FINDING A CONSENSUS ON CREDIBLE FEATURES AMONG SEVERAL PALEOCLIMATE RECONSTRUCTIONS

BY PANU ERÄSTÖ, LASSE HOLMSTRÖM, ATTE KORHOLA AND
JAN WECKSTRÖM

*National Institute for Health and Welfare and University of Oulu, University
of Oulu, University of Helsinki and University of Helsinki*

We propose a method to merge several paleoclimate time series into one that exhibits a consensus on the features of the individual times series. The paleoclimate time series can be noisy, nonuniformly sampled and the dates at which the paleoclimate is reconstructed can have errors. Bayesian inference is used to model the various sources of uncertainty and smoothing of the posterior distribution of the consensus is used to capture its credible features in different time scales. The technique is demonstrated by analyzing a collection of six Holocene temperature reconstructions from Finnish Lapland based on various biological proxies. Although the paper focuses on paleoclimate time series, the proposed method can be applied in other contexts where one seeks to infer features that are jointly supported by an ensemble of irregularly sampled noisy time series.

1. Introduction. Paleoclimatological proxy data, such as pollen, tree rings or ice cores, considered to be sensitive to past surface temperature variations can provide a continuous and long record of climatic changes where long-term instrumental data are lacking [Jansen et al. (2007)]. Paleoclimatological data are essential to place limited instrumental records in perspective and to assess the importance of forcing factors. However, it is important to realize that proxy records are indirect measures of climate change that often reflect changes in multiple aspects of climate [e.g., Legrande et al. (2006); Tingley et al. (2012)]. Each proxy inevitably has its advantages and limitations, and different proxies may yield information on different aspects of climate. For example, they may be sensitive to different seasonal signals, have different response times, and respond directly or indirectly to climate. It is therefore not surprising that, for example, temperature reconstructions based on different proxies can produce somewhat different results, despite the fact that they reflect a common underlying truth. One would therefore like to have a method that could capture, in a principled manner, those aspects of different reconstructions that find strongest support among most of them, that is, establish a "consensus" on the underlying features of the reconstructions.

To demonstrate the method suggested in this paper, we will find a consensus among the six Holocene, that is, post Ice Age mean air July temperature recon-
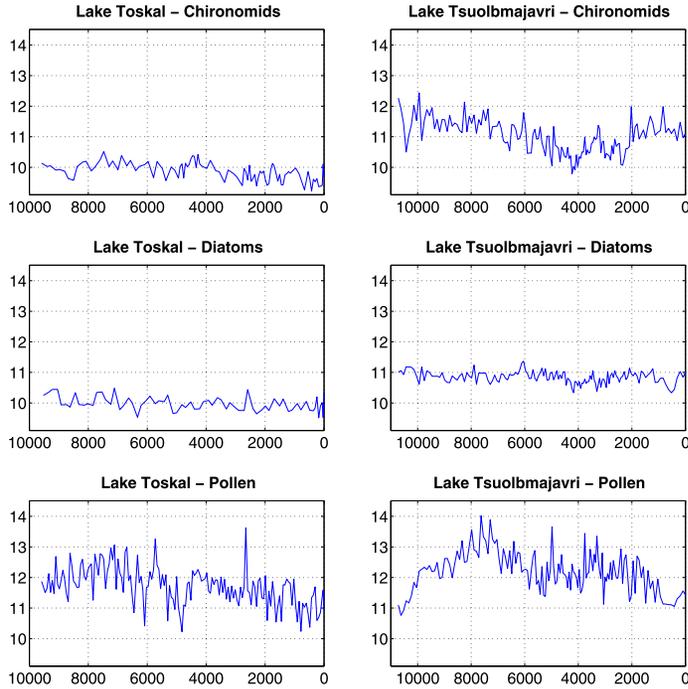
FIG. 1. *The six Holocene mean air July temperature reconstructions for Northern Fennoscandia used in the consensus analysis. The vertical axes show temperature in centigrade* (°C) *and the horizontal axes are calibrated years before present.*

structions shown in Figure 1. The reconstructions are based on three biological proxies analyzed from two lakes in Finnish Lapland and, as one can see, they differ from one another considerably, both in the overall temperature levels and in the details. The data behind the reconstructions and the consensus features the proposed method finds will be discussed in detail in Section 3, but let us first consider here some ad hoc methods that are often used to combine information across these types of paleoclimate time series. Such straightforward analyses are demonstrated in Figure 2. In the upper panel the reconstructions have been centered and then stacked into a single plot. A smooth has also been computed and it can be interpreted to represent the consensus temperature anomaly, that is, deviation from mean. In the lower panel the centered reconstructions have been averaged after first interpolating them with cubic splines or, alternatively, by smoothing them with local linear regression. While simple plots like these may reveal some features of the consensus anomaly, they clearly leave many questions unanswered. Individual time series are noisy, as both the reconstructed temperatures and the dates they are thought to correspond to contain errors. Such simple methods also tell us nothing about the uncertainty in the suggested consensus features that the presence of noise inevitably introduces. Further, the underlying signal may exhibit interesting
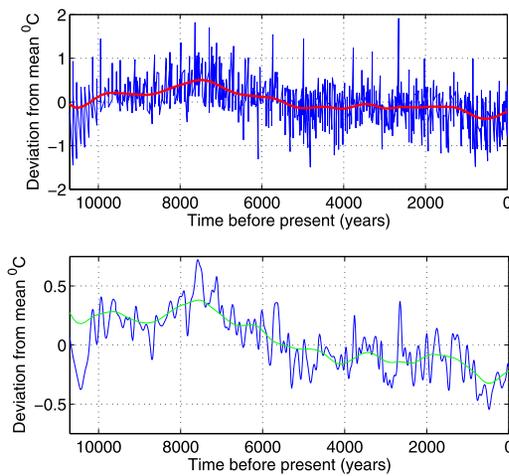
FIG. 2. *Simple methods to establish a consensus between temperature reconstructions. Upper panel*: *all six reconstructions of Figure* 1 *centered and stacked together* (*blue*) *and a local linear regression smooth* (*red*). *Lower panel*: *averages of cubic spline interpolants* (*blue*) *and local linear regression smooths* (*green*) *of the centered reconstructions. Local linear regression smooths employ a Gaussian kernel and bandwidths computed using a method from* Ruppert, Sheather and Wand (1995).

features in many different time scales and a single smooth or mean probably cannot capture all of them well.

In climate science, a popular approach to reconstruct large-scale past climate variation is to combine a number of individual proxy records using the so-called Composite Plus Scaling (CPS) method [e.g., Jones et al. (2009) and the references therein)]. In this method, a collection of proxy records is standardized and averaged after which the average is recalibrated against an available instrumental record of a particular environmental variable, such as temperature. In the calibration process, various regression techniques can be used to match an average of annually resolved proxy records with modern instrumental data. The method proposed in this paper works differently in that the individual reconstructions are not explicitly standardized or averaged and their consensus is found using an estimation process that does not directly rely on a modern instrumental record. Note that, contrary to the situation with annually resolved proxies such as tree rings, in the case of biological proxy records considered here only a few of the reconstructed temperatures would fall in a period for which instrumental measurements might be available, making regression based calibration unfeasible.

Our proposal to consensus analysis is a Bayesian approach that consists of two steps. First, given a set of reconstructions, we find their consensus by viewing the reconstructions as data in a hierarchical model that takes into account the uncertainties involved. In the second step we use scale space smoothing to reveal the salient features of the consensus in different time scales. The proposed approach was first outlined in Korhola et al. (2006) and Holmström et al. (2008) and it can be

viewed as an extension to multiple time series of the BSiZer methodology that has already found use in quantitative paleoecological analyses [Erästö and Holmström (2005, 2006, 2007); Holmström (2010a); Weckström et al. (2006)].

It can be argued that a better way to model the propagation of errors into the consensus would be to work directly with Bayesian temperature reconstructions instead of using a Bayesian model to combine non-Bayesian reconstructions, as is done here. However, while Bayesian models may be becoming more common-place, the vast majority of existing reconstructions are in fact non-Bayesian, based on various regression techniques, both parametric and nonparametric. See, for example, Birks (1995) and Birks et al. (2010) for extensive reviews of the kind of methods typically used in connection with diatoms, pollen, chironomids and other biological proxies. The method proposed here is therefore immediately widely applicable as a significant improvement over the simplistic ad hoc summaries commonly used to represent a consensus of such reconstructions.

To our knowledge, the first papers to describe a detailed Bayesian modeling approach to biological proxy based paleoclimate reconstruction are Vasko, Toivonen and Korhola (2000), Toivonen et al. (2001) and Korhola et al. (2002), who all used chironomid taxon abundances in lake sediments as temperature proxy. Their approach was further analyzed by Erästö and Holmström (2006) and more recently by Salonen et al. (2012). Bayesian reconstruction based on pollen abundances was described in Haslett et al. (2006). All these papers model explicitly the response of a biological proxy to temperature changes and reconstruct the temperature from taxon fossil abundance data in a single proxy record. More recently, a Bayesian hierarchical model was used by Brynjarsdóttir and Berliner (2011) to reconstruct climate for the past 400 years from several bore hole temperature profiles.

The approach suggested in Li, Nychka and Ammann (2010) is perhaps closer to the one proposed here in that a number of local reconstructions are combined to create a single temperature reconstruction, in their case for the whole northern hemisphere and the last 1000 years. As in the present paper, a biological proxy (pollen) enters the reconstruction process only as a temperature time series and not as raw taxon abundances, which would constitute the original data. In addition to pollen, tree rings and bore hole temperatures are also used in their model and external forcings are accounted for as well. However, no real proxy data are used and instead the proxy records are simulated on the basis of numerical climate model outputs. The reconstructions we aim to combine were obtained using taxon abundance data from actual sediment cores. Note that the same climate model simulation that was used in Li, Nychka and Ammann (2010) is employed also in the present paper but only to elicit a prior density for the consensus reconstruction. Other differences include the somewhat more general error models considered here, explicit modeling of dating uncertainty and the scale space approach to inference.

In Section 2 we describe our method, assuming first fixed dates for the reconstructed temperatures (Section 2.1) and then allowing dating errors in the analysis (Section 2.2). The idea of using multi-scale smoothing to capture temperature

variation in different time scales is explained in Section 2.3. The analysis of the consensus features in the six Holocene temperature reconstructions is presented in Section 3 and Section 4 offers a discussion of the main points of the paper. The Matlab functions used in the main computations are provided in Erästö et al. (2011b).

2. The method.

2.1. *Fixed dates.*   The method that we will describe can be used to analyze reconstructions of any continuous variable, but as our main interest is in the Holocene temperature, we frame the following description in terms of temperature reconstructions. Thus, consider m reconstructions $\mathbf{y}_1, \ldots, \mathbf{y}_m$ of past temperatures, where $\mathbf{y}_k = [y_{k1}, \ldots, y_{kj_k}]^T$ are the estimated past temperatures from the kth proxy series and let $\mathbf{t}_k = [t_{k1}, \ldots, t_{kj_k}]$ be the associated radiocarbon dating based chronology. Here $t_{k1} < \cdots < t_{kj_k}$ so that y_{k1} and y_{kj_k} are the reconstructions for the oldest and the youngest dates, respectively. We assume that the reconstructions are from a relatively limited geographical area so that they can be thought to reflect common underlying temperature variation and it is this common variation that we seek to capture.

In the example we will consider the reconstructions are based on fossil records in sediment cores obtained from subarctic lakes. Even when the cores come from a limited area, due to, for example, different lake altitudes, the overall temperature levels and therefore the mean temperatures in the reconstructions can vary considerably. We therefore consider only temperature anomalies, centering each reconstruction $\mathbf{y}_k$ by subtracting its mean $(1/j_k) \sum_l^{j_k} y_{kl}$ from all components y_{kl}. These centered time series represent reconstructions of past temperature anomalies (variation about the mean) and we attempt to capture the statistically significant (or "credible") features in what can be interpreted as the consensus of these anomalies in the general area where the core lakes are located. The features in the consensus that we are interested in are locations of maxima, minima and trends, all of which are not affected by centering. To avoid the introduction of new notation, we denote the centered reconstructions still by $\mathbf{y}_k$.

The consensus anomaly is modeled as a curve $\mu(t)$, where $t \in [a, b]$ is a time interval that includes all chronologies from all proxy records. We actually assume that μ can be described by a natural cubic spline with knots at the points t_{kj_l}. Such a spline is uniquely determined by its values at the knots because they determine the interpolating spline uniquely [Green and Silverman (1994)]. The fact that this spline space is finite dimensional greatly simplifies our analysis.

Let

$$(1) \qquad \mathbf{t} = \{t_1, \ldots, t_n\} = \bigcup_{k}^{m} \{t_{k1}, \ldots, t_{kj_k}\}$$

be the set of distinct dates, in increasing order, in all chronologies $\mathbf{t}_k$. Since all t_{kl}'s need not be different, we have that $n \leq j_1 + \cdots + j_m$. The anomaly curve is modeled as a natural cubic spline with values $\mu_i = \mu(t_i)$ at the knots t_i. Thus, instead of μ, we can from now on work with the finite dimensional vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$ of past anomalies at times t_i.

Now, let $\boldsymbol{\mu}_k$ be the part of $\boldsymbol{\mu}$ that corresponds to the chronology $\mathbf{t}_k$ of the kth reconstruction $\mathbf{y}_k$. We assume that

(2)
$$\mathbf{y}_k = \boldsymbol{\mu}_k + \boldsymbol{\varepsilon}_k,$$

where $\boldsymbol{\varepsilon}_k$ has the multivariate normal distribution $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ with an unknown covariance matrix $\boldsymbol{\Sigma}_k$. Our model therefore allows time-varying, correlated reconstruction errors that can also have different magnitudes for different proxies and cores. Such a model is supported by the exploratory analysis reported in Erästö et al. (2011a). We further assume that the anomalies are conditionally independent given the parameters $\boldsymbol{\mu}$ and $\{\boldsymbol{\Sigma}_k\} = \{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m\}$ so that the likelihood of the data $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T]^T$, given these parameters, is

(3)
$$p(\mathbf{y}|\boldsymbol{\mu}, \{\boldsymbol{\Sigma}_k\}) \propto \prod_{k=1}^{m} |\boldsymbol{\Sigma}_k|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}_k - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_k - \boldsymbol{\mu}_k)\right].$$

As a prior distribution for $\boldsymbol{\Sigma}_k$ we use an Inverse Wishart distribution,

(4)
$$p(\boldsymbol{\Sigma}_k|\mathbf{W}_k, \nu_k) \propto |\boldsymbol{\Sigma}_k|^{-(\nu_k+j_k+1)/2} \exp\left[-\frac{1}{2} \operatorname{tr}(\mathbf{W}_k \boldsymbol{\Sigma}_k^{-1})\right],$$

a standard choice in connection with a multivariate normal likelihood. As there seldom is any prior knowledge of a particular error correlation structure, we typically use a diagonal prior scale matrix $\mathbf{W}_k$ and select the degrees of freedom ν_k so that the prior (4) is rather vague, allowing nondiagonal posterior covariances. The relative magnitudes of the diagonal elements of $\mathbf{W}_k$ could also be used to model the increased level of difficulty of temperature reconstruction for the older sediment layers [Erästö and Holmström (2007)]. The $\boldsymbol{\Sigma}_k$'s are assumed to be independent a priori so that

(5)
$$p(\{\boldsymbol{\Sigma}_k\}) = p(\{\boldsymbol{\Sigma}_k\}|\{\mathbf{W}_k, \nu_k\}) = \prod_{k=1}^{m} p(\boldsymbol{\Sigma}_k|\mathbf{W}_k, \nu_k).$$

We have also experimented with a more complex model that allows reconstruction error correlations between different proxy records. Let again $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T]^T$ be the vector of length $j_1 + \cdots + j_m$ that contains all reconstructions. The more complex model considered assumes that

(6)
$$p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})\right],$$

where $\mathbf{G}\boldsymbol{\mu}$ is a modification of the consensus $\boldsymbol{\mu}$ where some components μ_i appear several times to account for the fact they correspond to dates in the joint chronology that appear in more than one reconstruction. The covariance matrix $\boldsymbol{\Sigma}$ again

has an inverse-Wishart prior

$$(7) \qquad p(\mathbf{\Sigma}|\mathbf{W}, \nu) \propto |\mathbf{\Sigma}|^{-(\nu+j+1)/2} \exp[-\tfrac{1}{2}\operatorname{tr}(\mathbf{W}\mathbf{\Sigma}^{-1})],$$

where now $j = j_1 + \cdots + j_m$ and $\mathbf{W}$ is the diagonal matrix whose diagonal elements are those of the matrices $\mathbf{W}_1, \ldots, \mathbf{W}_m$. The results reported in the paper all pertain to the model (3) and the more complex model (6) is discussed in Erästö et al. (2011a).

For the consensus anomaly $\boldsymbol{\mu}$ we use a smoothing prior that penalizes for roughness as measured by the variability of its components,

$$(8) \qquad p(\boldsymbol{\mu}|\lambda_0, \mathbf{t}) \propto \lambda_0^{(n-2)/2} \exp\left(-\frac{\lambda_0}{2}\boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu}\right).$$

In this formula, $\mathbf{K}$ is a symmetric positive semidefinite matrix such that

$$(9) \qquad \boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu} = \int_a^b [\mu''(t)]^2 \, dt$$

and $\lambda_0 > 0$. Thus, the roughness in the prior (8) is measured by the second derivative of the natural cubic spline that interpolates the values $\boldsymbol{\mu}$ at the knots t_i and the level of roughness penalty is controlled by λ_0 [Green and Silverman (1994)]. The power $(n-2)/2$ in the scaling factor reflects the rank of the matrix $\mathbf{K}$ which is $n-2$. Note that the smoothing prior (8) imposes dependence between the temperature anomalies $\boldsymbol{\mu}_k$ derived from these proxies. This is natural because the reconstructions are assumed to reflect common underlying temperature variation.

The parameter λ_0 describes our prior beliefs about the smoothnesss of μ. We consider it unknown with prior uncertainty described by a Gamma distribution. In principle, point estimation such as cross-validation can be used to choose suitable values for the prior distribution parameters [Erästö and Holmström (2005)], but we prefer here a choice that produces a posterior mean of μ of reasonable roughness. The important thing is to avoid choosing λ_0 too large because then the finest details of μ might be lost [Erästö and Holmström (2005, 2007)].

The joint posterior distribution of all the unknown parameters in the model is now obtained from the Bayes' formula,

$$(10) \qquad p(\boldsymbol{\mu}, \{\mathbf{\Sigma}_k\}, \lambda_0|\mathbf{y}, \mathbf{t}) \propto p(\lambda_0)p(\{\mathbf{\Sigma}_k\})p(\boldsymbol{\mu}|\lambda_0, \mathbf{t})p(\mathbf{y}|\boldsymbol{\mu}, \{\mathbf{\Sigma}_k\}),$$

where all the distributions on the right-hand side were defined above. Gibbs sampling can be used to generate a sample from this posterior distribution. An estimate of the consensus anomaly that is consistent with the data and our prior beliefs, together with its uncertainty, is described by the marginal posterior distribution $p(\boldsymbol{\mu}|\mathbf{y})$, which then can be approximated by the $\boldsymbol{\mu}$-component of this sample. The model (6) is handled similarly.

2.2. *Random dates.* In the previous section we assumed that the reconstructed temperature anomalies y_{kl} could be associated precisely with the dates t_{kl}. In reality, however, the core chronologies are derived from radiocarbon dating based estimates, a process that is not error-free. Taking into account this source of uncertainty can be important when one tries to make inferences about the common features in several temperature time series with different associated chronologies.

Let $\mathbf{t}_k = [t_{k1}, \ldots, t_{kj_k}]$ again be the radiocarbon dating based chronology for the kth reconstruction. Allowing for the fact that the dates t_{kl} have errors, we assume that they and the dates τ_{kl} in the true, unobserved chronology, satisfy $t_{kl} = \tau_{kl} + \delta_{kl}$, where δ_{kl} represents an error. Denote the true chronology for the kth reconstruction by $\boldsymbol{\tau}_k = [\tau_{k1}, \ldots, \tau_{kj_k}]$. We assume that both sequences $\mathbf{t}_k$ and $\boldsymbol{\tau}_k$ are strictly increasing. Note that, for $k \neq k'$, $\boldsymbol{\tau}_k$ and $\boldsymbol{\tau}_{k'}$ may well contain some dates that are known to be the same. This is the case, for example, when k and k' correspond to two different proxies analyzed from the same core and using the same sediment samples for both. Let

$$(11) \qquad \boldsymbol{\tau} = \{\tau_1, \ldots, \tau_n\} = \bigcup_k^m \{\tau_{k1}, \ldots, \tau_{kj_k}\}$$

be the set of distinct dates in all chronologies $\boldsymbol{\tau}_k$, $k = 1, \ldots, m$ [cf. (1)]. As with the dates t_{kl} in the previous section, since all τ_{kl}'s need not be different, we have in general that $n \leq j_1 + \cdots + j_m$. The observed dates t_{kl} for equal τ_{kl}'s are assumed to be also equal and we denote by $\mathbf{t} = \{t_1, \ldots, t_n\}$ the set of t_{kl}'s corresponding to $\boldsymbol{\tau}$. Our model for these distinct dates now is

$$(12) \qquad t_i = \tau_i + \delta_i,$$

$i = 1, \ldots, n$, and we assume that, given the parameters τ_i, the δ_i's are independent zero mean normal variables with known variances $\psi_i^2 > 0$. The variances that we will use are based on the standard errors associated with the chronologies (cf. Section 3.2). The likelihood of the observed dates $\mathbf{t}$ from (12) is

$$(13) \qquad p(\mathbf{t}|\boldsymbol{\tau}) = p(\mathbf{t}|\boldsymbol{\tau}, \{\psi_i^2\}) \propto \prod_{i=1}^n \psi_i^{-1} \exp\left[-\frac{1}{\psi_i^2}(t_i - \tau_i)^2\right],$$

where $\{\psi_i^2\} = \{\psi_1^2, \ldots, \psi_n^2\}$. We set a prior distribution on the τ_i's that enforces the correct temporal order of the chronology within each reconstruction,

$$(14) \qquad p(\boldsymbol{\tau}) \propto \prod_{k=1}^m \mathbb{1}(\tau_{k1} < \tau_{k2} < \cdots < \tau_{kj_k}).$$

Let now $\tau_{(1)} < \cdots < \tau_{(n)}$ be a permutation of $\boldsymbol{\tau}$ into an ascending order. The consensus anomaly is then modeled as natural cubic spline $\mu(\tau)$ with knots at the points $\tau_{(i)}$, uniquely determined by the vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^T$, $\mu_i = \mu(\tau_{(i)})$. The subsequent model details are exactly the same as in the previous section with

the exception that in the prior (8) of $\boldsymbol{\mu}$, the matrix $\mathbf{K}$ now depends on $\boldsymbol{\tau}$. The joint posterior (10) becomes

(15)
$$p(\boldsymbol{\mu}, \{\boldsymbol{\Sigma}_k\}, \lambda_0, \boldsymbol{\tau} | \mathbf{y}, \mathbf{t}) \propto p(\lambda_0) p(\{\boldsymbol{\Sigma}_k\}) p(\boldsymbol{\tau}) p(\boldsymbol{\mu} | \lambda_0, \boldsymbol{\tau})$$
$$\times\, p(\mathbf{y} | \boldsymbol{\mu}, \{\boldsymbol{\Sigma}_k\}) p(\mathbf{t} | \boldsymbol{\tau}).$$

A hybrid algorithm that uses Gibbs and Metropolis–Hastings Monte Carlo sampling can be used to generate a sample from this posterior distribution [e.g., Robert and Casella (2005)]. The proposal density for τ_i is $N(0, 10^{-2}\psi_i^2)$. Again, the model (6) can be handled similarly. For easy reference, Table 1 summarizes the quantities defined in this and the previous section.

2.3. *Scale space feature analysis.* The two previous sections showed how to estimate the consensus of several temperature reconstructions. This section explains how to find its credible features in different time scales. The key idea is that of a scale space. This concept has its roots in computer vision, but it has recently inspired a host of new statistical data analysis tools based on multi-scale smoothing. For an overview of these methods we refer to Holmström (2010b).

In the context of this article, the scale space approach amounts to using smoothing to make inferences about the credible, or "statistically significant," features of the consensus anomaly μ underlying the data. Thus, suppose that S_λ is a smoothing operator associated with a smoothing level $\lambda > 0$ and let $\mu_\lambda = S_\lambda \mu$ be the corresponding smooth of μ. In the classical scale space literature [e.g., Lindeberg (1994)], the smoother S_λ would typically be a Gaussian convolution (moving average with Gaussian weights) with convolution kernel width (the averaging window) determined by λ. However, in the statistical literature other smoothers are often used.

The idea is to make inferences about the features of μ_λ for a range of smoothing levels λ. Each μ_λ is interpreted to reveal features of μ at a certain time scale, little smoothing (small λ) showing the short time scale variation and heavy smoothing (large λ) revealing the coarsest features, such as the overall trend. We are, in particular, interested in the maxima and minima of μ_λ and therefore base our inferences on the derivative μ_λ' because its sign tells where the local trend is positive or negative. For Bayesian reasoning we need the posterior $p(\mu_\lambda' | \mathbf{y}, \mathbf{t})$. However, as the spline μ is uniquely represented by the vector $\boldsymbol{\mu}$ of its values at the knots, we may instead consider a smoothing matrix $\mathbf{S}_\lambda$, the smooth $\boldsymbol{\mu}_\lambda = \mathbf{S}_\lambda \boldsymbol{\mu}$, and then use another matrix $\mathbf{D}$ [e.g., Green and Silverman (1994)] to evaluate the derivative μ_λ' at some fixed dense set of time points $s_1 < \cdots < s_r$,

(16)
$$\mathbf{D}\boldsymbol{\mu}_\lambda = [\mu_\lambda'(s_1), \ldots, \mu_\lambda'(s_r)]^T.$$

The smoothing matrix used in our scale space feature analysis is defined as $\mathbf{S}_\lambda = (\mathbf{I} + \lambda \mathbf{K})^{-1}$ and it actually smooths a discrete set of points $\boldsymbol{\mu}$ by fitting a smoothing spline [Green and Silverman (1994)]. Instead of $p(\mu_\lambda' | \mathbf{y}, \mathbf{t})$, one can now analyze

TABLE 1

Glossary of symbols used, their associated likelihoods or priors and the full conditional posteriors of the estimated parameters. The multivariate normal distribution $\mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ in the conditional posterior of $\boldsymbol{\mu}$ is obtained as the product of (3) and (8) and it is discussed in Appendix B. In the conditional posterior of $\boldsymbol{\tau}$ we denote $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1^2, \ldots, \psi_n^2)$ [cf. (13)] and the proposal density for τ_i is $\mathrm{N}(0, 10^{-2}\psi_i^2)$

Symbol	Meaning	Likelihood or prior	Full conditional posterior
$\mathbf{y}_k$	reconstructed anomaly for proxy record k	(3)	
$\mathbf{y}$	$[\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T]^T$	(6)	
$\boldsymbol{\mu}$	consensus anomaly	(8)	$\boldsymbol{\mu}\|\{\boldsymbol{\Sigma}_k\}, \lambda_0, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$
$\boldsymbol{\mu}$	consensus anomaly (extended model)	(8)	$\boldsymbol{\mu}\|\{\boldsymbol{\Sigma}\}, \lambda_0, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{N}((\mathbf{G} + \lambda_0\boldsymbol{\Sigma}^{-1}(\mathbf{G}^T)^{-1}\mathbf{K})\mathbf{y}, (\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} + \lambda_0\mathbf{K})^{-1})$
λ_0	prior smoothing parameter of $\boldsymbol{\mu}$	Gamma(η, β)	$\lambda_0\|\boldsymbol{\mu}, \{\boldsymbol{\Sigma}_k\}, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{Gamma}((n-2)/2 + \eta, \boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu}/2 + \beta)$
$\boldsymbol{\mu}_k$	part of $\boldsymbol{\mu}$ corresponding to proxy record k		
$\boldsymbol{\varepsilon}_k$	$\mathbf{y}_k - \boldsymbol{\mu}_k$		
$\boldsymbol{\Sigma}_k$	covariance of $\boldsymbol{\varepsilon}_k$	(4)	$\boldsymbol{\Sigma}_k\|\boldsymbol{\mu}, \lambda_0, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{Inv\text{-}Wishart}_{\nu_k+1}([(\mathbf{y}_k - \boldsymbol{\mu}_k)(\mathbf{y}_k - \boldsymbol{\mu}_k)^T + \mathbf{W}_k]^{-1})$
$\boldsymbol{\Sigma}$	covariance of $[\boldsymbol{\varepsilon}_1^T, \ldots, \boldsymbol{\varepsilon}_m^T]^T$	(7)	$\boldsymbol{\Sigma}\|\boldsymbol{\mu}, \lambda_0, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{Inv\text{-}Wishart}_{\nu+1}([(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})^T + \mathbf{W}]^{-1})$
$\mathbf{t}_k$	chronology for proxy record k		
$\mathbf{t}$	set of distinct dates in the chronologies $\mathbf{t}_k$	(13)	
$\boldsymbol{\tau}_k$	true chronology for proxy record k		
$\boldsymbol{\tau}$	set of distinct dates in the true chronologies $\boldsymbol{\tau}_k$	(14)	$\boldsymbol{\tau}\|\boldsymbol{\mu}, \{\boldsymbol{\Sigma}_k\}, \lambda_0, \mathbf{y}, \mathbf{t} \propto \exp(-\frac{1}{2}((\boldsymbol{\tau} - \mathbf{t})^T\boldsymbol{\Psi}^{-1}(\boldsymbol{\tau} - \mathbf{t}) + \lambda_0\boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu}))p(\boldsymbol{\tau})$

the posterior distribution $p(\mathbf{DS}_\lambda \boldsymbol{\mu} | \mathbf{y}, \mathbf{t})$. For fixed dates, a large sample can first be generated from $p(\boldsymbol{\mu} | \mathbf{y}, \mathbf{t})$ and then transformed by multiplying the sample vectors by the matrix $\mathbf{DS}_\lambda$. Inference about the features of μ at the time scale λ is then based on this sample. With random dates, the scale space analysis needs samples from both $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$, as the smoothing matrix $\mathbf{S}_\lambda$ depends on $\boldsymbol{\tau}$ through $\mathbf{K}$.

Note here the difference between the parameter λ_0 used in constructing the consensus and the parameter λ in scale space feature analysis: λ_0 describes our prior beliefs about the underlying consensus μ, whereas different values of λ are used to explore the features of μ in different time scales. The choice of prior distribution for λ_0 is discussed in Section 3.3.2. We also emphasize that all inferences on the features of μ are made in a simultaneous fashion, over all time points s_j in (16). Therefore, instead of just examining the statistical significance of individual slopes $\mu(s_j)$, the credibility of whole patterns of trends are established. For more details on the inference procedures used we refer to Erästö and Holmström (2005).

3. Holocene temperature variation in Finnish Lapland.

3.1. *The data used.* We demonstrate the proposed method by finding the consensus among six temperature reconstructions based on high resolution lake sedimentary data (50–70 year intervals) of three biological proxies from two sites (Figure 1). The two lakes, Toskal and Tsuolbmajavri, selected for analysis are located at a climatically sensitive tree-line region of Finnish Lapland. They both contain fossil records of three fundamental climate proxies, pollen, chironomids (nonbiting midges) and diatoms (unicellular micro-algae) from the same sediment cores. The sediments of such remote lakes at high altitudes and latitudes are perhaps one of the few systems where a continuous, high resolution record of terrestrial environmental variability, uninfluenced by human impact throughout the post-glacial, can be found.

Past temperatures were reconstructed using regional training sets of lakes for pollen, chironomids and diatoms (304, 62 and 64 lakes, resp.) and a regression based reconstruction technique referred to as weighted averaging partial least squares (WA-PLS) [ter Braak and Juggins (1993)]. The model relates the modern mean July temperatures at the training lakes to the abundances of various proxy taxa preserved in the top (0–1 cm) surface sediments that represent the last few years of sediment accumulation. The past air temperatures are reconstructed by substituting in the regression model the taxon abundances found in the sediment cores from the two lakes selected for analysis. This approach is based on the assumption that each taxon has a certain optimal temperature at which it fares particularly well and that, therefore, the relative abundances of taxon fossils in a sediment layer reflect the temperature at the time the sediment layer was formed. For more details regarding the training sets and reconstruction models, see Seppä and Birks (2001), Seppä et al. (2002) and Weckström et al. (2006).

The sediment records are supported by chronologies based on multiple AMS 14C determinations [Seppä and Birks (2001); Seppä et al. (2002)]. As the chronology inevitably contains errors, an attempt is made to take this uncertainty into account by using the model described in Section 2.2. Table S.2 in Erästö et al. (2011a) gives all the data used in our consensus analysis: the sediment depths, calibrated ages and their standard errors as provided by the dating laboratory, as well as pollen-, chironomid- and diatom-based July mean temperature reconstructions for the lakes Toskal and Tsuolbmajavri.

3.2. *Chronology errors, prebinning.* The combined chronology (1) includes several pairs of dates with only a few years apart. The spline interpolant used in representing the consensus temperature anomaly as a continuous function $\mu(t)$ can exhibit unnatural wiggles between such nearby dates and we therefore aggregated the dates into 15 year wide bins. The chronology standard errors of aggregated dates could then be averaged, but we actually decided to smooth all of them as shown in Figure 3 and computed the parameters ψ_i in (13) from the values of this smooth. It retains the most important feature of the dating errors, namely, that they increase considerably when older sediment layers are considered. These approximations seem reasonable given the large standard errors associated with the dates and the rather simplistic dating error model (12) used.

3.3. *Priors for reconstruction errors and roughness.*

3.3.1. *Reconstruction error.* The prior distribution (4) of Σ_k has the mean $\mathbb{E}(\Sigma_k) = (v_k - j_k - 1)^{-1}\mathbf{W}_k$, where j_k is the dimension of the kth reconstruction $\mathbf{y}_k$. We use a diagonal scale matrix $\mathbf{W}_k = w_k\mathbf{I}_{j_k}$ such that $\mathbb{E}(\Sigma_k) = \bar{\sigma}_k^2\mathbf{I}_{j_k}$,
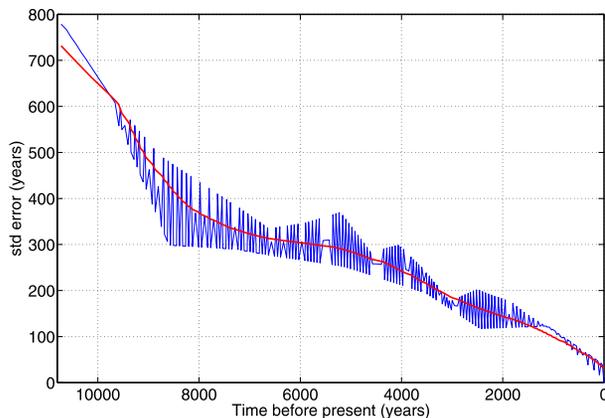

FIG. 3. *Standard errors of the combined binned chronology of the two sediment cores* (blue). *Average standard error is plotted when two or more dates coincide after binning. Also shown is a local linear smooth that was used in defining the parameters ψ_i of the dating model likelihood* (13).

where $\bar{\sigma}_k^2$ is an estimate for the upper bound of reconstruction error variance. Appendix A suggests a method to derive such upper bound estimates and the values obtained are given in Table 2. Since now $\bar{\sigma}_k^2 \mathbf{I}_{j_k} = (\nu_k - j_k - 1)^{-1} w_k \mathbf{I}_{j_k}$, we must have $w_k / \bar{\sigma}_k^2 = \nu_k - j_k + 1$. We set $w_k = 0.5$ for all k which corresponds to degrees of freedom ν_k between 77.9 and 163.1 and makes the priors rather vague.

The posterior values of the diagonal elements of the matrices $\boldsymbol{\Sigma}_k$ turned out to be significantly smaller than their prior values. As this may suggest that the values $\bar{\sigma}_k$ are too large (and thus truly only upper bounds), we also included in our analyses a second set of error covariance priors by using the value $\bar{\sigma}_k = 0.2$ for all reconstructions. In this case we opted for a tighter prior by taking $w_k = 50$ which corresponds to between 1319 and 1410 degrees of freedom in the priors.

Assuming smaller errors naturally leads to more features in the consensus analysis being flagged as credible. However, the independent evidence for some of these features discussed in Section 3.5 can be interpreted as lending some credence to these smaller reconstruction errors. Trying out different error sizes makes sense also because it probably is not possible to estimate them very reliably in the first place. Exploring temperature features for different error levels could also be thought as a form of scale space analysis where increasing error levels corresponds to more smoothing. In the following we refer to these two prior settings as "large" and "small" errors.

3.3.2. *Roughness.*   The parameter λ_0 in (8) is used to describe our prior belief about the variability or "roughness" of the time series of past temperatures. In choosing a prior for λ_0, very long instrumental records going back hundreds of years might be useful. However, the longest records in Finland span only about 150 years, a period that includes only 2–4 chronology dates for the six reconstructions considered, thus making roughness estimation impossible. We therefore decided to use a numerical climate model simulation in setting the prior roughness level.

A 1150 year long annual mean July temperature series for Northern Finland, extending from AD 850 to 1999, was extracted from the NCAR Climate System Model simulation described in Ammann et al. (2007). The time series is shown in Figure 4 (blue curve). The six reconstructions should actually be thought of as 30-year averages of mean July temperatures, sampled at dates included in their associated chronologies. For visual comparison between the simulation and the reconstructions we therefore applied a 30-year moving average to the simulated anomaly (red curve in Figure 4) and then sampled the average at the dates in the reconstruction chronologies. The results are shown in Figure 5. As one can see, the reconstructions are at least as rough as the simulation. It therefore appears that at least some prior smoothing indeed is required in the consensus analysis which motivates the use of a smoothing prior (8) for the consensus. Further, if the simulation is taken to represent the actual temperature variation, the reconstruction errors are not very large. The light blue band around each reconstruction is based on error bars of size $\pm 2\bar{\sigma}_k$, where the $\bar{\sigma}_k$'s are given in Table 2 of the Appendix.
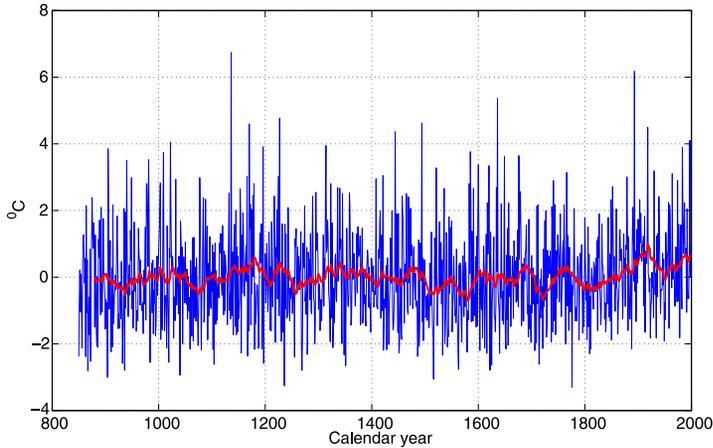
FIG. 4.   *Simulated mean July temperature anomaly for Northern Finland between AD* 850 *and* 1999 (*blue curve*) *together with the 30-year running mean* (*red curve*). *The vertical axis is the temperature anomaly in centigrade* (°C) *and the horizontal axis is the calendar year.*

To design a prior for λ_0, one can use the simulated time series also for more formal roughness estimation. Given a time series $\boldsymbol{\mu}$, one can measure its roughness by the quantity $R(\boldsymbol{\mu}) = \boldsymbol{\mu}^T \mathbf{K} \boldsymbol{\mu}$ in the exponent of (8). For the simulated 30-year running mean, evaluated at the joint chronology dates (1) contained in the interval from AD 850 to 1999, we have $R(\boldsymbol{\mu}) = 2.1 \cdot 10^{-4}$. Using the prior Gamma(20, 0.5) for λ_0, the posterior mean of $R(\boldsymbol{\mu})$ is $2.2 \cdot 10^{-4}$ and $2.5 \cdot 10^{-4}$ for the large and small prior errors, respectively. In both cases the mean posterior roughness of the consensus is therefore slightly larger than that of the simulations which, as indicated in Section 2.1, is desirable in order not to smooth too much before scale space analysis is carried out. We therefore used Gamma(20, 0.5) as the prior distribution for λ_0. Figure 6 shows the posterior distribution of $R(\boldsymbol{\mu})$ for both large and small prior error settings with the roughness of the simulation depicted as a dashed line. By testing other reasonable alternatives we also concluded that neither the mean nor the width of the prior distribution of λ_0 has a major effect on the estimated consensus features.

3.4. *The consensus and its credible features.*   Scale space analyses of the consensus anomaly with large and small prior reconstruction errors are shown in Figures 8 and 9, respectively. The top panel shows the reconstructed temperature anomalies (dots) together with the posterior mean of the consensus (blue curve). The middle panel shows the posterior mean again together with three smooths $\mathbb{E}(\mu_\lambda | \mathbf{y}, \mathbf{t})$ of the posterior consensus corresponding roughly to multidecadal (light blue), centennial (purple) and millennial (yellow) time scales (cf. Section 2.3). Comparing with the ad hoc methods discussed in the Introduction, we observe that there is a qualitative correspondence between the smoothing based
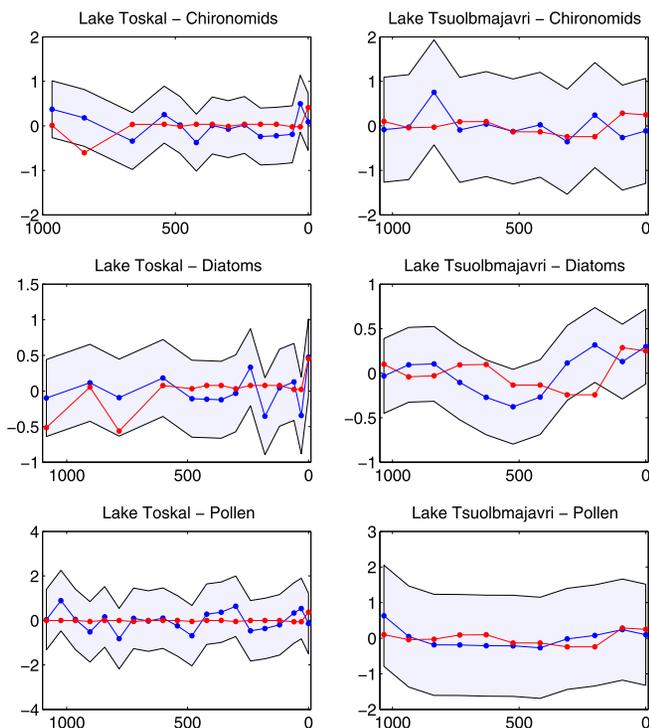
FIG. 5. *The six Holocene mean July temperature reconstructions for Northern Fennoscandia restricted to the time interval from AD* 850 *to* 1999 (*blue curves*) *together with the simulated* 30*-year means computed at the same time points* (*red curves*). *The light blue band around each reconstruction is based on error bars of size* $\pm 2\bar{\sigma}_k$, *where the* $\bar{\sigma}_k$'s *are given in Table* 2 *of the Appendix. The vertical axes show temperature anomaly in centigrade* (°C) *and the horizontal axes are time before present in years. Note the different temperature scales in the figures.*

curves of Figure 2 (red and green curves) and the centennial level posterior means of our scale space analyses as well as between the mean of the spline interpolants (lower panel, blue curve) and our multi-decadal posterior mean.

The bottom panel is a feature credibility map where the vertical axis represents the smoothing level λ (in logarithmic units), that is, the time scale at which the features are examined. The smoothing levels corresponding to the three smooths of the middle panel are indicated by horizontal lines of the same color. A pixel at a location (s_j, λ) is colored blue or red depending on whether the slope of the smoothed anomaly μ_λ is credibly negative or positive. Thus, blue and red indicate cooling and warming, respectively, at the particular time s_j and scale λ considered. Flagging of negative and positive slopes is based on their joint posterior probability which is required to exceed a given threshold α, typical values used being in the range [0.8, 0.95]. Gray color indicates that the sign of the slope is not credibly different from zero.
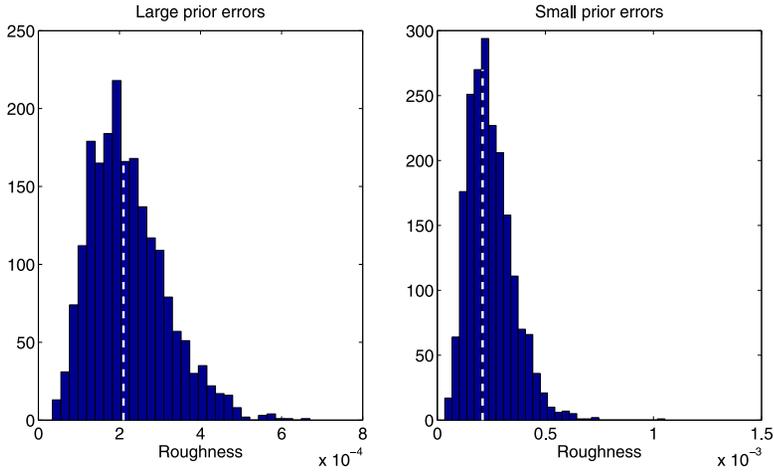
FIG. 6.   *Posterior distribution of the roughness measure $R(\boldsymbol{\mu}) = \boldsymbol{\mu}^T \mathbf{K} \boldsymbol{\mu}$ for large* (*left panel*) *and small* (*right*) *prior errors. The histograms are based on* 2000 *sample values and the dashed line indicates the roughness of the simulation.*

Figure 7 is a schematic illustration of how the map is drawn, focusing on the interval from 2729 to 2604 years before present and a multi-decadal smoothing level λ. In the upper panel, a few sample curves of μ_λ (green) together with the posterior mean $\mathbb{E}(\mu_\lambda | \mathbf{y}, \mathbf{t})$ (blue) are shown. The lower panel shows the corre-
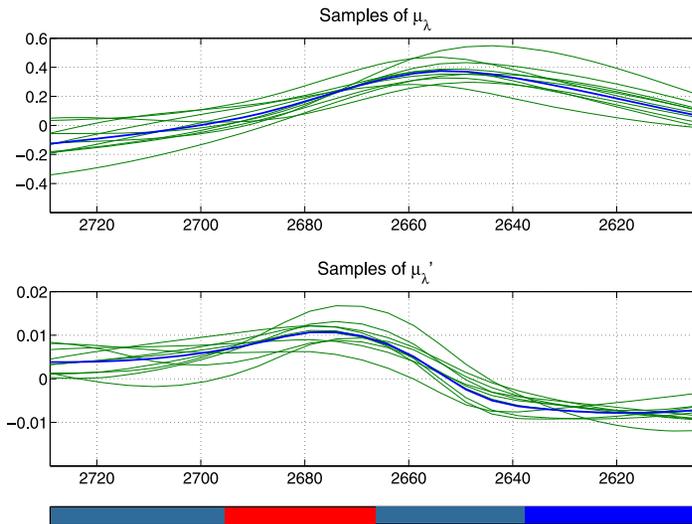

FIG. 7.   *Upper panel*: *sample curves of μ_λ* (*green*) *together with the posterior mean $\mathbb{E}(\mu_\lambda | \mathbf{y}, \mathbf{t})$* (*blue*). *Lower panel*: *corresponding samples of μ_λ' and the posterior mean $\mathbb{E}(\mu_\lambda' | \mathbf{y}, \mathbf{t})$. The color bar on the bottom depicts posterior sample based inference on the sign of μ_λ'. For more information, see the text.*

sponding samples of μ_λ' and the posterior mean $\mathbb{E}(\mu_\lambda'|\mathbf{y}, \mathbf{t})$. The color bar on the bottom depicts posterior sample based inference for the chosen fixed value of λ, where, with posterior probability at least α, the derivative of μ_λ is positive or negative on the intervals indicated by red and blue, respectively, and the probability is computed jointly over all time points s_j in these intervals. The full map, such as in the middle panels of Figures 8 and 9, is obtained by stacking such color bars, for the whole Holocene and for all scales λ considered.

As in our earlier scale space analyses of the paleoclimate, the credibility level was chosen as $\alpha = 0.8$ [e.g., Erästö and Holmström (2005, 2007, 2006); Weckström et al. (2006)]. Increasing the level, say, to 0.95, slightly shrinks the credible features (blue and red areas) but does not affect much the interpretation given in Section 3.5. The $\alpha = 0.95$ versions of all consensus credibility maps are included in the supplement [Erästö et al. (2011a)].

It is interesting to study also the effects on the consensus of the two lakes and the three proxies separately. Such an analysis is presented in Figure 10, where credibility maps for the lakes and the proxies based on large reconstruction errors are displayed. One can also analyze the role of each of the six reconstructions more quantitatively by considering their mean contributions to the posterior consensus. Appendix B proposes such an approach and to demonstrate the idea, we examined more closely the early Holocene warming suggested in the credibility map of Figure 9. The bottom panel of the second column of Figure 10 shows the mean contribution of each reconstruction to the slope of the consensus at a millennial time scale (yellow curve in Figure 9), from the beginning of the Holocene to 7000 years before present. Such a plot can be useful when one wants to focus the analysis on a particular feature in a limited time window.

The results of Figures 8–10 are based on $\boldsymbol{\mu}$-samples of size 4000 where the first 2000 were used for burn-in. Generating such a sample on a standard PC takes about 10 hours. A uniform grid of about 2000 time points s_j and a logarithmic grid of 200 smoothing levels λ were used in the scale space analyses. With random dates it takes about 10 hours to process a batch of 10 smoothing levels. Computations can be sped up by allocating the batches to different processors. Parameter convergence was checked visually. Initial values were picked from the priors for those parameters that are updated by Gibbs sampling and the carbon dating based values were used to initialize the chronologies. The posterior error covariances were almost diagonal but heteroskedastic with small off-diagonal elements. The chronologies changed only little in the simulation. The standard error of a radiocarbon date is commonly interpreted as a standard deviation of a normal distribution center at the date [cf. (13)]. To test the robustness of dating error assumptions, we repeated some of our analyses assuming either a much smaller (down to zero) or a much larger (up to several times the value used in the reported analyses) standard error, but the features proposed by the maps stayed the same. For very large standard errors this is due to proposals in the MCMC simulation being mostly rejected.
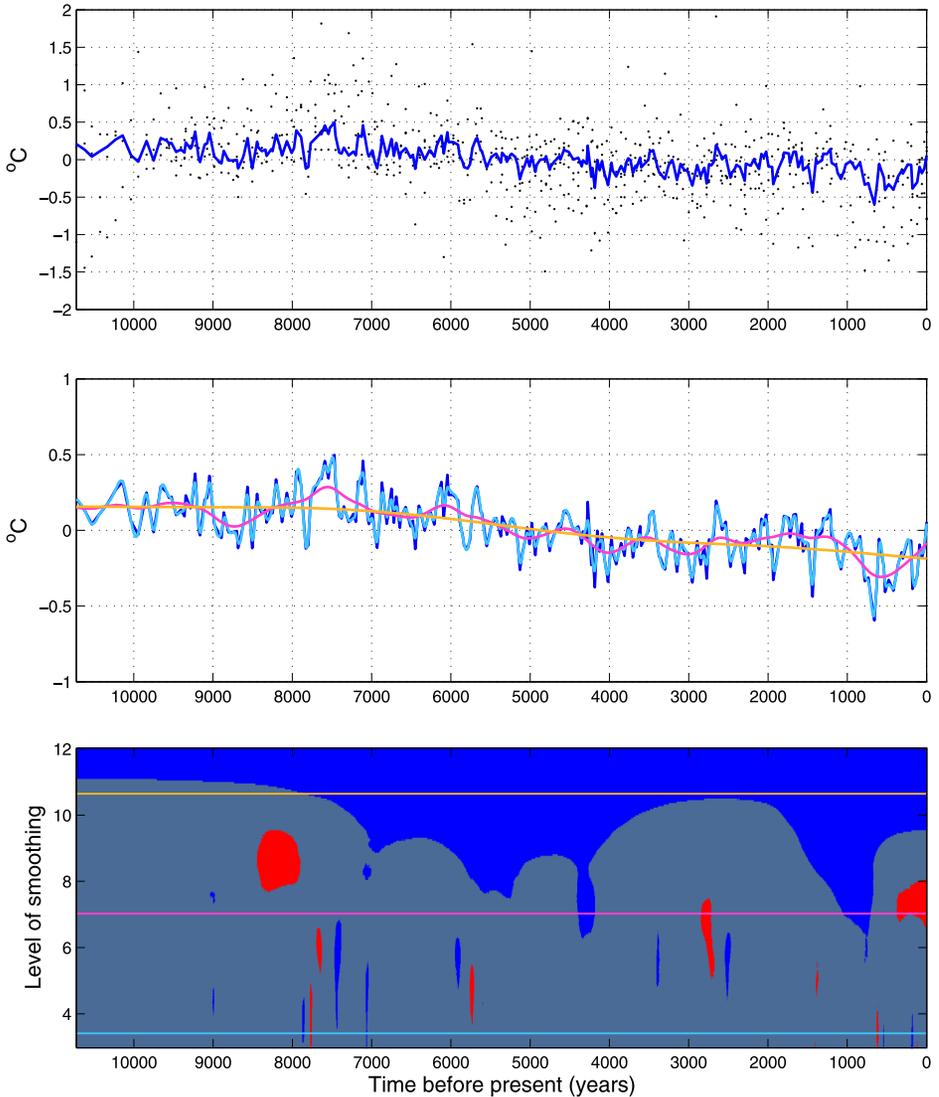
FIG. 8. *Scale space analysis of the consensus of six temperature reconstructions. The top panel shows the reconstructions (dots) and the posterior mean of the consensus (blue curve). Large reconstruction errors were assumed and the credibility level $\alpha = 0.8$. The middle panel shows the posterior mean of the consensus together with three smooths of the posterior consensus corresponding roughly to multi-decadal (light blue), centennial (purple) and millennial (yellow) time scales. The bottom panel is the credibility map where blue and red indicate credible cooling and warming, respectively. For more information see the text.*
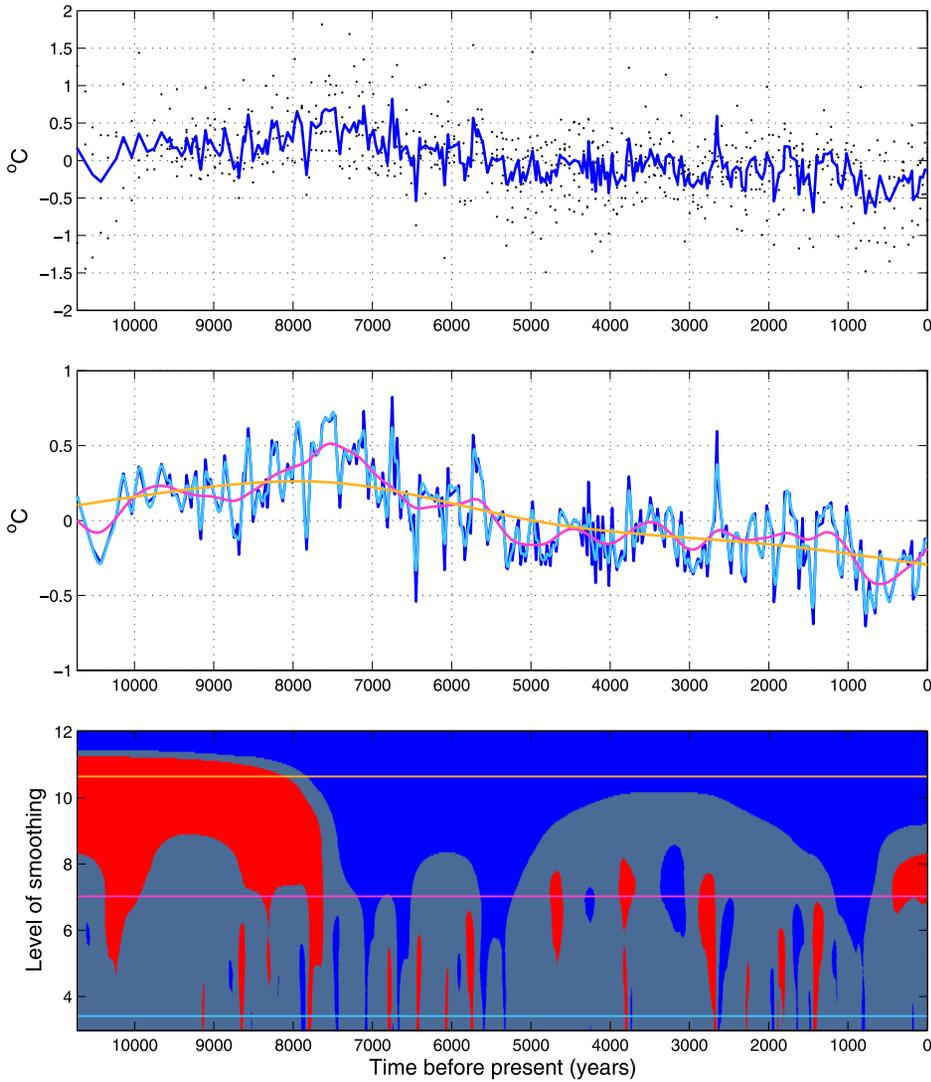
FIG. 9.   *Scale space analysis of the consensus of six temperature reconstructions. Small reconstruction errors were assumed and the credibility level α = 0.8. For more information see the caption of Figure* 8 *and the text.*

3.5. *Interpretation of results.*

3.5.1. *Consensus features.*   According to the credibility maps of Figures 8 and 9, overall cooling is the longest time scale feature of Holocene summer temperature in northern Finland, indicated by the continuous blue color in the topmost part of the maps. This is thought to be mostly due to the earth's changing orbital geometry around the sun. At millennial scales (yellow lines in the maps), the consensus
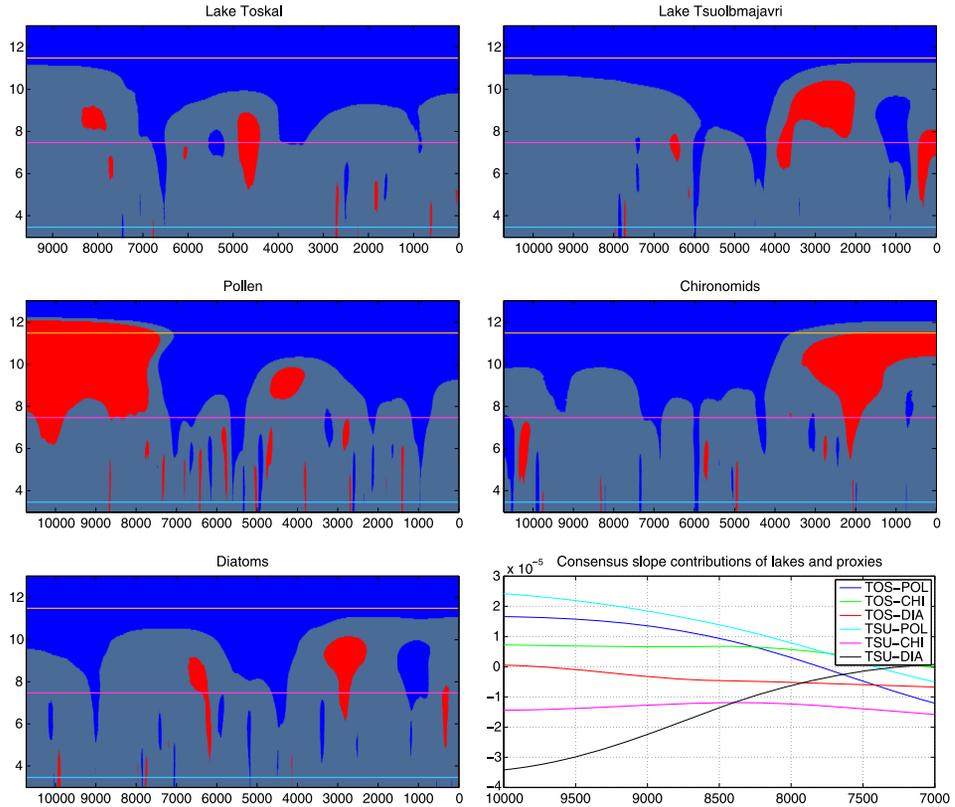
FIG. 10. *Consensus based on subgroups of the six temperature reconstructions considered. Large reconstruction errors are assumed and the credibility level is* 0.8. *In the top row, the Lake Toskal map is based on all three proxy records obtained from that lake and similarly for Lake Tsuolbmajavri. The other three maps show the consensus according to each proxy when the corresponding proxy records from each lake have been combined. The bottom panel of the second column is a more detailed analysis of how each reconstruction affects the overall consensus within a particular time interval on a millennial time scale. For more information see the caption of Figure* 8 *and the text.*

summer temperatures exhibit some other key aspects of Holocene climate evolu-
tion, such as an early Holocene warming trend shown strongly in Figure 9 and
weakly in Figure 8, together with a peak warming at around 8 kyr BP (8000 years
before present) indicated by red changing to blue, followed by a monotonic cool-
ing trend (blue color) until the present time. This overall pattern is predominantly
driven by annual mean and summer orbital forcing at the high northern latitudes
[Berger and Loutre (1991)]. In the Northern Hemisphere summer months the in-
coming solar radiation (insolation) peaked between 11 and 9 kyr BP [Kutzbach
(1981)], when insolation was approximately 7–9% higher than at present at 70°N,
and gradually declined since then. The relatively cool summer temperatures in the
early Holocene (rising trend before 9 kyr BP) in the consensus hence refer to a

slightly delayed timing of the Holocene Thermal Maximum (HTM) relative to this peak summer insolation, suggesting that the climate response to the orbital forcing must also be affected by some extra forcings and internal feedbacks in the climate system [Chapin III et al. (2000)]. The cool conditions in the earliest Holocene were apparently heavily influenced by the last substantial remnants of the large Fennoscandian and Laurentide continental ice sheets that trigged changes in ocean heat transportation and surface albedo [Kaplan and Wolfe (2006); Renssen et al. (2009)].

According to our consensus reconstruction, HTM in northern continental Europe occurred at around 8–9 kyr BP, when the inferred summer temperature values clearly exceeded the modern levels. This early peaking of Holocene warmth contradicts several earlier studies that place the timing of peak warming across a wide area of northern Europe closer to mid-Holocene at around 6 kyr BP [Davis et al. (2003); MacDonald et al. (2000); Kaufman et al. (2004)]. Evidence for the mid-Holocene thermal maximum in northern Europe comes largely from a northward and upward expansion of northern treelines, as well as from retreating glaciers [Jansen et al. (2007)]. However, a recent global assessment of treeline response to climate warming suggests that treeline advance may be more strongly associated with winter, rather than summer, warming [Harsch et al. (2009)]. In addition, in many parts of Scandinavia, glaciers started to retreat in the early Holocene, soon after the transient cooling event, termed the Finse event [8.5–8.0 kyr BP; Nesje et al. (2008)]. The early expression of peak summer warming identified in the present study is further consistent with a recent model simulation study [Renssen et al. (2009)], where maximum summer warmth in the northeast of Europe was placed closer to 8 kyr BP.

At multi-decadal to centennial scales (light blue and purple lines in the maps), climate variability as highlighted in our small-error analysis (less so with large reconstruction errors) shows a complex picture with indications of repeated warm and cold climate episodes, the specific causes of which are not fully understood. Some of the peaks found in our record seem to be coherent with the Holocene series of North Atlantic ice-rafting events defined by Bond et al. (1997) within the dating uncertainties (± 100 to 200 years). These include the weak temperature minima at around 1.4, 2.8, 4.2 and around 10.3 kyr BP, whereas the remaining mid- and early Holocene "Bond events" are not evident in our record. Neither can we find any event-like feature around the classical 8.2 kyr BP cooling event [Alley et al. (1997)], although the most pronounced decline in overall Holocene summer temperatures started in our record around this time (see above). Examination of the maps at the smallest smoothing levels shows credible fluctuations in summer temperature, in particular, between 7.0 and 5.0 kyr BP and from 3.0 kyr BP to the present, while more stable conditions occurred between 5.0 to 3.0 kyr BP and in the early Holocene. Solar variability is the most plausible explanation for the temporal dynamics of these short-term changes. Indeed, recent work utilizing spectral analysis of radionuclide records suggests that the solar cycles were particularly

prominent during the time intervals 6.0–4.5 kyr BP and 3.0–2.0 BP, whereas this periodic behavior faded during other time intervals [Knudsen et al. (2009)]. Hence, the high-variability intervals in our record coincide with the periods of intensive solar cycles, which in turn correlate with periods of significant reorganization of the ocean and atmospheric circulation in the North Atlantic region [Mayewski et al. (2004); Seidenkrantz et al. (2007)].

Our scale space consensus analysis (in particular, the credibility map of Figure 9) indicates that the Northern Fennoscandia summer climate experienced a succession of warming and cooling events during the most recent part of the Holocene, broadly similar to those documented earlier in Northern Hemisphere temperature reconstructions, including the Current Warm Period (CWP), Little Ice Age (LIA) and Medieval Climate Anomaly (MCA) [Jansen et al. (2007); Mann et al. (2008)]. The MCA commenced around 1.3 kyr BP and terminated around 0.8 kyr BP when temperatures started to decrease toward the LIA. Conditions slightly warmer than those of the 20th century may have prevailed in the North Atlantic climate regime during the MCA as deduced on the basis of our analysis. The peak medieval warmth is around 1.2 kyr BP in our record, which is earlier than in many previous published reconstructions, but is in accordance with Mann et al. (2008) who place the MCA between AD 1450 and AD 700. The LIA in our consensus reconstruction occurred perhaps between ca. 0.5 and 0.15 kyr BP (about AD 1500–1850), in agreement with the recent Arctic-wide synthesis of proxy temperature records [Kaufman et al. (2009)]. The recent warming (CWP) shows as a credibly positive temperature trend in centennial scales.

3.5.2. *Contributions from the proxies and the lakes.* Looking at the lake- and proxy-specific credibility maps of Figure 10, we note first that, of the three proxies, the pollen-based reconstructions suggest most features with somewhat fewer credible features exhibited by the chironomid and the diatom records. All three agree on a Holocene-wide cooling trend which therefore becomes part of the overall consensus. Still, on millennial scales (yellow line), the cooling trend after about 4 kyr BP in the chironomid record is a bit less certain than in the two other proxies. It is notable that evidence for early Holocene warming and the HTM in the overall consensus appears to come from the pollen record only. The millennial scale detail analysis shown in the bottom panel of the second column of Figure 10 clearly confirms this. The fact that in the large-error analysis of Figure 8 these show only weakly is probably due to the relatively large pollen reconstruction error upper bounds used for this analysis (cf. Table 2). The LIA is clearly visible as a credible temperature minimum only in the diatom record. However, combined with the cooling trend immediately prior to it, which is present also in pollen and chironomid reconstructions, the LIA signal in diatoms is strong enough to show in the consensus, too. The Bond events (cf. Section 3.5.1) are supported in varying degrees by different proxies. The warm MCA appears to be better reconstructed by

chironomids than pollen. The recent centennial-scale rise in temperatures exhibited in the consensus is driven mostly by the diatom record with the chironomids showing millennial scale warming during the last 2000–3000 years.

Considering the credibility maps in the first row of Figure 10, we notice that the records from the two lakes both support overall Holocene cooling and the LIA (although only barely for Toskal), whereas only Lake Toskal shows weak evidence for early Holocene warming. In light of the detail analysis of Figure 10 (lower right-hand corner panel), it appears that the strong millennial scale warming signal in the Lake Tsuolbmajavri pollen record is drowned by negative contributions from the chironomid and diatom reconstructions. Still, as noted above, when evidence in all records is included, the warming signal is strong enough to show in the overall consensus. Finally, we observe that only the Lake Tsuolmbajavri record suggests the presence of the MCA and that opposite features in the lake records at around 4 kyr BP may be the source of centennial-scale oscillations in the consensus during 5–3 kyr BP (purple curve in the middle panel of Figure 8).

4. Discussion. Given a collection of noisy reconstructions, the proposed method uses Bayesian inference to find those features of past climate variation that are supported by their consensus. Although only temperature was considered, other climate variables could be handled similarly. Further, while the reconstructions considered in this paper were based on radiocarbon dated sediments samples, the method is conceivably applicable to other proxy types that use different dating methods such as tree rings, varved lake sediments, ice cores and speleothem archives, where estimates of dating errors are available [see Jones et al. (2009) for a discussion of these and other proxy types]. In case of annually resolved records such as tree rings, the fixed dates version of the method might suffice. Also, although the paper focuses on an application to paleoclimate reconstruction, the method developed is likely to find use also in other contexts where a combination of information across several noisy time series is of interest.

Handling of dating errors in our consensus model could probably be considerably improved. A sophisticated Bayesian dating error model, BChron, was introduced in Haslett and Parnell (2008). Other recent proposals include, for example, Blaauw et al. (2003), Blaauw and Christen (2005) and Bronk Ramsey (2008). The problem of modeling the relationship between sediment depth and age was also analyzed in Telford, Heegaard and Birks (2004) and Heegaard, Birks and Telford (2005), and aligning multiple varve chronologies was considered in Auestad et al. (2008). Dating error models developed for spatial problems could also be useful; see, for example, Fanshawe and Diggle (2011) and Cressie and Kornak (2003). Still, while we readily acknowledge that the error model described in Section 2.2 may be too crude to reflect all aspects of uncertainty in the dating process, it nevertheless can serve as a first approximation that allows, in principle, the effect of dating errors to enter the posterior uncertainty of the consensus anomaly. In future work we hope to incorporate in the analysis ideas from more sophisticated error

models such as Bchron. Such an improvement in the analysis might be incorporated also in a system that uses Bayesian reconstructions to begin with. We leave these ideas for future work.

Another direction of development would be to include the spatial dependencies between the proxy records in the model. With only two core locations considered in our example, this is not relevant, but it might be useful when more locations are included in the consensus analysis.

We proposed to use climate simulations to gain insight into the variability of the past temperature. Of course, the simulation we used covers only a fraction of the approximately 10,000 years considered in the reconstructions and, therefore, in the analyses described in Section 3.3.2, one considers temperature roughness only for about 10% of the whole Holocene period. Still, although the mean temperature levels for the last 1150 years may be different from those during the rest of the Holocene, it may not be unreasonable to assume that the inter-annual temperature variation has not changed dramatically. By studying the simulated 30-year mean for the last 1150 years we may therefore gain at least some idea of its roughness during the whole Holocene. In a sense, such an assumption could be viewed as being somewhat analogous to the basic premise underlying proxy-based paleoclimate reconstructions, namely, that the relationship between the proxy records and the climate has not changed over thousands of years.

To summarize, the method described in this paper provides a means to estimate the consensus temperature variation in heterogenic time series and also to capture its salient features, such as maxima, minima and trends in different time scales in a statistically principled manner. Our model allows dating uncertainties, distinct or overlapping core chronologies, as well as time-varying, correlated reconstruction errors that can also have different magnitudes for different proxies and cores. We believe that the method has also wider applicability potential in data mining of various types of climate records and compiled time series. When applied to lake data series from northern Finland, a millennial-scale cooling trend was found since the Holocene thermal maximum at around 8 kyr BP associated with the decrease in orbitally driven summer insolation. Superimposed on the millennial-scale trends, the summer climate in northern Finland was punctuated by several quasicyclical climate events, the forcing mechanisms of which are not yet fully understood. Our scale space analysis also suggests that inconsistencies in climate reconstructions and their interpretations may be at least partly spurious; there is probably no single narrative that counts as the canonical version of Holocene climate change. Instead, there are many interpretations depending on the proxy and the resolution at which the data are gained and examined. Finally, while the paper focuses on paleoclimate time series, the proposed method can be applied in other contexts where one seeks to infer features that are jointly supported by an ensemble of irregularly sampled noisy time series.

TABLE 2
*Estimates of upper bounds of reconstruction
errors for the 6 proxy records considered*

Proxy record	$\bar{\sigma}_k$
Lake Toskal chironomids	0.32
Lake Toskal diatoms	0.27
Lake Toskal pollen	0.68
Lake Tsuolbmajavri chironomids	0.59
Lake Tsuolbmajavri diatoms	0.21
Lake Tsuolbmajavri pollen	0.71

APPENDIX A: ESTIMATION OF THE RECONSTRUCTION ERROR

We explain here how the temperature anomalies $\mathbf{y}_k$ were used to estimate upper bounds for the reconstruction error variances.

Assuming that $\mathbf{y}_k \sim \mathrm{N}(\boldsymbol{\mu}_k, \sigma_k^2 \mathbf{I}_{j_k})$, the distribution of the random variable $V_k = \|\mathbf{y}_k\|^2 = \mathbf{y}_k^T \mathbf{y}_k$ is determined by the parameter $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \sigma_k)$. We consider a fixed value $\bar{\sigma}_k > 0$ and the null hypothesis

$$H_0 : \Theta_0 = \{\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \sigma_k) \mid \boldsymbol{\mu}_k \in \mathbb{R}^m, \sigma_k \geq \bar{\sigma}_k\}$$

against the alternative

$$H_1 : \Theta_1 = \{\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \sigma_k) \mid \boldsymbol{\mu}_k \in \mathbb{R}^m, \sigma_k < \bar{\sigma}_k\}.$$

The null hypothesis is rejected if $V_k \leq \bar{v}_k$, where $\bar{v}_k$ is some fixed value. It is shown in Holmström and Erästö (2001) that the significance level of this test is given by

(17) $$\beta = \mathbb{P}(\chi_{j_k-1}^2 \leq \bar{v}_k / \bar{\sigma}_k^2),$$

where $\chi_{j_k-1}^2$ is a chi-square variable with $j_k - 1$ degrees of freedom. Setting $\beta = 0.05$, an upper bound for σ_k can therefore be estimated as

$$\bar{\sigma}_k = \sqrt{V_k / \chi_{j_k-1,0.05}^2},$$

where $\chi_{j_k-1,0.05}^2$ is the 5th percentile of the χ^2-distribution with $j_k - 1$ degrees of freedom. These values are listed in Table 2 for the six proxy records and they were used to define the large-error prior scale matrices $\mathbf{W}_k$ in the consensus analysis.

APPENDIX B: CONTRIBUTIONS OF INDIVIDUAL PROXY RECORDS TO THE CONSENSUS

It follows from (3) and (8) that

$$\boldsymbol{\mu} \mid \{\boldsymbol{\Sigma}_k\}, \lambda_0, \boldsymbol{\tau}, \mathbf{y}, \mathbf{t} \sim \mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0),$$

where

$$\boldsymbol{\Sigma}_0 = \left(\sum_{k=1}^{m} \boldsymbol{\Sigma}_k^{-1} + \lambda_0 \mathbf{K} \right)^{-1}$$

and

$$\boldsymbol{\mu}_0 = \boldsymbol{\Sigma}_0 \left(\sum_{k=1}^{m} \boldsymbol{\Sigma}_k^{-1} \mathbf{y}_k \right) = \sum_{k=1}^{m} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_k^{-1} \mathbf{y}_k,$$

where it is understood that $\boldsymbol{\Sigma}_k$ and $\mathbf{y}_k$ are extended to an $n \times n$ matrix and an n-dimensional vector, respectively, by putting zero entries to locations that correspond to those time points in the full joint chronology $\mathbf{t}$ that do not appear in the chronology $\mathbf{t}_k$ of proxy record k. It follows that the components of the posterior mean vector $\mathbb{E}(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_k^{-1} \mathbf{y}_k | \mathbf{y}, \mathbf{t})$ can be used to quantify the contribution of record k to the posterior of $\boldsymbol{\mu}$ at the time points $\tau_1, \ldots, \tau_n$. If $\mathbf{S}_\lambda$ and $\mathbf{D}$ are the matrices defined in Section 2.3, the contribution of record k to the slope of the smooth μ_λ' at the time points $s_1, \ldots, s_r$ [cf. (16)] can then be analyzed by considering the mean of $\mathbb{E}(\mathbf{DS}_\lambda \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_k^{-1} \mathbf{y}_k | \mathbf{y}, \mathbf{t})$, instead. This is the quantity depicted for each reconstruction in the bottom panel of the second column of Figure 10.

SUPPLEMENTARY MATERIAL

Supplement A: Additional analyses and the data used (DOI: 10.1214/12-AOAS540SUPPA; .pdf). The document (a pdf-file) reports exploratory analyses of the estimated reconstruction errors, shows additional credibility maps, and provides the data analyzed in the article.

Supplement B: The Matlab code (DOI: 10.1214/12-AOAS540SUPPB; .zip). The Matlab code (in a zip-file) used to compute the results of the article.

REFERENCES

ALLEY, R. B., MAYEWSKI, P. A., SOWERS, T., STUIVER, M., TAYLOR, K. C. and CLARK, P. U. (1997). Holocene climatic instability: A prominent, widespread event 8200 yr ago. *Geology* **25** 483–486.

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.

AUESTAD, B. H., SHUMWAY, R. H., TJØSTHEIM, D. and VEROSUB, K. L. (2008). Linear and nonlinear alignment of time series with applications to varve chronologies. *Environmetrics* **19** 409–427. MR2440040

BERGER, A. and LOUTRE, M. F. (1991). Insolation values for the climate of the last 10 million years. *Quaternary Science Reviews* **10** 297–317.

BIRKS, H. J. B. (1995). Quantitative palaeoenvironmental reconstructions. In *Statistical Modelling of Quaternary Science Data*, *Technical Guide* 5 (D. Maddy and J. S. Brew, eds.) 161–254. Quaternary Research Association, Cambridge.

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 Journal* **3** 68–110.

BLAAUW, M. and CHRISTEN, J. A. (2005). Radiocarbon peat chronologies and environmental change. *J. Roy. Statist. Soc. Ser. C* **54** 805–816. MR2196152

BLAAUW, M., HEUVELINK, G. B. M., MAUQUOY, D., VAN DER PLICHT, J. and VAN GEEL, B. (2003). A numerical approach to ^{14}C wiggle-match dating of organic deposits: Best fits and confidence intervals. *Quaternary Science Reviews* **22** 1485–1500.

BOND, G. et al. (1997). A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science* **278** 1257–1266.

BRONK RAMSEY, C. (2008). Deposition models for chronological records. *Quaternary Science Reviews* **27** 42–60.

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

CHAPIN III, F. S. et al. (2000). Arctic and boreal ecosystems of western North America as components of the climate system. *Global Change Biology* **6** 211–223.

CRESSIE, N. and KORNAK, J. (2003). Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statist. Sci.* **18** 436–456. MR2059325

DAVIS, B. A. S., BREWER, S., STEVENSON, A. C. and GUIOT, J. (2003). The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews* **22** 1701–1716.

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). Prior selection and multiscale analysis in Bayesian temperature reconstruction based on species assemblages. *Journal of Paleolimnology* **36** 69–80.

ERÄSTÖ, P. and HOLMSTRÖM, L. (2007). Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors. *J. Stat. Comput. Simul.* **77** 421–431. MR2395958

ERÄSTÖ, P., HOLMSTRÖM, L., KORHOLA, A. and WECKSTRÖM, J. (2011a). Supplement A to "Finding a consensus on credible features among several paleoclimate reconstructions." DOI:10.1214/12-AOAS540SUPPA.

ERÄSTÖ, P., HOLMSTRÖM, L., KORHOLA, A. and WECKSTRÖM, J. (2011b). Supplement B to "Finding a consensus on credible features among several paleoclimate reconstructions." DOI:10.1214/12-AOAS540SUPPB.

FANSHAWE, T. R. and DIGGLE, P. J. (2011). Spatial prediction in the presence of positional error. *Environmetrics* **22** 109–122. MR2843341

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*: *A Roughness Penalty Approach*. *Monographs on Statistics and Applied Probability* **58**. Chapman & Hall, London. MR1270012

HARSCH, M. A., HULME, P. E., MCGLONE, M. S. and DUNCAN, R. P. (2009). Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecol. Lett.* **12** 1040–1049.

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

HEEGAARD, E., BIRKS, H. J. B. and TELFORD, R. J. (2005). Relationships between calibrated ages and depth in stratigraphical sequences: Estimation procedure by mixed-effect regression. *The Holocene* **15** 612–618.

HOLMSTRÖM, L. (2010a). BSiZer. *Wiley Interdisciplinary Reviews*: *Computational Statistics* **2** 526–534.

HOLMSTRÖM, L. (2010b). Scale space methods. *Wiley Interdisciplinary Reviews*: *Computational Statistics* **2** 150–159.

HOLMSTRÖM, L. and ERÄSTÖ, P. (2001). Using the SiZer method in Holocene temperature reconstruction. Research Report A36, Rolf Nevanlinna Institute.

HOLMSTRÖM, L., ERÄSTÖ, P., WECKSTRÖM, J., NYMAN, M. and KORHOLA, A. (2008). A Bayesian reconstruction of Holocene temperature variation in Northern Fennoscandia. In *2008 Joint Statistical Meetings, Abstract Book* 256. Denver, CO.

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. et al. (2009). High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *The Holocene* **19** 3–49.

KAPLAN, M. R. and WOLFE, A. P. (2006). Spatial and temporal variability of Holocene temperature in the North Atlantic region. *Quaternary Research* **65** 223–231.

KAUFMAN, D. S. et al. (2004). Holocene thermal maximum in the western Arctic (0–180°W). *Quaternary Science Reviews* **23** 529–560.

KAUFMAN, D. S. et al. (2009). Recent warming reverses long-term arctic cooling. *Science* **325** 1236–1239.

KNUDSEN, M. F., RIISAGER, P., JACOBSEN, B. H., MUSCHELER, R., SNOWBALL, I. and SEIDENKRANTZ, M. S. (2009). Taking the pulse of the sun during the Holocene by joint analysis of ^{14}C and ^{10}Be. *Geophysical Research Letters* **36** L16701.

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

KORHOLA, A., WECKSTRÖM, J., HOLMSTRÖM, L. and ERÄSTÖ, P. (2006). Reconstructing climate from palaeolimnological archives using multiple proxy indicators and sites simultaneously. In 10*th International Paleolimnology Symposium. Abstract Volume*: 94. Duluth, MI.

KUTZBACH, L. E. (1981). Monsoon climate of the early Holocene: Climate experiment with the earth's orbital parameters for 9000 years ago. *Science* **214** 59–61.

LEGRANDE, A. N. et al. (2006). Consistent simulations of multiple proxy responses to an abrupt climate change event. *Proc. Natl. Acad. Sci. USA* **103** 837–842.

LI, B., NYCHKA, D. W. and AMMANN, C. M. (2010). The value of multiproxy reconstruction of past climate. *J. Amer. Statist. Assoc.* **105** 883–895. MR2752583

LINDEBERG, T. (1994). *Scale–Space Theory in Computer Vision*. Kluwer Academic Publishers, Dordrecht.

MACDONALD, G. M. et al. (2000). Holocene treeline history and climate change across Northern Eurasia. *Quaternary Research* **53** 302–311.

MANN, M. E., ZHANG, Z., HUGHES, M. K., BRADLEY, R. S., MILLER, S. K., RUTHERFORD, S. and NI, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci. USA* **105** 13252–13257.

MAYEWSKI, P. A. et al. (2004). Holocene climate variability. *Quaternary Research* **62** 243–255.

NESJE, A., BAKKE, J., DAHL, S. O., LIE, Ø. and MATTHEWS, J. A. (2008). Norwegian mountain glaciers in the past, present and future. *Global and Planetary Change* **60** 10–27.

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. *Nature Geoscience* **2** 411–414.

ROBERT, C. P. and CASELLA, G. (2005). *Monte Carlo Statistical Methods*. Springer, New York.

RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270. MR1379468

SALONEN, S., ILVONEN, L., SEPPÄ, H., HOLMSTRÖM, L., TELFORD, R. J., GAIDAMAVIČIUS, A., STANČIKAITE, 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 climate reconstruction. *The Holocene* **22** 413–424.

SEIDENKRANTZ, M. S., AAGAARD-SØRENSEN, S., SULSBRÜCK, H., KUIJPERS, A., JENSEN, K. G. and KUNZENDORF, H. (2007). Hydrography and climate of the last 4400 years in a SW Greenland fjord: Implications for Labrador Sea palaeoceanography. *The Holocene* **17** 387–401.

SEPPÄ, H. and BIRKS, H. J. B. (2001). July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: Pollen-based climate reconstructions. *The Holocene* **11** 527–539.

SEPPÄ, H., NYMAN, M., KORHOLA, A. and WECKSTRÖM, J. (2002). Changes of treelines and alpine vegetation in relation to post-glacial climate dynamics in northern Fennoscandia based on pollen and chironomid records. *Journal of Quaternary Science* **17** 287–301.

TELFORD, R. J., HEEGAARD, E. and BIRKS, H. J. B. (2004). All age–depth models are wrong: But how badly? *Quaternary Science Reviews* **23** 1–5.

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

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

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

WECKSTRÖM, J., KORHOLA, A., ERÄSTÖ, P. and HOLMSTRÖM, L. (2006). Temperature patterns over the past eight centuries in Northern Fennoscandia inferred from sedimentary diatoms. *Quaternary Research* **66** 78–86.

P. ERÄSTÖ
NATIONAL INSTITUTE FOR HEALTH
    AND WELFARE
P.O. BOX 30, FIN-00271 HELSINKI
FINLAND
E-MAIL: panu.erasto@helsinki.fi

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

A. KORHOLA
J. WECKSTRÖM
ENVIRONMENTAL CHANGE RESEARCH UNIT (ECRU)
DEPARTMENT OF ENVIRONMENTAL SCIENCES
UNIVERSITY OF HELSINKI
P.O. BOX 65, FIN-00014
FINLAND
E-MAIL: atte.korhola@helsinki.fi
        jan.weckstrom@helsinki.fi