



Technical Note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model

J. P. Werner¹ and M. P. Tingley²

¹Department for Earth Science and Bjerknes Centre for Climate Research, University of Bergen, P.O. Box 7803, 5020 Bergen, Norway

²Department of Meteorology and Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

Correspondence to: J. P. Werner (johannes.werner@geo.uib.no)

Received: 17 October 2014 – Published in Clim. Past Discuss.: 5 December 2014

Revised: 16 February 2015 – Accepted: 2 March 2015 – Published: 24 March 2015

Abstract. Reconstructions of the late-Holocene climate rely heavily upon proxies that are assumed to be accurately dated by layer counting, such as measurements of tree rings, ice cores, and varved lake sediments. Considerable advances could be achieved if time-uncertain proxies were able to be included within these multiproxy reconstructions, and if time uncertainties were recognized and correctly modeled for proxies commonly treated as free of age model errors.

Current approaches for accounting for time uncertainty are generally limited to repeating the reconstruction using each one of an ensemble of age models, thereby inflating the final estimated uncertainty – in effect, each possible age model is given equal weighting. Uncertainties can be reduced by exploiting the inferred space–time covariance structure of the climate to re-weight the possible age models. Here, we demonstrate how Bayesian hierarchical climate reconstruction models can be augmented to account for time-uncertain proxies. Critically, although a priori all age models are given equal probability of being correct, the probabilities associated with the age models are formally updated within the Bayesian framework, thereby reducing uncertainties. Numerical experiments show that updating the age model probabilities decreases uncertainty in the resulting reconstructions, as compared with the current de facto standard of sampling over all age models, provided there is sufficient information from other data sources in the spatial region of the time-uncertain proxy. This approach can readily be generalized to

non-layer-counted proxies, such as those derived from marine sediments.

1 Introduction

Large-scale climate reconstructions over the last two millennia rely heavily on climatic proxies that are annually resolved, assumed to be precisely dated by layer counting, and overlap with instrumental climate data – including tree rings, varved sediments, and annually layered ice cores (e.g., NRC, 2006; Jones et al., 2009). There would be considerable advantages if time-*uncertain* proxy records were able to be included within robust reconstruction methodologies that allowed for accurate propagation of uncertainties. Such methodologies would allow for radiometrically dated proxies, such as many marine or lacustrine sediment archives and speleothems, to be included in high-resolution reconstructions. These lower-resolution, time-uncertain records may preserve more low-frequency climate variability than annually resolved proxies (e.g., Jones et al., 2009), and therefore aid in identifying and characterizing multi-centennial to millennial scale climate variability over the late Holocene. Furthermore, lower-frequency, time-uncertain proxies have the potential to extend the time span of reconstructions, as proxies like tree rings and corals are abundant for the last several centuries but relatively sparse prior to the most recent millennium (e.g., Mann et al., 2008). Finally, inclusion of time-

uncertain terrestrial and nearshore marine sediment records will expand the spatial coverage of the proxy network.

There is well-established literature on modeling time-uncertainty in $\delta^{18}\text{O}$ observations from marine sediment cores (e.g., Imbrie et al., 1984; Shackleton et al., 1990; Lisiecki and Lisiecki, 2002; Huybers and Wunsch, 2004; Lisiecki and Raymo, 2005). A common approach in this context is to establish the age–depth relationship by maximizing some similarity metric, such as a correlation, between time-uncertain records (e.g., Lisiecki and Lisiecki, 2002; Lisiecki and Raymo, 2005) or between a time-uncertain record and a known, often orbital, signal (e.g., Imbrie et al., 1984; Shackleton et al., 1990). The maximization is often conducted under constraints or penalties designed to reflect the scientific understanding of the sedimentation process, and to guard against over-fitting (e.g., Martinson et al., 1982; Brüggemann, 1992; Lisiecki and Lisiecki, 2002; Lisiecki and Raymo, 2005). Alternatively, depth-derived age modeling emphasizes sedimentation or accumulation models in establishing an age–depth relationship (e.g., Shaw, 1964; Huybers and Wunsch, 2004). These studies generally involve averaging numerous, time-uncertain proxy series to explore temporal variability (Lisiecki and Raymo, 2007; Huybers and Wunsch, 2004), or to explore spatial variability at distinct time slices (e.g., Curry and Oppo, 2005; Hughes et al., 2014). In contrast, late-Holocene climate field reconstruction aims to infer climate variability in both space and time from both time-certain and time-uncertain proxy observations.

Time-uncertain proxies have in some cases been included in reconstructions of large-scale, late-Holocene climate indices or fields (e.g., Moberg et al., 2005; PAGES2k Consortium, 2013). In general, however, the age model errors associated with the time-uncertain proxies has have been explicitly addressed: a “best” age model is selected and then time uncertainties are ignored. Recent attempts to account for age model uncertainties (Anchukaitis and Tierney, 2013; Tierney et al., 2013; Comboul et al., 2014) generally involve repeating an analysis over possible age models, thereby inflating the uncertainty associated with the reconstruction.

Constraints on the age–depth models for time-uncertain observations can be achieved by exploiting information about the spatial and temporal covariance of the climate system. For example, the possible age models for a set of proxy records can be constrained by assuming that well-documented global-scale events, such as magnetic reversals, glacial terminations, or tephra layers (Haflidason et al., 2000), are simultaneous within a specified tolerance across the different time-uncertain records. Such assumptions reflect the intuition that large-scale climate features are recorded at different locations at “close” to the same time, where the extent to which events are permitted to be asynchronous is given by estimates of dating uncertainties, and reflected in tolerances or penalty functions.

Here we formalize the use of time-uncertain proxies within a Bayesian hierarchical model for climate field reconstruction.

Our approach is a data-derived compromise between the two end members defined by the selection of an optimal age model by some metric, or by treating all age models as equally likely. The statistical model is hierarchical, and model fitting exploits conditional dependencies by sequentially updating estimates of the climate conditional on the currently selected age model, and then updating the probabilities associated with members of an ensemble of age–depth models based on the current estimate of the climate. That is, inference on the spatiotemporal covariance structure of the climate process informs the extent to which similar signals in proxy records at different locations should be constrained to occur at similar times, while these time-uncertain observations likewise inform estimates of both the climate process and parameters describing the space–time covariance structure of the climate.

Our development focuses on banded climate archives that feature dating errors caused by skipping or over-counting layers. Measurements on tree rings are the best known and most used of these climate archives is tree ring records, where replication and cross-dating is generally adequate to address possible miscounting of rings. Other climate archives that form annual bands, such as ice cores, varved lake sediments, corals, and some speleothems, are not sampled with sufficient replication to safely assume that the age models are free of errors. We will assume throughout that corresponding to each banded, time-uncertain record there is a set of M age–depth models (ADMs) that, a priori, are treated as equally likely. We do not consider how this ensemble is constructed, but note that our results are generally applicable. Likewise, although we will work with the established method BARCAST (Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time), as described by Tingley and Huybers (2010a, b) and used subsequently in Tingley (2012), Werner et al. (2013), and Tingley and Huybers (2013), our results are readily transferable to any other Bayesian hierarchical model for inferring past climate.

Section 2 provides background information on BARCAST, and Sect. 3 describes the technical modifications to BARCAST required to update the probabilities of the ADMs. Results of numerical experiments characterizing and illustrating the advantages of our approach are presented in Sect. 4. Although the development below is specific for simplicity, the core idea of updating the probabilities associated with time-uncertain proxy data by including an ADM level within a Bayesian hierarchical model is general, and discussion provided in Sect. 5 focuses on how the core ideas can be extended, including to radiometrically derived ADMs.

2 Bayesian hierarchical models for climate field reconstructions

Bayesian hierarchical modeling is a natural framework for inferring past climate from proxy observations (Tingley

et al., 2012), and involves disentangling assumptions made about the climate system from assumptions made about the distribution of observations consistent with a given climate state. More generally, hierarchical modeling allows sophisticated models to be developed via the specification of a series of simpler, interlinked conditional probability statements (Berliner et al., 2000; Wikle et al., 2001; Gelfand et al., 2003; Gelman et al., 2003).

In the paleoclimate context, Bayesian hierarchical modeling proceeds by first specifying, at the process level, a simple, parametric model describing the stochastic variability of the target climate process. The data level of the hierarchy then describes the observations conditional on the state of the climate process – that is, forward models (Evans et al., 2013) are the natural way to describe the relationship between proxies and climate within the Bayesian framework. Finally, prior distributions, encoding pre-analysis beliefs, must be specified for all unknown parameters to provide mathematical closure. These priors can be more or less informative, as dictated by the analyst (e.g., Gelman et al., 2003; Tingley and Huybers, 2010a, b).

BARCAST (Tingley and Huybers, 2010a, b) models the target climate process as a first-order autoregressive [AR(1)] process in time, with multivariate normal innovations featuring exponentially decaying spatial covariance. Although based on a relatively simple process-level model, numerous studies have shown that BARCAST works well in practice for reconstructing temperature variations (e.g., Tingley and Huybers, 2010a; Werner et al., 2013, 2014; Tingley and Huybers, 2013). Mathematically, the processes level takes the form

$$\begin{aligned} C_{t+1} - \mu &= \alpha (C_t - \mu) + \epsilon_t \\ \epsilon_t &\sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (\text{independent}) \\ \Sigma_{i,j} &= \sigma^2 \exp(-\phi|x_i - x_j|), \end{aligned} \tag{1a}$$

where C_t is a vector of the climate values at N spatial locations at time step t , μ is the overall mean of the process, and the AR(1) coefficient α models temporal persistence. The innovations capture spatial persistence in the form of an exponential decrease of correlation as a function of separation between locations x_i and x_j , with e -folding distance $1/\phi$. As reflected in the process-level specification, the climate is persistent in both time and space, and the spatial and temporal sharing of information permitted by this persistence is critical in constraining age models for time-uncertain proxies. The climate process C is latent, in the sense that it is never observed without error.

At the data level, BARCAST specifies a separate linear forward model for each type of observation:

$$\begin{aligned} O_t &= \beta_0 + \beta_1 \cdot H_t \cdot C_t + e_t \\ e_t &\sim \mathcal{N}(\mathbf{0}, \tau^2 \cdot I) \quad (\text{independent}). \end{aligned} \tag{1b}$$

The parameters $(\beta_0, \beta_1, \tau^2, H_t)$ are assumed to be different for each type of observation (e.g., tree ring widths, ice

cores), but are often taken to be common for all observations of a given type. Furthermore, the instrumental observations are assumed to be unbiased and on the correct scale, so that, for this type of observation, $\beta_0 = 0$ and $\beta_1 = 1$. The selection matrix H_t is composed of zeros and ones, and selects out at time step t the locations for which there are proxy observations of a given type. That is, each proxy observation is assumed to be linear in the corresponding local, in time and space, value of the climate. Inference on the parameters and the latent climate process, proceeds via Markov chain Monte Carlo (MCMC; e.g., Gelman et al., 2003). Although for a discussion of technical details we refer the reader to Tingley and Huybers (2010a), we note that a core principle of MCMC is to estimate the joint probability distribution of all unknowns by iteratively sampling from each unknown conditional on the current values of all other unknowns. For example, we draw from the distribution of the climate process, conditional on the parameters, and then update the parameters conditional on the climate

One of the main shortcomings of BARCAST, shared by most other climate field reconstruction methods (Schneider, 2001; Smerdon et al., 2010, 2011; Luterbacher et al., 2004; Pauling et al., 2006; Guillot et al., 2014), is the inability to incorporate data with dating uncertainty in a statistically rigorous manner. As hierarchical models such as BARCAST are specified through a series of simple probability statements (Eq. 1a and b), they are naturally modular and amenable to modification. In particular, the description of BARCAST above can be viewed as conditional on the correct age model, and can be generalized to include updating of the probabilities associated with an ensemble of age models, conditional on the climate.

3 Extending BARCAST to include time-uncertain proxies

We focus on augmenting the basic BARCAST framework of Tingley and Huybers (2010a) to permit inclusion of annual-resolution, layer-counted proxies that feature ADM errors. Layers may be miscounted when the annual bands are weak, as can potentially occur in corals (Hendy et al., 2003), or when hiatuses in the record are misdated (e.g., speleothem data; Osete et al., 2012). An ensemble of possible ADM errors is shown in Fig. 1: after a top section that is perfectly dated, counting errors accumulate, leading to a mismatch between actual and estimated age. These timing errors accumulate further down the core, leading to a substantial spread in possible dates for the lowest layers. For the purposes of exposition, and to simplify notation, we consider the case of a single time-uncertain, layer-counted proxy; the formalism can then be repeated for each time-uncertain proxy.

3.1 Addressing miscounted layers

Associated with a single time-uncertain proxy record is an ensemble of possible ADMs, denoted by $\{\mathcal{T}_k, k = 1, \dots, M\}$. The ensemble of ADMs is generated based on understanding of the proxy archive, laboratory standards, and so forth, and should reflect an honest assessment of possible uncertainties. Here, we take the ensemble as given, and do not consider how it was generated. Although there may be some climate archives where the true ADM depends on the climatic state, we ignore this potential complication. We assume the M ADMs are a priori equally likely, and then seek to update the conditional posterior probabilities associated with the ADMs given the current draw of the climate and parameters, and use these probabilities at each iteration of the MCMC to randomly select one of the ADMs. That is, we define a categorical variable that corresponds to the possible age models, and draw from its conditional posterior distribution within the MCMC (Gelman et al., 2003; Tingley and Huybers, 2010a).

We first re-write the data-level model of BARCAST for the time-uncertain proxy. Whereas Eq. (1b) relates the set of observations available throughout space at time t to the concurrent climate field, here it is more convenient to relate \mathbf{O}_s , the time series of the time-uncertain proxy at location s , to \mathbf{C}_s , the co-located time series of the estimated climate process. In addition, we explicitly condition on the currently selected ADM. The distribution of the vector of proxy observations conditional on the climate time series and a particular ADM \mathcal{T} then takes the form

$$\begin{aligned} \mathbf{O}_s | \mathcal{T}, \mathbf{C}_s &= \beta_0 + \beta_1 \cdot \mathbf{\Lambda}_s^T \cdot \mathbf{C}_s + \mathbf{e}_s \\ \mathbf{e}_s &\sim \mathcal{N}(\mathbf{0}, \tau^2 \cdot \mathbf{I}) \quad (\text{independent}). \end{aligned} \quad (2)$$

The vector \mathbf{e}_s is a time series of independent normal errors at location s (cf. \mathbf{e}_t from Eq. 1b). Analogous to \mathbf{H}_t in Eq. (1b), $\mathbf{\Lambda}_s^T$ is a selection matrix of zeros and ones that picks out the elements of the vector \mathbf{C}_s corresponding to elements of \mathbf{O}_s , and is dependent on the ADM \mathcal{T} .

Importantly, $\mathbf{\Lambda}_s^T$ represents the only dependence of either the process or data levels on the ADM. The conditional posterior required to update the probabilities on members of the ensemble $\{\mathcal{T}_k\}$ is therefore the product of the likelihood of the time series of proxy observations at location s conditional on the climate and the selected ADM, and the prior probabilities of the possible ADMs. The likelihood is proportional to a multivariate normal probability density function, with diagonal covariance:

$$\begin{aligned} \mathcal{L}(\mathbf{O}_s | \mathcal{T}, \mathbf{C}_s) &\propto \exp\left(-\frac{1}{2\tau^2} \left[\mathbf{O}_s - \left(\beta_0 + \beta_1 \cdot \mathbf{\Lambda}_s^T \cdot \mathbf{C}_s \right) \right]^T \right. \\ &\quad \left. \times \left[\mathbf{O}_s - \left(\beta_0 + \beta_1 \cdot \mathbf{\Lambda}_s^T \cdot \mathbf{C}_s \right) \right] \right). \end{aligned} \quad (3)$$

Assuming equal prior probabilities for the ADMs, $\pi(\mathcal{T} = \mathcal{T}_k) = 1/M$, the conditional posterior probabilities

corresponding to the candidate ADMs $\{\mathcal{T}_k\}$ are then

$$p(\mathcal{T} = \mathcal{T}_k | \mathbf{C}_s, \mathbf{O}_s) \propto \mathcal{L}(\mathbf{O}_s | \mathcal{T}_k, \mathbf{C}_s) \cdot \pi(\mathcal{T}_k). \quad (4)$$

At this point, the problem is in theory solved. At each iteration of the MCMC, we sample the climate process and scalar parameters as described in Tingley and Huybers (2010a), and then select an ADM according to the conditional posterior probabilities in Eq. (4). This results in a Bayesian age model selection (BARCAST+AMS), whereby at each iteration of the MCMC, an ADM is selected conditional on the data and the current estimate of the climate state. In practice, the sampler is “sticky”, in the sense that it is slow to explore the full range of the joint climate and ADM probability distribution. Intuitively, the selected ADM exerts a strong control on the subsequent estimate of the climate, and the conditional posterior for the ADMs subsequently gives large probability to retaining the current ADM. As a result, the MCMC quickly hones in on the small subset of the ADMs that are most compatible with the initial estimate of the climate. The MCMC then wanders around this local optimum, and does not efficiently explore other local optima in the joint posterior. This well-known problem in MCMC methods, termed the “waiting time dilemma” (Wong and Liang, 1997), occurs when the MCMC sampler cannot escape a local optimum in reasonable time.

3.2 Parallel tempering for ADM selection

To produce more rapid mixing of the BARCAST+AMS – that is, a more rapid exploration of the full probability distribution – we turn to parallel tempering and Metropolis-Coupled MCMC (Altekar et al., 2004; Earl and Deem, 2005; Li et al., 2009). In parallel tempering, several instances of the MCMC sampler are run in parallel, each at a different “temperature”. Intuitively (and in two dimensions), MCMC seeks to explore the topography of a mountain range via a random walk, spending time at each location in proportion to the elevation. “Heating” of a chain promotes the more rapid exploration of the topography by reducing differences in elevation. Formally, a heated chain requires a modification to the likelihood term in the posterior (Altekar et al., 2004):

$$p_\theta(\mathcal{T} = \mathcal{T}_k | \mathbf{C}_s, \mathbf{O}_s) \propto \mathcal{L}(\mathbf{O}_s | \mathcal{T}_k, \mathbf{C}_s)^\theta \cdot \pi(\mathcal{T}_k), \quad (5)$$

where $\theta \in [0, 1]$ is the “inverse temperature” parameter. For $\theta = 1$ (no heating) the normal posterior is recovered. Decreasing θ reduces differences in the likelihood, and in the limit as $\theta \rightarrow 0$, the likelihood is unity for each ADM. In other words, heating increases the influence of the prior, and as we assume a flat prior, chains with smaller values of θ (more heating) will more rapidly explore the probability space.

A crucial drawback of heating a chain is that the stationary distribution of the MCMC is then not correct, in the sense that it does not match the target posterior probability distribution. Metropolis-Coupled MCMC (MC³; Altekar et al.,

2004; Earl and Deem, 2005) offers a solution by coupling several MCMC chains with different levels of heating to an unheated chain. The heated chains permit rapid exploration of the probability space, while the unheated chain benefits from increasing mixing while retaining the correct limiting distribution (Wong and Liang, 1997; Li et al., 2009). The chains are usually organized by increasing temperature, and only neighboring chains are coupled to one another (Earl and Deem, 2005).

Coupling is achieved by permitting state-swapping between the parallel chains. After a predetermined number of iterations of the Gibbs sampler, the states (current values of all parameters, including the climate process and selected ADM) of two neighboring chains are permitted to swap in a Metropolis step (e.g., Gelman et al., 2003). In the current context, we are most interested in encouraging mixing of the ADMs, so we heat the chains solely in the likelihoods of the ADM selection. As a result, only the posterior probabilities of the ADMs for each chain enter into the Metropolis ratio. The probability of swapping the states of two chains j and k , with heatings θ_j and θ_k , conditional on the current estimates of the climate, parameters, and age models is (Altekar et al., 2004)

$$p(j \leftrightarrow k | \mathcal{T}_{j,k}, \mathbf{C}_{j,k}, \mathbf{O}) = \min \left(1, \prod_s \frac{p_{\theta_k}(\mathcal{T}_j | \mathbf{C}_j, \mathbf{O}) \cdot p_{\theta_j}(\mathcal{T}_k | \mathbf{C}_k, \mathbf{O})}{p_{\theta_j}(\mathcal{T}_j | \mathbf{C}_j, \mathbf{O}) \cdot p_{\theta_k}(\mathcal{T}_k | \mathbf{C}_k, \mathbf{O})} \right). \quad (6)$$

The product is over all locations with time-uncertain proxies, and note that $\mathcal{T}_{j,k}, \mathbf{C}_{j,k}, \mathbf{O}$ are all dependent on the spatial location s . Although formally the full state of the MCMC, including all estimates for the climate and model parameters (including the selected ADM) is swapped between the two chains, it is equivalent and more efficient to instead swap their temperatures.

The heating and coupling procedure allows more diverse ADMs to be selected by the heated chains, while the unheated chain retains the correct stationary distribution. The result is a more rapid exploration of the probability space by the unheated chain, as coupling with the heated chains results in more rapid movement between local optima in the probability space. All posterior summaries are then based solely on the unheated chain.

The coupling requires specification of a number of parameters. To set the smallest of the heating parameters in the experiments below, we first ran several chains at a wide range of fixed heatings and without switching to identify θ_{\min} , the largest value of θ that still permits rapid exploration of the possible ADMs. We observe that the transition from almost no diversity in the selected ADM to sampling many of the ADMs is abrupt, and the exact value is not critical provided it is below this threshold. We then run at least six chains (depending on computational resources) with swapping enabled, using θ_{\min} for the hottest chain. Following Kofke (2002), intermediate values are specified to follow a geomet-

ric series. For six chains and a minimum inverse temperature $\theta_{\min} = 0.05$, we set $\theta \approx (1, 0.55, 0.30, 0.17, 0.09, 0.05)$.

A final technical issue concerns the convergence of the unheated chain after a successful swap with a heated chain. The heated chains do not have the same stationary distributions (Li et al., 2009) as the unheated chain, and the MCMC requires several iterations to converge to the correct (unheated) stationary distribution. We therefore discard from the analysis a number of samples of the unheated chain following each successful switch. For the numerical experiments below, we find by monitoring convergence of the error variance for the proxy, τ_p^2 , that discarding the first fifty iterations suffices (cf. Sect. 2.b.3 of Tingley and Huybers, 2010a). Parameter swaps are then proposed every 100 iterations. Since not all proposed swaps are accepted, this leaves enough samples for evaluation even when discarding the first fifty MCMC steps after a switch of the unheated chain. As with other parameters controlling the Metropolis coupling, applications in other settings may require different values.

There is a tradeoff between the allocated resources (number of chains) and the convergence of the unheated chain following a successful swap, as the closer the temperatures of the two swapped states are, the faster they will converge to their respective stationary distributions following a swap. Although implementation using the MPI (Message Passing Interface, www.mpi-forum.org) protocol could in principle be used on a distributed system, we have not tested this, and interprocess communication could induce an extra overhead caused by network latency.

Measuring convergence for Metropolis-Coupled parallel tempering is non-trivial, due to the additional computation burden of the numerous coupled chains. We therefore split the single, long unheated chain into a number of segments, and assess convergence using the potential scale reduction factor (\hat{R} from Gelman et al., 2003) between these segments for each parameter. For the experiments presented below, \hat{R} was usually below 1.05 for all parameters. Further evidence that the chains converge in reasonable time is provided by the fact that process-level parameters do indeed approach the true values specified in the simulations below. Finally, simulation experiments that include the true ADM within the candidate ensemble feature rapid convergence of the MCMC towards the true ADM and ADMs similar to it, regardless of the initial ADM, suggesting a reasonable exploration of the probability space.

4 Simulation experiments

We use two sets of numerical experiments to explore our ability to simultaneously constrain the ADMs of time-uncertain proxies and estimate the climate using BARCAST+AMS. Data are constructed using the BARCAST process and data-level models (Eqs. 1a and 1b); parameter values (Table 1) are constant across all experiments. To prove the concept while

Table 1. Model parameters for the experiments. Proxy 1 is time-certain and Proxy 2 features an uncertain age–depth model. The signal-to-noise ratio is set to about 0.25 (cf. Smerdon, 2012). The measurement noise for the instrumental data is given by τ_I .

Process	Proxy 1		Proxy 2	
α	0.6	$\tau_{P,1}^2$ 0.4	$\tau_{P,2}^2$ 0.4	0.4
μ	0.0	$\beta_{0,1}$ 0	$\beta_{0,2}$ 0	0
σ^2	0.9	$\beta_{1,1}$ 0.5	$\beta_{1,2}$ 0.5	0.5
ϕ^{-1}	1000 km			
τ_I	0.05			

limiting computational burden, we consider a small spatial domain and a limited number of proxy records. Experiments are run over 1300 time steps (years), with high-precision “instrumental” data available over the most recent 150 years. A mix of time-certain and time-uncertain “proxies” are available over the entire time span of the experiments. The time-uncertain proxies consists of annually resolved data, but with a 2 % probability (Comboul et al., 2014; McKay and Kaufmann, 2014) of miscounted layers. An ensemble of 499 age–depth models is created by randomly over- or undercounting each layer from the top of the proxy archive. In contrast to real world reconstructions, it is possible to explicitly include the correct ADM within the ensemble for these experiments.

The experiments are evaluated on the annual timescale, using the average cross correlation and root mean square error (RMSE) between the target and the reconstruction ensemble. In addition, we compare nominal and empirical coverage rates of estimated uncertainty intervals (Li et al., 2010; Tingley and Huybers, 2013) and calculate the continuous ranked probability score (CRPS; Gneiting and Raftery, 2007; Herbach, 2000), as they are more suitable evaluations of ensemble estimates. Positive (negative) differences between nominal and empirical coverage rates are indicative of uncertainty intervals that are too narrow (wide). CRPS, in contrast, is a combined measure of the sharpness (uncertainty) of the reconstruction and the actual coverage rates of the uncertainty intervals. The spatially or temporally averaged CRPS can be broken into two parts: the average reliability score ($\overline{\text{Reli}}$), which estimates how well the nominal coverage rates of the ensemble reconstructions correspond to the empirical ones, and the potential average CRPS ($\overline{\text{CRPS}}_{\text{pot}}$), which characterizes the spread of the ensemble or the width of the uncertainty intervals. All three measures – CRPS, $\text{avgCRPS}_{\text{pot}}$ and avgReli – are in the units of the evaluated variables, and thus enable quick comparison with the reconstructed signal. They thus do not offer a simple threshold to discern between success and failure, in contrast to the coefficient of efficiency and the reduction of error (Cook et al., 1994), where positive values indicate a “skillful” reconstruction.

Although the coefficient of efficiency and the reduction of error are often used to assess reconstructions, they are not

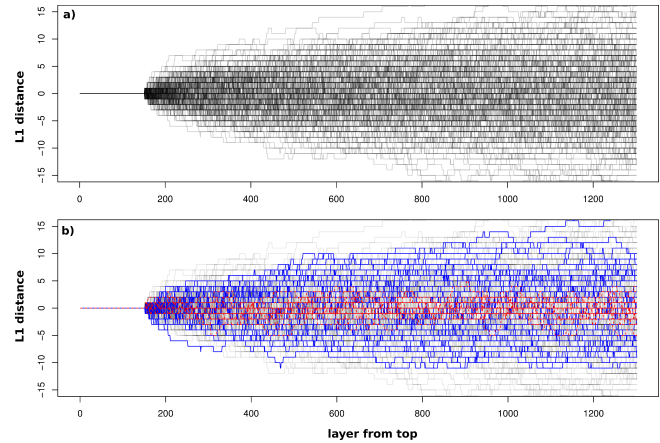


Figure 1. Trace plots of ADMs used in Sect. 4.1. Shown is the mismatch of layer number (experimental date) vs. true date, for each of the ensemble of ADMs. A perfect ADM would be a straight horizontal line at zero. (a) All ADMs. (b) All ADMs (black), ADMs with posterior probability > prior probability (blue), 20 ADMs closest to target (red).

proper scoring rules (Gneiting and Raftery, 2007), and are therefore not suitable for evaluating ensemble predictions – even if they are convenient and widely used.

4.1 One time-certain and one time-uncertain proxy

The first set of experiments features instrumental data over the 150 most recent time points, one time-certain proxy, and one proxy that is time-uncertain save for the overlap with the instrumental records. The set of ADMs associated with the time-uncertain proxy is shown in Fig. 1. We run three sets of experiments, with the two proxies collocated, or separated by roughly three-fourths, or twice the spatial de-correlation length scale of 1000 km. Reconstructions are carried out using both the algorithm outlined above (BARCAST+AMS) and randomly sampling from the prior distributions of the ADMs within the original BARCAST algorithm.

To evaluate how the heating (Sect. 3) influences the posterior distribution of the ADMs, we calculate the distribution of L1 distances between the true ADM and each of the 500 ADMs that form the ensemble. In the left panel of Fig. 3 we show the distribution of L1 distances for all ADMs (“weighted” by the flat prior), and in the right panel the L1 distances of the drawn ADMs, weighted by the posterior distribution of the ADMs. The cooler chains each converge to a single (but different) ADM that is close to optimal, but fail to adequately explore the probability space. The two warmest chains, in contrast, switch almost constantly, leading to a flatter posterior distribution of selected ADMs across L1 distance. Note that the L1 distance to the target ADM is not a quantity that the BARCAST+AMS algorithm evaluates directly – instead BARCAST+AMS selects an ADM by evaluating the likelihood of the observations under each possible

ADM and the current estimate of the climate, leveraging the signal that is common to the two proxy series.

Coupling the chains permits a more reasonable exploration of the ensemble of ADMs (Fig. 4). The unheated chain explores ADMs with progressively larger L1 deviations from the true ADM as the separation between the proxies increases. When the two proxies are co-located, the unheated chain puts most of the probability mass on a small number of ADMs, but successfully wanders between them. The posterior distribution of L1 distances is narrower and centered on lower values as compared with the prior, indicating that the estimated climate signal provides a strong constraint on the distribution of possible ADMs consistent with the data. As the separation between the proxies increases, there is less information to constrain the ADMs for the time-uncertain proxy, and the posterior distribution of L1 distances for the unheated chain becomes broader. These features are similar whether or not the correct ADM is included in the ensemble of possible ADMs for the time-uncertain proxy.

Results of the three experiments are summarized in Table 2, using both BARCAST+AMS (implemented using MC³) and randomly selecting from the ADMs within BARCAST. As expected, all measures for both analyses indicate a better reconstruction at the location of the time-certain proxy, where results are comparable between the two analysis choices.

The value of updating the probabilities associated with the ADMs within BARCAST+AMS can be seen when comparing performance metrics at the location of the time-uncertain proxy (Table 2). When the two proxies are separated by about three-fourths of the decorrelation length scale, the more targeted ADM selection afforded by the Metropolis-Coupled MCMC results in a larger cross correlation and reduced RMSE as compared with random ADM selection, while CRPS and its two components are each smaller. Metropolis-Coupled MCMC therefore results in narrower ensemble reconstructions with better coverage probabilities. The empirical coverage rate of 90 % uncertainty intervals is in all cases marginally high at the location of the time-uncertain proxy, and the width of 90 % uncertainty intervals at the time-uncertain proxy location is narrower when estimated using BARCAST+AMS as compared with random sampling of the ADMs.

A visual representation of the ability of BARCAST+AMS to learn about the ADMs is provided by trace plots of the prior ensemble of ADMs, color-coded to indicate those ADMs that feature larger posterior probability than prior probability. When the two proxies are separated by about three-fourths the decorrelation length scale, only a small subset of the ADMs feature increased posterior probability, and their spread is narrow as compared with the original ensemble (Fig. 1). These results clearly demonstrate the improvements afforded by formally updating the probabilities associated with the ADMs. We note that the ability to learn about the posterior distribution of the ADMs is a strong function

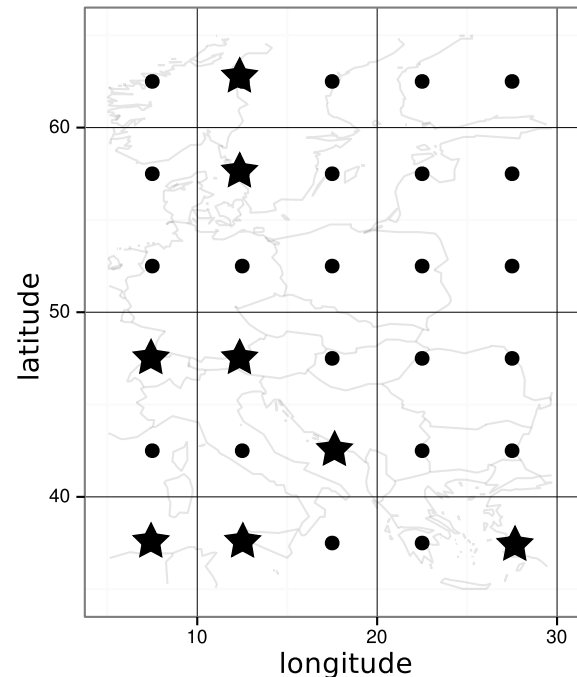


Figure 2. Spatial domain of the reconstruction experiments, with a superimposed map of Europe to provide a scale. Experiments are purely simulated, and are not related to the climate of this region. Stars mark locations of proxies used in Sect. 4.2.

of the amount of nearby (as measured by the spatial decorrelation length scale) information that is available to constrain the climate at the location of the time-uncertain proxy. Indeed, when the time-certain proxy is more than twice the spatial correlation length from the time-uncertain proxy there is little gain over random ADM selections (Table 2). This illustrates the need for a shared signal between the proxies that can be used to correct for the misdating in either. The exact source of the shared signal is not important, be it through spatial closeness as constructed here, or through a long-range teleconnection.

4.2 No instrumental data and all proxies time-uncertain

The second experiment features no instrumental data, and all eight proxy records are time-uncertain (Fig. 2). Furthermore, the data-level parameters of Eq. (1b), τ_p^2 , β_0 , and β_1 , are treated as known and fixed at their true values. The goal of the experiment is to demonstrate the feasibility of simultaneously estimating, from strictly time-uncertain information, the process-level parameters that govern the space–time covariance structure of the climate field (α , μ , ϕ , σ^2), the climate field itself, and the probabilities associated with the ADMs. That is, we seek to estimate the climate field and its covariance properties from time-uncertain information, while simultaneously exploiting the space–time covariance

Table 2. Experimental results with one time-certain and one time-uncertain proxy record, with instrumental data present, for both BARCAST+AMS and random ADM selection (rand). Results are shown at the location of the time-certain proxy (“Certain”) and the location of the time-uncertain proxy (“ADM”). The spatial decorrelation length scale is 2000 km, and the separation between the proxies varies between experiments. The reconstructions are evaluated over the pre-instrumental period only. The standard deviation of the local target climate signal is about 1.3.

	Dating	Separation	Crosscorr	RMSE	$\overline{\text{CRPS}}$	$\overline{\text{CRPS}}_{\text{pot}}$	Reli	90 % cov	90 % width
BARCAST+AMS	Both	0	0.64	1.04	0.39	0.39	6.17 E-3	0.93	2.55
	ADM	751 km	0.45	1.29	0.49	0.49	3.05 E-3	0.92	3.05
	Certain		0.61	1.06	0.43	0.42	4.03 E-3	0.89	2.44
	ADM	2084 km	0.23	1.46	0.56	0.56	3.03 E-3	0.92	3.47
	Certain		0.57	1.11	0.43	0.42	6.18 E-3	0.91	2.62
	rand	Both	0	0.63	1.02	0.38	0.38	6.23 E-3	0.92
rand	ADM	751 km	0.33	1.42	0.54	0.54	3.28 E-3	0.93	3.44
	Certain		0.61	1.05	0.43	0.42	4.65 E-3	0.89	2.42
	ADM	2084 km	0.24	1.53	0.58	0.57	4.07 E-3	0.94	3.75
	Certain		0.57	1.11	0.43	0.43	3.92 E-3	0.91	2.64

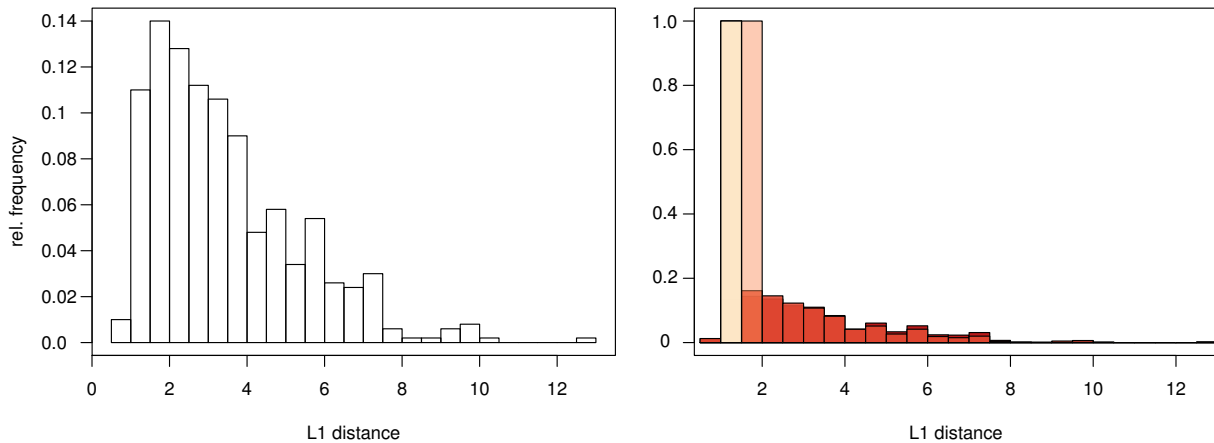


Figure 3. Left panel: distribution of all available ADMs by L1 distance. Right panel: distribution of drawn ADMs with heated chains without state swapping. The beige (dark red) corresponds to the coolest (hottest) chain. Results are for the two-proxy experiment (Sect. 4.1) with intermediate spatial separation.

of the climate to constrain the ADMs. Although idealized, this experimental setup is useful in assessing the plausibility of reconstructing annually resolved climate from time-uncertain proxies.

Both BARCAST+AMS and random sampling of ADMs provide adequate estimates of the mean, μ , and the AR(1) persistence parameter, α (Fig. 5). Although both methods overestimate σ^2 , the partial sill (Banerjee et al., 2004) of the spatially covarying innovations, and underestimate the inverse-range parameter ϕ , the distributions resulting from BARCAST+AMS are marginally closer to the true values (Fig. 5). As discussed in Zhang (2004), Tingley and Huybers (2010a), and Mannshardt et al. (2013), separate inference of these two parameters is notoriously difficult, and their posterior draws are tightly correlated (see Fig. 9 of Tingley and

Huybers, 2010a). In the current setting, with only eight noisy, time-uncertain proxies, the noise is such that separate, tight estimates on the range and partial sill are not possible. We note, however, that predictive performance in spatial settings tends to be only weakly impacted when using an inverse-range parameter that is smaller than optimal (Kaufman and Shaby, 2013).

Even in the absence of time-certain observations, BARCAST+AMS allows the spatial covariance of the climate system to update the probabilities of the possible ADMs. Indeed, the posterior distribution of L1 distances to the true ADM is narrower and peaked at lower values as compared with the prior distribution (Fig. 6), indicating that the estimated climate in fact places a strong constraint on the ensemble of possible ADMs.

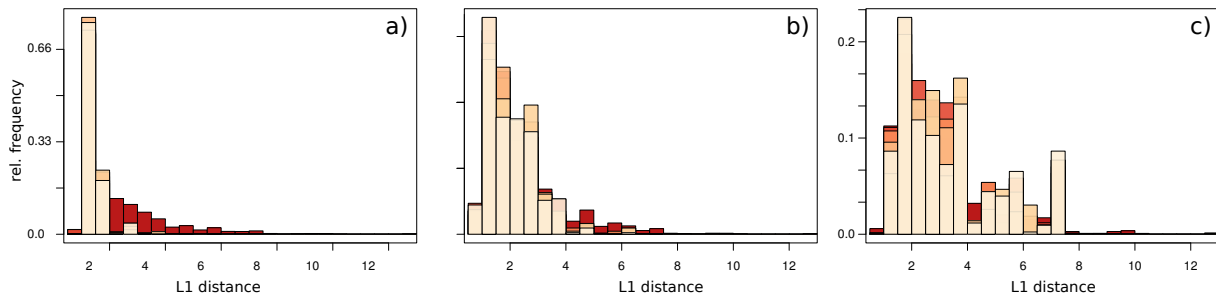


Figure 4. Draws of ADMs (by mismatch) for one perfectly dated and one time-uncertain proxy. Colors from beige (unheated) chain to red (warmest) chain. (a–c) The separation between the two proxies is 0, 750, 2084 km. Compare panel (b) to Fig. 3.

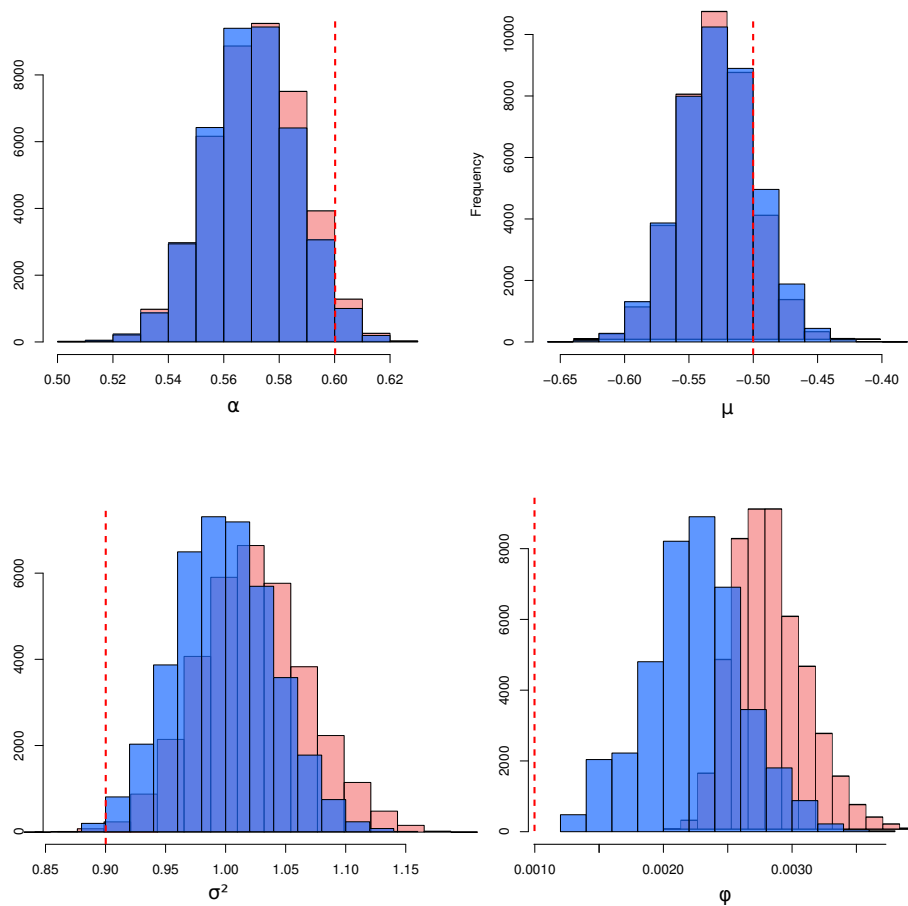


Figure 5. Histograms of the draws of the process-level parameters (α , μ , σ^2 , ϕ) for the experiment with only time-uncertain proxies. Blue histograms: results using BARCAST+AMS. Red histograms: results using random age model selection. Vertical red lines denote the true parameter values from Table 1.

Evaluations of the reconstructions carried out using BARCAST+AMS and random ADM selection within BARCAST provide further evidence of the benefits of updating the probabilities for the ADMs (Fig. 7). All six panels show the results for BARCAST+AMS (left, blue) and random ADM selection (red, right) for all grid cells. Results at the proxy locations are highlighted by crosses, and the results from the rest of the grid are summarized using a box plot. For BAR-

CAST+AMS, the cross correlation and RMSE at the proxy locations are close to the values expected for time-certain data, given the signal-to-noise ratio of the proxies, and are superior to values under random ADM selection. That is, BARCAST+AMS successfully counteracts some of the increased uncertainty that results from dating errors at the proxy locations (crosses in Fig. 7). At locations where the reconstruction requires spatial interpolation, the distinction between the

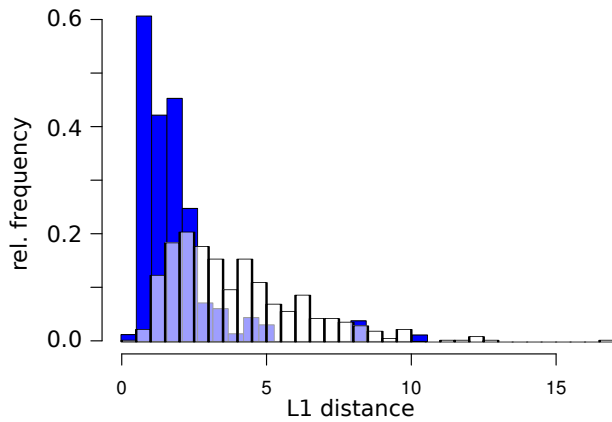


Figure 6. Histogram of prior (white) and posterior (blue) distribution of L1 distances to the true ADMs, for the experiment with only time-uncertain proxies. The histograms bin distances over all eight time-uncertain proxies.

two analyses in terms of correlations and RMSE is less pronounced. At these locations, a larger fraction of the total uncertainty arises from the spatial mapping, as opposed to dating uncertainty, so the benefits afforded by constraining the ADMs via BARCAST+AMS are less pronounced.

At all locations, both BARCAST+AMS and random ADM selection result in reasonable empirical coverage rates for nominally 90 % uncertainty intervals, and the reliability scores (Reli) are comparable between the two analyses. That is, both methods result in uncertainty intervals with reasonable coverage probabilities. However, BARCAST+AMS results in lower $\overline{\text{CRPS}}_{\text{avg}}$ and $\overline{\text{CRPS}}_{\text{pot}}$, with the distinction more pronounced at the proxy locations, indicating that the ensemble inferences from BARCAST+AMS are more sharply peaked than those from random ADM selection, despite featuring similar coverage probabilities.

5 Discussions and extensions

We have described and implemented an extension of a Bayesian hierarchical model for climate field reconstructions that accounts for uncertainties due to misspecified age–depth models in annually resolved proxy records. Although we have focused on a particular type of time-uncertainty (miscounting of annual layers) and the BARCAST reconstruction algorithm of Tingley and Huybers (2010a), the methodology we outline is broadly applicable. To achieve adequate mixing in the MCMC, we make use of Metropolis-Coupled MCMC with parallel tempering (Altekar et al., 2004; Earl and Deem, 2005). These techniques increase computational demands, by a factor given roughly by the number of coupled chains, but are necessary to ensure adequate exploration of the probability space.

As demonstrated with simulation experiments, our method places higher posterior probability on ADMs with low L1

distance to the correct ADM (Figs. 1 and 4). Moreover, reconstructions that update the probabilities associated with the ADMs based on the current draw of the climate feature better score and skill metrics than reconstructions that feature random ADM selection (Fig. 7, Table 2).

A number of useful extensions to the general framework we have proposed are possible. Many proxy archives, such as marine sediment cores, do not form annual layers. Proxies derived from these archives, such as $\delta^{18}\text{O}$, TEX_{86} , $\text{U}_{37}^{K'}$, and Mg/Ca , are generally measured as an average over a depth increment of sediment. The time boundaries of each increment are, in turn, determined by an ADM that is generally constrained using radiometric dating. As Tingley and Huybers (2010a) discuss how to extend BARCAST to account for proxies that average the climate process in time and/or space, we here focus on how to adapt the updating of the age models in this case. In the context of annually resolved, layer-counted proxies, the matrix $\mathbf{\Lambda}_s^T$ in Eq. (3) features a single one in each row that picks out the climate value corresponding to each proxy observation, and is otherwise composed of zeros. If each proxy observation instead represents an average over some number of time points, then each row of $\mathbf{\Lambda}_s^T$ features a segment composed of the corresponding averaging weights. Both the weights themselves, and their positions within the rows of $\mathbf{\Lambda}_s^T$, are determined by the ADM. Provided that the averaging inherent to each proxy observation can be specified from the selected ADM, rows of the selection matrices $\mathbf{\Lambda}_s^T$ can be altered on the basis of the currently selected ADM, and the analysis then proceeds as above.

The procedure can likewise be adapted if the natural time increments of the problem are not annual, e.g., for marine sediment data that often feature multi-decadal to centennial resolution. The underlying process-level model (for the evolution of the climate variables) would then need to be adapted to the new question of interest, and we note that the stationarity assumption of the currently specified process level will not hold, for example, when considering glacial–interglacial timescales.

Our presentation has assumed the a priori existence of an ensemble of possible ADMs for each time-uncertain proxy. In many cases, such an ensemble of ADMs is not available with the proxy record – though recent efforts to define standards for proxy metadata suggest that the original dating information, such as radiometric ages and uncertainties, be included along with the proxy observations (see the selection criteria outlined in PAGES2k Consortium, 2014). Recent work in radiocarbon dating has focused on stochastic modeling of the sedimentation process, and development of Bayesian models to constrain possible depositional histories, and therefore ADMs, conditional on a set of imperfect age control points (e.g., Ramsey, 2008; Blockley et al., 2007; Blaauw and Christen, 2011). Given dating information, any of the existing ADM construction algorithms, such as BACON (Blaauw and Christen, 2011), could then be used to produce an ADM ensemble.

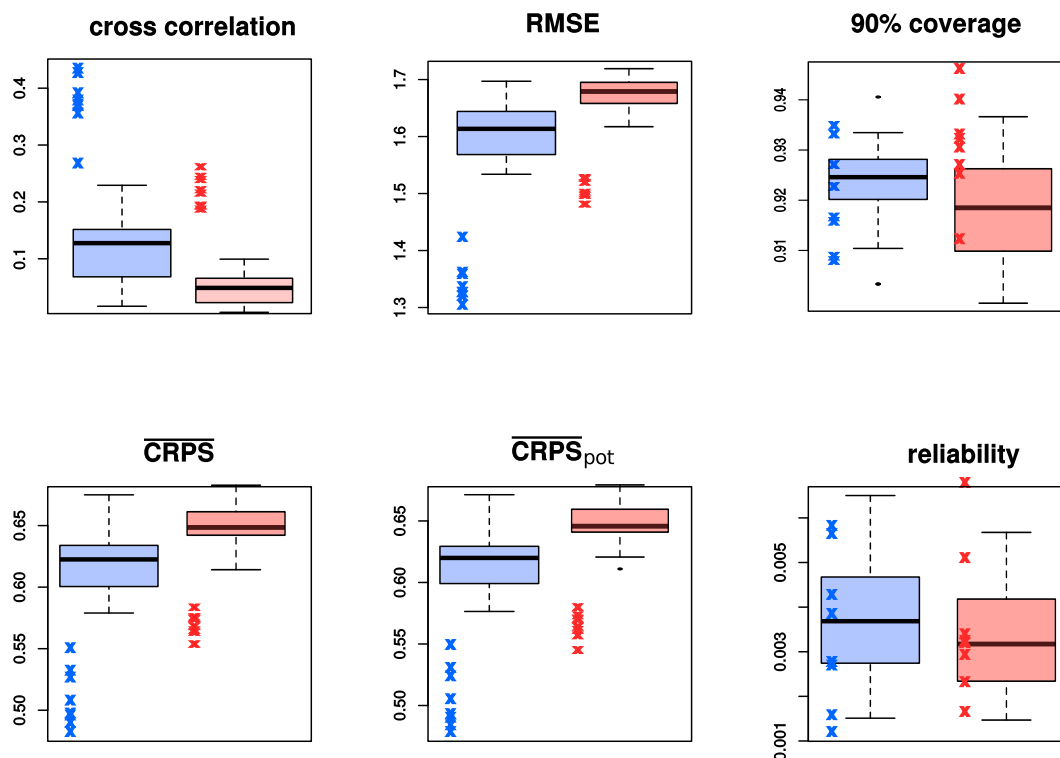


Figure 7. Experimental results with eight time-uncertain proxies and no instrumental data. Blue: results using BARCAST+AMS. Red: results using random age model selection. Results for the proxy locations are shown as crosses, while results at other locations, where there are no observations, are displayed with box plots.

Alternatively, the actual algorithm that generates ADMs from the age constraints could be embedded within the climate reconstruction algorithm. As opposed to updating the probabilities associated with each member of the ADM ensemble at each iteration of the MCMC, the analysis would instead proceed by generating a new ADM at each iteration of each coupled chain, and then accepting or rejecting this proposal based on a Metropolis (Gelman et al., 2003) step. One advantage of embedding the ADM generating mechanism within the reconstruction is an increase in the number of candidate ADMs included in the model. As a new ADM is generated at each iteration of the MCMC for each of the parallel chains, a larger number of possible age models are included in the analysis.

There are many potential benefits to including time-uncertain proxy records within annually resolved, late-Holocene climate field reconstructions. Inclusion of time-uncertain observations would increase the number of proxy records and the diversity of proxy types available to such analyses, and increase the spatial coverage of the proxy network. Furthermore, lower-resolution, time-uncertain records may permit improved inference on low-frequency climate variability (e.g., Jones et al., 2009; Kaufman, 2014), and have the potential to extend the time span of reconstructions.

Climate constructions are only as reliable as the appropriateness of the assumptions they rely upon. Advances in climate reconstructions and uncertainty quantification may be achievable via the development of simple, flexible, and scientifically reasonable forward models for commonly used proxies (Evans et al., 2013). Efforts in that regard so far have generally focused on annually resolved time-certain proxies (Tolwinski-Ward et al., 2010, 2014). Our results point to the benefits of extending this line of inquiry to time-uncertain proxies as we have demonstrated that they can be included in multi-proxy reconstructions in a statistically rigorous manner that fully propagates uncertainties.

The Supplement related to this article is available online at [doi:10.5194/cp-11-533-2015-supplement](https://doi.org/10.5194/cp-11-533-2015-supplement).

Acknowledgements. Discussions at the PAGES Advances in Climate Field Reconstructions workshop, held at Woods Hole Oceanographic Institute, 15–16 April 2014 provided the impetus for this work. J. P. Werner is supported by the Centre for Climate Dynamics at the Bjerknes Centre. M. P. Tingley is supported in part by NSF P2C2 grant 1304309.

Edited by: V. Masson-Delmotte

References

- Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F.: Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference, *Bioinformatics*, 20, 407–415, 2004.
- Anchukaitis, K. J. and Tierney, J. E.: Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, *Clim. Dynam.*, 41, 1291–1306, 2013.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E.: *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, New York, 2004.
- Berliner, L., Wikle, C., and Cressie, N.: Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling, *J. Climate*, 13, 3953–3968, 2000.
- Blaauw, M. and Christen, J. A.: Flexible paleoclimate age-depth models using an autoregressive gamma process, *Bayesian Analysis*, 6, 457–474, 2011.
- Blockley, S., Blaauw, M., Ramsey, C. B., and van der Plicht, J.: Building and testing age models for radiocarbon dates in lateglacial and early Holocene sediments, *Quaternary Sci. Rev.*, 1915–1926, 2007.
- Brüggemann, W.: A minimal cost function method for optimizing the age-depth relation of deep-sea sediment cores, *Paleoceanography*, 7, 467–487, 1992.
- Comboul, M., Emile-Geay, J., Evans, M. N., Mirnategui, N., Cobb, K. M., and Thompson, D. M.: A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually banded coral archives, *Clim. Past*, 10, 825–841, doi:10.5194/cp-10-825-2014, 2014.
- Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology – a review and comparison of 2 techniques, *Int. J. Climatol.*, 14, 379–402, 1994.
- Curry, W. and Oppo, D.: Glacial water mass geometry and the distribution of $\delta^{13}\text{C}$ of ΣCO_2 in the western Atlantic Ocean, *Paleoceanography*, PA1017, doi:10.1029/2004PA001021, 2005.
- Earl, D. J. and Deem, M. W.: Parallel tempering: theory, applications, and new perspectives, *Phys. Chem. Phys.*, 7, 3910–3916, doi:10.1039/B509983H, 2005.
- Evans, M. N., Tolwinski-Ward, S., Thompson, D., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Sci. Rev.*, 76, 16–28, 2013.
- Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S.: Spatial modeling with spatially varying coefficient processes, *J. Am. Stat. Assoc.*, 98, 387–396, 2003.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D.: *Bayesian Data Analysis*, 2nd edn., Chapman & Hall, 2003.
- Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, 102, 359–378, 2007.
- Guillot, D., Rajaratnam, B., and Emile-Geay, J.: Statistical paleoclimate reconstructions via Markov random fields, *Annals of Applied Statistics*, in review, 2014.
- Hafliðason, H., Eiriksson, J., and Kreveld, S. V.: The tephrochronology of Iceland and the North Atlantic region during the Middle and Late Quaternary: a review, *J. Quaternary Sci.*, 15, 3–22, 2000.
- Hendy, E. J., Gagan, M. K., and Lough, J. M.: Chronological control of coral records using luminescent lines and evidence for non-stationary ENSO teleconnections in northeast Australia, *Holocene*, 13, 187–199, 2003.
- Herbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, 15, 559–570, 2000.
- Hughes, A. L. C., Clark, C. D., and Jordan, C. J.: Flow-pattern evolution of the last British Ice Sheet, *Quaternary Science Reviews*, 89, 148–168, doi:10.1016/j.quascirev.2014.02.002, 2014.
- Huybers, P. and Wunsch, C.: A depth-derived Pleistocene age model: uncertainty estimates, sedimentation variability, and nonlinear climate change, *Paleoceanography*, 19, PA1028, doi:10.1029/2002PA000857, 2004.
- Imbrie, J., Hays, J., Martinson, D., McIntyre, A., Mix, A., Morley, J., Pisias, N., Prell, W., and Shackleton, N.: The orbital theory of Pleistocene climate: support from a revised chronology of the marine $\delta^{18}\text{O}$ record, in: *Milankovitch and Climate: Understanding the Response to Astronomical Forcing*, vol. 1, Springer, New York, 269 pp., 1984.
- Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., van Ommen, T. D., Vinther, B. M., Luterbacher, J., Wahl, E. R., Zwiars, F. W., Mann, M. E., Schmidt, G. A., Ammann, C. M., Buckley, B. M., Cobb, K. M., Esper, J., Goosse, H., Graham, N., Jansen, E., Kiefer, T., Kull, C., Küttel, M., Mosley-Thompson, E., Overpeck, J. T., Riedwyl, N., Schulz, M., Tudhope, A. W., Villalba, R., Wanner, H., Wolff, E., and Xoplaki, E.: High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects, *Holocene*, 19, 3–49, 2009.
- Kaufman, C. and Shaby, B.: The role of the range parameter for estimation and prediction in geostatistics, *Biometrika*, 100, 473–484, 2013.
- Kaufman, D.: A community-driven framework for climate reconstructions, *Eos, Transactions American Geophysical Union*, 7 October 2014, 95, 361–362, 2014.
- Kofke, D. A.: On the acceptance probability of replica-exchange Monte Carlo trials, *J. Chem. Phys.*, 117, 6911–6914, 2002.
- Li, B., Nychka, D., and Ammann, C.: The value of multiproxy reconstruction of past climate, *J. Am. Stat. Assoc.*, 105, 883–895, 2010.
- Li, Z., Protopopescu, V. A., Arnold, N., Zhang, X., and Gorin, A.: Hybrid parallel tempering and simulated annealing method, *Appl. Math. Comput.*, 212, 216–228, 2009.
- Lisiecki, L. and Raymo, M.: A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records, *Paleoceanography*, 20, PA1071, doi:10.1029/2004PA001071, 2005.
- Lisiecki, L. and Raymo, M.: Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics, *Quaternary Sci. Rev.*, 26, 56–69, 2007.
- Lisiecki, L. E. and Lisiecki, P. A.: Application of dynamic programming to the correlation of paleoclimate records, *Paleoceanography*, 17, 1049, doi:10.1029/2001PA000733, 2002.
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, 303, 1499–1503, 2004.
- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *P. Natl. Acad. Sci. USA*, 105, 13252–13257, 2008.

- Mannshardt, E., Craigmile, P., and Tingley, M. P.: Statistical modeling of extreme value behavior in North American tree-ring density series, *Clim. Change*, 117, 843–858, 2013.
- Martinson, D., Menke, W., and Stoffa, P.: An inverse approach to signal correlation, *J. Geophys. Res.*, 87, 4807–4818, 1982.
- McKay, N. P. and Kaufmann, D. S.: An extended Arctic proxy temperature database for the past 2000 years, *Scientific Data*, 1, 140026, doi:10.1038/sdata.2014.26, 2014.
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlen, W.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, 433, 613–617, 2005.
- NRC: Surface Temperature Reconstructions for the Last 2000 Years, The National Academies Press, Washington DC, 2006.
- Osete, M.-L., Martín-Chivelet, J., Rossi, C., Edwards, R. L., Egli, R., Muñoz-García, M. B., Wang, X., Pavón-Carrasco, F. J., and Heller, F.: The Blake geomagnetic excursion recorded in a radiometrically dated speleothem, *Earth Planet. Sc. Lett.*, 353, 173–181, 2012.
- PAGES2k Consortium: Continental-scale temperature variability over the common era, *Nat. Geosci.*, 6, 339–346, 2013.
- PAGES2k Consortium: PAGES 2k Proxy Database, Tech. rep., IGBP Pages 2k, available at: <http://www.pages-igbp.org/>, 2014.
- Pauling, A., Luterbacher, J., Casty, C., and Wanner, H.: 500 years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation, *Clim. Dynam.*, 26, 387–405, 2006.
- Ramsey, C.: Deposition models for chronological records, *Quaternary Sci. Rev.*, 27, 42–60, 2008.
- Schneider, T.: Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values, *J. Climate*, 14, 853–871, 2001.
- Shackleton, N., Berger, A., and Peltier, W.: An alternative astronomical calibration of the lower Pleistocene timescale based on ODP Site 677, *Trans. R. Soc. Edinburgh Earth Sci.*, 81, 251–261, 1990.
- Shaw, A.: *Time in Stratigraphy*, McGraw-Hill, New York, 1964.
- Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, *Wiley Interdisciplinary Reviews (WIREs) Clim. Change*, 3, 63–77, doi:10.1002/wcc.149, 2012.
- Smerdon, J. E., Kaplan, A., Chang, D., and Evans, M. N.: A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium, *J. Climate*, 23, 4856–4880, 2010.
- Smerdon, J. E., Kaplan, A., Zorita, E., González-Rouco, J. F., and Evans, M.: Spatial performance of four climate field reconstruction methods targeting the Common Era, *Geophys. Res. Lett.*, 38, L11705, doi:10.1029/2011GL047372, 2011.
- Tierney, J. E., Smerdon, J. E., Anchukaitis, K. J., and Seager, R.: Multidecadal variability in East African hydroclimate controlled by the Indian Ocean, *Nature*, 493, 389–392, 2013.
- Tingley, M. P.: A bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record, *J. Climate*, 25, 777–791, 2012.
- Tingley, M. P. and Huybers, P.: 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, 2010a.
- Tingley, M. P. and Huybers, P.: 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, 2010b.
- Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, *Nature*, 496, 201–205, 2013.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt-Shamseldin, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Sci. Rev.*, 35, 1–22, 2012.
- Tolwinski-Ward, S., Tingley, M., Evans, M., Hughes, M., and Nychka, D.: Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response, *Clim. Dynam.*, 44, 1–16, 2014.
- Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, *Clim. Dynam.*, 36, 2419–2439, 2010.
- Werner, J. P., Smerdon, J., and Luterbacher, J.: A pseudoproxy evaluation of bayesian hierarchical modelling and canonical correlation analysis for climate field reconstructions over Europe, *J. Climate*, 851–867, 2013.
- Werner, J. P., Toreti, A., and Luterbacher, J.: Stochastic Models for Climate Reconstructions – How wrong is too wrong?, *NOLTA Proc.*, IEICE, Tokyo, 2014.
- Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M.: Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds, *J. Am. Stat. Assoc.*, 96, 382–397, 2001.
- Wong, W. H. and Liang, F.: Dynamic weighting in Monte Carlo and optimization, *P. Natl. Acad. Sci. USA*, 94, 14220–14224, 1997.
- Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, *J. Am. Stat. Assoc.*, 99, 250–261, 2004.