

A Pseudoproxy Evaluation of Bayesian Hierarchical Modeling and Canonical Correlation Analysis for Climate Field Reconstructions over Europe*

JOHANNES P. WERNER AND JUERG LUTERBACHER

Department of Geography, Justus-Liebig-University, Giessen, Germany

JASON E. SMERDON

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

(Manuscript received 31 December 2011, in final form 26 July 2012)

ABSTRACT

A pseudoproxy comparison is presented for two statistical methods used to derive annual climate field reconstructions (CFRs) for Europe. The employed methods use the canonical correlation analysis (CCA) procedure presented by Smerdon et al. and the Bayesian hierarchical model (BHM) method adopted from Tingley and Huybers. Pseudoproxy experiments (PPEs) are constructed from modeled temperature data sampled from the 1250-yr paleo-run of the NCAR Community Climate System Model (CCSM) version 1.4 model by Ammann et al. Pseudoproxies approximate the distribution of the multiproxy network used by Mann et al. over the European region of interest. Gaussian white noise is added to the temperature data to mimic the combined signal and noise properties of real-world proxies. Results indicate that, while both methods perform well in areas with good proxy coverage, the BHM method outperforms the CCA method across the entire field and additionally returns objective error estimates.

1. Introduction

A concerted research effort over the last decade has focused on reconstructing global or regional climate during the Common Era (CE) using networks of climate proxies [see, e.g., Jones et al. (2009) for a review]. A significant area of focus has been over Europe and the North Atlantic where instrumental, documentary, and proxy data are abundant (e.g., Luterbacher et al. 1999, 2000, 2002, 2004, 2007; Pauling et al. 2003, 2006; Xoplaki et al. 2005; Küttel et al. 2009, 2010; Riedwyl et al. 2009; Guiot et al. 2005, 2010). These regional reconstructions employ the same or similar methods used to reconstruct global or hemispheric climatic fields, and therefore are subject to many of the same challenges that have been widely discussed for the latter group (e.g., Jones et al.

2009; Smerdon et al. 2011). For example, outstanding methodological questions are tied to the impact of proxy distributions and abundance (e.g., Pauling et al. 2003; Küttel et al. 2007; Smerdon et al. 2011, and references therein), the connections between climate and proxy responses across different spectral domains and multiple environmental variables (e.g., Evans et al. 2006; D'Arrigo et al. 2008), the role of teleconnections and noise in the calibration data (e.g., Christiansen et al. 2009; Smerdon et al. 2011), and the impact of methodological choices on derived reconstructions (e.g., Mann et al. 2007; Hegerl et al. 2007; Lee et al. 2008; Christiansen et al. 2009; Tingley et al. 2012; Smerdon et al. 2011; Wahl and Smerdon 2012). The answers to these questions are ultimately fundamental to successful reconstructions of past climatic variability (e.g., North et al. 2006; Jansen et al. 2007; Jones et al. 2009).

To address the existing challenges and improve CE climate field reconstructions, multiple methodological approaches have been emerging recently as alternatives to the more traditional multivariate linear regression schemes that have been widely used for reconstruction problems. For instance, paleoclimatic data assimilation schemes recently have been proposed and explored

* Lamont-Doherty Earth Observatory Contribution Number 7621.

Corresponding author address: Johannes P. Werner, Department of Geography, Climatology, Climate Dynamics and Climate Change Justus-Liebig-University, Senckenbergstrasse 1, D-35390 Giessen, Germany.
E-mail: johannes.werner@geogr.uni-giessen.de

(Widmann et al. 2010; Goussé et al. 2010; Luterbacher et al. 2010a). The paleoclimatic reconstruction problem also can be formulated in Bayesian frameworks (Tingley and Huybers 2010a,b; Li et al. 2010). Both assimilation and Bayesian approaches generally have the benefit of incorporating physical or process-based information about the climate and the climate–proxy connection as constraints on the reconstruction problem, while providing more comprehensive uncertainty estimates for the derived reconstructions. These benefits alone justify further application of these methods, as well as robust comparisons between established methods and the emerging efforts.

One important tool for assessing CE reconstruction methods is millennium-length, forced transient simulations with fully coupled general circulation models (GCMs) (e.g., González-Rouco et al. 2003, 2006; Gómez-Navarro et al. 2011; Ammann et al. 2007; Schmidt et al. 2011). These model simulations are used to derive controlled and systematic reconstruction experiments for methodological comparisons and evaluations—an approach known as pseudoproxy experiments (PPEs); see Smerdon (2012) for a review. The motivation for PPEs stems from the fact that real-world reconstructions are derived from many different methods, calibration choices, and proxy networks. Uncertainty in any given real-world reconstruction is therefore a combined result of the employed method, the adopted calibration data and calibration time interval, the spatial and temporal sampling of the proxy network, and the actual climate–proxy connection of each proxy record used for the reconstruction. If the objective is to isolate the impact of one of these factors, it is difficult to do so from comparisons between available real-world reconstructions. PPEs have allowed some of the above challenges to be circumvented by adopting a common framework that can be systematically altered and evaluated, and thus test reconstruction methods and their dependencies.

Here we build on previous work to apply and evaluate Bayesian algorithms for paleoclimate reconstructions using PPEs for methodological evaluation. We specifically use the Bayesian Algorithm for Reconstructing Climate in Space and Time (BARCAST) developed by Tingley and Huybers (2010a), which was evaluated in PPEs using instrumental data over North America. BARCAST is further evaluated herein for the first time in a European PPE framework built on output from a millennium-length simulation from the National Center for Atmospheric Research (NCAR) Community Climate System Model version 1.4 (CCSM) (Ammann et al. 2007). The longer time scale provided by the subsequent PPEs based on the millennium-length simulation, relative to the shorter time interval allowed by PPEs that use the

instrumental data (Tingley and Huybers 2010a), allows us to expand the BARCAST evaluation to lower frequencies and makes our results more directly comparable to the wider array of methodological studies that have used millennium-length simulations for PPEs. Our focus on Europe builds on multiple other studies that have evaluated reconstruction methods with millennial simulations (Riedwyl et al. 2009; Küttel et al. 2007), and the use of data derived from the same global simulation experiments used by Smerdon et al. (2011) further couches our efforts in a larger experimental context. We also compare the BARCAST reconstructions to experiments that employ the canonical correlation analysis (CCA) method applied by Smerdon et al. (2010b).

2. Data

The employed data are based on the transient paleoclimate simulation described by Ammann et al. (2007) using the NCAR CCSM 1.4 driven by natural and anthropogenic forcings estimated from 850 to 1999 Common Era (CE). The resulting annual surface temperature field output has been interpolated to a 5° longitude–latitude grid using bilinear interpolation (Smerdon et al. 2008; Rutherford et al. 2008; Smerdon et al. 2010a). We selected an area covering the northeastern Atlantic Ocean, Europe, and North Africa (30°–80°N, 20°W–45°E). The field is a subset of the global domain used in earlier studies (e.g., Smerdon et al. 2010b, 2011).

From this dataset we select two subsets of data from the CCSM field: one for the pseudoinstrumental data and a second for the pseudoproxy network. Throughout the article we refer to the model world, unless explicitly stated, and thus drop the prefix “pseudo” in relation to the simulated data. To mimic spatial data availability in the instrumental period we approximate the Jones et al. (1999) dataset by selecting only those grid points that have less than 30% missing annual data, based on a global analysis by Mann and Rutherford (2002). No effort was made to duplicate the changing data coverage in time; that is, all instrumental data were assumed to be available for all calibration years at the selected grid cells. The annual temperature data at these locations were directly used as the instrumental data for the climate field reconstruction (CFRs).

The employed pseudoproxy network approximates spatially the proxy network used by Mann et al. (1998) restricted to the study area. However, the proxy network remains stable in time: in contrast to real world CFRs all proxies are available throughout the full reconstruction period. Note that in Mann et al. about half the employed temperature sensitive proxy data in our reconstruction area of interest comprise long instrumental time series

(10 of 21 time series) and some were originally used as predictors for precipitation. The spatial distribution of pseudoproxy data is shown as dots in Fig. 2 (and subsequent figures showing spatial characterizations of the CFRs). Even though more proxies have become available through national and international projects and programs such as the European Union (EU) sixth framework program MILLENNIUM or Past Global Changes (PAGES) (Newman et al. 2010), there still are few millennium-length annually resolved temperature proxy time series available over the area of study (e.g. Büntgen et al. 2011; Esper et al. 2012). The regional European–Mediterranean subset of proxy data also used in earlier pseudoproxy experiments (Smerdon et al. 2008; Rutherford et al. 2008; Smerdon et al. 2010b, 2011) can therefore be seen as a best-case scenario when employing only highly resolved proxy data available through the full reconstruction period. This selection also allows for consistent comparisons between our results and published experiments that have used other methods in various reconstruction areas.

The proxy time series are constructed by adding white Gaussian noise to the temperature data at the selected proxy sites. The proxy signal-to-noise ratios (SNRs) in terms of standard deviation used for this study were 0.5 and 0.25, roughly spanning the range of estimated SNRs in real-world climate field reconstructions (CFRs) [cf. also Smerdon (2012) for a review; a more detailed description of the data is given by Smerdon et al. (2010b)]. Note that the variety of different proxy types is ignored in this study and the employed proxy response function is simpler than encountered in the real world. However, as a linear response function with white Gaussian noise has been standard in previous pseudoproxy studies, it is useful to use this traditional construction. We also primarily aim to test and compare the general skill of the adopted reconstruction methods, especially the ability of the employed models to capture and reconstruct the spatiotemporal evolution of the temperature field.

3. Reconstruction methods

Many different methods have been used to reconstruct past climate during the Common Era. In principle the reconstruction methods consist of two different parts: a (usually statistical) model and an inference mechanism. The inference mechanisms range from simple linear regressions (Bürger et al. 2006; Luterbacher et al. 2004, 2007; Xoplaki et al. 2005; Riedwyl et al. 2009) or so-called inverse regression (Mann et al. 1998) by minimizing an error measure or maximizing a likelihood function (or a combination thereof) or through application of neural networks (Guiot et al. 2005, 2010) to scalings of composite predictors (e.g., cf. Esper et al. 2005). A more complex

method is Bayesian inference, where a likelihood function is combined with a prior probability density function (PDF) to yield a posterior PDF for the fields and also the parameters; for example, see Gelman et al. (2003) or Tingley and Huybers (2010a). As the full joint (multivariate) probability density functions are often complicated, they can be estimated using the Gibbs sampler and the Metropolis–Hastings algorithm. The employed statistical models can be either localized descriptions of the climate field, such as the one presented by Tingley and Huybers (2010a,b) and used herein, or based on spatiotemporal eigenfunctions of the climate field and the proxy network, similar to the approaches in multivariate regressions such as CCA or principal component regression (PCR) (see, e.g., Cook et al. 1994; Luterbacher et al. 2000, 2002; Riedwyl et al. 2009).

a. Pointwise hierarchical model with Bayesian inference

Many dynamical systems can be modeled using statistical descriptions (Gardiner 1990; Risken 1989); in fact, stochastic modeling of deterministic dynamics is the foundation of modern thermodynamics as shown in the seminal papers of Einstein (1905, 1906). Stochastic modeling can be employed to describe the evolution of slowly varying characteristics of a dynamical system with a distinct time scale separation: the fast, often high dimensional degrees of freedom should be on a time scale much shorter than the slowly varying quantity of interest. In such cases, the effect of the fast degrees of freedom on the slow variations can be replaced by a suitable noise process [see Just et al. (2001) and Kantz et al. (2004)]. The parameters of the stochastic description often can be derived by careful analysis of the time series to be modeled (Just et al. 2003; Stemler et al. 2007; Anishchenko et al. 2002). Specifically with regard to continental temperature fields, the driving processes of annual temperature anomalies are on time scales of months (Rossby waves), weeks (cyclonic activity), or faster (convection); the evolution of ocean temperatures and circulation patterns, however, extends to time scales of years and longer. Our study area consists mainly of the European landmass, thus—while of course being influenced by the Atlantic Ocean—a time scale separation should nevertheless be present (cf. Hasselmann 1976), especially for annual temperature anomalies. One additional challenge when modeling extended spatiotemporal systems is the nonseparability of spatial and temporal dynamics of many systems. Tingley et al. (2012) give the annual temperature anomalies as an example for a nonseparable system: nonuniformity of the temporal autocorrelation (persistence) leads to nonseparability of the spatiotemporal cross-covariance matrix. In our area, the persistence

is in fact mostly uniform, and spatiotemporal dynamics (such as frontal systems) typically occur on shorter time scales and are removed by the averaging process.

In contrast to the usual methods in stochastic modeling, where the model is derived by careful analysis of the data, reasonable a priori assumptions, verified through preliminary analysis of the data, about the processes are used by Tingley and Huybers (2010a) and revisited below to create a simple model. The model is then verified to work reasonably well by checking diagnostics such as the convergence of the posteriors or predictive experiments. Those predictions can be made by using the derived set of parameters to estimate, for example, the temperatures at locations where available data were withheld from the initial experiment. In the context of pseudoproxy experiments, the reconstructions can be interpreted as predictive experiments.

1) THE BAYESIAN HIERARCHICAL MODEL

To actually employ a Bayesian hierarchical model (BHM) in climate field reconstructions, the climate field as well as the response of the different types of proxies must be modeled as a hierarchy of stochastic processes (e.g., Tingley and Huybers 2010a; Tingley et al. 2012; Li et al. 2010). Another level of hierarchy is represented by the model parameters that are not set to a fixed value, but rather by postulating a probability density function estimated from the data. The corresponding parameters are called “hyperparameters” and they are used to represent the prior knowledge about the system derived either from an understanding of the processes themselves or through initial analyses of the data. These parameters are discussed in detail by Tingley and Huybers (2010a) and the specific selections for our model are given in the appendix. Ultimately, the BHM provides estimates for the posterior PDFs of field variables and process parameters. These posterior PDFs can be used to evaluate the derived results; failure to converge can hint at problematic model assumptions, both in the model/likelihood and prior specifications, and/or insufficient amounts of data. Similar conclusions are implied by discrepancies between the posteriors and expert knowledge entering through the prior PDFs.

The employed statistical model was originally proposed by Tingley and Huybers (2010a). We will briefly show and motivate the chosen model equations for the temperature field at the N locations $\mathbf{T}_t \in \mathbb{R}^N$ at time step $t \in [850, 1980]$ and the instrumental and proxy response $\mathbf{W}_{P,t} \in \mathbb{R}^N$, $\mathbf{W}_{I,t} \in \mathbb{R}^N$:

$$\begin{aligned} \mathbf{T}_{t+1} - \mu &= \alpha(\mathbf{T}_t - \mu) + \epsilon_{T,t}, & \mathbf{W}_{I,t} &= \mathbf{H}_{I,t}(\mathbf{T}_t + \epsilon_{I,t}), \\ \mathbf{W}_{P,t} &= \mathbf{H}_{P,t}(\beta_1 \mathbf{T}_t + \beta_0 + \epsilon_{P,t}). \end{aligned} \quad (1)$$

The matrices $\mathbf{H}_{I/P,t} \in \mathbb{R}^{N \times N}$ are diagonal with one at position (i, i) when an observation in year t at location (i) was made and zero otherwise. The stochastic terms denoted by $\epsilon_{P,t}$ and $\epsilon_{I,t}$ are multivariate normal with a diagonal covariance matrix $\mathbf{I}\tau_P^2$ and $\mathbf{I}\tau_I^2$. They are used to model the local noise in the proxy response function and the errors in the instrumental observations. The interannual climate variability is described by the multivariate normal term $\epsilon_{T,t} \sim N(0, \mathbf{\Sigma})$, where the spatial covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$ is given by

$$(\mathbf{\Sigma})_{ij} = \sigma^2 \exp(-\phi|x_i - x_j|). \quad (2)$$

In other words, a uniform local variance of the climate field is assumed with a covariance matrix decreasing isotropically with distance along the loxodrome connecting locations x_i and x_j . A temperature anomaly at some location thus depends on its past value through the spatially uniform persistence term α , but has a stochastic component corresponding to interannual variability. The temperature anomalies at two locations are related to each other through the covariance matrix $\mathbf{\Sigma}$ if they are close together in space. A similar assumption is also used by Cook et al. (1999) where the spatial covariance structure is convex instead of concave. However, this means that teleconnections caused by large-scale atmospheric circulations, such as the Greenland temperature seesaw [cf., e.g., Loewe (1937) and van Loon and Rogers (1978), first described by Cranz (1770)], are ignored entirely. This of course leads to reduced skill when trying to reconstruct the climate field in a data sparse region. In contrast to this, EOF-based methods regress patterns of climatic fields and patterns of proxies, leaving some of the spatial covariance intact. Thus they rely on the temporal stationarity of identified spatial patterns. While we use a very simple spatial covariance matrix, the de facto nonuniformity of the spatial covariance should be addressed in future studies; in fact Tingley and Huybers (2010a,b) and Tingley et al. (2012) already address possible extensions. Throughout the article, however, the simple model will be used and evaluated.

The proxies \mathbf{W}_P are modeled as a linear response function distorted by additive white noise. Inclusion of more elaborate proxy response functions are of course possible within the BHM framework. Process-based proxy models for tree-ring growth (Tolwinski-Ward et al. 2010), pollen/habitat description (Ohlwein and Wahl 2012), and forward modeling of coral $\delta^{18}\text{O}$ (Thompson et al. 2011) represent potential future data model improvements.

2) BAYESIAN INFERENCE AND PRIOR SELECTION

From the above stochastic descriptions we can derive the probability density function of the data conditional

on the process parameters and values—the so-called likelihood function. The probability density function of the process θ conditional on the data x is then derived using Bayes formula:

$$P(\theta | x) \propto L(x | \theta)P(\theta). \quad (3)$$

Here $L(x | \theta)$ is the likelihood function and $P(\theta | x)$ is the resulting posterior probability density function. The term $P(\theta)$ denotes the prior: knowledge about the process enters the description here. For a more detailed description on Bayesian inference, see, for example, Gelman et al. (2003). A purported advantage over a purely maximum likelihood estimation is the ability to include expert knowledge through the prior, which can be partly overcome by the data. Nevertheless, an incorrectly chosen prior can still have a detrimental effect on the overall results of the method, especially in cases of limited data availability.

The prior PDFs of the parameters were selected to be conjugate to the likelihood, as described by Tingley and Huybers (2010a), with the exception of the prior of ϕ . The stochastic terms are Gaussian processes: thus the conjugate priors for α , μ are normal, while the priors for σ^2 and $\tau_{I,P}^2$ are inverse gamma. The selected hyperparameters are shown in Table A1 in the appendix in the discussion of the convergence of the algorithm: cf. Gelman (2006) for a discussion on the subtleties of prior selection in hierarchical models.

The draws from the PDFs are created using a Gibbs sampler with one Metropolis step for the draws of the spatial covariance parameter ϕ , as drawing directly from its posterior is more complicated. The first steps of the Gibbs sampler are iterated over the climate field only in order to speed up convergence prior to running the full Gibbs sampler. If the model fits the data reasonably well, the sampler will converge to a final full probability density function of field and parameters. The final set of parameters then can be verified, for example, by predictive experiments using withheld data for validation. We use an initial ensemble of four chains with 5000 iterations for the full Gibbs sampler each. After discarding the first half of the runs, convergence of the parameters is checked using the measure \hat{R} (cf. Gelman et al. 2003; see Table A2). Usually all chains converge during this first run of the code, scarcely needing a second (longer) run of the estimation process. Following Gelman et al., each subsequent run doubles the number of steps. To facilitate data storage and processing the second half of each chain is then thinned out to a mere 250 steps, leaving a total of 1000 draws to estimate the posterior PDFs. The convergence also has been checked by running the code with different initial conditions, and

no significant deviation in the resulting estimates was detected. The code employed is an adaption of the algorithm by Tingley and Huybers (2010a). It has been converted from Matlab to FORTRAN. Additionally, the prior selection was changed to reflect knowledge of the data that was acquired during preprocessing.

Note that the input data are standardized prior to applying all reconstruction methods. This is standard practice in multiproxy reconstructions that use proxy records with variable units in calibration. This simplifies implementation of the stochastic description given above. The data are standardized to have zero mean and uniform variance in the calibration period (simulation years 1856–1980). As remarked by Tingley (2012), standardization of autocorrelated data over a limited time interval leads to variance inflation outside the standardization interval. As the autocorrelation coefficient of the data is on the order of 0.2 only and the interval is 125 years long, the effect is negligible when compared to the other uncertainties in the data. The resulting reconstructions must subsequently be rescaled. At locations without any data, here mainly the region north of 70°N, the calibration mean and standard deviation are of course unknown. The values at these locations are estimated as weighted averages of the nearest neighbors; reconstructions at these locations therefore contain additional uncertainties.

b. Multivariate linear regression

In contrast to the BHM, multivariate linear regression is the underlying formalism of most climate field reconstruction methods used to date. The fundamental approach relates a matrix of climate proxies to a matrix of climate data during a common time interval of overlap (generally termed the calibration interval) using a linear model. For instance, let $\mathbf{P} \in \mathbb{R}^{m \times n}$ and $\mathbf{T} \in \mathbb{R}^{r \times n}$ be the matrices of the pseudoproxy network and the instrumental temperature records, where m is the number of proxies, r is the number of spatial locations in the instrumental field, and n is the temporal dimension corresponding to the period of overlap between the proxy and instrumental data. Here \mathbf{P} and \mathbf{T} can be written in terms of their standardized (centered and normalized; denoted with a prime) form as

$$\mathbf{P} = \mathbf{M}_p + \mathbf{S}_p \mathbf{P}', \quad \mathbf{T} = \mathbf{M}_t + \mathbf{S}_t \mathbf{T}',$$

where \mathbf{M}_p is a matrix of identical columns with each row corresponding to the across-column time average of the matrix \mathbf{P} , and \mathbf{S}_p is a diagonal matrix with elements equal to the standard deviations of the rows of \mathbf{P} ; \mathbf{M}_t and \mathbf{S}_t are similarly defined for \mathbf{T} . Multivariate linear regression derives a relationship between \mathbf{P}' and \mathbf{T}' of the form

$$\mathbf{T}' = \mathbf{B}\mathbf{P}' + \epsilon,$$

where \mathbf{B} is a matrix of regression coefficients and ϵ is the residual error. The mean squared error is minimized if \mathbf{B} is chosen as

$$\mathbf{B} = (\mathbf{T}'\mathbf{P}'^T)(\mathbf{P}'\mathbf{P}'^T)^{-1},$$

where the superscript T denotes the matrix transpose.

The above formalism works best when the temporal dimension is larger than the spatial dimension of both matrices. In global CFR applications, this condition is almost never met; specifically, the time dimension in the calibration interval is often an order of magnitude smaller than the number of spatial grid points that are targeted for reconstruction. The inversion above is therefore underdetermined and the problem requires regularization. For the European case considered herein, however, the number of grid cell locations is 101 and the number of years in the calibration interval is 124. Regularization is therefore not strictly required, but is still applied here to filter noise and weigh the most strongly correlated target and proxy patterns.

c. Canonical correlation analysis

Our general approach employs the canonical correlation analysis (CCA) formalism described in Smerdon et al. (2011). CCA employs singular value decomposition (SVD) to perform three matrix reductions of 1) the standardized proxy matrix, 2) the standardized instrumental matrix, and 3) the cross-covariance matrix that emerges in the expression for \mathbf{B} above. The decomposition of the normalized proxy and instrumental matrices using SVD is written as

$$\mathbf{P}' = \mathbf{U}_p \boldsymbol{\Sigma}_p \mathbf{V}_p^T, \quad \mathbf{T}' = \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{V}_t^T, \quad (4)$$

where the columns of \mathbf{U}_p and \mathbf{U}_t are spatial patterns (empirical orthogonal functions) and the rows of $\boldsymbol{\Sigma}_p \mathbf{V}_p^T$ and $\boldsymbol{\Sigma}_t \mathbf{V}_t^T$ are corresponding orthogonal time series [principal components (PCs)]. The diagonal matrices $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_p$ contain the ordered nonnegative singular values, the squares of which are proportional to the percent variance explained by each principal component. It is often the case in climatological data that the ordered singular values decrease quickly so that a subset of EOF-PC pairs accounts for most of the variance in the original climatic field. Thus we can find sets of d_p and d_t leading EOF-PC pairs that are good approximations of the \mathbf{P}' and \mathbf{T}' matrices, respectively. These reduced-rank versions are written

$$\mathbf{P}'^r = \mathbf{U}_p^r \boldsymbol{\Sigma}_p^r \mathbf{V}_p^{rT}, \quad \mathbf{T}'^r = \mathbf{U}_t^r \boldsymbol{\Sigma}_t^r \mathbf{V}_t^{rT}, \quad (5)$$

where \mathbf{P}'^r and \mathbf{T}'^r are the rank-reduced estimates of \mathbf{P}' and \mathbf{T}' and the superscript r denotes the truncation of a matrix so that only the first r singular values are retained. Substituting the matrices \mathbf{P}'^r and \mathbf{T}'^r into the expression for \mathbf{B} given above yields

$$\mathbf{B} = \mathbf{U}_t^r \boldsymbol{\Sigma}_t^r \mathbf{V}_t^{rT} \mathbf{V}_p^r (\boldsymbol{\Sigma}_p^r)^{-1} \mathbf{U}_p^{rT}.$$

Applying the classical CCA regularization yields

$$\mathbf{B}_{\text{cca}} = \mathbf{U}_t^r \boldsymbol{\Sigma}_t^r \mathbf{O}_t^r \boldsymbol{\Sigma}_{\text{cca}}^r \mathbf{O}_p^{rT} (\boldsymbol{\Sigma}_p^r)^{-1} \mathbf{U}_p^{rT}, \quad (6)$$

where $\mathbf{O}_t^r \boldsymbol{\Sigma}_{\text{cca}}^r \mathbf{O}_p^{rT}$ is the truncated SVD of the cross-covariance matrix $\mathbf{V}_t^{rT} \mathbf{V}_p^r$ that retains a number d_{cca} of canonical coefficients. Note that the upward limit of d_{cca} is given by the dimensions of $\mathbf{V}_t^{rT} \mathbf{V}_p^r$, or $d_{\text{cca}} \leq \min(d_p, d_t)$.

The application of CCA thus requires the selection of three truncation parameters d_p , d_t , and d_{cca} for each reconstruction. Following Smerdon et al. (2010b), we employ a “leave half out” cross-validation technique to optimize the selection of the three CCA dimensions. To perform the leave-half-out cross-validation procedure, the target period is split into two temporal halves. Two sets of reconstructions are generated using all possible parameter combinations and calibrated on each half of the target data. Cross-validation RMSE is calculated on the left-out halves of the target data. These validation statistics from both experiments are combined to yield the statistics for the entire target interval, from which optimal parameter combination are determined. In this manner, full-rank representations of T' , P' and the canonical coefficient matrix are allowed and can in principle be selected based on the cross-validation statistics.

Recent studies (e.g., Smerdon et al. 2010b, 2011) showed that the overall error associated with different multivariate linear regression method is quite similar. It should therefore be sufficient to use one of them as a benchmark for the performance of multivariate regression methods. Additionally, principal component regression, which has been the preferred method of reconstruction over the EU domain (e.g., Luterbacher et al. 2004; Pauling et al. 2006; Riedwyl et al. 2009), is similar to CCA. In PCR the regression matrix is left at full rank while CCA not only truncates the singular value spectra of the instrumental and proxy matrices \mathbf{T}' and \mathbf{P}' but also of the regression matrix \mathbf{B} . Thus, CCA is a suitable and applicable method for evaluating performance of multivariate linear regression based CFRs over the European/Mediterranean domain.

4. Results

In this section we compare the climate field reconstructions from CCA and BHM to the known CCSM

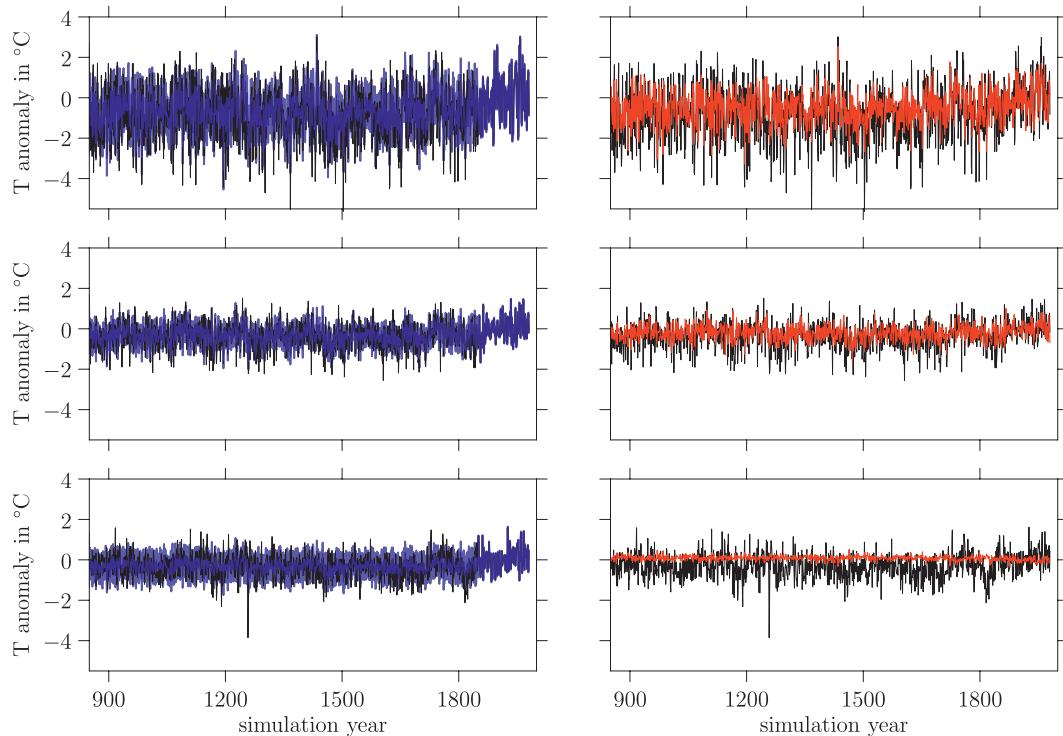


FIG. 1. Comparison of reconstructed gridpoint time series—(left) column BHM, blue line: median, shaded area: 80% uncertainty band; (right) column CCA, red lines—at a proxy SNR of 0.5 for three different locations: (top) 62.5°N, 17.5°E, a location surrounded by proxy data; (middle) 47.5°N, 12.5°E, directly at a proxy location; and (bottom) 32.5°N, 42.5°E, in the southeastern corner of the reconstruction area. Black lines are the target field.

model target during the reconstruction interval. We focus on pointwise error measures in order to assess the performance of the two methods discussed above. Additionally we show qualitative results from selected locations to illustrate some properties of the CFRs.

a. Qualitative comparison

Figure 1 shows a comparison of annual reconstructions from the BHM and CCA methods using a SNR of 0.5. We sample from three points with different amounts of proxy information: 1) the top row (62.5°N, 17.5°E) is a location in Fenno-Scandia that has no local proxy data but several proxy sites close by; 2) the middle row (47.5°N, 12.5°E) is collocated with a proxy site in the Alps; and 3) the bottom row (32.5°N, 42.5°E) is the grid cell in the southeastern corner of the reconstruction area, remote from any proxy information. We plot the annual anomaly data with respect to the 1856–1980 period for CCA using red lines (right column); for the BHM CFR, we choose the mean of the posterior PDFs (heavy blue lines in the left column) as the best estimate and the uncertainty band (light blue area) is the area between the upper and the lower 10% quantiles in the corresponding years. The black lines in all figures show the CCSM target.

The CCA CFR reconstructs, at least for the central and northern sites (top and center row in Fig. 1), much of the variability of the CCSM target time series, albeit with some (small) bias and a reduced variance. For the bottom row, however, the reconstructed temperatures in the grid cell in the southeastern corner of the reconstruction area (bottom row of Fig. 1) yield a variance that is greatly reduced and a larger bias.

For the BHM CFR, the reconstructed annual temperature anomalies encompass most of the target variance. The target remains close to the uncertainty range, which is slightly wider in the top row, corresponding to a slightly higher uncertainty as the proxy information is not collocated with the reconstructed field at this point. For the southeastern corner of the area (bottom row in Fig. 1), the trajectories returned by the Gibbs sampler cover a very high temperature range, fluctuating around the regional reconstruction mean. The estimate provided by the algorithm for this location is very uncertain, although the reconstructed trajectory is still close to the target time series. This result is not surprising, because the model selected for the BHM-based CFR has a simple spatial covariance structure; namely, the dependence of temperature anomalies at two locations decreases exponentially with distance between them. The spatial

TABLE 1. Summary of the spatial skill measures (correlation coefficient, rms error, mean bias, and standard deviation ratio). The median of the skills depicted in Figs. 2–5 for CCA and BHM and both noise levels.

Method	Correlation	RMSE	Bias	SDR
BHM (SNR 0.5)	0.49	0.81	0.10	0.58
BHM (SNR 0.25)	0.29	0.98	0.21	0.55
CCA (SNR 0.5)	0.28	0.86	0.32	0.32
CCA (SNR 0.25)	0.15	0.85	0.41	0.17

correlation length is estimated by the algorithm to be on the order of some 1000 km. The considered point (32.5°N, 42.5°E) is far away from the closest proxy site, so any estimate for the annual temperature anomaly is uncertain, even if the returned estimate looks very promising.

b. Pointwise

To assess the spatial skill of the reconstructions, we use several local error measures used by Smerdon et al. (2011, 2010b). All skill measures are evaluated over the reconstruction period unless explicitly stated. A summary

of the skill measures can be found in Table 1, in which the median values of the spatial skill over the full reconstruction region are shown.

Local cross-correlation coefficients are first calculated between the reconstructions provided by both methods and the target field. The result is displayed in Fig. 2 for both reconstruction methods and noise strengths (top: SNR = 0.5, bottom: SNR = 0.25). All reconstructions exhibit substantially higher cross-correlation coefficients in areas with dense proxy sampling. The Atlantic Ocean and the southeastern target areas where there is no proxy sampling yield correlation coefficients below 0.5, even for the larger SNR case. This leads to an overall decrease of the correlation coefficient shown in Table 1. The BHM CFR performs better for both noise levels, even when including areas that are severely limited by the choice of the model; locations distant from proxy information by design cannot be reconstructed with good skill as information on the climate field exponentially decreases with distance. Note that the BHM returns estimates also for the locations where no instrumental information was available during the calibration period.

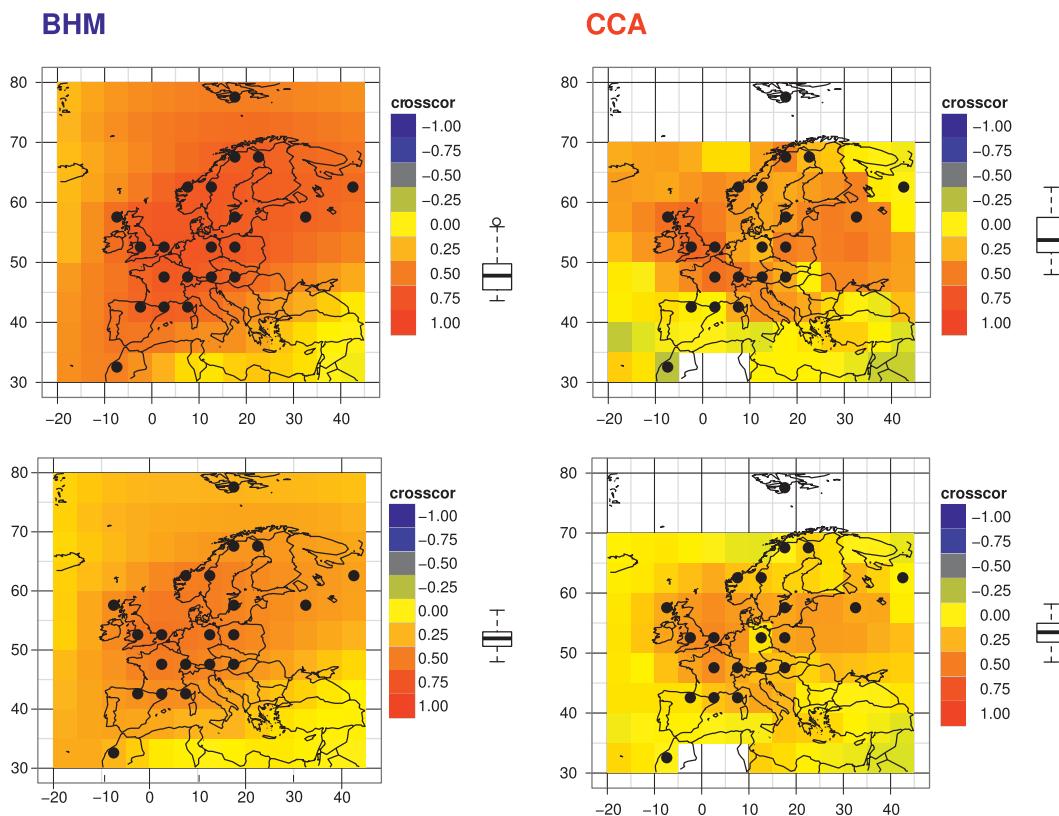


FIG. 2. The correlation coefficient between the reconstructed climate field and the target field (the simulation data) over the reconstruction period at time lag zero. The (left) BHM and (right) CCA climate field reconstructions are derived for two different noise strengths, (top) SNR = 0.5 and (bottom) SNR = 0.25. Points mark location of proxy data. The boxplots to the right of the color bars show the distribution of the grid point correlation coefficients.

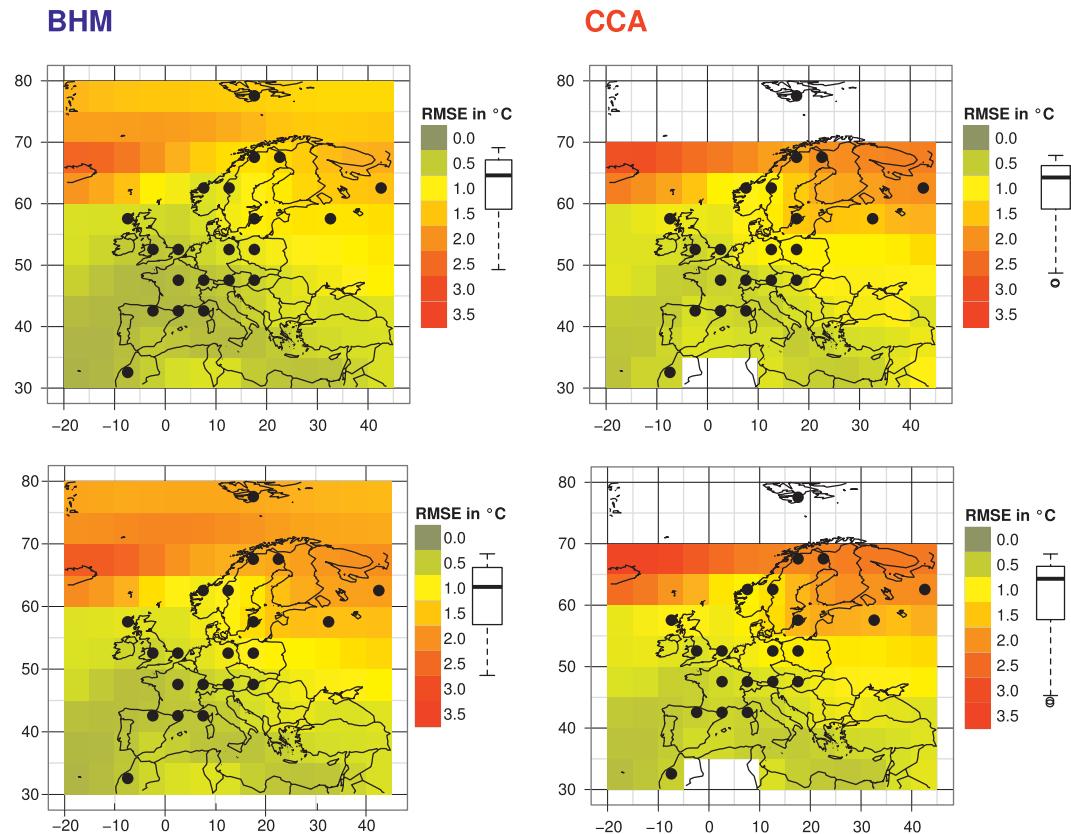


FIG. 3. The local rms error (RMSE) of the reconstruction during the reconstruction period only. The (left) BHM and (right) CCA climate field reconstructions are derived for two different noise strengths, (top) SNR = 0.5 and (bottom) SNR = 0.25. Points mark location of proxy data. The boxplots show the distribution of the RMSE for the corresponding map.

Those points, the area north of 70°N and three grid points in northern Africa, are not reconstructed by the CCA method since the regression needs target data during the instrumental period. No additional input data were used for the BHM-based method, but the algorithm uses the spatial covariance structure in the model equations (1) to fill the gaps. As discussed in section 3, results in these regions should be interpreted carefully owing to the unknown calibration period mean and variance in the grid cells. They are, however, included in the distributions shown as boxplots to the right of the color bars in Figs. 2–4.

Reconstruction errors are additionally measured using the rms error (RMSE). The overall picture plotted in Fig. 3 (again with BHM and CCA in the left and right columns, respectively, and stronger noise in the bottom row) is comparable to that shown by the cross correlation. The RMSE nevertheless can be large in some locations. The CCA CFR has an error of more than 3°C over Iceland, where the RMSE of the BHM CFR also approaches 3°C. Judging from the overview in Table 1, both reconstruction methods perform on average about

equally well. In general, central European temperature anomalies are again reconstructed more skillfully than northern European/Atlantic Ocean ones. In contrast to the median of the correlation coefficient, the median of the RMSE of the CCA CFR is lower than that of the BHM CFR in the stronger noise case, indicating better performance. Comparing the two box plots in Fig. 3 shows that the bulk of the distribution of RMSEs (the box) covers a similar range for both methods, the result for the CCA CFR is more skewed and has higher values at several points (marked by the outliers).

The correlation coefficient is calculated with respect to the mean of the time series and normalized by the standard deviation (e.g., cf. Taylor 2001). To decouple the errors in mean and amplitude of climate variability, we evaluate both the mean bias of the reconstruction relative to the target and the standard deviation ratio between the CFR and target fields. The spatial distribution of mean bias of both reconstructions is shown in Fig. 4. Note that the BHM-based reconstruction (left panels) exhibits an overall lower temperature bias than the CCA reconstruction (right panels) for both noise

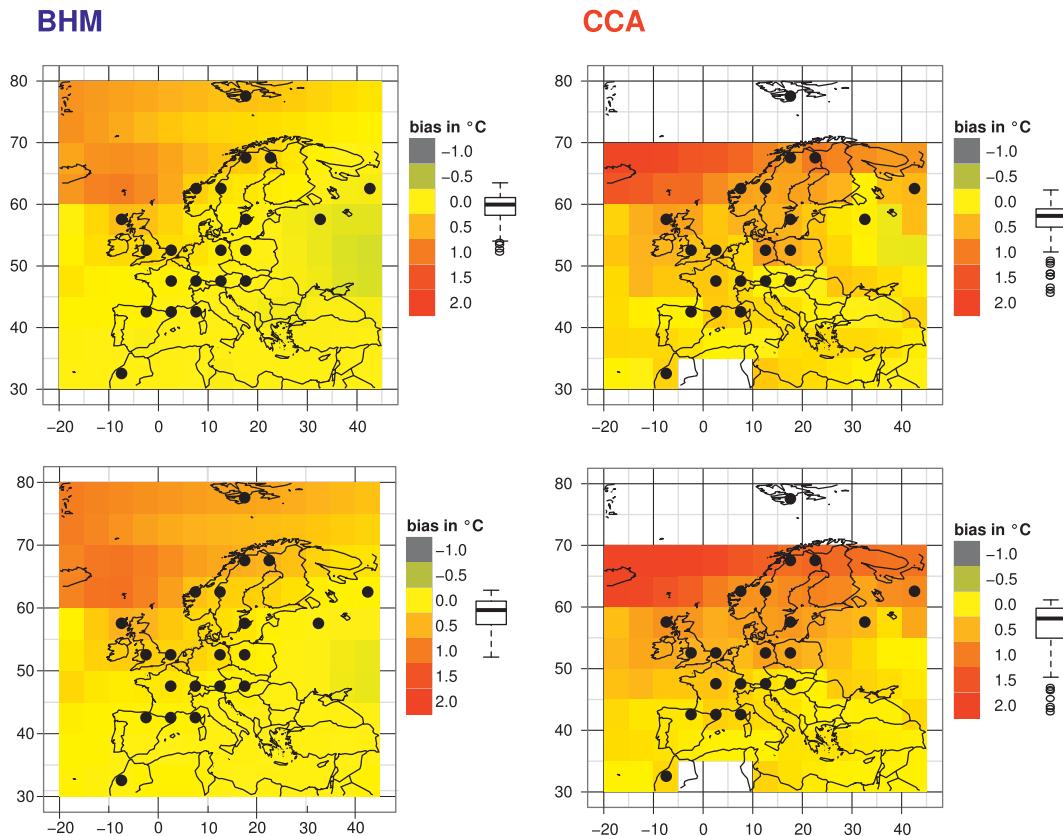


FIG. 4. The local mean bias of the reconstructions over the reconstruction period. The (left) BHM and (right) CCA CFRs are derived for two different noise strengths, (top) SNR = 0.5 and (bottom) SNR = 0.25. Points mark location of proxy data. The box plots to the right of the color bars show the distribution of the bias.

levels. This can also be seen in Table 1 in which the average mean bias is 0.08° and 0.21°C for the BHM CFR with a SNR of 0.5 and 0.25, respectively. The values are substantially higher for both CCA CFRs (0.32 and 0.41), although large biases are limited to the northernmost portion of the reconstruction area where it exceeds 2°C in some areas. The BHM CFR can deal much better with the recent warming, except for northern Europe—the higher rate of recent warming in that area cannot be reconstructed using a uniform mean μ . Clearly some refinement of the model is thus needed in the future.

We also present the standard deviation ratios (SDRs) of the temperature anomalies, indicating how well the interannual temperature variability is reconstructed. The standard deviation of the temperature reconstructions at the different grid points is calculated and then divided by the standard deviation of the target field's temperature anomalies at that point. The aforementioned feature of the BHM method, the drawing of several thousand trajectories of the climate field, can lead to problems when calculating the climate variability because an average of the returned trajectories for the field

reconstruction is used. The variability of this average can be substantially smaller than that of the different trajectories. This can be seen in the bottom panels of Fig. 1 when comparing the width of the uncertainty band made by the different reconstruction trajectories to the variability of the best guess (median). When comparing the variability of the reconstructions to that of the target, we therefore use the trajectory variability, which is in turn the stochastic term σ in Eq. (1) corrected by the normalization parameters of the input data.

The result is shown in Fig. 5. The CCA-based reconstruction (right panel) underestimates the variability for many locations in the targeted region. This is a common feature in regression-based CFRs (Smerdon et al. 2011; Christiansen et al. 2009). For northeastern Europe, however, the climate variability is overestimated in a few grid cells. The BHM-based method, shown in the left panel of Fig. 5, performs differently: a slight overestimation of the climate variability in the north and in the west can be identified. While we normalize the input instrumental and proxy data, we still attribute this outcome to the higher interannual temperature variability in the north in the

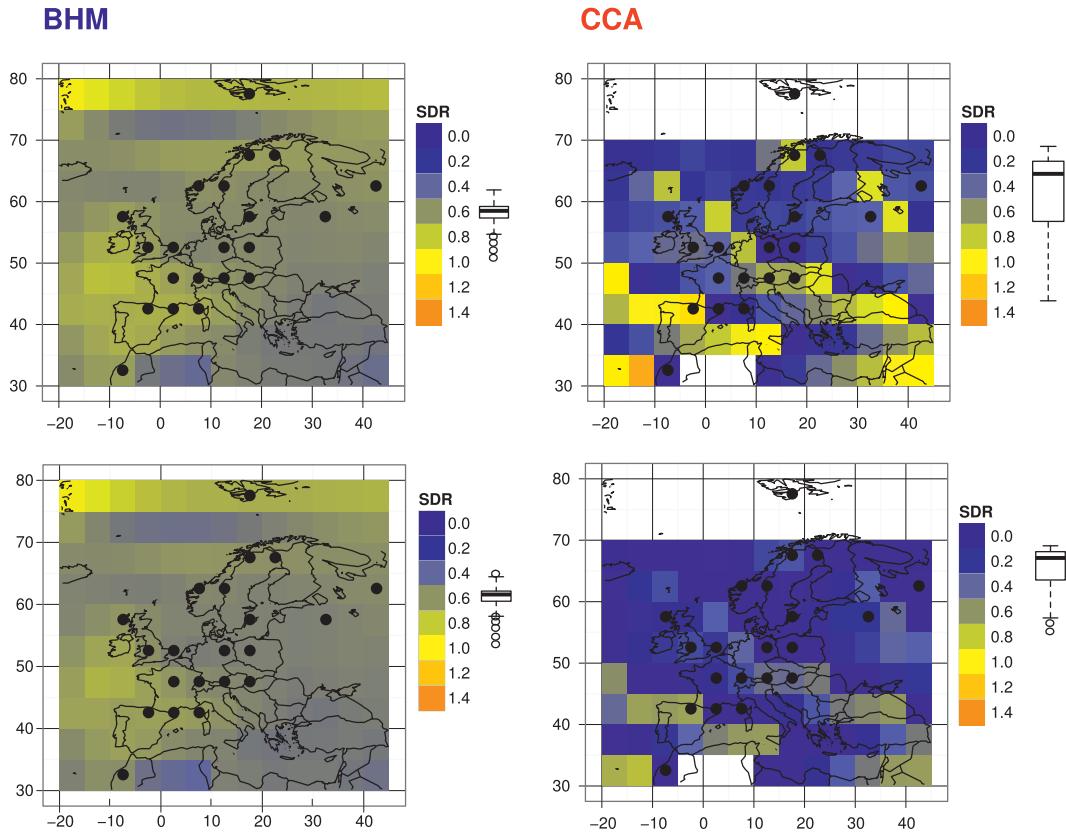


FIG. 5. The pointwise standard deviation ratio (SDR) of the reconstruction and the simulation (target field) over the reconstruction period. The (left) BHM and (right) CCA CFRs are derived for two different noise strengths, (top) SNR = 0.5 and (bottom) SNR = 0.25. Points mark location of proxy data. The box plots to the right of the color bars show the distribution of the SDR.

CCSM field and, for the northernmost cells, the estimation of the correct normalization parameters from the neighboring locations.

c. Area averages

Figure 6 (top row) compares the area mean of the two reconstruction methods (CCA in red, BHM in blue) for SNR = 0.5 with the model target field (black), smoothed using an 11-yr floating average; all the error measures for the averaged temperature anomalies over the full reconstruction region are shown in Table 2. While both reconstructions follow the general shape of the target quite well—the decadal variations of both show good agreement with the target—the CCA CFR exhibits both a higher temperature bias and reduced variability. These findings are also represented in the box plot (right panel in Fig. 6) for the annual averaged temperature anomalies. The BHM CFR shows a comparatively lower bias; however, interannual variability is inflated. The CCA CFR shows again a substantially higher bias and a reduced variability.

Using only data from the area (40°–60°N, 0°–20°E) results in an improvement for CCA (cf. Fig. 6, bottom row), while the interannual variability of the BHM-based results is slightly decreased. Omitting the proxy-sparse region, where performance of CCA was relatively poor, leads to substantially better performance. It is worth noting that some of the CCA performance on the fringes of the domain could potentially result from the constrained identification of the EOF patterns. These patterns in the European domain may be better identified in hemispheric or global reconstructions, and thus improve skill for EOF-based multivariate regression approaches such as CCA.

5. Conclusions and outlook

While BHM-based reconstructions perform well over areas with dense proxy networks, the performance of the model used herein decreases with spatial distance from proxy sites. The discussed stochastic model provides a mechanism to estimate pointwise climate field variables

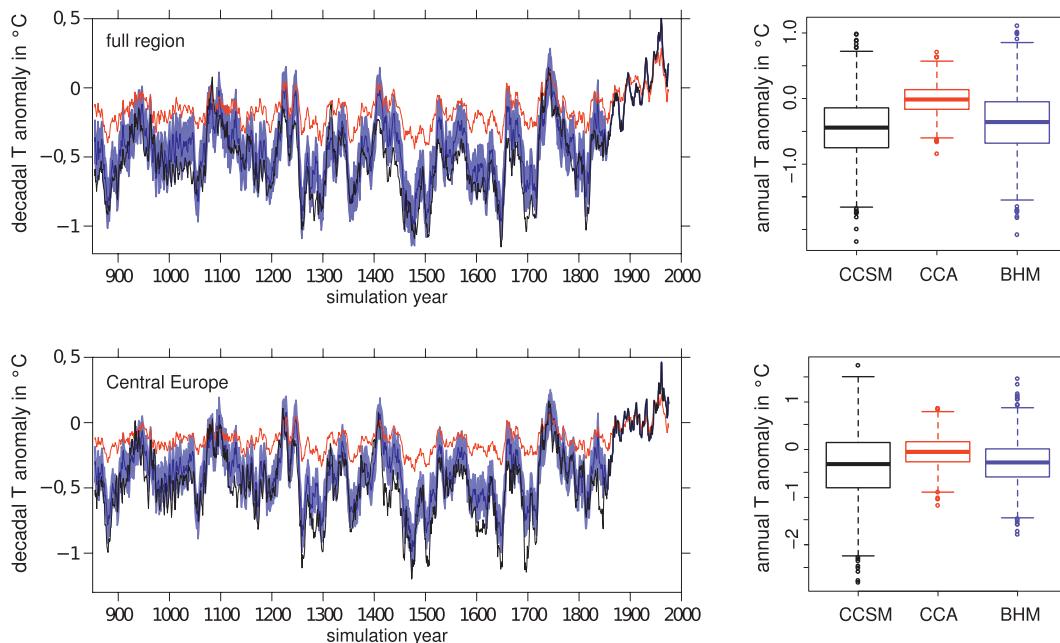


FIG. 6. Weighted area averages over (top) the whole reconstruction domain and (bottom) western central Europe (40° – 60° N, 0° – 20° E). (left) Time series: weighted area average of the target field (black), BHM reconstruction (blue) with 80% uncertainty band (shaded), and CCA reconstruction (red), all smoothed with an 11-yr running average. (right) Box plots of the weighted area average (annual data); note the different ranges of the ordinates.

from more than one data source, as the reconstruction takes into account not only the collocated data but also data that are modeled to share a common signal. The employed model does, however, rely on a stationary stochastic description of the climate field and a suitable model for each type of proxy considered. While efforts exist to provide such models, for example, for pollen–biome relations (Ohlwein and Wahl 2012), coral (Thompson et al. 2011), or tree-growth-based climate reconstructions (Tolwinski-Ward et al. 2010), inversion of these models is cumbersome and calculation of the full posterior PDFs in a computationally convenient form remains challenging. While it would be possible to formulate the entire problem using only Metropolis-like steps, the computational costs are currently significant.

In contrast to this, the EOF- and multivariate regression-based reconstruction methods do not rely on a definite proxy response function, facilitating inclusion of many

different proxy types. Skill in data-sparse regions nevertheless was found to be limited and CCA reconstructions suffer both from a substantial bias and variance loss. The returned CFRs, while not directly accompanied with a suitable error measure in CCA, should thus be considered with care, taking into account the behavior observed in the presented pseudoproxy results. However, one advantage is that the computational costs of regression-based methods are substantially lower when not factoring in elaborate estimation of error measures.

The results from this study indicate that a BHM CFR is generally superior to CCA CFRs over the European/Mediterranean area. This result is not limited to a single error measure: all of the measures considered in this article show better performance of the BHM CFR when compared to the CCA CFR with the exception of the field RMSE for $\text{SNR} = 0.25$. Additionally, a comparison

TABLE 2. Table of skill measures for the weighted average temperature anomalies over the reconstruction area, both for annual and decadal smoothed (11-yr floating average) anomalies.

Method	Annual				Decadal			
	Correlation	RMSE	Bias	SDR	Correlation	RMSE	Bias	SDR
BHM (SNR 0.5)	0.83	0.28	0.07	1.33	0.93	0.13	0.08	0.89
BHM (SNR 0.25)	0.53	0.52	0.21	1.11	0.70	0.28	0.21	0.92
CCA (SNR 0.5)	0.72	0.57	0.48	0.51	0.86	0.50	0.48	0.43
CCA (SNR 0.25)	0.46	0.71	0.60	0.32	0.65	0.63	0.60	0.21

of weighted area averages over both the full reconstruction area and an area with high proxy availability shows that the BHM CFR 1) has smaller warm biases, 2) recovers the interannual variability much better than the CCA CFR, and 3) the performance increase for higher proxy availability is more enhanced for the regression based method.

Over larger and less homogeneous areas (e.g., the full Northern Hemisphere), the BHM CFR based on the discussed model cannot be used. The model must be refined significantly to reflect, for example, the different behavior of landmasses versus oceans. Also, the information gain from long-range teleconnections that CCA CFRs rely on is likely much higher. Through long range atmospheric waves, the synoptic situation over the Atlantic Ocean does indeed influence Europe, a relationship that is exploited by using proxy information from that region. The stochastic model used in the BHM CFR needs to be adapted to make use of these teleconnections.

The results from our version of BARCAST are encouraging—even with a stochastic model that is far from optimal, as indicated by some of the posterior PDFs of the parameters, performance is superior to the multivariate regression based CCA CFRs used herein. These results, along with the added value of impartial error estimates, warrant both the additional scientific work needed to develop and invert appropriate stochastic models and the computational costs associated with the Bayesian inference used in this method.

Future work to use similar methods to reconstruct real-world seasonal temperature and precipitation variability over the European/Mediterranean area, using new high-resolution data from different archives and collaborations such as the PAGES 2K initiative (Newman et al. 2010) and the sixth EU framework program MILLENNIUM, is underway. While some parts of the stochastic model can remain as is, problems such as the time scale separation should be addressed more closely, both through careful analysis of observation and model data and theoretical consideration of the dynamical processes involved.

Acknowledgments. Supported in part by CIRCE (36961), ACQWA (212250), the Deutsche Forschungsgemeinschaft, SPP INTERDYNAMIK, projects PRIME (LU1608/1-1, ZO - 133/6-1) and PRIME2k (LU1608/1-2, ZO-133/6-2), “Historical climatology of the Middle East based on Arabic sources back to ad 800”; LU 1608/2-1), NSF Grant AGS0902436, and NOAA Grant NA10OAR4320137. JW gratefully acknowledges travel support by Columbia University, New York, and the University of Colorado, Boulder, Colorado, that facilitated workshop interactions that were important for the early development of this work.

TABLE A1. Overview of the prior PDFs and their parameters for the BHM-based CFR.

Parameter	Form	Hyperparameters
α	Normal	$N(\alpha_\mu, \alpha_\sigma), \alpha_\mu = 0.1, \alpha_\sigma = 0.1$
μ	Normal	$N(\mu_\mu, \mu_\sigma), \mu_\mu = 0, \mu_\sigma = 5$
σ^2	Inv-gamma	shape = 3.5, scale = 42
ϕ	Lognormal	$\log\phi \sim N(\Phi_\mu, \Phi_\sigma), \Phi_\mu = -7, \Phi_\sigma = 1.2$
τ_I^2	Inv-gamma	shape = 0.5, scale = 0.5
τ_P^2	Inv-gamma	shape = 0.5, scale = 5
β_0	Normal	$N(\beta_{0,\mu}, \beta_{0,\sigma}), \beta_{0,\mu} = 0, \beta_{0,\sigma} = 4$
β_1	Normal	$N(\beta_{1,\mu}, \beta_{1,\sigma}), \beta_{1,\mu} = 0.73, \beta_{1,\sigma} = 4$

APPENDIX

Performance of BARCAST in This Study

In Table A1 we show the selected priors for the model parameters and the hyperparameters. The priors were chosen to be conjugate to the likelihoods to facilitate formulation of the problem. The hyperparameters were chosen after analyzing the input data. The mean for the persistence parameter, α_μ , was estimated through Kramers–Moyal expansion of the instrumental data, also verifying the uniform persistence (except for some grid cells over the Atlantic Ocean). The standard deviation α_σ is relatively wide, indicating the uncertainty of this preliminary data analysis, as only 130 years of instrumental data are used. The prior of the mean temperature, μ , is also normally distributed with mean of the instrumental period $\mu_\mu = 0$ (as given by normalization of the data) and a large standard deviation ($\mu_\sigma = 5$). The prior of the interannual temperature variability σ^2 is inverse gamma, with shape = 3.5, scale = 35, as estimated from the temperature data of the instrumental period. The prior of the spatial correlation length ϕ is lognormal [$\log\phi \sim N(\Phi_\mu, \Phi_\sigma)$]. It is centered around $\Phi_\mu = -7$ (corresponding to about 1000 km) with a relatively wide standard deviation of $\Phi_\sigma = 1.2$ (corresponding to a range between some 100 and 3500 km; cf. Tingley and Huybers 2010a). The priors of both the instrumental measurement error and the proxy noise, $\tau_{I,P}^2$ respectively, are inverse gamma. Shape and scale for τ_I^2 are 0.5 each, corresponding to one observation with a standard deviation of 1. The shape and scale of the prior for τ_P^2 are 0.5 and 5, respectively, corresponding to one observation with a standard deviation of 2. This was also a selection by Tingley and Huybers (2010a). Both priors are relatively wide and uninformative. The priors of the parameters of the linear proxy response function $W_i = \beta_0 + \beta_1 T_i$ are both normal. As the data have been pre-processed to have zero mean and unit variance, the scaling β_1 is expected to be $[(1 - \tau_P^2)(1 - \alpha^2)/\sigma^2]^{(+1/2)}$, which was chosen as the prior mean with a rather large

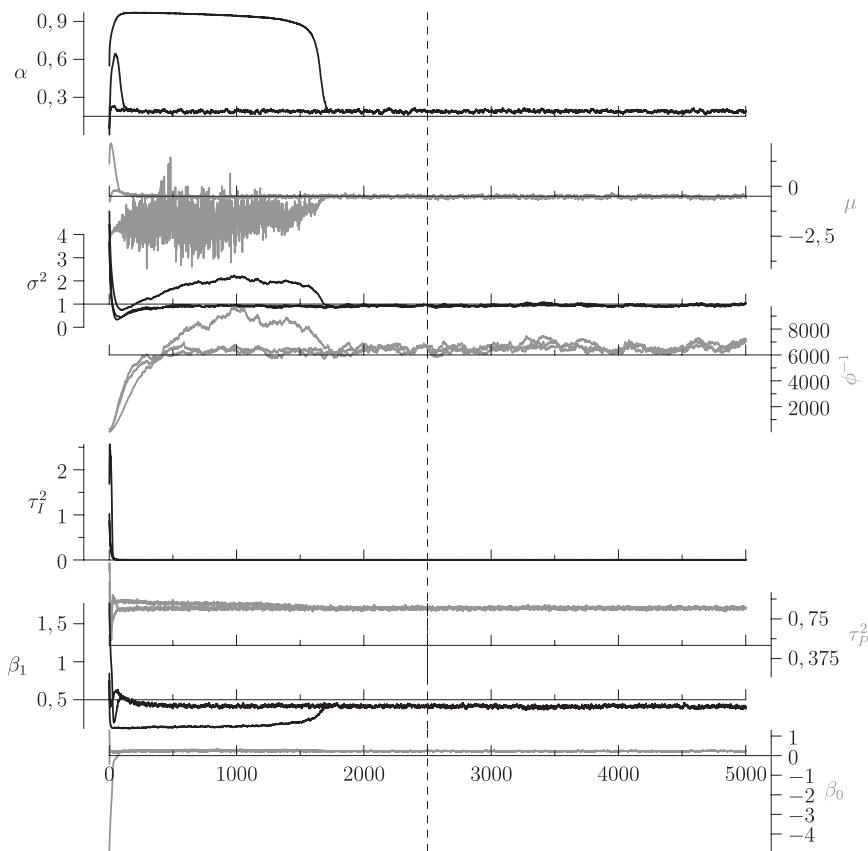


FIG. A1. Draws of the Gibbs sampler of the model parameters vs the step number of the sampler. After about 2500 steps (dashed vertical line), the algorithm converges to a final distribution.

variance of 8. The offset β_0 is related to the mean temperature and the mean proxy value in the instrumental period: due to the normalization we expect it to be zero but, as with the scaling, add a substantial uncertainty to this estimate.

We now evaluate convergence of the parameters. A purely qualitative and often misleading (cf. Gelman et al. 2003) way is the visual inspection of the draws. In Fig. A1 we display the parameters versus the iteration step for all chains for a SNR of 0.5. All parameters stabilize around a mean value after about 1000 iteration steps. Discarding the first 2500 steps, we now evaluate the measure \hat{R} , following the definition by Gelman et al. (2003, chapter 11). The measure essentially compares the within chain variance to the between chain variance. If all chains fluctuate around the same mean with a similar variability, the value of \hat{R} is close to one. The values of \hat{R} for both noise strengths and all eight parameters are displayed in Table A2. Convergence of the parameters can indeed be achieved after a few hundred iterations. Note that convergence of the parameters does not imply convergence of all temperature values. As there are 1132×130 temperature values to be

checked we do not display results for those. In Fig. A2 we show histograms of the draws (steps 2501 to 5000 of the Gibbs sampler) for the parameters. The prior probability density functions are omitted: As discussed above and shown in Table A1 the priors are relatively wide and uninformative and no insights could be gained from visual comparison of prior and estimated posterior PDFs. The resulting estimates for the posterior PDFs are all of the expected form as derived from the model with the exception of the interannual variability σ^2 and the spatial correlation length ϕ . Note that this can be attributed to the strong conditional dependence between those two parameters (cf. Tingley and Huybers 2010a,b). This can also be readily recognized in Fig. A3. The expected logarithmic interdependence from the

TABLE A2. Overview of the convergence measures \hat{R} for all parameters and both noise levels.

	Parameter							
	σ	μ	σ^2	ϕ	τ_I^2	τ_P^2	β_0	β_1
$\hat{R}_{\text{SNR } 0.5}$	1.0044	0.9999	0.9999	1.0007	1.0006	0.9999	1.0044	1.0015
$\hat{R}_{\text{SNR } 0.25}$	1.0057	0.9999	0.9999	1.0008	1.0008	0.9999	1.0007	1.0161

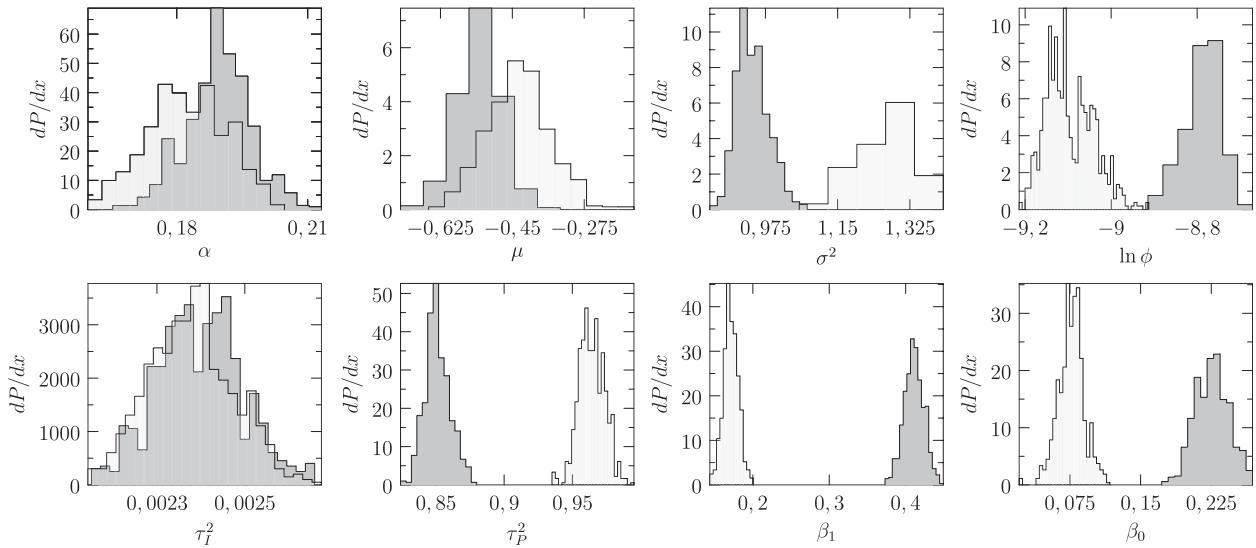


FIG. A2. Histogram of the draws of the model parameters of steps 2501–5000 of the Gibbs sampler: SNR = 0.5 (dark shading) and SNR = 0.25 (light shading).

model indeed seems to be present in the subsequent draws. Additionally the resulting spatial correlation length is rather large. This can probably in part be attributed to the recent, almost uniform warming over most of the reconstruction area in the instrumental period. Also, when evaluating correlation patterns over the target area, the first patterns (the warming and two large-scale dipole patterns) already explain far in excess of 85% of the variability. This is also expressed by the strong link between temperature anomalies in Poland and mean European temperature anomalies discussed by Luterbacher et al. (2010b). Additionally, the assumption of an isotropic spatial variance–covariance structure is not optimal.

The resulting modes of the posterior PDFs of the other parameters do match the outside knowledge. The persistence is slightly overestimated due to the few Atlantic Ocean grid cells as discussed above. Estimates for the interannual variability in low and high noise case differ a bit, as does the spatial correlation length: the stronger noise decreases spatial correlation in the reconstruction period and increases the resulting variability in the posterior PDFs of the temperatures. Draws for the proxy noise are close to the true constructed values of 80% noise variance and 94% noise variance for SNR = 0.5 and SNR = 0.25, respectively. The draws for the linear proxy response function roughly match the expected values.

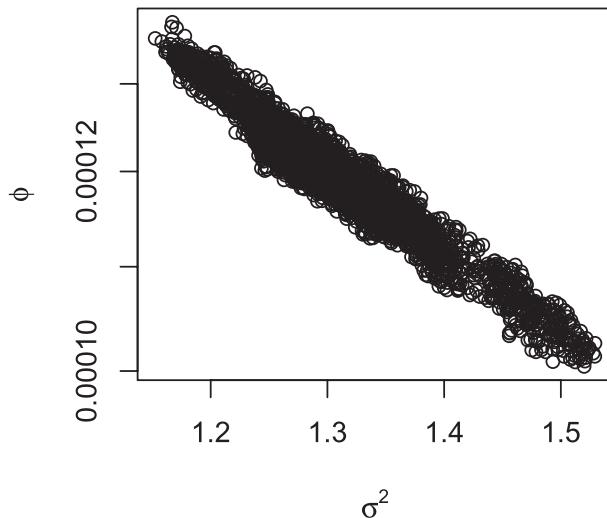


FIG. A3. Draws of ϕ vs corresponding draws of σ^2 .

REFERENCES

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas, 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.

Anishchenko, V., V. Astakhov, A. Neiman, T. Vadisova, and L. Shimansky-Geier, 2002: *Nonlinear Dynamics of Chaotic and Stochastic Systems*. 1st ed. Springer-Verlag, 374 pp.

Büntgen, U., and Coauthors, 2011: 2500 years of European climate variability and human susceptibility. *Science*, **331**, 578–582.

Bürger, G., I. Fast, and U. Cubasch, 2006: Climate reconstruction by regression—32 variations on a theme. *Tellus*, **58A**, 227–235.

Christiansen, B., T. Schmith, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, **22**, 951–976.

Cook, E. R., K. R. Briffa, and P. D. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *Int. J. Climatol.*, **14**, 379–402.

- , D. M. Meko, D. W. Stahle, and M. K. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate*, **12**, 1145–1162.
- Cranz, D., 1770: Historie von Grönland: Enthaltend die Beschreibung des Landes und der Einwohner ec., insbesondere die Geschichte der dortigen Mission der Evangelischen Brüder zu Neu-Herrnhut und Lichtenfels: mit 8 Kupfertafeln und einem Register. 2nd ed. Heinrich Detlef Ebers, Barby and Weidmanns Erben und Reich, 1128 pp.
- D'Arrigo, R. D., R. Allan, R. Wilson, J. Palmer, J. Sakulich, J. E. Smerdon, S. Bijaksana, and L. O. Ngkoimani, 2008: Pacific and Indian Ocean climate signals in a tree-ring record of Java monsoon drought. *Int. J. Climatol.*, **28**, 1889–1901, doi:10.1002/joc.1679.
- Einstein, A., 1905: Über die von der Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, **17**, 549–560.
- , 1906: Zur Theorie der Brownschen Bewegung. *Ann. Phys.*, **19**, 371–381.
- Esper, J., D. C. Frank, R. J. S. Wilson, and K. R. Briffa, 2005: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. *Geophys. Res. Lett.*, **32**, L07711, doi:10.1029/2004GL021236.
- , and Coauthors, 2012: Orbital forcing of tree-ring data. *Nat. Climate Change*, **2**, 862–866, doi:10.1038/NCLIMATE1589.
- Evans, M. N., B. K. Reichert, A. Kaplan, K. J. Anchukaitis, E. A. Vaganov, M. K. Hughes, and M. A. Cane, 2006: A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J. Geophys. Res.*, **111**, G03008, doi:10.1029/2006JG000166.
- Gardiner, C. W., 1990: *Handbook of Stochastic Methods*. 2nd ed. Springer-Verlag, 447 pp.
- Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.*, **1**, 515–533.
- , J. Carlin, H. Stern, and D. Rubin, 2003: *Bayesian Data Analysis*. 2nd ed. Chapman & Hall, 668 pp.
- Gómez-Navarro, J. J., J. P. Montávez, S. Jerez, P. Jiménez-Guerrero, R. Lorente-Plazas, J. F. González-Rouco, and E. Zorita, 2011: A regional climate simulation over the Iberian Peninsula for the last millennium. *Climate Past*, **7**, 451–472, doi:10.5194/cp-7-451-2011.
- González-Rouco, F., H. von Storch, and E. Zorita, 2003: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. *Geophys. Res. Lett.*, **30**, 2116, doi:10.1029/2003GL018264.
- , H. Beltrami, E. Zorita, and H. Von Storch, 2006: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling. *Geophys. Res. Lett.*, **33**, L01703, doi:10.1029/2005GL024693.
- Goosse, H., E. Crespin, A. de Montety, M. Mann, H. Renssen, and A. Timmermann, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J. Geophys. Res.*, **115**, D09108, doi:10.1029/2009JD012737.
- Guiot, J., A. Nicault, C. Rathgeber, J. L. Edouard, E. Guibal, G. Pichard, and C. Till, 2005: Last-millennium summer-temperature variations in western Europe based on proxy data. *Holocene*, **15**, 489–500.
- , and Coauthors, 2010: Growing season temperatures in Europe and climate forcings over the past 1400 years. *PLoS ONE*, **5**, e9972, doi:10.1371/journal.pone.0009972.
- Hasselmann, K., 1976: Stochastic climate models. Part I: Theory. *Tellus*, **28**, 473–485.
- Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new, validated 1500-year temperature reconstruction. *J. Climate*, **20**, 650–666.
- Jansen, E. J., and Coauthors, 2007: Palaeoclimate. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 433–497.
- Jones, P. D., M. New, D. Parker, S. Martin, and I. Rigor, 1999: Surface air temperature and its changes over the last 150 years. *Rev. Geophys.*, **37**, 173–199.
- , and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *Holocene*, **19**, 3–49.
- Just, W., H. Kantz, C. Rödenbeck, and M. Helm, 2001: Stochastic modelling: Replacing fast degrees of freedom by noise. *J. Phys.*, **34A**, 3199, doi:10.1088/0305-4470/34/15/302.
- , —, M. Ragwitz, and F. Schmüser, 2003: Nonequilibrium physics meets time series analysis: Measuring probability currents from data. *Europhys. Lett.*, **62**, 28–34.
- Kantz, H., W. Just, N. Baba, K. Gelfert, and A. Riegert, 2004: Fast chaos versus white noise: Entropy analysis and a Fokker–Planck model for the slow dynamics. *Physica D*, **187**, 200–213, doi:10.1016/j.physd.2003.09.006.
- Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner, 2007: Testing a European winter surface temperature reconstruction in a surrogate climate. *Geophys. Res. Lett.*, **34**, L07710, doi:10.1029/2006GL027907.
- , and Coauthors, 2009: The importance of ship log data: Reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750. *Climate Dyn.*, **34**, 1115–1128, doi:10.1007/s00382-009-0577-9.
- , J. Luterbacher, and H. Wanner, 2010: Multidecadal changes in winter circulation–climate relationship in Europe: Frequency variations, within-type modifications, and long-term trends. *Climate Dyn.*, **36**, 957–972, doi:10.1007/s00382-009-0737-y.
- Lee, T. C. K., F. W. Zwiers, and M. Tsao, 2008: Evaluation of proxy-based millennial reconstruction methods. *Climate Dyn.*, **31**, 263–281.
- Li, B., D. Nychka, and C. Ammann, 2010: The value of multiproxy reconstruction of past climate. *J. Amer. Stat. Assoc.*, **105**, 883–895.
- Loewe, F., 1937: A period of warm winters in western Greenland and the temperature see-saw between western Greenland and central Europe. *Quart. J. Roy. Meteor. Soc.*, **63**, 365–372.
- Luterbacher, J., C. Schmutz, D. Gyalistras, E. Xoplaki, and H. Wanner, 1999: Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys. Res. Lett.*, **26**, 2745–2748.
- , and Coauthors, 2000: Monthly mean pressure reconstruction over Europe for the Late Maunder Minimum period. *Int. J. Climatol.*, **20**, 1049–1066.
- , and Coauthors, 2002: Reconstruction of sea level pressure fields over the eastern North Atlantic and Europe back to 1500. *Climate Dyn.*, **18**, 545–561.
- , D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner, 2004: European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, **303**, 1499–1503.
- , M. A. Liniger, A. Menzel, N. Estrella, P. M. Della-Marta, C. Pfister, T. Rutishauser, and E. Xoplaki, 2007: Exceptional European warmth of autumn 2006 and winter 2007: Historical context, the underlying dynamics, and its phenological impacts. *Geophys. Res. Lett.*, **34**, L12704, doi:10.1029/2007GL029951.

- , and Coauthors, 2010a: Circulation dynamics and its influence on European and Mediterranean January–April climate over the past half millennium: Results and insights from instrumental data, documentary evidence and coupled climate models. *Climatic Change*, **101**, 201–234.
- , and Coauthors, 2010b: Climate change in Poland in the past centuries and its relationship to European climate: Evidence from reconstructions and coupled climate models. *The Polish Climate in the European Context: A Historical Overview*, R. Przybylak et al., Eds., Springer, 3–39.
- Mann, M. E., and S. Rutherford, 2002: Climate reconstruction using ‘pseudoproxies.’ *Geophys. Res. Lett.*, **29**, 1501, doi:10.1029/2001GL014554.
- , R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392**, 779–787.
- , S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.
- Newman, L., T. Kiefer, B. Otto-Bliessner, and H. Wanner, 2010: The science and strategy of the Past Global Changes (PAGES) project. *Curr. Opin. Environ. Sustainability*, **2**, 193–201, doi:10.1016/j.cosust.2010.04.004.
- North, G., and Coauthors, 2006: *Surface Temperature Reconstructions for the Last 2,000 Years*. The National Academies Press, 145 pp.
- Ohlwein, C., and E. Wahl, 2012: Review of probabilistic pollen-climate transfer methods. *Quat. Sci. Rev.*, **31**, 17–29, doi:10.1016/j.quascirev.2011.11.002.
- Pauling, A., J. Luterbacher, and H. Wanner, 2003: Evaluation of proxies for European and North Atlantic temperature field reconstructions. *Geophys. Res. Lett.*, **30**, 1787, doi:10.1029/2003GL017589.
- , —, C. Casty, and H. Wanner, 2006: 500 years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Climate Dyn.*, **26**, 307–405.
- Riedwyl, N., M. Küttel, J. Luterbacher, and H. Wanner, 2009: Comparison of climate field reconstruction techniques: Application to Europe. *Climate Dyn.*, **32**, 381–395.
- Risken, H., 1989: *The Fokker–Planck Equation: Methods of Solution and Applications*. Springer-Verlag, 472 pp.
- Rutherford, S., M. Mann, E. Wahl, and A. C., 2008: Reply to comment by Jason E. Smerdon et al. on “Robustness of proxy-based climate field reconstruction methods.” *J. Geophys. Res.*, **113**, D18107, doi:10.1029/2008JD009964.
- Schmidt, G. A., and Coauthors, 2011: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.*, **4**, 33–45, doi:10.5194/gmd-4-33-2011.
- Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *WIREs Climate Change*, **3**, 63–77, doi:10.1002/wcc.149.
- , A. Kaplan, and D. Chang, 2008: On the origin of the standardization sensitivity in RegEM climate field reconstructions. *J. Climate*, **21**, 6710–6723.
- , —, and D. E. Amrhein, 2010a: Erroneous model field representations in multiple pseudoproxy studies: Corrections and implications. *J. Climate*, **23**, 5548–5554.
- , —, D. Chang, and M. N. Evans, 2010b: A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J. Climate*, **23**, 4856–4880.
- , —, E. Zorita, J. F. González-Rouco, and M. N. Evans, 2011: Spatial performance of four climate field reconstruction methods targeting the Common Era. *Geophys. Res. Lett.*, **38**, L11705, doi:10.1029/2011GL047372.
- Stemler, T., J. P. Werner, H. Benner, and W. Just, 2007: Stochastic modeling of experimental chaotic time series. *Phys. Rev. Lett.*, **98**, 044102, doi:10.1103/PhysRevLett.98.044102.
- Taylor, K., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106** (D7), 7183–7192.
- Thompson, D. M., T. R. Ault, M. N. Evans, J. E. Cole, and J. Emile-Geay, 2011: Comparison of observed and simulated tropical climate trends using a forward model of coral $\delta^{18}\text{O}$. *Geophys. Res. Lett.*, **38**, L14706, doi:10.1029/2011GL048224.
- Tingley, M. P., 2012: A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. *J. Climate*, **25**, 777–791.
- , and P. Huybers, 2010a: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *J. Climate*, **23**, 2759–2781.
- , and —, 2010b: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: Comparison with the regularized expectation–maximization algorithm. *J. Climate*, **23**, 2782–2800.
- , P. F. Craigmile, M. Haran, B. Li, E. Mannshardt-Shamseldin, and B. Rajaratnam, 2012: Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.*, **35**, 1–22.
- Tolwinski-Ward, S. E., M. N. Evans, M. K. Hughes, and K. J. Anchukaitis, 2010: An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dyn.*, **36**, 2419–2439.
- van Loon, H., and J. C. Rogers, 1978: The seesaw in winter temperatures between Greenland and northern Europe. Part I: General description. *Mon. Wea. Rev.*, **106**, 296–310.
- Wahl, E. R., and J. E. Smerdon, 2012: Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. *Geophys. Res. Lett.*, **39**, L06703, doi:10.1029/2012GL051086.
- Widmann, M., H. Goosse, G. van der Schrier, R. Schnur, and J. Barkmeijer, 2010: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate Past*, **6**, 627–644, doi:10.5194/cp-6-627-2010.
- Xoplaki, E., J. Luterbacher, N. S. H. Paeth, D. Dietrich, M. Grosjean, and H. Wanner, 2005: European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophys. Res. Lett.*, **32**, L15713, doi:10.1029/2005GL023424.