climate process from incomplete proxy and instrumental time series. The space–time covariance is assumed separable and exponential in space. The prior model describes the evolution of the true surface temperatures as a multivariate autoregressive process with spatially correlated innovations. The authors test their model by reconstructing North American surface temperatures using an instrumental surface temperature data set, after corrupting a number of time series to mimic proxy observations. The results are also compared with those obtained using the regularized expectation–maximization algorithm (RegEM) and it is concluded that a Bayesian algorithm produces more skillful reconstructions as measured by the coefficient of efficiency and the length of the uncertainty intervals.

Li, Nychka and Ammann (2010) use Bayesian hierarchical modeling to reconstruct past Northern Hemisphere mean temperatures. Their model combines information from proxies with different temporal resolution and forcings which act as external drivers of large-scale temperature evolution. However, no real proxy data are used and, instead, the proxy records are simulated on the basis of numerical climate model outputs. Further, the model does not include a spatial component. The results of the paper emphasize the importance of information that reflects climate on a variety of frequencies.

Brynjarsdóttir and Berliner (2011) reconstruct ground surface temperature histories with uncertainty estimates for the past 400 years from nine borehole temperature records using Bayesian hierarchical modeling. Temperature histories and heat flow parameters for boreholes in the same region share the same mean and variance. To find out whether the sharing of information across groups of data

has any influence, they also fit single-site models to each of the nine boreholes and conclude that combining the boreholes in two subregions allows the ground surface temperature history parameters to borrow strength across boreholes.

Tingley et al. (2012) present an overview of the challenges in inferring with uncertainties a climate process through space and time. The authors propose a unifying Bayesian modeling and notational framework for the paleoclimate reconstruction problem. As one advantage of hierarchical modeling they view the possibility of constructing and testing each component independently of the others before they are incorporated into the hierarchy.

The method proposed in this article is directly related to the work of Toivonen et al. (2001), Vasko, Toivonen and Korhola (2000) and Korhola et al. (2002), who were the first to use detailed Bayesian modeling for paleoclimate reconstruction from assemblage data. Toivonen et al. (2001) introduced a Bayesian response model called Bum based on a unimodal model for an organism's response to temperature. Vasko, Toivonen and Korhola (2000) then further developed the Bum model and introduced a Bayesian hierarchical multinomial regression model that takes into account dependency between species. This model was called Bummer and it was further analyzed and modified in Erästö and Holmström (2006) and Salonen et al. (2012).

The starting point of our approach is the Bummer model that we extend in several important ways. As opposed to Bummer, our model handles multiple proxy records and also takes into account their spatial correlations. The cores can have different chronologies and the reconstruction is performed on a common chronology obtained as their union. Finally, instead of the simple i.i.d. normal model used in Bummer, the temporal part of the temperature field prior is defined by a multivariate Gaussian smoothing prior with the smoothing parameter hyperprior elicited using numerical climate model simulation. The shortcomings of the simple i.i.d. model Bummer model were demonstrated in Erästö and Holmström (2006).

Haslett et al. (2006) also used hierarchical Bayesian modeling to reconstruct the prehistoric climate at Glendalough in Ireland from fossil pollen data. A single core is used for reconstruction and, as in Erästö and Holmström (2006), a temporally smoothing temperature prior is used to reflect the fact that climate change can be assumed to exhibit a degree of smoothness. The use of European-wide pollen-vegetation-climate relationships led at some time intervals to multimodality in the posterior distribution of the reconstructed environmental variables. To avoid the multimodality typical to continental training sets with multiple strong climatic gradients, we limited the training set to Scandinavian and Baltic State environments where a simple south–north temperature gradient is dominant.

Finally, Ohlwein and Wahl (2012) interpret the model of Haslett et al. (2006) as a Bayesian version of the so-called Modern Analog Technique (MAT) and their general framework has similarities with tour approach, too. The authors also discuss the special challenges in pollen-based environmental reconstructions as well

as the appropriateness of the unimodal response model. Some of the model elements used in Paciorek and McLachlan (2009) are also similar to our proposal.

The rest of the paper is structured as follows. The model and its various components are described in Section 2. The data used in the example reconstructions and the results obtained are presented in Section 3, and Section 4 summarizes our main conclusions. An online supplement [Holmström et al. (2015a)] includes an analysis of the Gaussian taxon response model used, reference reconstructions from Greenland ice cores and Scandinavian records, additional reconstructions based on our pollen data, the core chronologies, and charts of the sediment core pollen abundances for the most important taxa used in temperature reconstructions. All data used in this work are available in the online supplement Holmström et al. (2015b) and the Matlab code used in reconstructions is in the online supplement Holmström et al. (2015c).

## 2. The model.

2.1. *The Bayesian method.* Bayesian inference is based on Bayes' theorem, which in its simplest form can be written as

$$p(\boldsymbol{\Theta}|\text{data}) = \frac{p(\boldsymbol{\Theta})p(\text{data}|\boldsymbol{\Theta})}{p(\text{data})} \propto p(\boldsymbol{\Theta})p(\text{data}|\boldsymbol{\Theta}). \tag{1}$$

Here "data" consists of the available observations and in our case includes training lake and sediment core pollen abundances as well as modern temperatures at the training lakes. The model parameters as well as the past unknown temperatures are included in $\boldsymbol{\Theta}$. The density $p(\text{data}|\boldsymbol{\Theta})$ is the likelihood of the data, the prior distribution $p(\boldsymbol{\Theta})$ describes our prior beliefs about the model parameters, and $p(\boldsymbol{\Theta}|\text{data})$ is the posterior distribution of $\boldsymbol{\Theta}$. Using the posterior distribution, the investigator can in principle answer any question about the probabilities of the unknown quantities of interest. Additional levels of hierarchy can be added to the model by assuming that the prior of $\boldsymbol{\Theta}$ depends on another parameter $\boldsymbol{\psi}$ which in turn has its own prior $p(\boldsymbol{\psi})$, etc. For more information on Bayesian modeling, see, for example, Banerjee, Carlin and Gelfand (2004) and Gelman et al. (2004).

2.2. *Notation.* In the following, the symbols for "modern" (training) and sediment fossil quantities have the superscript $m$ and $f$, respectively. We assume $n$ training lakes with known modern temperatures and $C$ cores with $l$ pollen taxa counted from the training lakes and $l_c$ taxa counted from core $c = 1, \ldots, C$. All core taxa are present also in the training lakes. For core $c$, the number of depths sampled is $n_c$, indexed according to increasing sediment age. The term "site" refers either to a training lake or to a depth in a core. Therefore, there are $n + n_1 + \cdots + n_C$ sites altogether.

*Training lakes.*

| | |
|---|---|
| $\mathbf{x}^m = [x_1^m, \ldots, x_n^m]^T$ | modern training temperatures (30-year annual means); |
| $\mathbf{y}_i^m = [y_{i1}^m, \ldots, y_{il}^m]^T$ | scaled modern pollen taxon abundances ($y_{i\cdot}^m \equiv \sum_{j=1}^{l} y_{ij}^m = 100$) at the training lake $i$, $i = 1, \ldots, n$; |
| $\mathbf{Y}^m = [\mathbf{y}_1^m, \ldots, \mathbf{y}_n^m]$ | $l \times n$ matrix of modern taxon abundances. |

*Cores.*

| | |
|---|---|
| $\mathbf{x}_c^f = [x_{c1}^f, \ldots, x_{cn_c}^f]^T$ | unknown past temperatures for core $c$, $c = 1, \ldots, C$; |
| $\mathbf{X}^f = \{\mathbf{x}_1^f, \ldots, \mathbf{x}_C^f\}$ | set of all past temperatures; |
| $\mathbf{y}_{ci}^f = [y_{ci1}^f, \ldots, y_{cil_c}^f]^T$ | scaled pollen taxon abundances ($y_{ci\cdot}^f \equiv \sum_{j=1}^{l_c} y_{cij}^f = 100$) for core $c$ at site (depth) $i$, $i = 1, \ldots, n_c$, $c = 1, \ldots, C$; |
| $\mathbf{Y}_c^f = [\mathbf{y}_{c1}^f, \ldots, \mathbf{y}_{cn_c}^f]$ | $l_c \times n_c$ matrix of taxon abundances for core $c$, $c = 1, \ldots, C$; |
| $\mathbf{Y}^f = \{\mathbf{Y}_1^f, \ldots, \mathbf{Y}_C^f\}$ | set of all core taxon abundances. |

*Reconstruction times*: *Chronologies.* The past temperature $x_{ci}^f$ for core $c$ at depth $i$ corresponds to a time $t_{ci}$ determined using radiocarbon or other dating technique. The sequence $t_{c1} > \cdots > t_{cn_c}$ is referred to as the chronology of core $c$. We will reconstruct the past temperature on a time grid defined by the union of all such chronologies,

$$\mathbf{t} = \{t_1, \ldots, t_N\} = \bigcup_{c=1}^{C} \bigcup_{i=1}^{n_c} \{t_{ci}\},$$

where $t_1 > \cdots > t_N$. Note that one may have $N < n_1 + \cdots + n_c$, because different core chronologies may include identical dates. The same grid is used for each core, which means that for a given core, pollen abundance data will not be available for all time points. However, intra- and inter-core temperature correlations will help estimate the corresponding past temperatures in a reasonable manner. We use the notation $\tilde{\mathbf{x}}_c^f = [\tilde{x}_{c1}^f, \ldots, \tilde{x}_{cN}^f]^T$ for the past temperatures at core $c$ on this union chronology and $\tilde{\mathbf{X}}^f = [(\tilde{\mathbf{x}}_1^f)^T, \ldots, (\tilde{\mathbf{x}}_C^f)^T]^T$ for the $NC$ dimensional vector that contains the past temperatures on the union chronology grid for all cores. Thus, $\mathbf{x}_c^f$ is a subset of $\tilde{\mathbf{x}}_c^f$ and $\mathbf{X}^f$ is a subset of $\tilde{\mathbf{X}}^f$.

2.3. *A Bayesian multinomial Gaussian response model for multiple cores.* Our starting point is the Bummer model introduced in Vasko, Toivonen and Korhola (2000). We will first generalize it to multiple cores and then propose a further extension that takes into account spatial and temporal correlations among the cores (Sections 2.4 and 2.5).

Our aim is to find the posterior density $p(\widetilde{\mathbf{X}}^f|\mathbf{Y}^f, \mathbf{x}^m, \mathbf{Y}^m)$ of past temperatures $\widetilde{\mathbf{X}}^f$ given data $\mathbf{Y}^f, \mathbf{x}^m$ and $\mathbf{Y}^m$. If $\boldsymbol{\theta}$ contains the parameters of the model, taking $\boldsymbol{\Theta} = \{\mathbf{X}^f, \boldsymbol{\theta}\}$ and conditioning the probabilities on $\mathbf{x}^m$, we get from (1) that

$$
\begin{aligned}
p(\widetilde{\mathbf{X}}^f|\mathbf{Y}^f, \mathbf{x}^m, \mathbf{Y}^m) &= \int p(\widetilde{\mathbf{X}}^f, \boldsymbol{\theta}|\mathbf{Y}^f, \mathbf{x}^m, \mathbf{Y}^m)\, d\boldsymbol{\theta} \\
&\propto \int p(\widetilde{\mathbf{X}}^f, \boldsymbol{\theta}|\mathbf{x}^m) p(\mathbf{Y}^f, \mathbf{Y}^m|\mathbf{x}^m, \mathbf{X}^f, \boldsymbol{\theta})\, d\boldsymbol{\theta}.
\end{aligned} \tag{2}
$$

In practice, posterior inference on past temperatures is performed by generating a sample from $p(\widetilde{\mathbf{X}}^f, \boldsymbol{\theta}|\mathbf{Y}^f, \mathbf{x}^m, \mathbf{Y}^m)$ and keeping the part corresponding to $\widetilde{\mathbf{X}}^f$.

Sites are assumed to be conditionally independent given the temperatures and model parameters and, therefore, the likelihood term can be expanded as

$$
\begin{aligned}
&p(\mathbf{Y}^f, \mathbf{Y}^m|\mathbf{x}^m, \mathbf{X}^f, \boldsymbol{\theta}) \\
&\quad = p(\mathbf{Y}^f|\mathbf{X}^f, \boldsymbol{\theta}) p(\mathbf{Y}^m|\mathbf{x}^m, \boldsymbol{\theta}) = \prod_{c=1}^{C} p(\mathbf{Y}_c^f|\mathbf{x}_c^f, \boldsymbol{\theta}) p(\mathbf{Y}^m|\mathbf{x}^m, \boldsymbol{\theta}) \\
&\quad = \prod_{c=1}^{C} \prod_{i=1}^{n_c} p(\mathbf{y}_{ci}^f|x_{ci}^f, \boldsymbol{\theta}) \prod_{i=1}^{n} p(\mathbf{y}_i^m|x_i^m, \boldsymbol{\theta}),
\end{aligned} \tag{3}
$$

where in the second equality the conditional independence of the cores was assumed. This is one of the assumptions made in the original Bummer model and may well be an oversimplification. We decided to adopt it in order to limit the complexity of the model.

Each site is assumed to have its own set of taxon occurrence probabilities that reflects the probability of observing the various taxa at that site. Let $\mathbf{p}_1^m, \ldots, \mathbf{p}_n^m \in \mathbb{R}^l$ be the taxon probabilities at the modern sites and let $\mathbf{p}_{c1}^f, \ldots, \mathbf{p}_{cn_c}^f \in \mathbb{R}^{l_c}$ be the corresponding probabilities for core $c$. Denote

$$
\mathbf{P}^m = [\mathbf{p}_1^m, \ldots, \mathbf{p}_n^m], \qquad \mathbf{P}^f = \bigcup_{c=1}^{C} \{\mathbf{p}_{c1}^f, \ldots, \mathbf{p}_{cn_c}^f\}.
$$

Following Vasko, Toivonen and Korhola (2000), we use a Gaussian function to model how pollen abundance responds to temperature. The unimodal shape of the response is intended to reflect the fact that each pollen taxon host plant is assumed to have an optimum temperature at which it fares particularly well and that the favorability of the temperature declines symmetrically around this optimum [cf. Korhola et al. (2002)]. For taxon $j$ at modern site $i$ the response is characterized by

$$
\lambda_{ij}^m = \alpha_j \exp\left[-\left(\frac{\beta_j - x_i^m}{\gamma_j}\right)^2\right], \qquad i = 1, \ldots, n,\ j = 1, \ldots, l, \tag{4}
$$

where $\alpha_j$ is a scaling factor, $\beta_j$ models the optimum temperature for taxon $j$, and $\gamma_j$ is a tolerance parameter. These parameters are assumed to be the same for both training data and the cores. Therefore, the response for core $c$ is described by

$$\lambda_{cij}^f = \alpha_{k(c,j)} \exp\left[-\left(\frac{\beta_{k(c,j)} - x_{ci}^f}{\gamma_{k(c,j)}}\right)^2\right], \tag{5}$$

where $i = 1, \ldots, n_c$, $j = 1, \ldots, l_c$ and the indices $k(c,1), \ldots, k(c,l_c)$ correspond to the taxa counted from core $c$. Let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_l]^T$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_l]^T$, $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_l]^T$ and define $\boldsymbol{\alpha}_c^f = [\alpha_{k(c,1)}, \ldots, \alpha_{k(c,l_c)}]^T$, and similarly for $\boldsymbol{\beta}_c^f$ and $\boldsymbol{\gamma}_c^f$. Thus, $\boldsymbol{\alpha}_c^f$, $\boldsymbol{\beta}_c^f$ and $\boldsymbol{\gamma}_c^f$ are the subvectors of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ that correspond to those taxa that appear in core $c$. All modern Gaussian response model parameters are now denoted by $\boldsymbol{\vartheta}^m = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}]$ and the corresponding parameters for core $c$ by $\boldsymbol{\vartheta}_c^f = [\boldsymbol{\alpha}_c^f, \boldsymbol{\beta}_c^f, \boldsymbol{\gamma}_c^f]$. The parameter vector $\boldsymbol{\theta}$ above is then defined as $\boldsymbol{\theta} = \{\mathbf{P}^m, \mathbf{P}^f, \boldsymbol{\vartheta}^m\}$.

Other environmental factors besides the temperature can affect pollen taxon abundances and this is modeled by treating the taxon probabilities as random variables that follow a Dirichlet distribution,

$$\begin{aligned} &\mathbf{p}_i^m | x_i^m, \boldsymbol{\vartheta}^m \sim \text{Dirichlet}(\boldsymbol{\lambda}_i^m), \qquad i = 1, \ldots, n, \\ &\mathbf{p}_{ci}^f | x_{ci}^f, \boldsymbol{\vartheta}_c^f \sim \text{Dirichlet}(\boldsymbol{\lambda}_{ci}^f), \qquad i = 1, \ldots, n_c, c = 1, \ldots, C, \end{aligned} \tag{6}$$

where $\boldsymbol{\lambda}_i^m = [\lambda_{i1}^m, \lambda_{i2}^m, \ldots, \lambda_{il}^m]$ and $\boldsymbol{\lambda}_{ci}^f = [\lambda_{ci1}^f, \lambda_{ci2}^f, \ldots, \lambda_{cil_c}^f]$. Considering the full conditional distributions of the probability vectors $\mathbf{p}_i^m$ and $\mathbf{p}_{ci}^f$, the components of $\boldsymbol{\lambda}_i^m$ and $\boldsymbol{\lambda}_{ci}^f$ can be interpreted as "pseudo counts" that are added to the actual observed taxon relative abundances (cf. Appendix B). The observed scaled taxon abundances are assumed to follow multinomial distributions with the probabilities $\mathbf{p}_i^m$, $\mathbf{p}_{ci}^f$,

$$\begin{aligned} &\mathbf{y}_i^m | x_i^m, \boldsymbol{\theta} \sim \text{Mult}(y_{i\cdot}^m, \mathbf{p}_i^m), \qquad i = 1, \ldots, n, \\ &\mathbf{y}_{ci}^f | x_{ci}^f, \boldsymbol{\theta} \sim \text{Mult}(y_{ci\cdot}^f, \mathbf{p}_{ci}^f), \qquad i = 1, \ldots, n_c, c = 1, \ldots, C. \end{aligned} \tag{7}$$

We note that because of the Dirichlet distribution used, the average taxon probabilities (6) are determined by the relative size of the responses $\lambda_{ij}^m$ (or $\lambda_{cij}^f$). As a result, the temperature dependent taxon probabilities in the model can assume much more general shapes than just a simple Gaussian. This also means that the interpretation of the parameters $\alpha_j$, $\beta_j$ and $\gamma_j$ is not straightforward. This is discussed in more detail in Holmström et al. (2015a).

The prior term in (2) can be factored as

$$p(\widetilde{\mathbf{X}}^f, \boldsymbol{\theta} | \mathbf{x}^m) = p(\mathbf{P}^m | \mathbf{x}^m, \boldsymbol{\vartheta}^m) p(\mathbf{P}^f | \mathbf{X}^f, \boldsymbol{\vartheta}^m) p(\widetilde{\mathbf{X}}^f, \boldsymbol{\vartheta}^m | \mathbf{x}^m), \tag{8}$$

and further, by (6),

$$p(\mathbf{P}^m|\mathbf{x}^m, \boldsymbol{\vartheta}^m) = \prod_{i=1}^{n} p(\mathbf{p}_i^m|x_i^m, \boldsymbol{\vartheta}^m) = \prod_{i=1}^{n} \text{Dirichlet}(\mathbf{p}_i^m|\boldsymbol{\lambda}_i^m), \tag{9}$$

$$p(\mathbf{P}^f|\mathbf{X}^f, \boldsymbol{\vartheta}^m) = \prod_{c=1}^{C}\prod_{i=1}^{n_c} p(\mathbf{p}_{ci}^f|x_{ci}^f, \boldsymbol{\vartheta}_c^f) = \prod_{c=1}^{C}\prod_{i=1}^{n_c} \text{Dirichlet}(\mathbf{p}_{ci}^f|\boldsymbol{\lambda}_{ci}^f). \tag{10}$$

Here conditional independence of the probabilities, given the temperatures, was assumed. Assuming that the taxon-specific parameters are mutually independent *a priori*, the third factor on the right-hand side of (8) can be written as

$$p(\widetilde{\mathbf{X}}^f, \boldsymbol{\vartheta}^m|\mathbf{x}^m) = p(\widetilde{\mathbf{X}}^f)\prod_{j=1}^{l} p(\alpha_j)\prod_{j=1}^{l} p(\beta_j)\prod_{j=1}^{l} p(\gamma_j). \tag{11}$$

The above model reduces to the original Bummer if only a single core is considered ($C = 1$) and the priors in (11) are specified appropriately. In particular, Bummer uses an i.i.d. Gaussian prior for $\widetilde{\mathbf{X}}^f$ that does not model temporal correlation between past temperatures. In the next section, a spatio-temporal prior for $\widetilde{\mathbf{X}}^f$ is described.

2.4. *A spatio-temporal model for past temperatures.* We now define the prior distributions on the right-hand side of (11). The priors of the scaling factor $\alpha_j$ and the tolerance parameter $\gamma_j$ are specified analogously to Vasko, Toivonen and Korhola (2000) and Korhola et al. (2002),

$$\alpha_j \sim \text{Unif}(0.1, 60), \qquad \gamma_j \sim \text{Gamma}(9, 1/3), \qquad j = 1, \ldots, l.$$

For the prior of the optimum taxon temperature $\beta_j$ of taxon $j$, our approach is different from the original Bummer specification which used a Gaussian prior centered on the modern temperature of the single core lake used in temperature reconstruction. With several cores involved and all of them located at one end of a large training set temperature gradient (cf. Section 3.1), it makes more sense to work in the spirit of empirical Bayes analysis and define reasonable priors with the help of information gleaned from the training data. Thus, following the weighted-averaging partial least squares (WA-PLS) modeling idea of ter Braak and Juggins (1993), we first estimate the optimal temperature by $\hat{\beta}_j = (\sum_{i=1}^n y_{ij}^m)^{-1}\sum_{i=1}^n y_{ij}^m x_i^m$, where $x_i^m$ and $y_{ij}^m$ are the modern temperature and the abundance of taxon $j$ for training lake $i$, respectively. The $\hat{\beta}_j$'s thus estimated vary between $-2.7$°C and 6.2°C. A vague prior for $\beta_j$ is then defined as

$$\beta_j \sim \text{N}(\hat{\beta}_j, (1.5\sqrt{3})^2), \qquad j = 1, \ldots, l.$$

For more discussion on the choice of this particular prior, see Salonen et al. (2012).

It remains to describe the prior distribution of the vector $\widetilde{\mathbf{X}}^f$ that consists of the unknown past temperatures $\widetilde{\mathbf{x}}_c^f = [\widetilde{x}_{c1}^f, \ldots, \widetilde{x}_{cN}^f]^T$ for all cores, defined on the union chronology time grid $t_1 > \cdots > t_N$. The prior is a multivariate Gaussian with a separable covariance matrix obtained as the Kronecker product of spatial and temporal covariances,

$$\boldsymbol{\Sigma} = \mathbf{C}_S \otimes \mathbf{C}_T \in \mathbb{R}^{CN \times CN}. \tag{12}$$

To estimate the $C \times C$ spatial covariance matrix $\mathbf{C}_S$, two different approaches were tried. In the first, non-Bayesian approach, we applied Estimated Generalized Least Squares to fit a continuously indexed isotropic covariance function $C_S(\mathbf{s}, \mathbf{s}')$ to the training temperature residuals obtained after subtracting a linear trend and then defined $\mathbf{C}_S = [C_S(\mathbf{s}_c, \mathbf{s}_{c'})]$, where $\mathbf{s}_c, \mathbf{s}_{c'}$ are the core locations [e.g., Cressie (1993)]. In the Bayesian approach, an isotropic spatial covariance of temperatures was included in the hierarchical model as an additional parameter with its own prior distribution (Section 2.5). The two methods lead to rather similar reconstructions and in the following we will report only results obtained with the latter approach.

While the training data can be expected to inform us of the correlations between temperatures at different locations, we do not have any such direct knowledge of the *past* temperatures that could be used to specify the temporal covariance $\mathbf{C}_T$. We therefore believe that one should not use a temporal prior that makes too restrictive assumptions about the actual past temperature values and instead use a prior that basically only describes their internal variability or "roughness" [cf. Erästö and Holmström (2006)].

The matrix $\boldsymbol{\Sigma}$ in (12) is a block matrix with blocks $C_S(\mathbf{s}_c, \mathbf{s}_{c'})\mathbf{C}_T$ so that the temporal covariance for each core is $C_S(\mathbf{s}_c, \mathbf{s}_c)\mathbf{C}_T$ and, because of stationarity, $C_S(\mathbf{s}_c, \mathbf{s}_c)$ actually does not depend on $c$. To define $\mathbf{C}_T$, we assume that it defines the dependence structure of a process

$$\widetilde{x}_{c(i+1)}^f = \widetilde{x}_{ci}^f + \frac{1}{\sqrt{\kappa}}(t_{i+1} - t_i)\varepsilon_i, \tag{13}$$

where the $\varepsilon_i$'s are independent standard Gaussian variables and $\kappa > 0$. Thus, if $\widetilde{\mathbf{x}}_c^f = [\widetilde{x}_{c1}^f, (\widetilde{\mathbf{x}}_{c*}^f)^T]^T$, so that $\widetilde{\mathbf{x}}_{c*}^f = [\widetilde{x}_{c2}^f, \ldots, \widetilde{x}_{cN}^f]^T$, we have for a fixed $\widetilde{x}_1^f$ that

$$p(\widetilde{\mathbf{x}}_{c*}^f | \widetilde{x}_{c1}^f, \kappa) \propto \kappa^{(N-1)/2} \exp\left[ -\frac{\kappa}{2} \sum_{i=2}^{N} \left( \frac{\widetilde{x}_{ci}^f - \widetilde{x}_{c(i-1)}^f}{t_i - t_{i-1}} \right)^2 \right]. \tag{14}$$

Assuming that $\widetilde{x}_{c1}^f \sim \mathrm{N}(\mu_c, 1)$, where $\mu_c$ is the modern temperature at core lake $c$, we have after some matrix algebra (cf. Appendix A) that

$$\begin{aligned} p(\widetilde{\mathbf{x}}_c^f | \kappa) &= p(\widetilde{\mathbf{x}}_{c*}^f | \widetilde{x}_{c1}^f, \kappa) p(\widetilde{x}_{c1}^f) \\ &\propto \kappa^{(N-1)/2} \exp\left[ -\tfrac{1}{2} (\widetilde{\mathbf{x}}_c^f - \boldsymbol{\mu}_c)^T \boldsymbol{\Gamma}^{-1} (\widetilde{\mathbf{x}}_c^f - \boldsymbol{\mu}_c) \right], \end{aligned} \tag{15}$$

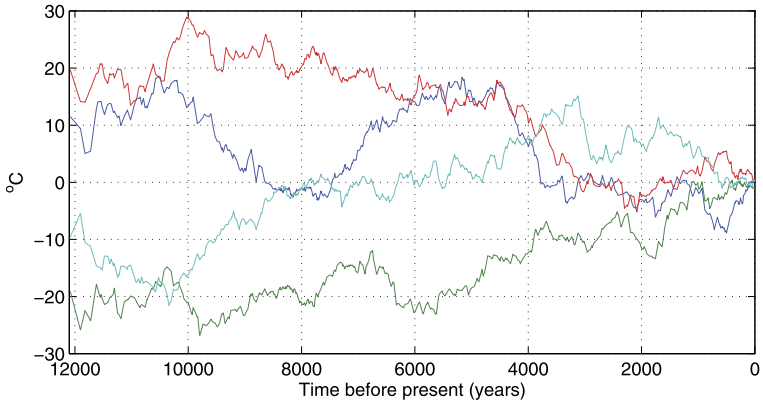


FIG. 1. *Four realizations of past temperatures generated from their prior distribution when the modern temperature is set at* $\mu_c = 0$.

where

$$\boldsymbol{\mu}_c = [\mu_c, \ldots, \mu_c]^T, \qquad \boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{B}_{21}^T \mathbf{B}_{22}^{-1} \mathbf{B}_{21} \kappa + 1 & \kappa \mathbf{B}_{21}^T \\ \kappa \mathbf{B}_{21} & \kappa \mathbf{B}_{22} \end{pmatrix}^{-1}$$

and where the matrices $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ are defined in Appendix A. Thus, for core $c$, $\widetilde{\mathbf{x}}_c^f | \kappa \sim \mathrm{N}(\boldsymbol{\mu}_c, \boldsymbol{\Gamma})$ and $\mathbf{C}_T = \mathbf{C}_T(\kappa) = \boldsymbol{\Gamma}$ in (12).

We see from (13) or (14) that $\kappa$ is a smoothing parameter that controls the roughness of past temperature variation. A prior distribution will be specified for $\kappa$ in Section 2.6. By definition,

$$\widetilde{\mathbf{X}}^f | \boldsymbol{\Sigma} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{16}$$

with

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_C]^T$$

and $\boldsymbol{\mu}_c = [\mu_c, \ldots, \mu_c]^T \in \mathbb{R}^N$. To get an idea of the nature of this prior, Figure 1 shows sample paths from the marginal distribution of $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ that correspond to a single core when $\kappa = 306$, a point estimate suggested for spatio-temporal reconstruction by the method of Section 2.6. As one can see, going back in time, the variance for past temperatures grows rapidly, making the prior very vague. We also note that the prior favors rather slowly varying temperature time series, limiting physically unreasonable fluctuations. This is considered reasonable because the Holocene climate has been relatively stationary.

2.5. *The spatial covariance model.* Let $x^m(\mathbf{s})$, $\mathbf{s} \in D$, denote the field of modern temperatures in the region $D$ where the training lakes are located. We assume that

$$x^m(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}),$$

where $\mu(\mathbf{s})$ is a trend and $\delta(\mathbf{s})$ is zero-mean and isotropic. The covariance function of $x^m(\mathbf{s})$ is then

$$C_S(\mathbf{s}, \mathbf{s}') = \mathrm{Cov}(x^m(\mathbf{s}), x^m(\mathbf{s}')) = \mathrm{Cov}(\delta(\mathbf{s}), \delta(\mathbf{s}')).$$

In general, the trend can be modeled as $\mu(\mathbf{s}) = \boldsymbol{\xi}(\mathbf{s})\boldsymbol{\omega}$, where $\boldsymbol{\xi}(\mathbf{s}) = [\xi_1(\mathbf{s}), \ldots, \xi_q(\mathbf{s})]$ are fixed covariate functions and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_q]^T$ are unknown parameters. We found that a linear trend is a plausible assumption and therefore took $\boldsymbol{\xi}(\mathbf{s}) = [1, \xi_2(\mathbf{s}), \xi_3(\mathbf{s})]$, $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$, where $\xi_2(\mathbf{s})$ and $\xi_3(\mathbf{s})$ are the latitude and longitude of the location $\mathbf{s}$, respectively. Further, of the various parametric models considered, an exponential covariance appeared to reflect spatial dependence in the data best and we therefore assume that

$$C_S(\mathbf{s}, \mathbf{s}') = C(r, \boldsymbol{\nu}) = \begin{cases} \nu_1 \exp(-r/\nu_2), & \text{if } r > 0, \\ \nu_3 + \nu_1, & \text{otherwise,} \end{cases} \tag{17}$$

where $r$ is the great circle distance between $\mathbf{s}$ and $\mathbf{s}'$ (in kilometers), and $\boldsymbol{\nu} = [\nu_1, \nu_2, \nu_3]$ with $\nu_1, \nu_2, \nu_3 \geq 0$.

By the construction described in Section 2.4, the first diagonal element of the temporal covariance matrix $\mathbf{C}_T$ is equal to 1. It follows that, in the prior model (16) based on the separable covariance (12), the marginal covariance of the modern temperatures $[\tilde{x}^f_{11}, \ldots, \tilde{x}^f_{C1}]^T$ at the core lakes is equal to $\mathbf{C}_S$. We therefore take $\mathbf{C}_S = \mathbf{C}_S(\boldsymbol{\nu}) = [C_S(\mathbf{s}_c, \mathbf{s}_{c'})]$, where $\mathbf{s}_c$ and $\mathbf{s}_{c'}$ are the core locations and $C_S(\mathbf{s}_c, \mathbf{s}_{c'})$ is computed from (17).

Denote then by $\mathbf{s}_1, \ldots, \mathbf{s}_n$ the locations of the training lakes where the modern mean temperature is known and let

$$\mathbf{x}^m = \begin{bmatrix} x^m(\mathbf{s}_1) \\ \vdots \\ x^m(\mathbf{s}_n) \end{bmatrix}, \qquad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}(\mathbf{s}_1) \\ \vdots \\ \boldsymbol{\xi}(\mathbf{s}_n) \end{bmatrix}, \qquad \boldsymbol{\delta} = \begin{bmatrix} \delta(\mathbf{s}_1) \\ \vdots \\ \delta(\mathbf{s}_n) \end{bmatrix}.$$

Including the spatio-temporal model in the hierarchy, the formula (11) is replaced by

$$\begin{aligned} &p(\tilde{\mathbf{X}}^f, \boldsymbol{\vartheta}^m, \boldsymbol{\nu}, \boldsymbol{\omega}, \kappa | \mathbf{x}^m) \\ &\quad = p(\tilde{\mathbf{X}}^f | \boldsymbol{\vartheta}^m, \boldsymbol{\nu}, \boldsymbol{\omega}, \kappa, \mathbf{x}^m) p(\boldsymbol{\nu}, \boldsymbol{\omega} | \boldsymbol{\vartheta}^m, \kappa, \mathbf{x}^m) p(\boldsymbol{\vartheta}^m, \kappa | \mathbf{x}^m) \\ &\quad = p(\tilde{\mathbf{X}}^f | \boldsymbol{\nu}, \kappa) p(\boldsymbol{\nu}, \boldsymbol{\omega} | \mathbf{x}^m) p(\boldsymbol{\vartheta}^m) p(\kappa), \end{aligned} \tag{18}$$

where $p(\boldsymbol{\vartheta}^m)$ can be further factored as in (11). The first factor on the right-hand side is defined by (16) and the second factor can be further developed as

$$p(\boldsymbol{\nu}, \boldsymbol{\omega} | \mathbf{x}^m) \propto p(\boldsymbol{\nu}, \boldsymbol{\omega}) p(\mathbf{x}^m | \boldsymbol{\nu}, \boldsymbol{\omega}).$$

We assume that, given the parameters $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$, the modern temperatures $\mathbf{x}^m$ (or the residuals $\boldsymbol{\delta}$) follow a multivariate normal distribution,

$$\mathbf{x}^m | \boldsymbol{\nu}, \boldsymbol{\omega} \sim \mathrm{N}(\boldsymbol{\xi}\boldsymbol{\omega}, \mathbf{C}^m_S(\boldsymbol{\nu})),$$

where $\mathbf{C}_S^m(\boldsymbol{\nu}) = [C_S(\mathbf{s}_i, \mathbf{s}_j)]$, $\mathbf{s}_i$ and $\mathbf{s}_j$ are training lake locations and $C_S(\mathbf{s}_i, \mathbf{s}_j)$ is computed from (17). The prior distributions for parameters $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$ are assumed to be independent,

$$p(\boldsymbol{\nu}, \boldsymbol{\omega}) = \prod_{i=1}^{3} p(\nu_i) \prod_{i=1}^{3} p(\omega_i).$$

Considering the notation of Section 2.3, when spatial and temporal dependence is included in the model, the vector $\boldsymbol{\theta}$ of all model parameters is expanded to $\boldsymbol{\theta} = \{\mathbf{P}^m, \mathbf{P}^f, \boldsymbol{\vartheta}^m, \boldsymbol{\nu}, \boldsymbol{\omega}, \kappa\}$.

The linear trend in the model is

$$\mu(\mathbf{s}) = \boldsymbol{\xi}\boldsymbol{\omega} = \omega_1 + \omega_2\xi_2(\mathbf{s}) + \omega_3\xi_3(\mathbf{s}), \tag{19}$$

where priors for the parameters $\omega_i$ can be elicited by considering known mean annual temperatures in the part of northern Europe where the training lakes are located. Inari in northern Finland (68°39′N, 27°32′E) and Tartu in Estonia (58°18′N, 26°44′E) are located approximately on the same longitude and the difference in their annual mean temperatures (years 1981–2010) is about −7°C, or about −0.7°C per a degree of latitude. We therefore assume that $\omega_2 \sim \mathrm{N}(-1, 0.5^2)$. It is natural to assume that the temperature changes much less in the east-west direction and, therefore, we take $\omega_3 \sim \mathrm{N}(0, 0.5^2)$. Then, setting $\omega_2 = -0.7$, $\omega_3 = 0$, and using the fact that the mean annual temperature in Helsinki (60°10′N, 24°56′E) is $\mu(\mathbf{s}) = 5.9$ °C, one gets from (19) that $\omega_1 = 47.1$°C, which suggests that a reasonable prior is $\omega_1 \sim \mathrm{N}(47, 3^2)$.

Following Tingley and Huybers (2010a), the prior of $\nu_1$ (partial sill) is an Inverse-gamma distribution, $\nu_1 \sim$ Inverse-gamma$(0.5, 0.2)$, where the parameters were selected so that the prior is rather vague with mode near a point estimate of $\nu_1$ (cf. Section 2.4). For the range parameter $\nu_2$ we took $\nu_2 \sim$ Inverse-gaussian$(200, 500)$. This conforms to the rule of thumb suggested in Journel and Huijbregts (1978), page 194, since the prior density essentially vanishes when $\nu_2$ exceeds 800, half the maximum distance over the field of our data. The nugget parameter $\nu_3$ is assumed to be small, $\nu_3 \sim$ Gamma$(0.01, 10)$.

2.6. *Prior of the temporal smoothing parameter.* We still need to specify the temporal smoothing parameter $\kappa$ that encodes our prior beliefs about the variability of past temperatures $\widetilde{\mathbf{X}}^f$. Denoting by $\rho = C_S(\mathbf{s}_c, \mathbf{s}_c)$ the diagonal element of the spatial covariance $\mathbf{C}_S$ in (12), the marginal prior density of past temperatures at core $c$ is given by

$$p\big(\widetilde{\mathbf{x}}_c^f \,|\, \boldsymbol{\nu}, \kappa\big) = \kappa^{(N-1)/2} \exp\left[ -\frac{(\widetilde{x}_{c1}^f - \mu_c)^2}{2\rho} - \frac{\kappa}{2\rho} \sum_{i=2}^{N} \left( \frac{\widetilde{x}_{ci}^f - \widetilde{x}_{c(i-1)}^f}{t_i - t_{i-1}} \right)^2 \right] \tag{20}$$

[cf. (14) and (15)]. If $\widetilde{\mathbf{x}}_c^f$ were known, using a point estimate $\hat{\rho}$ for $\rho$ (cf. Section 3.2), the "best" $\kappa$ in the sense of maximizing (20) would be

$$\hat{\kappa} = \hat{\rho}\left[\frac{1}{N-1}\sum_{i=2}^{N}\left(\frac{\widetilde{x}_{ci}^f - \widetilde{x}_{c(i-1)}^f}{t_i - t_{i-1}}\right)^2\right]^{-1}. \tag{21}$$

In principle, one could try to employ here an existing long instrumental temperature record but, given that the longest records cover only the last couple of hundred years, this is not a viable option. Instead of a real instrumental record, we therefore used an 1150 year long time series of simulated annual mean temperatures from AD 850 to 1999 for the area where the cores are located, extracted from the NCAR Climate System Model simulation described in Ammann et al. (2007). As the reconstructed temperatures should be thought of as 30-year annual means (because the training temperatures are such) at the union chronology time points $t_i$, and the start of the chronology is commonly taken to be AD 1950, we restricted the simulated time series to the interval [AD 850, AD 1950], computed its 30-year moving average from AD 1950 backward, and then sampled the resulting time series at the times $t_i$. The original simulation, the moving average and the subseries corresponding to the union chronology are shown in Figure 2.

However, besides reconstructions for the union chronology times $t_i$, we will also be interested in reconstructions for individual core chronologies (Section 3). The problem is that, for some cores, only a small number of dates between AD 850 and AD 1950 correspond to actual sediment slices (cf. the online supplement),

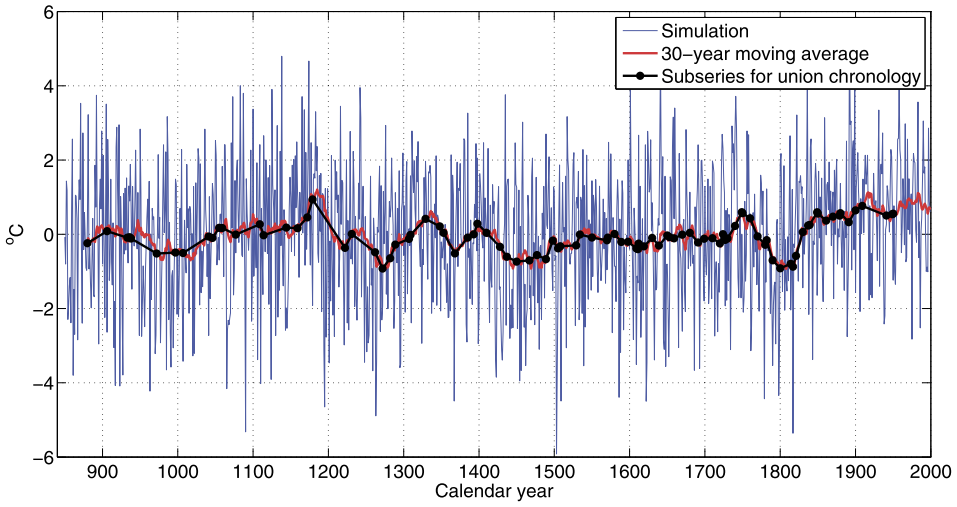


FIG. 2. *Blue curve*: *NCAR Climate System Model simulation of mean annual temperature anomaly for the area where the cores are located. Red curve*: 30-*year moving average of the simulated anomaly. Black curve*: *the* 30-*year mean evaluated at the union chronology time points.*

TABLE 1
*Parameters of the prior distribution* Gamma$(a, b)$ *of the temperature smoothing parameter* $\kappa$ *for independent, spatially independent and spatio-temporal reconstructions, as well as the corresponding estimate* $\hat{\kappa}$ *from* (21) *and the posterior mean of* $\kappa$

| **Reconstruction** | $a, b$ | $\hat{\kappa}$ | $\mathbb{E}(\kappa \mid \text{data})$ |
|---|---|---|---|
| Arapisto | 22, 215 | 4650 | 7284 |
| Flarken | 412, 49 | 20,309 | 20,861 |
| Raigastvere | 536, 43 | 23,148 | 23,313 |
| Rõuge | 19, 230 | 4343 | 10,336 |
| Union (spatially independent) | 4, 240 | 1041 | 9439 |
| Union (spatio-temporal) | 1.1, 274 | 306 | 4374 |

making estimation of temperature time series roughness dubious. We therefore extrapolated the roughness information in the simulated time series to the whole Holocene as follows. From the moving average $z_i$, $i = 880, \ldots, 1999$, we computed for each time difference $k$ the mean value of $(z_i - z_{i-k})^2$ and imputed this value for $(\tilde{x}^f_{ci} - \tilde{x}^f_{c(i-1)})^2$ in (21), when the interval $[t_{ci}, t_{c(i-1)}]$ is not contained in the range [AD 880, AD 1950] and $t_{ci} - t_{c(i-1)} = k$. The prior for $\kappa$ was then defined as Gamma$(a, b)$, with $a$ and $b$ selected so that the prior mean $(ab)$ is approximately equal to the estimate $\hat{\kappa}$ in (21) and the prior variance $(ab^2)$ is rather large (cf. Table 1).

To get an idea how well this procedure might capture the true characteristics of past temperature variation, we show in Figure 3 the centered 30-year annual

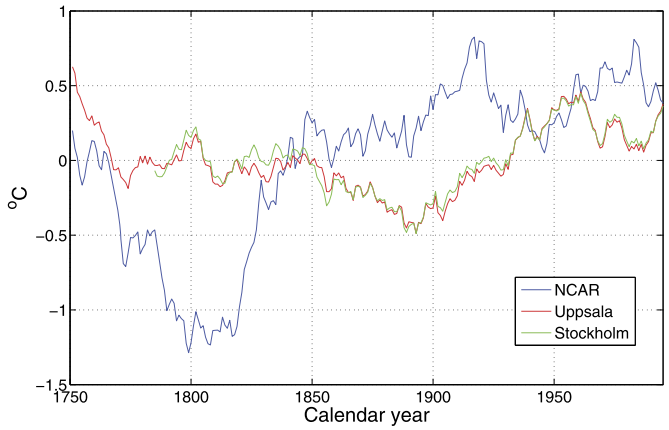


FIG. 3. 30-*year moving averages of the computer simulated time series and instrumental temperature records from Uppsala and Stockholm. All three records have been centered by subtracting the mean.*

means of a part of the simulated series and two instrumental records, one from Stockholm and one from Uppsala, two Swedish cities located close to each other and on approximately the same latitudes as the core lakes used for reconstruction [Moberg and Bergström (1997)]. It appears that the simulated time series could be a little too rough to mimic actual temperature variation, at least in the Stockholm–Uppsala area during the last couple of hundred years. As indicated in Table 1, the posterior mean of $\kappa$ actually tended to be larger than the estimate $\hat{\kappa}$ which appears to support this observation.

## 3. Example reconstructions.

3.1. *The data*. Our modern pollen-temperature training set includes $n = 173$ lakes with known 30-year modern annual mean temperatures ($\mu_c$) and surface sediments analyzed for relative abundances of a total of $l = 104$ pollen taxa. For more details on the training set, see Seppä et al. (2009) and Antonsson et al. (2006). Instead of using absolute numbers of pollen grains in sediment samples, we scaled all counts to the interval [0, 100]. Although this results in the loss of some information in the data, the changing environment is in fact thought to be reflected in the *relative* abundances of various pollen taxa and not in their absolute numbers. Further, the absolute total counts at different sites varied greatly (from 169 to 3654) and the Bummer model that underlies our reconstruction methods appears to work best when the total counts do not differ too much across sediment samples. A similar scaling of counts to a fixed interval was also suggested by Haslett et al. (2006) when, as is often the case, only the relative abundances pollen taxa are known.

Past temperature reconstructions were made from four sediment cores obtained from lakes Arapisto, Flarken, Raigastvere and Rõuge. The chronologies of Lakes Arapisto, Flarken and Raigastvere are based on radiocarbon dating. Conventional bulk radiocarbon datings were obtained from Flarken (13 datings) and Raigastvere (10 datings) because these cores were sampled before the use of AMS technique, while the Arapisto core was dated with 7 AMS datings. All datings were calibrated and the age–depth curves for all sites were constructed using the median values of the probability distributions of the calibrated ages. All three sites have generally stable sedimentation rates, which increases the reliability of the chronologies [Sarmaja-Korjonen and Seppä (2007), Seppä, Hammarlund and Antonsson (2005), Seppä and Poska (2004)]. Lake Rõuge is partly annually laminated but the varve chronology is floating. The chronology and age–depth model for the lake were derived by correlating the paleomagnetic secular variation (PSV) curve with the clear anchor points of the PSV curve of the Finnish varved lake Nautajärvi [Seppä et al. (2009)]. The obtained chronology is supported by AMS dates. Figure 4 shows the locations of the training lakes and cores on a map of northern Europe and Table 2 provides additional information on the core lakes. The four core chronologies consist of a total of 586 time points, but, as some of these are shared by more than one core, the total number of dates in the union chronology is only 572 (cf. Table 3).

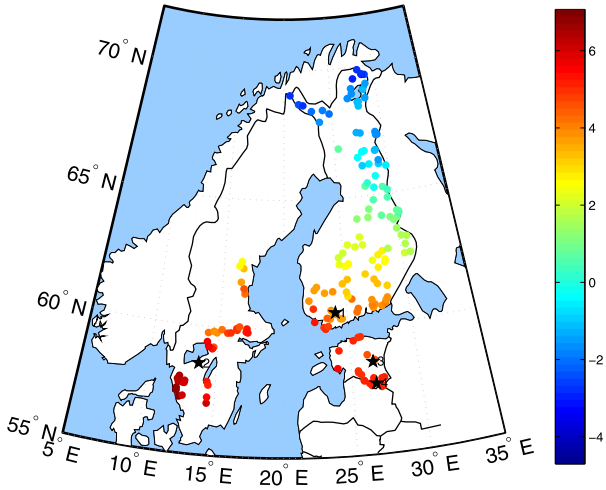


FIG. 4. *Locations of the training lakes and cores. The training lakes are marked with dots whose colors indicate the associated annual mean temperature (in °C). The core lakes are as follows*: 1 = *Arapisto*, 2 = *Flarken*, 3 = *Raigastvere*, 4 = *Rõuge*.

The full chronologies are listed in Holmström et al. (2015a). For more details, see Sarmaja-Korjonen and Seppä (2007) (Arapisto), Seppä, Hammarlund and Antonsson (2005) (Flarken), Seppä and Poska (2004) (Raigastvere) and Seppä et al. (2009) (Rõuge).

3.2. *The different reconstruction models used.* In addition to the spatio-temporal reconstruction described in Section 2, we also considered two additional approaches. First, reconstructions were made for each core separately. The model for each core is exactly the same as in the multi-core case ($C = 1$ in Section 2), but with the union chronology replaced by the actual chronology of the core and the spatial part $\mathbf{C}_S$ of $\boldsymbol{\Sigma}$ in (12) omitted. We refer to these reconstructions as "independent." The second variation was to perform multi-core reconstruction on the union chronology but to replace $\mathbf{C}_S$ by an identity matrix. This ignores distance-

TABLE 2
*The four core lakes used for the pollen-based temperature reconstruction. The modern temperature is* $\mu_c$

| Lake | Latitude | Longitude | $\mu_c$ (°C) | Country |
|---|---|---|---|---|
| Arapisto | 60°35′N | 24°05′E | 4.5 | Finland |
| Flarken | 58°33′N | 13°40′E | 5.9 | Sweden |
| Raigastvere | 58°35′N | 26°39′E | 5.0 | Estonia |
| Rõuge | 57°44′N | 26°45′E | 5.5 | Estonia |

TABLE 3

*Details about core chronologies and the union chronology. For each chronology, shown are its length as well as its youngest and oldest samples. For the four core lakes, in Section 2.2 these quantities are denoted by $n_c$, $t_{c1}$ and $t_{cn_c}$, respectively. Time is expressed as years before present, with 0 corresponding to AD 1950*

| Core | Chronology length | Youngest sample | Oldest sample |
|---|---|---|---|
| Arapisto | 98 | 0 | 10,852 |
| Flarken | 114 | 118 | 12,084 |
| Raigastvere | 115 | 0 | 11,594 |
| Rõuge | 259 | 0 | 11,821 |
| Union | 572 | 0 | 12,084 |

based spatial correlation between the cores but leaves intact interaction through the shared environmental response parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. We refer to this case as "spatially independent," which refers to a lack of an explicit spatial dependence component in the model.

The prior, the estimate $\hat{\kappa}$ from (21) and the posterior mean for the smoothing parameter $\kappa$ in each case is given in Table 1. In (20), for the independent and spatially independent models, $\rho = 1$, and for the spatio-temporal model, we took $\hat{\rho} = 0.2937$, the value obtained from a point estimate of the spatial covariance (cf. Section 2.4). We note that in some cases the posterior mean of $\kappa$ lies far in the right tail of the prior distribution. We therefore recomputed the reconstruction in these cases with vague $\kappa$ priors centered at the posterior means of Table 1. Now the new posterior means were quite close to the prior means and the reconstructions themselves changed little. We therefore believe that the priors of Table 1 are reasonable and result in reliable temperature reconstructions.

The posterior means for the parameters of the spatial covariance were

$$\mathbb{E}(\boldsymbol{\nu}|\text{data}) = [0.2108, 147.9279, 0.0698]^T,$$
$$\mathbb{E}(\boldsymbol{\omega}|\text{data}) = [47.4144, -0.7014, -0.0472]^T.$$

The posterior mean covariance matrix is

$$\mathbb{E}(\mathbf{C}_S|\text{data}) = \begin{bmatrix} 0.281 & 0.003 & 0.035 & 0.020 \\ 0.003 & 0.281 & 0.001 & 0.001 \\ 0.035 & 0.001 & 0.281 & 0.111 \\ 0.020 & 0.001 & 0.111 & 0.281 \end{bmatrix},$$

where the lakes appear in the order Arapisto, Flarken, Raigastvere and Rõuge. Thus, the elements in the first row from left to right show the variance for Arapisto, the covariance between Arapisto and Flarken, the covariance between Arapisto and Raigastvere, and the covariance between Arapisto and Rõuge, and so on.

For each time point $t_i$, the posterior mean temperature and its 95% highest posterior density interval were computed. As such point-wise credible intervals may

underestimate the uncertainty in the paleotemperature time series regarded as a whole curve, we also calculated a 95% simultaneous credible band employing the method of "Simultaneous Credible Intervals" suggested in Erästö and Holmström (2005). Using the generated posterior sample, the method first finds a $\Delta > 0$ such that

$$\mathbb{P}\left(\max_{i=1,\ldots,N}\left|\frac{\tilde{x}_{ci}^{f} - \mathbb{E}(\tilde{x}_{ci}^{f}|\text{data})}{\text{Std}(\tilde{x}_{ci}^{f}|\text{data})}\right| \le \Delta|\text{data}\right) = 0.95$$

and then defines the simultaneous credible band as

$$\mathbb{E}(\tilde{x}_{ci}^{f}|\text{data}) \pm \Delta\,\text{Std}(\tilde{x}_{ci}^{f}|\text{data}), \qquad i = 1, \ldots, N.$$

The point-wise and simultaneous credible intervals are probability intervals based on the posterior probability which itself is determined by the data and the model assumptions. The reconstruction accuracy of a simplified version of our single-lake model (essentially the Bummer model) was checked using training set cross-validation in Toivonen et al. (2001), Vasko, Toivonen and Korhola (2000) and Salonen et al. (2012), where it was found that, in terms of root mean square error of prediction, it performed competitively against standard methods, such as WA-PLS. The structure of the spatio-temporal prior prevents such validation for the more complex model considered here.

3.3. *The Gaussian response model.* The plausibility of the Gaussian response model of Section 2.3 is discussed extensively in the online supplement [Holmström et al. (2015a)] and we summarize here the main conclusions. First, based on comparisons with the training data, the model appears to describe the observed relative taxon abundances reasonably well. The overall character of predicted abundances as a function of temperature also seem plausible with nearby lakes and the multi-core reconstructions producing similar response curves. For most taxa the optimal temperature ranges suggested by the estimated response curves do not seem unreasonable. The similarity of the response curves of the two multi-core reconstructions (spatially independent and spatio-temporal models) is consistent with the similarity of their temperature reconstructions (see below). In the case of the more southern lakes (Flarken, Raigastvere and Rõuge), the estimated peak relative abundances of warmer temperature taxa exceeds the abundance seen in the training and core data, while for the northernmost lake (Arapisto), these abundances are considerably lower. This is not unexpected, considering the modest share of most of these warmer temperature taxa in the Arapisto core.

The posterior values of the optimal temperature parameter $\beta_j$ for the warmer temperature taxa substantially exceed their prior values. This may be explained by the fact that the prior is centered on the optimal value estimated from the training set and the training lake temperatures are likely to be considerably lower than many of the past temperatures at the core lakes. The posterior mean of $\beta_j$ only

roughly corresponds to the temperature at which the modeled abundance probability of taxon $j$ peaks, although this correspondence seems to be more robust for the multi-core reconstructions. Therefore, one should not interpret the parameter $\beta_j$ as representing a precise optimal taxon temperature. Also, the posterior values of the tolerance parameter $\gamma_j$ tend to be very large, making the Gaussian response function [(4) and (5)] flat, undermining $\beta_j$'s role as a clearly defined optimum temperature.

3.4. *Interpretation of reconstructed temperature histories*. The past temperature reconstructions for different models are shown in Figures 5–7. For comparison, we also include in all figures reconstructions made with the WA-PLS method [ter Braak and Juggins (1993)], one of the most popular calibration methods used in pollen-based reconstructions. The source for these reconstructions was Seppä et al. (2009).

Figure 5 displays the independent reconstructions for each lake. Figure 6 shows the spatially independent reconstructions and Figure 7 shows the reconstructions made with the full spatio-temporal model. In all figures, the thick curve is the posterior mean, the thin curve is the WA-PLS reconstruction and a dot at AD 1950 marks the current mean annual instrumental temperature. Lighter and darker gray show the point-wise and simultaneous 95% credible bands, respectively. In Figures 6 and 7, for each lake, the black line marks the oldest date in its own chronology.

Our first observation is that for each lake the general features of the reconstructions based on the spatially independent and spatio-temporal models are quite similar and both differ to some extent from the reconstructions made independently from single cores. Allowing the cores to interact, either through shared parameters or spatial correlation, also makes reconstructions for all lakes more similar. Further, the full spatio-temporal reconstructions generally have least posterior uncertainty, as exhibited by the smaller credibility intervals. From roughly 7700 years before present onward, the uncertainties are largest in the Arapisto record, which also has the lowest sediment sample resolution for this period (cf. the online supplement). The large uncertainties to the left of the black lines in Figures 6 and 7 are due to lack of pollen data for the lake in question, which causes the reconstructed temperatures to be supported only by the priors.

The distinct feature in the results obtained with independent reconstructions from single cores is the abrupt rise of temperature during the early Holocene. These single core reconstructions also show generally higher temperature values during the mid-Holocene (roughly 8000–4000 years ago) than in the WA-PLS-based reconstructions (Figure 5). In the spatially independent reconstructions, the temperature values during the mid-Holocene are more consistent with those generated with the WA-PLS technique (Figure 6). The reconstructions based on the full spatio-temporal model (Figure 7) show trends which are most compatible with the WA-PLS-based trends, with most gradual temperature rise in the early Holocene.

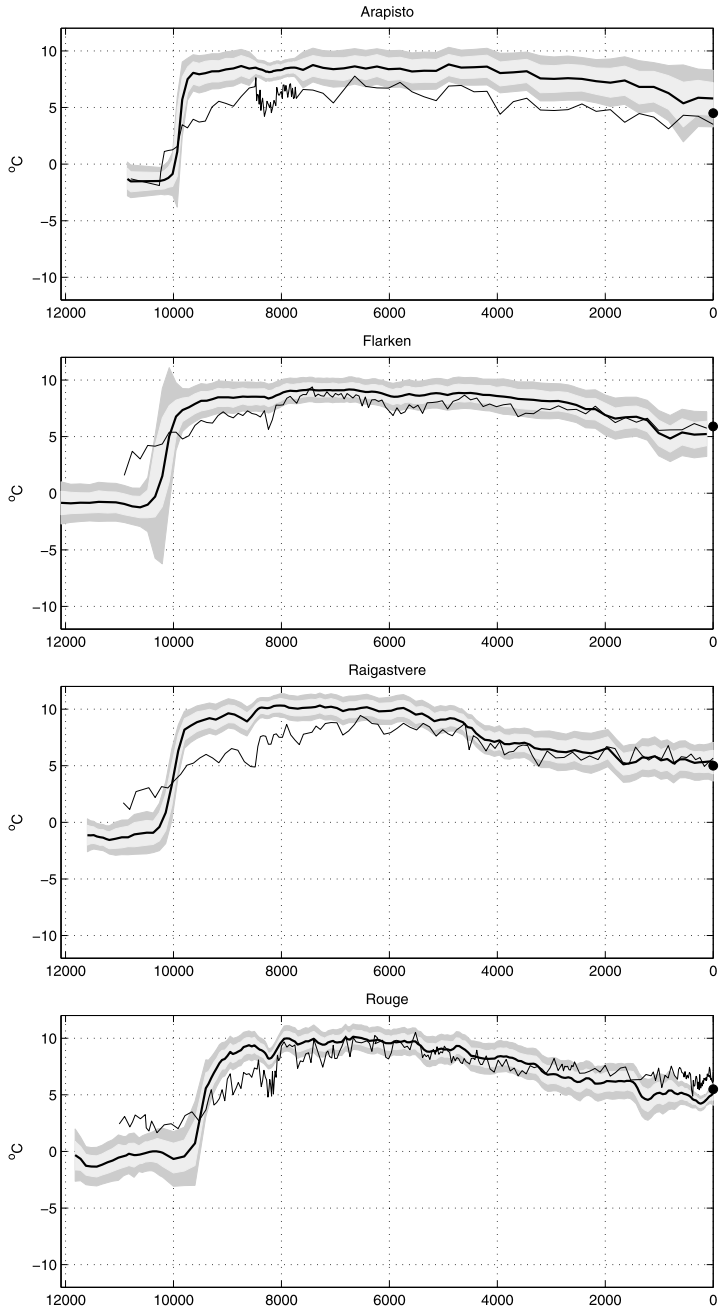


FIG. 5. *Temperature reconstructions made independently from each core. The thick curve is the posterior mean and the thin curve is the WA-PLS reconstruction. Light and dark gray show the point-wise and simultaneous* 95% *credible bands, respectively. Horizontal axis: time in years before present. Vertical axis: mean annual temperature in centigrades. The dot at AD* 1950 *marks the current mean annual instrumental temperature.*

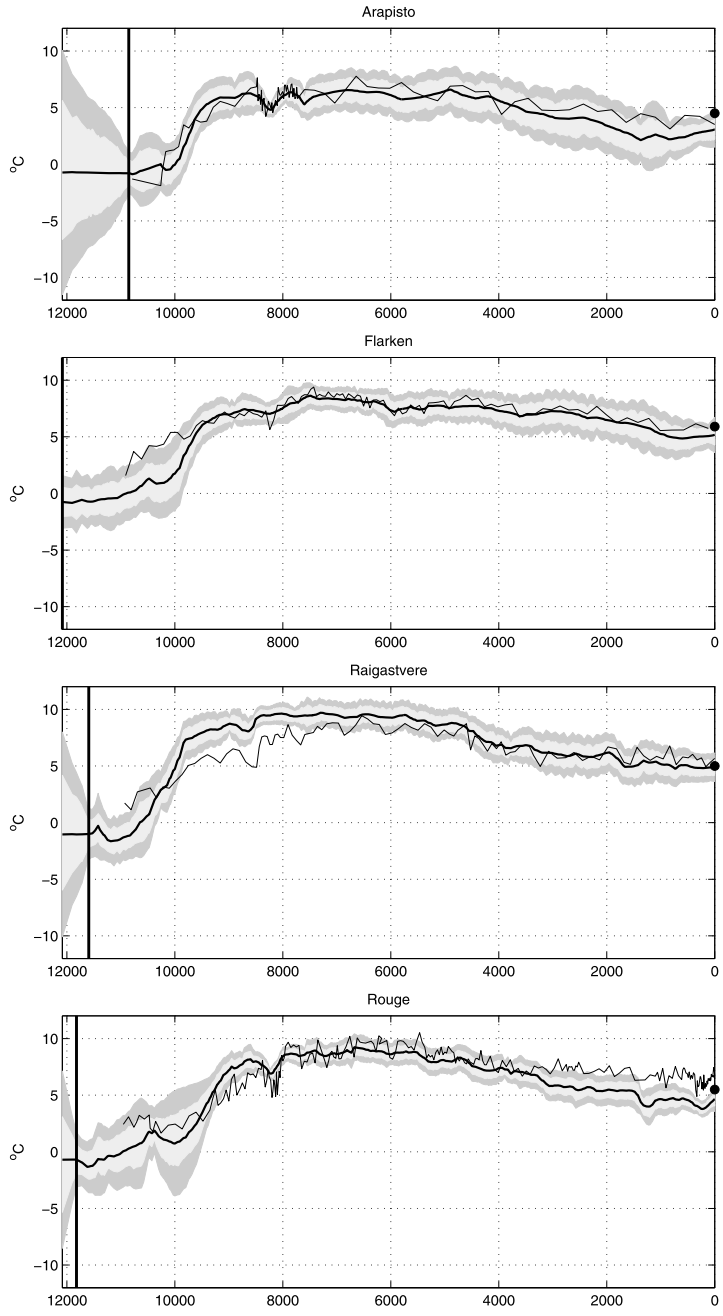


FIG. 6. *Temperature reconstructions based on the spatially independent model with no explicit spatial interaction. The thick curve is the posterior mean and the thin curve is the WA-PLS reconstruction. Light and dark gray show the point-wise and simultaneous* 95% *credible bands, respectively. For each lake, the black line marks the oldest date in its own chronology. Horizontal axis: time in years before present. Vertical axis: mean annual temperature in centigrades. The dot at AD* 1950 *marks the current mean annual instrumental temperature.*

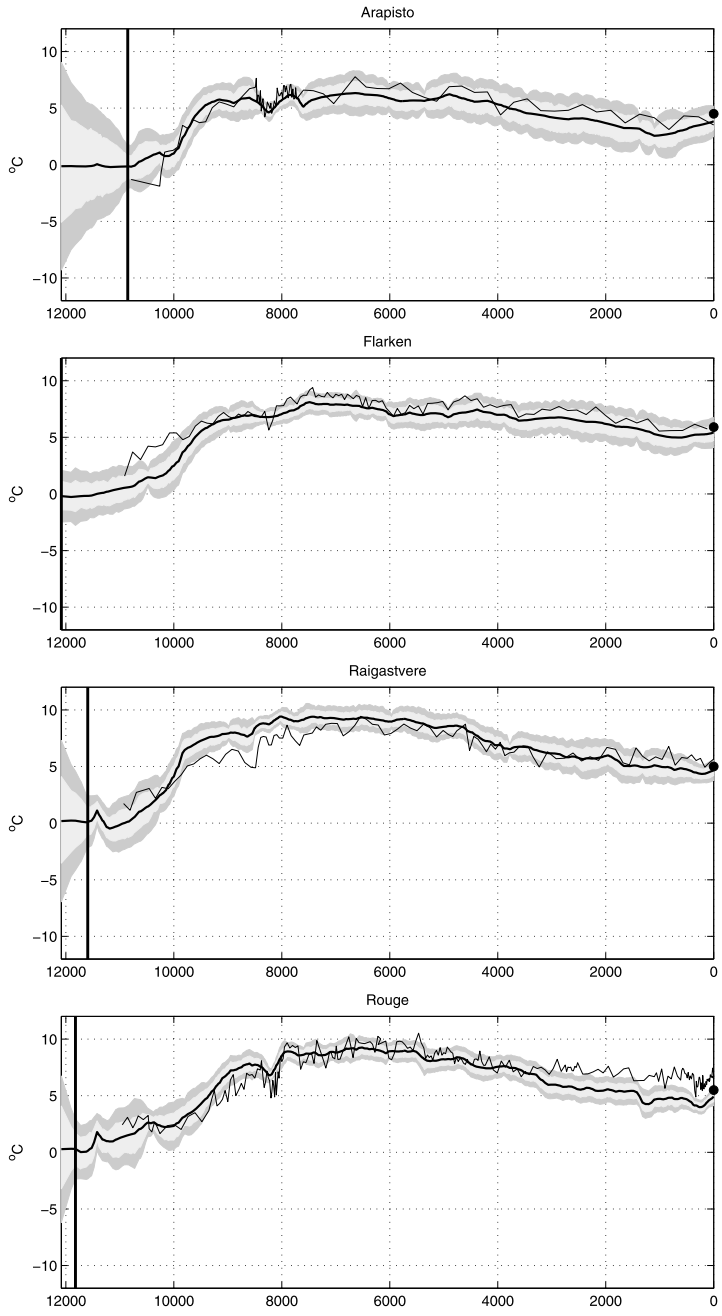


FIG. 7. *Temperature reconstructions based on the full spatio-temporal model. The thick curve is the posterior mean and the thin curve is the WA-PLS reconstruction. Light and dark gray show the point-wise and simultaneous* 95% *credible bands, respectively. For each lake, the black line marks the oldest date in its own chronology. Horizontal axis: time in years before present. Vertical axis: mean annual temperature in centigrades. The dot at AD* 1950 *marks the current mean annual instrumental temperature.*

The Holocene thermal maximum, the warmest period of the Holocene, strongly expressed in northern Europe in paleoclimatic data and model simulations [Renssen et al. (2009, 2012)], is observable in the spatially independent and full spatio-temporal model reconstructions at about 8000–5000 years ago, with temperature value of 8–9°C at the Rõuge, Raigastvere and Flarken, and about 6°C at the northernmost site Arapisto in Finland. These patterns are generally concordant with the WA-PLS results. Moreover, the results show that the high sample resolution helps decrease the uncertainty in the reconstructions. This is reflected particularly in the Arapisto and Rõuge records, where a higher number of pollen samples were analyzed between 8500–8000 years ago to detect possible indications of an abrupt cold event widely observed in northern Europe [Alley and Agústsdóttir (2005), Wiersma and Renssen (2006)]. This event is reflected in Rõuge and Arapisto data by a $\sim$1°C temperature dip, while the influence of high sample resolution is apparent by markedly smaller statistical uncertainties at 8000–8500 years ago (Figure 7). The cold event 8200 years ago is present also in the WA-PLS based reconstructions from Flarken, but it shows less clearly in the posterior means, presumably due to low temporal sample resolution of this record. The Little Ice Age (about AD 1550 to 1850) and the subsequent warming show best in the spatio-temporal reconstructions.

Two features in these reconstructions require further analysis. The first is the rate and magnitude of early Holocene warming that appears quite different for the independent reconstructions (Figure 5) and the joint, multi-core reconstructions (Figures 6 and 7). The second question concerns timing of the onset of warming.

An online supplement [Holmström et al. (2015a)] includes published reconstructions from Greenland ice cores that often are used as a reference when Holocene climate is reconstructed for Northern Europe and the North Atlantic region (Figures S.5 and S.6). The reconstruction in Figure S.6 suggests that the early rise in temperature has been 6–10°C (depending on the amount of smoothing applied) and our reconstructions are within that range. However, according to the ice core records, the rate at which the temperature rises in the individual, single-core reconstructions is too high. This view is also supported by the Scandinavian reference reconstructions (Figures S.7, S.8 and S.9 in the supplement) as well as the WA-PLS reconstructions displayed in Figures 5, 6 and 7. The global and hemispheric reconstructions in Marcott et al. (2013) and Shakun et al. (2012) also support this conclusion, although they may be less relevant than the more local Greenland and Scandinavian records. Looking at the pollen abundances in the four cores (Figure S.10), we notice that *Alnus* (alder), *Corylus* (hazel), *Ulmus* (elm) and *Tilia* (linden) are among the taxa whose growing abundance coincides with the onset of warming. These are all taxa with optimal temperatures that are likely to be higher than the past temperatures at the four core lakes (see also Figure S.2 in the online supplement). This can explain the timing of the temperature rise reconstructed for these lakes, but it does not clarify why the single-core Bayesian model seems to overestimate the rate of temperature change. If desired, this could be

remedied by increasing considerably the value of the temporal smoothing parameter $\kappa$ in (13), but such an ad hoc choice might be difficult to justify. Alternatively, one might use a time-dependent smoothing parameter that would smooth the onset of warming differently from the rest of the Holocene. Similarly to our present approach, such a choice could perhaps be based on numerical climate simulations. Such considerations are left for future research.

Compared with the single-core reconstructions, the rate of warming in the spatially independent and spatiotemporal joint reconstructions is in much better agreement with the Greenland and Scandinavian reference records as well as the WA-PLS reconstructions. It appears that the potential difficulty the single-core Bayesian model has in handling such a rapid rise can be alleviated by borrowing strength from other cores. Thus, sharing the abundance model parameters between the cores, and therefore effectively increasing the number of data available for their estimation, already makes a significant difference. Spatial smoothing then further tempers the reconstructed temperature rise.

One might, however, suspect that the apparent difference in the timing of the onset of warming in the four independent reconstructions alone when combined with correlations between the reconstructions explains the more gradual warming in the joint reconstructions. Indeed, while Arapisto, Flarken and Raigastvere temperatures start to rise almost simultaneously, the onset of warming for Rõuge appears to take place later (Figure 5). This is all the more problematic since Lake Rõuge is the southernmost of the four core lakes, and therefore would be expected to warm first. Such a discrepancy could be explained by the rather wide confidence intervals around 10,000 BP, but another possibility is the relative paucity of chronology dates for Lake Rõuge between 10,200 BP and 9400 BP (cf. the online supplement). We therefore made reconstructions also with Rõuge data before 9400 BP left out. The results are shown in Figures S.10, S.11 and S.12 in the online supplement. While the temperature rise in the joint reconstructions is now somewhat sharper than in Figures 6 and 7, it is still much more gradual than in the single-core reconstructions of Figure 5. We conclude that the main factor in decreasing the rate of early Holocene warming in the joint reconstructions is sharing of the taxon-specific response parameters. This, of course, does not exclude the possibility of additional smoothing in the joint reconstructions because of chronology misalignments caused by dating errors. The error in the radiocarbon dates varies between the four lakes and depends on the age of the sediment sample, being generally larger for the oldest samples. Thus, errors of 100–200 years are likely for the oldest samples of lakes Arapisto, Flarken and Raigastvere, but for Lake Rõuge they can be even larger. Even with the earliest Rõuge data left out, some smoothing may therefore result at the time of early Holocene warming because the reconstructions may not be correctly aligned. The best solution would be to let the dating errors influence the reconstructions and their posterior uncertainty by incorporating them in the hierarchical model. A simple additive error model

was proposed in Erästö et al. (2012), but a more satisfactory approach would include a sophisticated Bayesian chronology model such as the Bchron of Haslett and Parnell (2008) as a model component. We will consider this in future work.

3.5. *Computational details*. In all cases, a Metropolis-within-Gibbs sampler [e.g., Robert and Casella (2004)] was run for 30,000 iterations, the first 15,000 were used for burn-in and from the last 15,000, every 5th sample was kept for inference. Thus, each posterior analysis was based on a sample of size 3000. The relevant conditional posterior distributions are given in Appendix B.

In both spatially independent and spatio-temporal reconstructions, some chronology time points are not associated with corresponding pollen abundance data. Our strategy was to first update, one by one, the temperatures which do have associated pollen data and after that those without pollen data, conditioning them on those with pollen data. Adaptive simulation was used both for temperatures and the environmental response parameters [Gelman et al. (2004)]. The adaptive phase consisted of 10,000 iterations and the subsequent fixed phase of 20,000 iterations that used the proposal variances from the last adaptive step.

The initial values for the components of the temperature vector $\widetilde{x}_c^f$ were simulated from $\mathrm{N}(\mu_c, 1.5^2)$, where $\mu_c$ is the modern temperature at core $c$ (see Table 2). The initial values of $\kappa$, the trend parameters $\omega_i$ and the range parameter $\nu_2$ were generated from their priors. For partial sill $\nu_1$, the mode of the prior was used and the nugget $\nu_3$ was initialized at its prior mean.

The abundances among the taxa analyzed vary considerably, with many taxa appearing in the sediment samples only rarely and only some appearing in substantial abundance. We therefore thought it best to use taxon-specific initialization for the scaling factor $\alpha_j$. The iteration for $\alpha_j$ was started at $\alpha_{j,\max}/2$, where $\alpha_{j,\max}$ is the largest observed abundance of taxon $j$. Such initialization accelerated convergence substantially. The initial values for $\beta_j$ and $\gamma_j$ were generated from their prior distributions $\mathrm{N}(\hat{\beta}_j, (1.5\sqrt{(3)})^2)$ and Gamma(9, 1/3), respectively.

The algorithms were implemented in Matlab and run on a PC with an Intel Core i7 3770 CPU. Table 4 summarizes approximate computation times in different cases.

**4. Conclusions.** We propose a novel Bayesian approach for the reconstruction of past temperature variation during the Holocene, using fossil pollen data from multiple sediment cores. A spatio-temporal model was described that takes into account both temporal correlations within the cores and spatial correlations between them. Temporal correlations were modeled with a smoothing prior where the smoothing parameter hyperprior was elicited using numerical climate simulation. The temporal smoothing prior is very vague and favors rather slowly varying temperature time series, which is consistent with the relatively stationary climate conditions during the Holocene. An isotropic covariance was used to model the

TABLE 4
*CPU times for temperature reconstructions. For each lake, the CPU time is for reconstruction made independently using its own core chronology. The last two reconstructions are joint reconstructions using all four cores*

| **Reconstruction** | **CPU time (hours)** |
|---|---|
| Arapisto | 3 |
| Flarken | 3 |
| Raigastvere | 3 |
| Rõuge | 3 |
| Union (spatially independent) | 26 |
| Union (spatio-temporal) | 28 |

spatial dependence of the temperatures across the sites from which the sediment samples were obtained.

Taking into account spatial dependencies between reconstructions reduced uncertainty and made their overall shapes more similar. Given that the four cores considered are from a geographically restricted area and that the temperature history at the four sites therefore must have been similar, it can be argued that the spatio-temporal reconstructions are an improvement over the reconstructions made independently from each core or those without explicit spatial dependencies. The spatio-temporal reconstructions are also smoother, less uncertain and generally more realistic. In addition, they are more consistent with the results obtained with WA-PLS, a popular method for pollen-based reconstructions.

The proposed model is directly applicable to reconstructions from other biological proxies records, such as diatoms and chironomids. Other climate variables besides temperature could also be considered. It would also be interesting to consider a larger set of proxy records from a more extensive geographic area. In some situations a nonstationary spatial covariance might have to be used to model different types of correlations within and between distinctly different types of regions. Finally, the chronologies were assumed error-free, which of course is a simplification. Therefore, future work will need to also address the uncertainty related to the various sources of errors involved in constructing the chronologies.

## APPENDIX A: TEMPORAL COVARIANCE STRUCTURE

Computation of the covariance matrix $\mathbf{\Gamma}$ in (15) is needed for efficient implementation of the sampling procedures used in estimation. Following, for example, Kaipio and Somersalo (2005), the quadratic form in the exponent of (14) is first written as

$$\sum_{i=2}^{N}\left(\frac{\tilde{x}_{ci}^{f}-\tilde{x}_{c(i-1)}^{f}}{t_{ci}-t_{c(i-1)}}\right)^{2}=\|\mathbf{L}\tilde{\mathbf{x}}_{c}^{f}\|^{2},$$

where $\mathbf{L} = [L_{ij}] \in \mathbb{R}^{(N-1)\times N}$,

$$L_{ij} = \begin{cases} -(t_{c(i+1)} - t_{ci})^{-1}, & \text{when } j = i, \\ (t_{c(i+1)} - t_{ci})^{-1}, & \text{when } j = i+1, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\mathbf{L}^T\mathbf{L} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

where $\mathbf{B}_{11} \in \mathbb{R}$, $\mathbf{B}_{12} \in \mathbb{R}^{N-1}$, $\mathbf{B}_{21} \in \mathbb{R}^{(N-1)\times 1}$ and $\mathbf{B}_{22} \in \mathbb{R}^{(N-1)\times(N-1)}$. Then

$$p(\tilde{\mathbf{x}}^f_{c*}|\tilde{x}^f_{c1}, \kappa) \propto \kappa^{(N-1)/2} \exp\left[-\frac{\kappa}{2}(\tilde{\mathbf{x}}^f_{c*} + \mathbf{B}_{22}^{-1}\mathbf{B}_{21}\tilde{x}^f_{c1})^T \mathbf{B}_{22}(\tilde{\mathbf{x}}^f_{c*} + \mathbf{B}_{22}^{-1}\mathbf{B}_{21}\tilde{x}^f_{c1})\right],$$

where $\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\tilde{x}^f_{c1} = [\tilde{x}^f_{c1}, \ldots, \tilde{x}^f_{c1}] \in \mathbb{R}^{N-1}$. The formula (15) then follows readily when $\tilde{x}^f_{c1} \sim \mathrm{N}(\mu_c, 1)$.

## APPENDIX B: THE CONDITIONAL POSTERIORS

From (3)–(10) and Section 2.4,

$$\begin{aligned} p(\tilde{\mathbf{X}}^f, \boldsymbol{\theta}|\mathbf{Y}^f, \mathbf{x}^m, \mathbf{Y}^m) &\propto \prod_{c=1}^{C}\prod_{i=1}^{n_c} p(\mathbf{y}^f_{ci}|x^f_{ci}, \boldsymbol{\theta}) \prod_{i=1}^{n} p(\mathbf{y}^m_i|x^m_i, \boldsymbol{\theta}) \prod_{i=1}^{n} p(\mathbf{p}^m_i|x^m_i, \boldsymbol{\vartheta}^m) \\ &\quad \times \prod_{c=1}^{C}\prod_{i=1}^{n_c} p(\mathbf{p}^f_{ci}|x^f_{ci}, \boldsymbol{\vartheta}^f_c) p(\tilde{\mathbf{X}}^f|\boldsymbol{\nu}, \kappa, ) p(\kappa) p(\mathbf{x}^m|\boldsymbol{\nu}, \boldsymbol{\omega}) \prod_{i=1}^{3} p(\omega_i) \prod_{i=1}^{3} p(\nu_i) \\ &\quad \times \prod_{j=1}^{l} p(\alpha_j) \prod_{j=1}^{l} p(\beta_j|\mathbf{x}^m) \prod_{j=1}^{l} p(\gamma_j) \\ &= \prod_{c=1}^{C}\prod_{i=1}^{n_c} \mathrm{Mult}(\mathbf{y}^f_{ci}|y^f_{ci\cdot}, \mathbf{p}^f_{ci}) \prod_{i=1}^{n} \mathrm{Mult}(\mathbf{y}^m_i|y^m_{i\cdot}, \mathbf{p}^m_i) \\ &\quad \times \prod_{i=1}^{n} \mathrm{Dirichlet}(\mathbf{p}^m_i|\boldsymbol{\lambda}^m_i) \prod_{c=1}^{C}\prod_{i=1}^{n_c} \mathrm{Dirichlet}(\mathbf{p}^f_{ci}|\boldsymbol{\lambda}^f_{ci}) \\ &\quad \times \mathrm{N}(\tilde{\mathbf{X}}^f|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \mathrm{Gamma}(\kappa|a, b) \times \mathrm{N}(\mathbf{x}^m|\boldsymbol{\xi}\boldsymbol{\omega}, \mathbf{C}^m_S(\boldsymbol{\nu})) \times \mathrm{N}(\boldsymbol{\omega}|\boldsymbol{\mu}_{\boldsymbol{\omega}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}) \\ &\quad \times \text{Inv-gamma}(\nu_1|0.5, 0.2) \times \text{Inv-gaussian}(\nu_2|200, 500) \\ &\quad \times \mathrm{Gamma}(\nu_3|0.01, 10) \\ &\quad \times \prod_{j=1}^{l} \mathrm{Unif}(\alpha_j|0.1, 50) \prod_{j=1}^{l} \mathrm{N}(\beta_j|\hat{\beta}_j, (1.5\sqrt{3})^2) \prod_{j=1}^{l} \mathrm{Gamma}(\gamma_j|9, 1/3), \end{aligned}$$

where $\boldsymbol{\Sigma} = \mathbf{C}_S(\boldsymbol{\nu}) \otimes \mathbf{C}_T(\kappa)$, $\boldsymbol{\mu}_{\boldsymbol{\omega}_0} = [47, -1, 0]$, and $\boldsymbol{\Sigma}_{\boldsymbol{\omega}_0} = \mathrm{diag}(3^2, 0.5^2, 0.5^2)$.

Therefore, the full conditional posterior distributions of the unknown parameters are

$$p(\widetilde{\mathbf{X}}^f|\cdot) \propto \prod_{c=1}^{C}\prod_{i=1}^{n_c} \text{Dirichlet}(\mathbf{p}_{ci}^f|\boldsymbol{\lambda}_{ci}^f) \exp\Big[-\frac{1}{2}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu})\Big],$$

$$p(\kappa|\cdot) \propto \kappa^{(C(N-1)+2(a-1))/2} \times \exp\Big[-\frac{\kappa}{b}-\frac{1}{2}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu})^T\big(\mathbf{C}_S(\boldsymbol{\nu})\otimes\mathbf{C}_T(\kappa)\big)^{-1}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu})\Big],$$

$$p(\mathbf{p}_i^m|\cdot) = \text{Dirichlet}(\mathbf{p}_i^m|\mathbf{y}_i^m+\boldsymbol{\lambda}_i^m),$$

$$p(\mathbf{p}_{ci}^f|\cdot) = \text{Dirichlet}(\mathbf{p}_{ci}^f|\mathbf{y}_{ci}^f+\boldsymbol{\lambda}_{ci}^f),$$

$$p(\alpha_j|\cdot) \propto \prod_{i=1}^{n}\text{Dirichlet}(\mathbf{p}_i^m|\boldsymbol{\lambda}_i^m)\prod_{c=1}^{C}\prod_{i=1}^{n_c}\text{Dirichlet}(\mathbf{p}_{ci}^f|\boldsymbol{\lambda}_{ci}^f)\times\text{Unif}(\alpha_j|0.1,50),$$

$$p(\beta_j|\cdot) \propto \prod_{i=1}^{n}\text{Dirichlet}(\mathbf{p}_i^m|\boldsymbol{\lambda}_i^m)\prod_{c=1}^{C}\prod_{i=1}^{n_c}\text{Dirichlet}(\mathbf{p}_{ci}^f|\boldsymbol{\lambda}_{ci}^f)\times\text{N}(\beta_j|\hat{\beta}_j,(1.5\sqrt{3})^2),$$

$$p(\gamma_j|\cdot) \propto \prod_{i=1}^{n}\text{Dirichlet}(\mathbf{p}_i^m|\boldsymbol{\lambda}_i^m)\prod_{c=1}^{C}\prod_{i=1}^{n_c}\text{Dirichlet}(\mathbf{p}_{ci}^f|\boldsymbol{\lambda}_{ci}^f)\times\text{Gamma}(\gamma_j|9,1/3),$$

$$p(\boldsymbol{\omega}|\cdot) = \text{N}\big(\boldsymbol{\omega}|\boldsymbol{\Sigma}_{\boldsymbol{\omega}}(\boldsymbol{\xi}^T\mathbf{C}_S^m(\boldsymbol{\nu})^{-1}\mathbf{x}^m+\boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1}\boldsymbol{\mu}_{\boldsymbol{\omega}_0}),\boldsymbol{\Sigma}_{\boldsymbol{\omega}}\big),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \big(\boldsymbol{\xi}^T\mathbf{C}_S^m(\boldsymbol{\nu})^{-1}\boldsymbol{\xi}+\boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1}\big)^{-1},$$

$$p(\nu_i|\cdot) \propto \frac{1}{\sqrt{\det(\mathbf{C}_S(\boldsymbol{\nu}))^N\det(\mathbf{C}_S^m(\boldsymbol{\nu}))}} \times \exp\Big[-\frac{1}{2}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu})^T\big(\mathbf{C}_S(\boldsymbol{\nu})\otimes\mathbf{C}_T(\kappa)\big)^{-1}(\widetilde{\mathbf{X}}^f-\boldsymbol{\mu}) - \frac{1}{2}\big(\mathbf{x}^m-\boldsymbol{\xi}\boldsymbol{\omega}\big)^T\mathbf{C}_S^m(\boldsymbol{\nu})^{-1}\big(\mathbf{x}^m-\boldsymbol{\xi}\boldsymbol{\omega}\big)\Big]p(\nu_i).$$

Here $|\cdot$ denotes conditioning on the rest of the parameters and the data. Note in the above formulas that $\boldsymbol{\lambda}_{ci}^f$ depends on the past temperatures $\widetilde{x}_{ci}^f$ and both $\boldsymbol{\lambda}_i^m$ and $\boldsymbol{\lambda}_{ci}^f$ depend on the temperature response parameters $\alpha_j$, $\beta_j$, $\gamma_j$ (cf. Sections 2.3 and 2.4). In MCMC simulation, the probabilities $\mathbf{p}_i^m$ and $\mathbf{p}_{ci}^f$ as well as the spatial trend parameter $\boldsymbol{\omega}$ can be updated using Gibbs sampling while all other parameters are updated using the Metropolis–Hastings algorithm.

**Acknowledgments.** We are grateful to Dr. Caspar Ammann from NCAR who provided us with the simulated temperature times series used in Section 2.6.

## SUPPLEMENTARY MATERIAL

**Supplement A: Additional analyses, reconstructions and description of the data** (DOI: 10.1214/15-AOAS832SUPPA; .pdf). The document (a pdf-file) includes an analysis of the Gaussian response model and its parameters, reference records from Greenland ice cores and Scandinavian lake sediments, additional reconstructions, a list of the core chronologies for the four lakes used for temperature reconstruction, and charts of relative abundances of the ten most common pollen taxa in the samples.

**Supplement B: The data** (DOI: 10.1214/15-AOAS832SUPPB; .zip). The data used in the article (an Excel file).

**Supplement C: The Matlab code** (DOI: 10.1214/15-AOAS832SUPPC; .zip). The Matlab code used in reconstructions.

## REFERENCES

ALLEY, R. B. and AGÚSTSDÓTTIR, A. M. (2005). The 8k event: Cause and consequences of a major Holocene abrupt climate change. *Quat. Sci. Rev.* **24** 1123–1149.

AMMANN, C. M., JOOS, F., SCHIMEL, D. S., OTTO-BLIESNER, B. L. and TOMAS, R. A. (2007). Solar influence on climate during the past millennium: Results from transient simulations with the NCAR climate system model. *Proc. Natl. Acad. Sci. USA* **104** 3713–3718.

ANTONSSON, K., BROOKS, S. J., SEPPÄ, H., TELFORD, R. J. and BIRKS, H. J. B. (2006). Quantitative palaeotemperature records iferred from fossil chironomid and pollen assemblages from Lake Gilltjärnen, northern central Sweden. *J. Quat. Sci.* **21** 831–841.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, London.

BIRKS, H. J. B., HEIRI, O., SEPPÄ, H. and BJUNE, A. E. (2010). Strengths and weaknesses of quantitative climate reconstructions based on late-quaternary biological proxies. *The Open Ecology Jounal* **3** 68–110.

BRYNJARSDÓTTIR, J. and BERLINER, L. M. (2011). Bayesian hierarchical modeling for temperature reconstruction from geothermal data. *Ann. Appl. Stat.* **5** 1328–1359. MR2849776

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. MR1239641

DAHL, E. (1998). *The Phytogeography of Northern Europe*: *British Isles*, *Fennoscandia*, *and Adjacent Areas*. Cambridge Univ. Press, Cambridge.

ERÄSTÖ, P. and HOLMSTRÖM, L. (2005). Bayesian multiscale smoothing for making inferences about features in scatterplots. *J. Comput. Graph. Statist.* **14** 569–589. MR2170202

ERÄSTÖ, P. and HOLMSTRÖM, L. (2006). Selection of prior distributions and multiscale analysis in Bayesian temperature reconstructions based on fossil assemblages. *J. Paleolimnol.* **36** 69–80.

ERÄSTÖ, P., HOLMSTRÖM, L., KORHOLA, A. and WECKSTRÖM, J. (2012). Finding a consensus on credible features among several paleoclimate reconstructions. *Ann. Appl. Stat.* **6** 1377–1405. MR3058668

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

HASLETT, J. and PARNELL, A. (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *J. Roy. Statist. Soc. Ser. C* **57** 399–418. MR2526125

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. Roy. Statist. Soc. Ser. A* **169** 395–438. MR2236914

HOLMSTRÖM, L., ILVONEN, L., SEPPÄ, H. and VESKI, S. (2015a). Supplement A to "A Bayesian spatiotemporal model for reconstructing climate from multiple pollen records." DOI:10.1214/15-AOAS832SUPPA. An on line supplement.

HOLMSTRÖM, L., ILVONEN, L., SEPPÄ, H. and VESKI, S. (2015b). Supplement B to "A Bayesian spatiotemporal model for reconstructing climate from multiple pollen records." DOI:10.1214/15-AOAS832SUPPB. The data used in reconstructions.

HOLMSTRÖM, L., ILVONEN, L., SEPPÄ, H. and VESKI, S. (2015c). Supplement C to "A Bayesian spatiotemporal model for reconstructing climate from multiple pollen records." DOI:10.1214/15-AOAS832SUPPC. The Matlab code used in reconstructions.

JANSEN, E. et al. (2007). Palaeoclimate. 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* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, eds.) 433–497. Cambridge Univ. Press, Cambridge.

JONES, P. D., BRIFFA, P. D., OSBORN, T. J., LOUGH, J. M., VAN OMMEN, T. D., VINTHER, B. M., LUTERBACHER, J., 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. *Holocene* **19** 3–49.

JOURNEL, A. G. and HUIJBREGTS, C. J. (1978). *Mining Geostatistics*. Academic Press, San Diego.

JUGGINS, S. and BIRKS, H. J. B. (2012). Quantitative environmental reconstructions from biological data. In *Tracking Environmental Change Using Lake Sediments*, *Data Handling and Numerical Techniques* **5** (H. J. B. Birks, A. F. Lotter, S. Juggins and J. P. Smol, eds.) 431–494. Springer, Dordrecht.

KAIPIO, J. and SOMERSALO, E. (2005). *Statistical and Computational Inverse Problems. Applied Mathematical Sciences* **160**. Springer, New York. MR2102218

KORHOLA, A., VASKO, K., TOIVONEN, H. T. T. and OLANDER, H. (2002). Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling. *Qaternary Science Reviews* **21** 1841–1860.

LI, B., NYCHKA, D. W. and AMMANN, C. M. (2010). The value of multi-proxy reconstruction of past climate. *J. Amer. Statist. Assoc.* **105** 883–911.

MARCOTT, S. A., SHAKUN, J. D., CLARK, P. U. and MIX, A. C. (2013). A reconstruction of regional and global temperature for the past 11,300 years. *Science* **339** 1198–1201.

MASSON-DELMOTTE, V. et al. (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* (T. F. Stocker, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia and P. M. Midgley, eds.) 383–464. Cambridge Univ. Press, Cambridge.

MOBERG, A. and BERGSTRÖM, H. (1997). Homogenization of Swedish temperature data. Part III: The long temperature records from Uppsala and Stockholm. *Int. J. Climatol.* **17** 667–699.

NRC (2006). *Surface Temperature Reconstructions for the Last 2000 Years*. The National Academies Press, Washington.

OHLWEIN, C. and WAHL, E. R. (2012). Review of probabilistic pollen-climate transfer methods. *Quat. Sci. Rev.* **31** 17–29.

PACIOREK, C. J. and MCLACHLAN, J. S. (2009). Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. *J. Amer. Statist. Assoc.* **104** 608–622. MR2751442

RENSSEN, H., SEPPÄ, H., CROSTA, X., GOOSSE, H. and ROCHE, D. M. (2012). Global characterization of the Holocene thermal maximum. *Quat. Sci. Rev.* **48** 7–19.

RENSSEN, H., SEPPÄ, H., HEIRI, O., ROCHE, D. M., GOOSSE, H. and FICHEFET, T. (2009). The spatial and temporal complexity of the Holocene thermal maximum. *Nat. Geosci.* **2** 411–414.

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. MR2080278

SALONEN, J. S., ILVONEN, L., SEPPÄ, H., HOLMSTRÖM, L., TELFORD, R. J., GAIDAMAVIČIUS, A., STANČIKAITĖ, M. and SUBETTO, D. (2012). Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative cllimate reconstructions. *Holocene* **22** 413–424.

SARMAJA-KORJONEN, K. and SEPPÄ, H. (2007). Abrupt and consistent responses of aquatic and terrestrial ecosystems to the 8200 cal. yr BP cold event: A lacustrine record from Lake Arapisto, Finland. *Holocene* **17** 455–464.

SEPPÄ, H., HAMMARLUND, D. and ANTONSSON, K. (2005). Low-frequency and high-frequency changes in temperature and effective humidity during the Holocene in south-central Sweden: Implications for atmospheric and oceanic forcings of climate. *Clim. Dyn.* **25** 285–297.

SEPPÄ, H. and POSKA, A. (2004). Holocene annual mean temperature changes in Estonia and their relationship to solar insolation and atmospheric circulation patterns. *Quat. Res.* **61** 22–31.

SEPPÄ, H., BJUNE, A. E., TELFORD, R. J., BIRKS, H. J. B. and VESKI, S. (2009). Last nine-thousand years of temperature variability in Northern Europe. *Clim. Past* **5** 523–535.

SHAKUN, J. D. et al. (2012). Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484** 49–54.

TER BRAAK, C. J. F. and JUGGINS, S. (1993). Weighted averaging partial least squares regression (WA-PLS): An improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* **269/270** 485–502.

TINGLEY, M. P., CRAIGMILE, P. F., HARAN, M., LI, B., MANNSHARDT-SHAMSELDIN, E. and RAJARATNAM, B. (2012). Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.* **35** 1–22.

TINGLEY, M. P. and HUYBERS, P. (2010a). A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1: Development and applications to paleoclimate reconstruction problems. *J. Climate* **23** 2759–2781.

TINGLEY, M. P. and HUYBERS, P. (2010b). A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 2: Comparison with the regularized expectation-maximization algorithm. *Journal of Climate* **23** 2782–2800.

TOIVONEN, H. T. T., MANNILA, H., KORHOLA, A. and OLANDER, H. (2001). Applying Bayesian statistics to organism-based environmental reconstruction. *Ecol. Appl.* **11** 618–630.

VASKO, K., TOIVONEN, H. T. T. and KORHOLA, A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *J. Paleolimnol.* **24** 243–250.

WIERSMA, A. P. and RENSSEN, H. (2006). Model data comparison for the 8.2 ka BP event: Confirmation of a forcing mechanism by catastrophic drainage of Laurentide Lakes. *Quat. Sci. Rev.* **25** 63–88.

WOODWARD, F. I. (1987). *Climate and Plant Distribution*. Cambridge Univ. Press, Cambridge.

L. HOLMSTRÖM
L. ILVONEN
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF OULU
P.O. BOX 3000
FIN-90014
FINLAND
E-MAIL: lasse.holmstrom@oulu.fi
liisa.ilvonen@oulu.fi

H. SEPPÄ
DEPARTMENT OF GEOSCIENCES
AND GEOGRAPHY
UNIVERSITY OF HELSINKI
P.O.BOX 64
FIN-00014
FINLAND
E-MAIL: heikki.seppa@helsinki.fi

S. VESKI
INSTITUTE OF GEOLOGY
TALLINN UNIVERSITY OF TECHNOLOGY
EHITAJATE TEE 5
19086 TALLINN
ESTONIA
E-MAIL: siim.veski@ttu.ee