Climate
of the Past

Multi-timescale data assimilation for atmosphere–ocean state estimates

Nathan Steiger[1] and Gregory Hakim[2]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA
[2]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA

Correspondence to: Nathan Steiger (nsteiger@ldeo.columbia.edu)

Abstract. Paleoclimate proxy data span seasonal to millennial timescales, and Earth's climate system has both high- and low-frequency components. Yet it is currently unclear how best to incorporate multiple timescales of proxy data into a single reconstruction framework and to also capture both high- and low-frequency components of reconstructed variables. Here we present a data assimilation approach that can explicitly incorporate proxy data at arbitrary timescales. The principal advantage of using such an approach is that it allows much more proxy data to inform a climate reconstruction, though there can be additional benefits. Through a series of offline data-assimilation-based pseudoproxy experiments, we find that atmosphere–ocean states are most skillfully reconstructed by incorporating proxies across multiple timescales compared to using proxies at short (annual) or long (∼ decadal) timescales alone. Additionally, reconstructions that incorporate long-timescale pseudoproxies improve the low-frequency components of the reconstructions relative to using only high-resolution pseudoproxies. We argue that this is because time averaging high-resolution observations improves their covariance relationship with the slowly varying components of the coupled-climate system, which the data assimilation algorithm can exploit. These results are consistent across the climate models considered, despite the model variables having very different spectral characteristics. Our results also suggest that it may be possible to reconstruct features of the oceanic meridional overturning circulation based on atmospheric surface temperature proxies, though here we find such reconstructions lack spectral power over a broad range of frequencies.

1 Introduction

Paleoclimate proxies sample widely different timescales. High-resolution paleoclimate proxies such as tree rings or corals have annual or seasonal resolution, whereas lower resolution proxies such as sediment cores can provide anywhere from annual- to millennial-scale information depending on the core and its location (Bradley, 2014). Additionally, high-resolution proxies tend to be short, and are mostly limited to the past two millennia, whereas some low-resolution proxies can reach the Cenozoic (e.g., Zachos et al., 2008). In addition to the many timescales of proxies, the climate system itself varies across a large range of timescales: from atmospheric blocking to ocean overturning circulation to ice-age cycles. Thus, any faithful reconstruction of past climate must account for as many of these timescales, captured by both proxies and climate models, as possible.

Few paleoclimate reconstruction methods have been created that specifically incorporate multiple proxy timescales. Most reconstructions use either low- or high-resolution proxies alone. If multiple scales of proxy data are used together, researchers often resort to coarsening high-resolution proxies (e.g., PAGES 2k Consortium, 2013) or linearly interpolating low-resolution proxies to a "higher resolution" (e.g., Mann et al., 2008). One major reason for this is that many traditional multivariate regression methods are not constructed to easily calibrate in both low- and high-frequency domains; Mann et al. (2005), for example, present a such a modified method and discuss these and related challenges. Additionally, including multiple timescales is not entirely a methodological problem but partly a temporal sampling issue: given that instrumental temperature data only span the past 150 years or so, low-frequency reconstruction tech-

niques have few degrees of freedom on which to be calibrated and validated if the timescale is longer than about a decade. Specific methods that have been developed with multiple timescales in mind include the time-series reconstruction methods of Li et al. (2010) and Hanhijärvi et al. (2013). Li et al. (2010) use a Bayesian hierarchical model approach, while Hanhijärvi et al. (2013) use an approach based on pairwise comparisons that is particularly flexible and was used extensively, for example, in a recent high-profile paper that reconstructed continental-scale temperatures over the common era (PAGES 2k Consortium, 2013). Space–time reconstruction methods include Guiot and Corona (2010) and Carro-Calvo et al. (2013). Guiot and Corona (2010) developed a spectral analog approach, where analogs are drawn from an instrumental temperature data product, that they used to reconstruct European April–September temperatures over the past 1400 years. Carro-Calvo et al. (2013) developed a generalized neural network approach and used it to reconstruct winter precipitation in the Mediterranean over the past 300 years.

Data assimilation (DA) provides a flexible framework for combining information from paleoclimate proxies with the dynamical constraints of climate models. In principle, DA can provide reconstructions of any model variable, from surface temperature to sea water salinity to atmospheric geopotential height. Among DA techniques, we are unaware of any method specifically designed for the challenges of paleoclimate reconstructions that incorporates proxies across an arbitrary range of timescales. DA-based reconstructions have so far used only a single uniform timescale (e.g., Goosse et al., 2012) or have performed separate reconstructions at different uniform timescales (Mathiot et al., 2013). Traditional DA adjoint methods as applied in weather forecasting can and do include multiple timescales (Kalnay, 2003, p. 181–184), but these timescales are very short by comparison to the timescales involved in paleoclimate. Here we develop a DA-based algorithm for space–time climate reconstructions that can assimilate proxies at any time resolution. Because of the limited time span of observational data sets, we explore the features and skill of this technique within a synthetic, pseudoproxy framework. The ensuing pseudoproxy experiments use an "offline" DA implementation, wherein prior ensembles are drawn from a previously run climate model simulation and are not integrated forward in time after each DA update step. This allows us to test the algorithm over long time spans, perform carefully controlled experiments, and unambiguously define errors.

Multiproxy reconstructions can potentially overcome some limitations of single proxy reconstructions, such as filling in for the missing frequency components of a particular proxy (Li et al., 2010). However, the primary, pragmatic benefit to incorporating proxies at multiple temporal resolutions is that more information can inform the reconstruction. By comparison with current weather observations, paleoclimate proxies are more expensive and time-consuming to gather

and their spatial distribution is far less extensive. Therefore, including any additional unbiased information should meaningfully improve the reconstructions. In addition, it is possible that particular reconstruction methods could benefit from multi-scale proxy data. Within a coupled atmosphere–ocean DA framework, Tardif et al. (2014) suggest that assimilating time-averaged observations of atmospheric variables may improve present-day estimates of ocean circulation. They argue that these improvements arise from the fact that time averaging high-frequency observations improves the signal over noise in the covariance relationship between the atmosphere and the slowly varying ocean overturning circulation. We test this hypothesis within a paleoclimate context and assesses whether or not atmosphere–ocean state estimates can be improved by including proxies and climate states at multiple timescales. Therefore this test goes beyond the benefit of simply being able to include more proxies in climate reconstructions.

2 Assimilation technique

Data assimilation refers to a mathematical technique of optimally combining observations (or within this context, proxy data) with prior information, typically from a model. The model, in this case a climate model, provides an initial, or prior, state estimate that one can update in a Bayesian sense based on the observations and an estimate of the errors in both the observations and the prior. The prior contains any climate model variables of interest and the updated prior, called the posterior, is the best estimate of the climate state given the observations and the error estimates. The basic state update equations of DA (e.g., Kalnay, 2003) are given by

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K}[\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_b)], \qquad (1)$$

where $\mathbf{K}$ can be written as

$$\mathbf{K} = \mathrm{cov}(\boldsymbol{x}_b, \mathcal{H}(\boldsymbol{x}_b))[\mathrm{cov}(\mathcal{H}(\boldsymbol{x}_b), \mathcal{H}(\boldsymbol{x}_b)) + \mathbf{R}]^{-1}, \qquad (2)$$

and cov represents a covariance expectation. $\boldsymbol{x}_b$ is the prior (or "background") estimate of the state vector and $\boldsymbol{x}_a$ is the posterior (or "analysis") state vector. Observations (or proxies) are contained in vector $\boldsymbol{y}$. The observations are estimated by the prior through $\mathcal{H}(\boldsymbol{x}_b)$, which is, in general, a nonlinear vector-valued observation operator that maps $\boldsymbol{x}_b$ from the state space to the observation space. For example, tree-ring width may be estimated from grid-point values of temperature and moisture in the prior. Matrix $\mathbf{K}$, the Kalman gain, weights $\boldsymbol{y} - \mathcal{H}(\boldsymbol{x}_b)$ (which is called the innovation) and transforms it into state space. Matrix $\mathbf{R}$ is the error covariance matrix for the observations. We note that Eq. (1) assumes that $\boldsymbol{x}_b$ and $\mathcal{H}(\boldsymbol{x}_b)$ are Gaussian-distributed and that their errors are unbiased. The DA update process involves computing Eq. (1) to arrive at the posterior state; within the context of the climate reconstruction problem, the posterior state is the reconstructed state for a given time. Space–time reconstructions

are obtained by iteratively estimating the posterior state for each year or time segment of the reconstruction.

From the first covariance term of Eq. (2), we can interpret **K** as "spreading" the information contained in the observations through the covariance between the prior and the prior-estimated observations. This implies that, other things being equal, larger values of $cov(\boldsymbol{x}_b, \mathcal{H}(\boldsymbol{x}_b))$ will weight the innovation more heavily; thus, this new information not contained in the prior has a bigger influence. One way to improve this covariance relationship may be to use time-averaged observations, particularly if the model or climate system has better covariance relationships at longer timescales.

For the update calculations we employ an ensemble square-root Kalman filter with serial observation processing, applied to time averages (see Steiger et al., 2014, for a detailed DA algorithm and fuller discussion of DA terminology). We extend the technique of Dirren and Hakim (2005), Huntley and Hakim (2010), and Steiger et al. (2014) by iteratively applying the state-update equations across multiple timescales by leveraging the serial observation processing approach to the Kalman filter (Houtekamer and Mitchell, 2001). The previous related approaches have only considered annual timescales or less and cannot be trivially applied to proxies of arbitrary timescales: the prior must be constructed so as to provide meaningful time averages and the algorithm must be able to handle irregular proxy timescales.

The following general algorithm allows one to assimilate any collection of observation or proxy data, including time averages with irregular duration:

1. Construct a prior ("background") ensemble $\mathbf{x}_b$ at the highest temporal resolution of interest (e.g., monthly or annual), or a collection of them with one for each time step (e.g., monthly or annual ensembles assigned to particular months or years).

2. Loop over observations and assimilate each at their own timescale:

   (a) Decompose the prior ensemble(s) that overlap in time with the observation y into time averages (overbar) and deviations from this average (prime) via $\mathbf{x}_b = \bar{\mathbf{x}}_b + \mathbf{x}_b'$, such that the time average of $\bar{\mathbf{x}}_b$ matches the timescale of y.

   (b) Estimate the observation via the "proxy system model" (Evans et al., 2013) $\mathcal{H}(\bar{\mathbf{x}}_b)$, and update the time-averaged ensemble(s) $\bar{\mathbf{x}}_b \xrightarrow{\text{DA}} \bar{\mathbf{x}}_a$, with $\bar{\mathbf{x}}_a$ as the posterior (or "analysis") time-mean ensemble(s).

   (c) Add back the time deviations, $\mathbf{x}_b = \bar{\mathbf{x}}_a + \mathbf{x}_b'$, which can serve as the prior(s) for another observation.

   Note that if y shares the same timescale as $\mathbf{x}_b$, then the method is the same, but with $\mathbf{x}_b' = 0$.
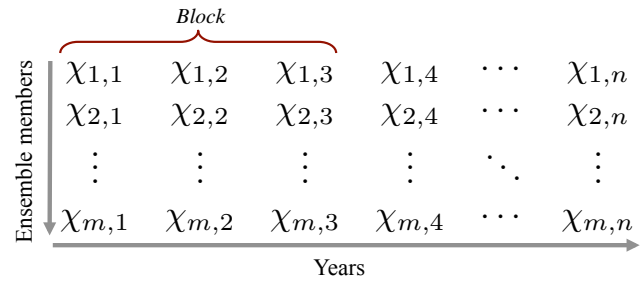

Figure 1. Schematic of the general reconstruction method using an offline approach. Prior ensembles of m state vectors, χ, are assigned to each of the n years. To retain some temporal coherency, the rows are composed of time-coherent blocks drawn from a climate model simulation (arbitrarily illustrated here as a 3-year block, or three consecutive annual states). The method updates prior ensembles for specific years corresponding to annual proxy data points, while for long-timescale proxies prior ensembles are computed by time averaging across the rows corresponding to the years of a proxy data point.

3. After all observations have been assimilated, the ensemble mean of $\mathbf{x}_b$ provides the best estimate of the state for each analysis time.

We now discuss an illustrative implementation of this general algorithm that we employ for the experiments in this paper. Consider a paleoclimatic situation where the observations are a collection of annual proxy data and also proxy data representing irregularly averaged climatic information. Let the prior ensembles be constructed, as we do in our experiments, by a random sample of annually averaged climate states from a long climate simulation and initially assigned to each year of a reconstruction (Fig. 1). Following the steps outlined above, proxies representing differently averaged time intervals can be assimilated by averaging over the prior ensembles for the time intervals defined by the proxy. For example, a proxy value representing information over the years 1700–1720 would update the prior ensembles averaged in time over that same interval. Annual proxies can simply be assimilated by updating the ensembles for each year of available proxy data (Fig. 1). This approach proceeds by assimilating each proxy over its full time extent, and after every proxy is assimilated, one is left with an updated version of what one started with: a time sequence of ensemble state estimates at annual resolution.

In the example above and all of the experiments shown here we use a "no-cycling" or "offline" DA approach, where the prior ensembles are drawn from existing climate model simulations. This approach has vast computational benefits over a "cycling" or "online" approach, where one must integrate an ensemble of climate model simulations forward in time after each DA update step. Indeed, for the paleoclimate reconstruction problem, it is infeasible to cycle an ensemble of tens to hundreds of CMIP5-class coupled-climate models (as used here) for hundreds or thousands of years.

Moreover, in the offline case one may use hundreds to thousands of ensemble members from multiple models and simulations, reducing the potential for model bias and sampling error. It is also advantageous to use an offline approach when the predictability time limit of the model is shorter than the timescale of the observations: for example, if observations are only available at annual resolution yet the model cannot skilfully forecast the climate state a year into the future, then no useful information is gained by cycling the model. Matsikaris et al. (2015) recently compared online and offline approaches to paleoclimate DA with a fully coupled Earth system model and found no improvement with the online method, suggesting that the model was unable to provide useful information at analysis times. Nevertheless, one way the approach outlined here can generalize to the online approach is by cycling on the shortest timescale (e.g., annual or seasonal) and updating longer timescales at the end of the appropriate interval without cycling.

Also note that for the sake of simplicity in the illustrative example and throughout the paper, we are assuming that an irregular, long-timescale proxy is just an average of some climate variable over a given time interval. Real proxies are nearly always more complex than this and would necessitate a more sophisticated proxy system model ($\mathcal{H}(x_b)$ in Eqs. 1 and 2); however, the algorithm described above is general and covers the case when such models are available.

An important detail of the algorithm outlined above is that when updating the long-timescale states, the deviations from the time mean are not also updated. For most applications of this algorithm, we presume that there would be essentially two types of proxy timescales involved, annual and decadal timescales. Thus, we assume here that the decadal proxies would not inform annual climate variability. However, there is no formal or practical obstacle that stops one from updating these deviations from the time mean (Huntley and Hakim, 2010). If, for example, one were working with proxies from two very similar timescales where there may be important shared covariance information, the deviations could also be updated by appending them to the mean states, $\overline{x}_b$, and then proceeding with the algorithm.

One key point about the method outlined above is that if a reconstruction uses the offline approach together with multiple timescales, then a random sample of annually averaged climate states will not have meaningful multi-year averages. Temporal consistency of the priors will need to be ensured in order to have coherent long-timescale covariance relationships. One way to account for this is to draw priors in random blocks of consecutive years from the employed climate model simulation (see Fig. 1). The length of these blocks can be determined based on the needs of the specific reconstruction problem (e.g., the length of the longest proxy timescale) and the length of available model simulations. If multiple long simulations are available (they need not be from the same model), different rows in Fig. 1 could be different model simulations and the block length could be

the length of the reconstruction; this option avoids any discontinuities in time that result from small block lengths.

The DA technique of state space augmentation (as discussed in, for example, Anderson, 2001) has long been used for handling arbitrary observation operators and can also be used for updating time-averaged quantities or parameter estimation (e.g., Annan et al., 2005). Such an approach does have some important similarities to what we present here, particularly the ability to update time-averaged information. However, we are not aware of any paper that has worked out a multi-scale state augmentation approach for the paleoclimate reconstruction problem. We think this could represent a viable approach, but it remains untested for paleoclimate reconstructions.

3 Experimental framework

3.1 Models and variable characterizations

For the experiments presented here, we are interested in (1) how the reconstruction methodology proposed in Sect. 2 performs in both the atmosphere and ocean, (2) how the differing timescales of the atmosphere and ocean may be leveraged in the reconstruction process, and (3) how these results vary with two different models having quite different spectral characteristics in their coupled-climate systems. To this end we choose two long preindustrial control simulations (part of the Coupled Model Intercomparison Project Phase 5, available for download at http://www.earthsystemgrid.org/), one from the climate model GFDL-CM3 (800 years in length) and the other from CCSM4 (1051 years in length). We also choose two illustrative reconstruction variables, global-mean 2 m air temperature and the Atlantic meridional overturning circulation (AMOC). Figures 2 and 3 characterize the global-mean temperature and an AMOC index for each simulation (defined here as the maximum value of the overturning stream function in the North Atlantic between 25 and 70° N and between depths of 500 and 2000 m), respectively. In these reconstructions the state vector, x_b, only contains global latitude–longitude gridded values of 2 m air temperature together with global-mean temperature and the AMOC index as single-dimension appended state variables (rather than deriving them from the state vector itself). $\mathcal{H}(x_b)$ simply uses the surface temperature values of the state vector at the proxy locations. Note that even though these are only single-dimension variables, the DA framework proposed here can trivially reconstruct spatial variables as well (Steiger et al., 2014). From Figs. 2 and 3 we see that these two models display different spectral characteristics for both global-mean temperature and the AMOC index.

We next assess whether there are strong covariance relationships between the observation variables and the reconstruction variables at different time averages. Recall that the key covariance relationship in the DA update equations is between the prior variables and the prior estimate of the ob-
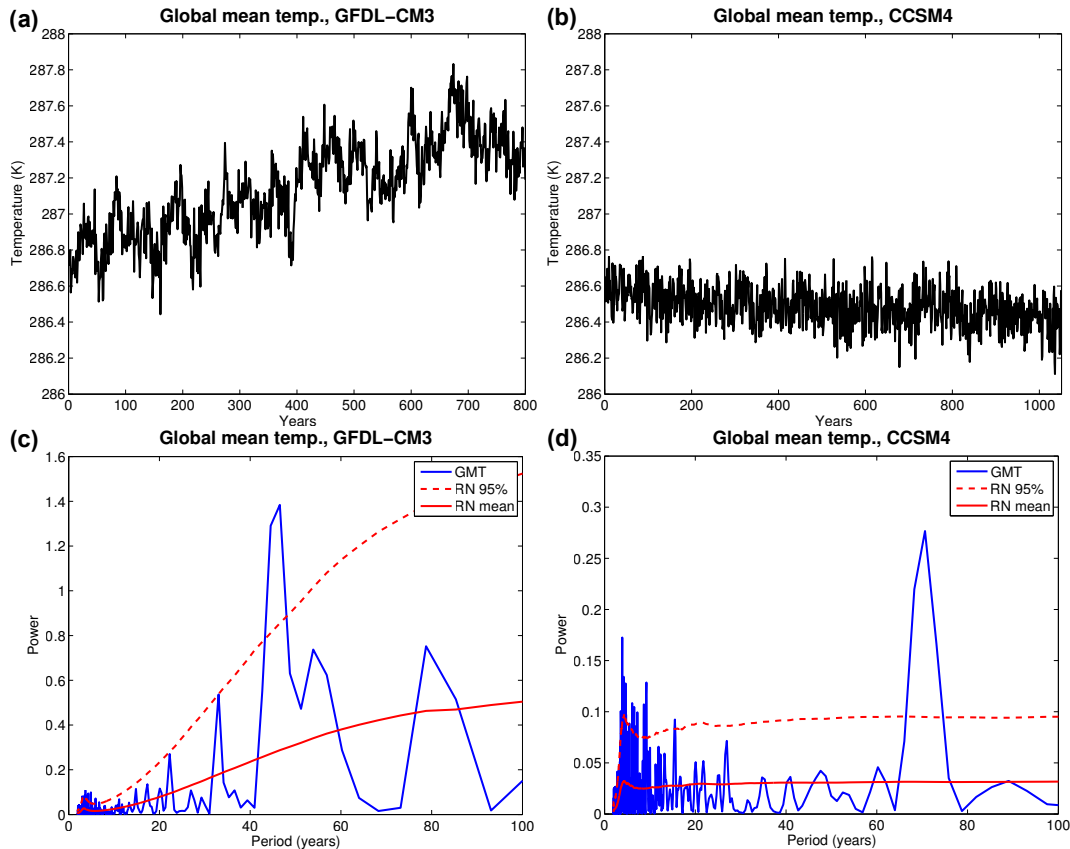
Figure 2. Characterization of the global-mean 2 m air temperature variables used in this paper. Panels **(a)** and **(b)** show the global-mean temperature time series for the preindustrial control simulations of GFDL-CM3 and CCSM4, respectively. Panels **(c)** and **(d)** show their respective power spectra (GMT) with a best-fit red noise (RN) spectrum (computed as in Schneider and Neumaier, 2001) and an estimated 95 % confidence interval.

servations (Eq. 2). A simple assessment of this is shown in Fig. 4, which shows the correlation between the prior variables and the surface temperature time series at the pseudo-proxy grid points for both climate simulations at a range of time averages. (Note that the correlation of two time series is simply the covariance normalized by the product of the standard deviations of the two time series.) Figure 4 indicates that there is increased covariance information (or more locations with higher correlations) between surface temperature and the prior variables at longer timescales. This information is leveraged by the equations of DA to potentially improve the low-frequency components of the reconstructed variables. An important point about computing correlations at increasing time averages is that the number of degrees of freedom in the time series is also reduced, potentially spuriously inflating the correlations in Fig. 4. Accounting for these reduced degrees of freedom by performing a test of statistical significance would not, however, be particularly germane: the DA equations do not "know" about 95 % confidence intervals, just the covariance information. If, after performing the reconstructions and computing several different skill metrics, we see an increase in reconstruction skill, then we can

infer that the information was in fact useful for the reconstructions.

3.2 Pseudoproxy construction

The pseudoproxy experiments employed here follow the general framework of many previous studies (see Smerdon, 2012, for a summary and review) but with some important modifications. Generally, after one or more climate model simulations are chosen to represent nature, a pseudoproxy network is chosen that mimics real-world proxy availability, similar to the network chosen here and shown in Fig. 5a; this particular network is composed of a spatially thinned version of the proxy collection of PAGES 2k Consortium (2013) (thinned over Asia and North America, where the proxy density is high) and all of the proxy locations in Shakun et al. (2012) and Marcott et al. (2013). Pseudoproxies are typically generated by adding random white noise to the chosen network of climate model temperature series. We note that the choice of white noise (as opposed to "red noise", for instance) is a simplification of the "real" noise in proxies. However, we consider this to be a reasonable choice be-
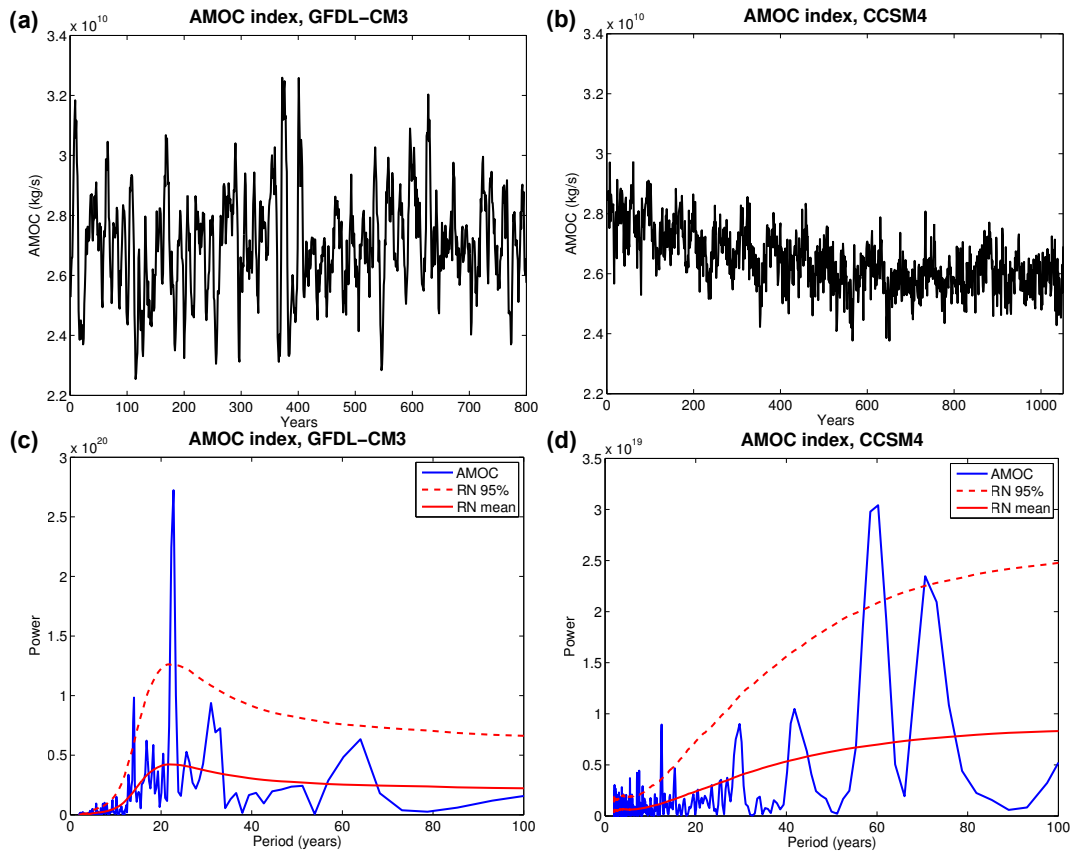
Figure 3. Characterization of the Atlantic meridional overturning circulation (AMOC) index variables used in this paper. Panels **(a)** and **(b)** show the AMOC index time series (defined in the text) for the preindustrial control simulations of GFDL-CM3 and CCSM4, respectively. Panels **(c)** and **(d)** show their respective power spectra with a best-fit red noise (RN) spectrum (computed as in Schneider and Neumaier, 2001) and an estimated 95 % confidence interval.

cause the purpose of the present work is primarily to illustrate a new reconstruction method. The added noise is usually assumed to be the same value for all proxy locations, with a common signal-to-noise ratio (SNR) being 0.5 (where $\mathrm{SNR} \equiv \sqrt{\mathrm{var}(X)/\mathrm{var}(N)}$, and where X is a grid-point temperature series drawn from the true state, N is an additive noise series, and var is the variance.). Following recent work by Wang et al. (2014), we instead randomly draw SNR values from a distribution characteristic of real proxy networks (Fig. 5b). This distribution is a shifted gamma distribution (shape parameter = 1.667, scale parameter = 0.18, shifted by 0.15) with a mean SNR of 0.45 and is modeled after Fig. 3 from Wang et al. (2014).

Also, in contrast to nearly all pseudoproxy experiments, we use pseudoproxies at two different timescales for each model. Importantly, we use the same SNR distribution for both timescales and add the noise to the time series *after* averaging. Within the DA framework, the additive error for each proxy is accounted for in the entries of the diagonal matrix $\mathbf{R}$. The SNR equation above is related to $\mathbf{R}$ in that each of these entries is equal to var(N) for a given proxy. The process of adding the noise after averaging ensures that $\mathbf{R}$

is statistically identical for each reconstruction. This process isolates the role of the covariance relationships in Eq. (2). By drawing from the same SNR distribution for all pseudoproxy timescales we are also assuming that the distribution is an appropriate characterization of the error in long-timescale proxies; we assume this for simplicity and also because we are not aware of a systematic assessment of SNR values for low-resolution proxies as Wang et al. (2014) have done for annual-resolution proxies.

We also note that an important idealization of the present pseudoproxy experiments, which we share with all pseudoproxy experiments heretofore published, is that we use a perfect model approximation where the pseudoproxies from one model simulation are used to reconstruct that same simulation; for example, pseudoproxies from the CCSM4 simulation are used to reconstruct the CCSM4 simulation. In a real DA-based reconstruction the climate model will never be a perfect description of the real climate system from which the assimilated observations are derived. Since the purpose of the present work is to illustrate a new algorithm, we have not considered this additional layer of complexity. This additional aspect could only be fully assessed within a study
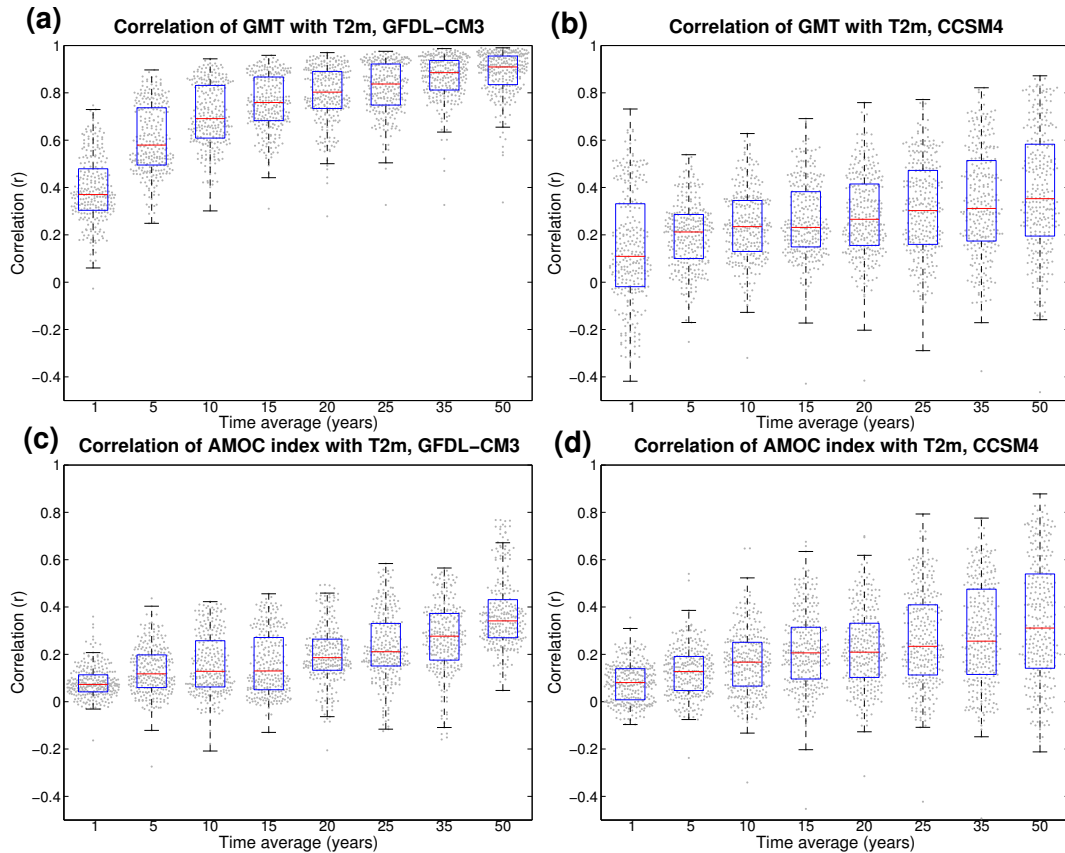
Figure 4. Panels **(a)** and **(b)** show the distribution of correlation values between the global-mean 2 m air temperature (GMT) and the 2 m air surface temperatures (T2m) at the proxy locations for GFDL-CM3 and CCSM4 at a range of time averages. Panels **(c)** and **(d)** show similar correlation distributions but for the correlation between the Atlantic meridional overturning circulation (AMOC) index and the 2 m air surface temperatures at the proxy locations. The correlations are computed for proxy grid points at a given time average, with the distribution of these correlations shown in gray dots and also summarized by box plots, where the blue boxes enclose the 25th and 75th percentiles and the whiskers extend to ± 1.5 times this interquartile range.
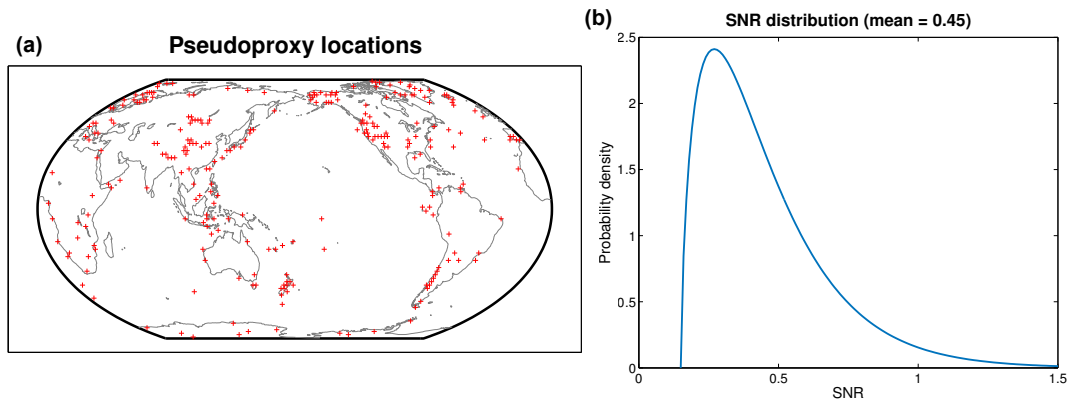

Figure 5. **(a)** Pseudoproxy locations used in this study ($n = 274$), drawn from the predominantly high-resolution (annual) proxy collection of PAGES 2k Consortium (2013) and all the comparatively low-resolution (decadal to centennial) proxy locations in Shakun et al. (2012) and Marcott et al. (2013). **(b)** The signal-to-noise ratio (SNR) distribution for the pseudoproxies, based on a real-world estimate of Wang et al. (2014). For a given Monte Carlo experiment, the SNR for each pseudoproxy was randomly drawn from this distribution.

of real proxy climate reconstructions; using one simulation to reconstruct another can assess inter-model differences, but it is unclear how these results would relate to model–nature differences.

3.3 Pseudoproxy experiments

The primary results of this paper are presented in a series of 12 experiments using only atmospheric surface temperature pseudoproxies to reconstruct the global-mean temperature and AMOC index of the two climate model simulations discussed previously. For each variable, and each model, three experiments are performed: (1) short (annual) pseudoproxies only, (2) long (5- or 20-year time averages) pseudoproxies only, and (3) both short and long time-averaged pseudoproxies. We have chosen the long timescale for the CCSM4 simulation to be 20 years, and we note that an alternative choice of one to several decades gives similar results (not shown). The situation is more complex with the GFDL-CM3 simulation because of the presence of an approximate 22-year periodic signal in the AMOC (Fig. 3a and c). A choice of 20 years for GFDL-CM3 would effectively undersample the AMOC variability, and so we have chosen a long timescale of 5 years for GFDL-CM3. Unfortunately, a long timescale of 5 years for CCSM4 shows little difference in the results over the annual timescale reconstructions (not shown), as would be suggested by the small difference in correlation (covariance) between 1 and 5 years (Fig. 4b, d).

Both the short-only and long-only reconstructions use 100 pseudoproxies randomly drawn from the network of 274 proxy locations shown in Fig. 5a. For the mixed-resolution reconstructions, 100 pseudoproxies are randomly drawn from the network for each timescale, giving a total of 200. This is an approximation of the real-world setting, where one usually has proxies at multiple timescales and would like to use all of them. Following the algorithm outlined in Sect. 2, for the multi-scale reconstructions, we assimilate the long-timescale pseudoproxies first, followed by the annual timescale pseudoproxies; we also performed these reconstructions by swapping which timescale was assimilated first and found statistically identical results (not shown), as would be expected from the linearity of this approach. For these mixed-resolution reconstructions, we have also ensured that there is no overlap between locations associated with the two timescales.

We have reconstructed the first 400 years of each simulation while drawing the priors from the following 400 years of the simulations. Each year had a prior size of 1000 (e.g., from Fig. 1, $m = 1000$), while the blocks were randomly drawn in 20-year continuous segments. This uniform block length was chosen because it was the longest timescale of the pseudoproxies and because the pseudoproxies were constructed over regular long intervals and thus discontinuities at block edges were not a concern (see Fig. 1 and the discussion in Sect. 2). Because the prior ensemble size was 1000, we did not employ covariance localization, a common DA practice for controlling sampling error. Each of the 12 reconstructions is repeated 100 times in a Monte Carlo fashion where new proxy networks and SNR values are randomly chosen each iteration; the new pseudoproxy networks are randomly drawn from the network shown in Fig. 5a and the SNR values are randomly drawn from the distribution shown in Fig. 5b. All the reconstruction figures show the mean of 100 of these Monte Carlo reconstruction iterations along with error bars indicating $\pm 2\sigma$ of the "grand ensemble" of analysis ensembles for all the Monte Carlo iterations (with an ensemble size of 1000 and 100 iterations, the grand ensemble has 1×10^5 members).

4 Reconstruction results

The reconstructions of global-mean temperature are shown in Fig. 6 along with their associated $\pm 2\sigma$ error estimates. In Fig. 6, panels (a) and (b) show the reconstructions with the annual pseudoproxies, (c) and (d) show the reconstructions with the long-timescale proxies, and (e) and (f) show the reconstructions for both timescales. Skill metrics, computed at annual resolution, are shown for each reconstruction: correlation (r), coefficient of efficiency (CE), and mean continuous ranked probability score (CRPS). The coefficient of efficiency for a data series comparison of length N is defined as

$$\mathrm{CE} = 1 - \frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \, ,$$

where x is the "true" time series, $\overline{x}$ is the true time-series mean, and $\hat{x}$ is the reconstructed time series. The metric CE has the range $-\infty < \mathrm{CE} \leq 1$, where $\mathrm{CE} = 1$ corresponds to a perfect match and $\mathrm{CE} < 0$ generally indicates no reconstruction skill or a bias in the reconstruction. The CRPS is a "strictly proper" scoring metric that accounts for the skill of both the mean and the spread of an ensemble forecast or state estimate (Gneiting and Raftery, 2007). The CRPS measures the area of the squared difference between the cumulative distribution functions of the posterior ensemble state estimate and the true state (a Heaviside function centered on the true value). A smaller area would indicate a more skillful reconstruction, so smaller values of CRPS are better. All CRPS values shown in the figures are temporal means of CRPS over the reconstruction interval. Comparing the reconstructions in Fig. 6 we see that the bottom panels with both timescales have the best skill and the smallest error bars, indicating greater confidence. For GFDL-CM3, there is a 25 % reduction in the mean standard deviation over the annual-only experiment and a 17 % reduction over the long timescale only; for CCSM4, there is a 2 % reduction in the mean standard deviation over the annual timescale only and a 28 % reduction over the long timescale only.
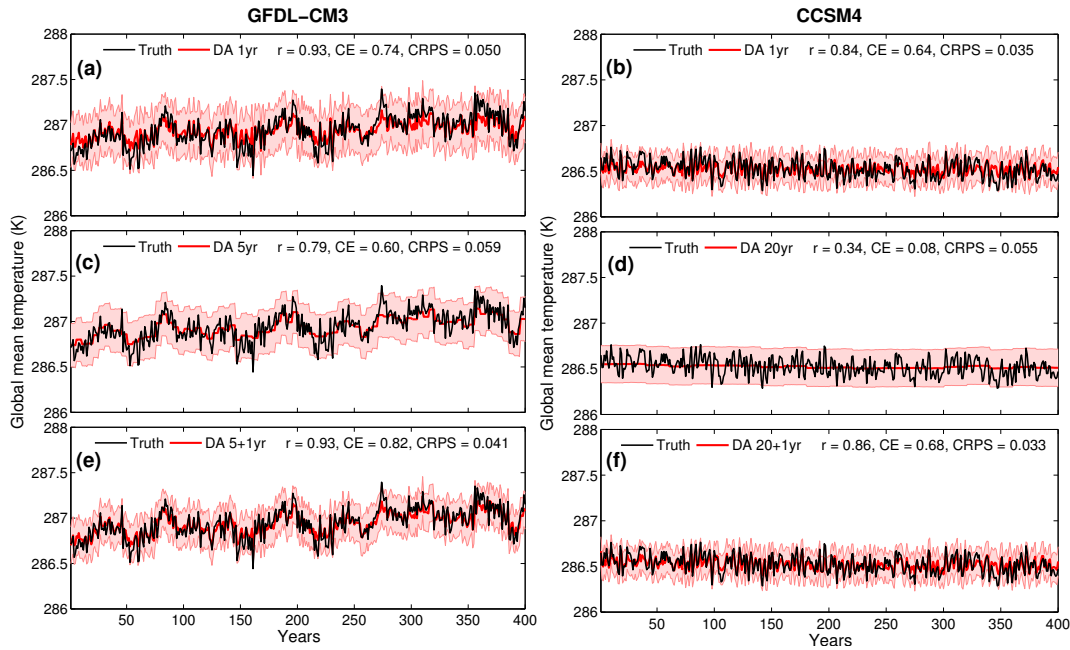
Figure 6. Global-mean temperature reconstructions (mean of 100 Monte Carlo iterations, with error bars indicating $\pm 2\sigma$ of the iterations and analysis ensembles) for the three types of experiments discussed in the text and for each climate model simulation. Black lines indicate the true time series, while red lines indicate the reconstructed time series for only short-timescale (annual) pseudoproxies, only long timescale (5 or 20 years) pseudoproxies, and both long- and short-timescale pseudoproxies. Skill metrics of the reconstructions, correlation (r), coefficient of efficiency (CE), and mean continuous ranked probability score (CRPS) are shown at the top of each subpanel.
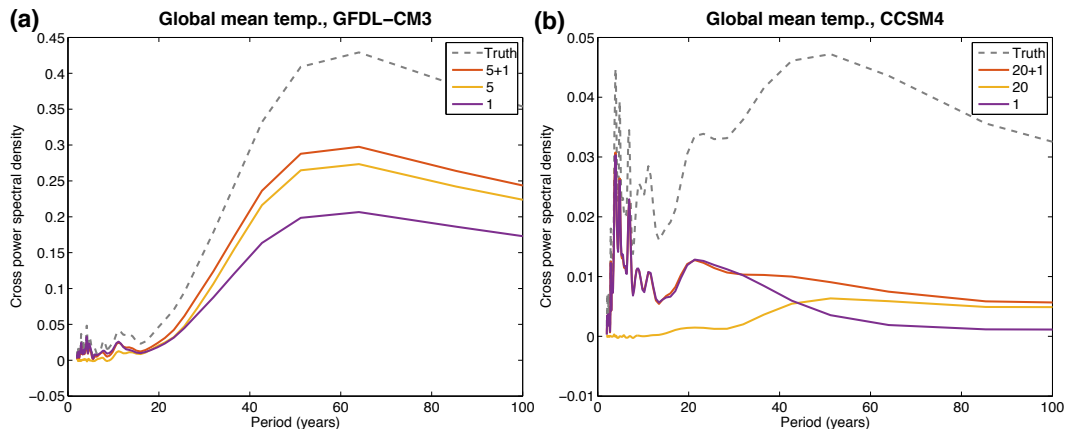

Figure 7. Cross spectra of the reconstructed global-mean temperature time series with the true global-mean temperature time series, for the reconstructions shown in Fig. 6. For reference, the dashed gray line indicates the cross spectra of the true time series with itself, or equivalently its own power spectrum.

Note that the long-timescale reconstructions shown in Fig. 6c and d have sharp edges at 5- (for GFDL-CM3) or 20-year (for CCSM4) intervals. This is due to the simplified experimental design we have employed where all the long-timescale pseudoproxies are averages over a given 5- or 20-year period. As discussed in Sect. 2, this experimental design is only a single illustrative example of the general algorithm. The data from real proxies are not always apportioned into specific time frames but can be scattered irregularly in

time (e.g., fossil coral records). Using many irregular proxies will act to smooth the long-timescale reconstructions. As long as the timescales can be estimated and an appropriate proxy system model is used, the algorithm of Sect. 2 can handle any real proxy data. While not dealt with explicitly here, real long-timescale (low-resolution) proxies also have dating uncertainty, which will also tend to smooth the reconstructions. The algorithm can account for dating uncertainty through the Monte Carlo framework by repeating the recon-
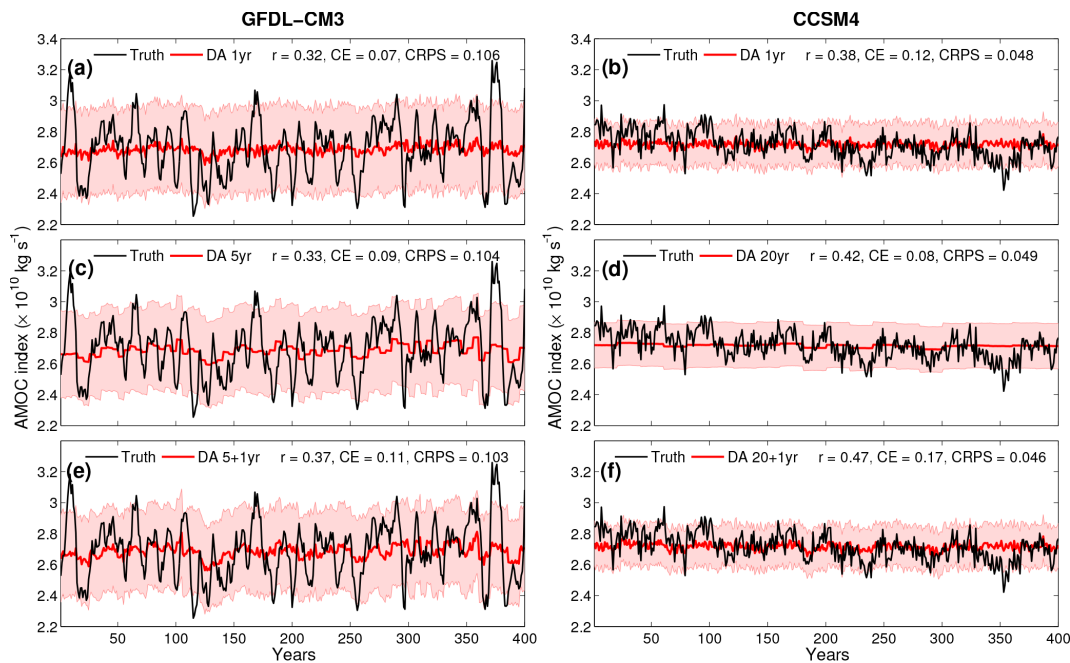
Figure 8. AMOC index reconstructions (mean of 100 Monte Carlo iterations, with error bars indicating $\pm 2\sigma$ of the iterations and analysis ensembles) for the three types of experiments discussed in the text and for each climate model simulation. Black lines indicate the true time series, while red lines indicate the reconstructed time series for only short-timescale (annual) pseudoproxies, only long-timescale (5 or 20 years) pseudoproxies, and both long- and short-timescale pseudoproxies. Skill metrics of the reconstructions, correlation (r), coefficient of efficiency (CE), and mean continuous ranked probability score (CRPS) are shown at the top of each subpanel.

structions many times and sampling from an age model for a given proxy.

One assessment of skill as a function of timescale is to compute the cross spectrum of the reconstructed time series with the true time series (Fig. 7). The cross spectra in this case reveal the relationship between the two time series as a function of frequency or period. As a point of reference, the dashed gray lines in this figure indicate the cross spectra of the true time series with itself, which is the same as its own power spectrum.[1] Considering Fig. 7b we see that the annual-only reconstruction does a better job of matching the power at short periods than the 20-year-only reconstruction; however, the 20-year-only reconstruction performs better at longer periods. The mixed timescale reconstruction, 20+1, does better or just as well as the single-timescale reconstructions at both short and long periods. This same general result holds for Fig. 7a, though it is more difficult to see because of the much larger power at longer periods in the GFDL-CM3 simulation.

Figure 8 shows three timescale reconstructions of the AMOC index for the two model simulations, similar to

Fig. 6. In these AMOC index reconstructions, we see the same general patterns as with the global-mean temperature reconstructions, where the multi-scale reconstructions provide the most skill (r, CE, CRPS) as well as the smallest error bars. However, while the multi-scale improvements in skill for the AMOC reconstructions are significant, these are generally smaller than for the global mean temperature reconstructions. For GFDL-CM3 there is a 3 % reduction in the mean standard deviation of the error bars over the annual-only experiment and a 1 % reduction over the long timescale only; for CCSM4, there is a 2 % reduction in the mean standard deviation over the annual timescale only and a 4 % reduction over the long timescale only. Figure 9 shows the corresponding cross spectra for the reconstructions shown in Fig. 8. Given that the pseudoproxies are of surface air temperature, it is not surprising that the absolute skill values of the AMOC reconstructions are reduced relative to the reconstructions of global-mean temperature. Though it is striking how much power is lost in the reconstructions (Fig. 9) considering that these are "perfect model" experiments. This result is most likely due to the fact that the covariances between surface temperatures and the AMOC index are quite small when compared with the covariances for temperature (Fig. 4); note that even though the CCSM4 mean correlation value for temperature, (Fig. 4b), is comparable to that for the AMOC, (Fig. 4d), the absolute correlation values are much larger, including a cluster of positive correlation val-

---

[1]Following a common technique to reduce noise in the cross spectra, they are computed using Welch's averaged periodogram method, which samples segments of the time series and averages the power spectra of these samples to arrive at the cross power spectral densities. As a result, the gray line spectra in Figs. 7 and 9 should not be expected to precisely match up with Figs. 2 and 3

ues between 0.6 and 0.4. Thus surface temperatures at the global proxy locations are relatively uninformative about the AMOC.

An additional result from Fig. 9 is the improved low-frequency components of the AMOC reconstructions when time-averaged surface temperature pseudoproxies are used. We argue that this result follows from the fact that the annual observations of atmospheric surface temperature are essentially noise to the slowly varying ocean. One may improve the information content relevant to the ocean by averaging over the atmospheric noise. This interpretation may also be seen in Fig. 4, where the correlation (covariance) information between the atmosphere and the ocean is particularly low at annual averages but improves at longer time averages (as also seen in Tardif et al., 2014). To further test this idea we performed mixed-resolution experiments in which the 100 pseudoproxies for each timescale were taken from the same locations. Therefore, here the long-timescale proxies are just the time-averaged versions of the annual proxies. We found essentially identical results with those shown in Figs. 6–9: all CRPS values were identical to three decimal places, r and CE were either identical to two decimal places or had changes of only ± 0.01, and there were no discernible differences in the cross spectra.

We note that all the cross spectra of the reconstructions shown in Figs. 7 and 9 show a decrease in power relative to the true state, though this need not always be the case. In additional experiments we performed using global ocean heat content, we found that this reconstructed variable tended to have more power than the true state and was thus higher than the respective dashed gray lines (not shown). Therefore the reduced power relative to the true state in Figs. 7 and 9 should not be interpreted as saying something general about the nature of DA-based reconstructions or the particular approach employed here.

As an approximation of a real reconstruction scenario, the experiments shown in Figs. 6 and 8 with two timescales use twice as many pseudoproxies as the single-timescale experiments (200 vs. 100). Therefore the improved skill might simply be a consequence of having more observation information. We tested this idea by repeating all the experiments shown here but instead increasing the number of observations to 200 for each experiment: the single-timescale reconstructions used 200 randomly drawn pseudoproxies and the multi-scale reconstructions used 100 randomly drawn pseudoproxies each for the two timescales (the same as in the previous multi-scale reconstructions). Figure 10 is a characteristic example of the results of these additional tests. Figure 10a shows the reconstructions of the AMOC index with the CCSM4 model output and Fig. 10b shows the respective cross spectra. In (a), the skill is best for the multi-scale reconstructions and in (b) the cross spectra show the same general result of improved low-frequency power for the time-averaged pseudoproxies. However, the cross spectra for the $20 + 1$ reconstruction are not always closest to the true spec-

trum, suggesting that the number of pseudoproxies does play a role in improving the spectrum of the reconstructions. Indeed, it should be the case that, as long as the proxies are unbiased, adding more of them will improve a DA-based reconstruction.

5 Conclusions

This paper presents a data assimilation approach for paleoclimate reconstructions that can explicitly incorporate proxy data on arbitrary timescales. This approach generalizes previous data assimilation techniques in the sense that many scales of both proxies and climate states can be included explicitly in a single reconstruction framework. The primary interest in such a reconstruction technique is that it allows for the inclusion of much more proxy data in climate reconstructions. Given the spatially sparse and noisy nature of proxies, more information will tend to improve the quality of the reconstructions. Besides this benefit, using multi-scale proxy data may be particularly useful given the many inherent timescales of the climate system, such as the fast timescales of the atmosphere and the slow timescales of the ocean. We performed three types of realistic atmosphere–ocean pseudoproxy reconstructions to assess the impact of using observations at multiple timescales: (1) short (annual) pseudoproxies only, (2) long ($\sim$ decadal) pseudoproxies only, and (3) both short and long time-averaged pseudoproxies. We found for both global-mean temperature and an index of the AMOC that the reconstructions that incorporated proxies across both short and long timescales were more skillful than reconstructions that used short or long timescales alone (Figs. 6 and 8). This result holds even when the number of pseudoproxies for the single-timescale reconstructions are doubled (Fig. 10a). Multi-scale reconstructions would be expected to perform better than single-scale reconstructions because they can include information at multiple timescales and because the prior can be better conditioned as it is used from one timescale to the next.

We found that reconstructions incorporating long-timescale pseudoproxies improve the low-frequency components of the reconstructions over reconstructions that only use annual timescale pseudoproxies (Figs. 7, 9, and 10b). This result may at first seem surprising because the annual pseudoproxies should contain the low-frequency information. It is helpful to recall that the data assimilation algorithm outlined here proceeds by sequentially finding the optimal state at each time segment of interest given the prior, the observations, and their respective errors. This state update critically relies on the covariance between the prior and the model estimate of the observations (Eq. 2). If, for example, surface temperature observations do not covary well with the AMOC at annual resolution, then the posterior AMOC estimate will be little changed compared to the prior (Tardif et al., 2014). But if the time average of surface temperatures
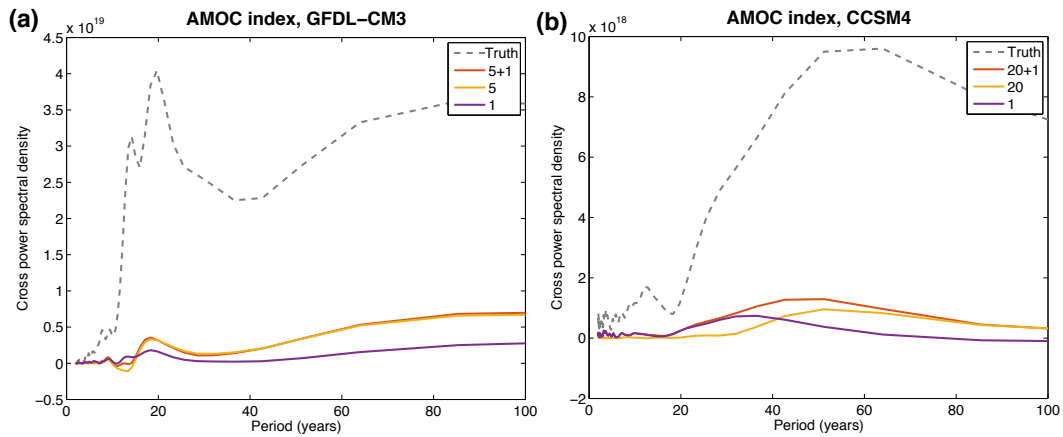
Figure 9. Cross spectra of the reconstructed AMOC index time series with the true AMOC index time series, for the reconstructions shown in Fig. 8. For reference, the dashed gray line indicates the cross spectra of the true time series with itself, or equivalently its own power spectrum.
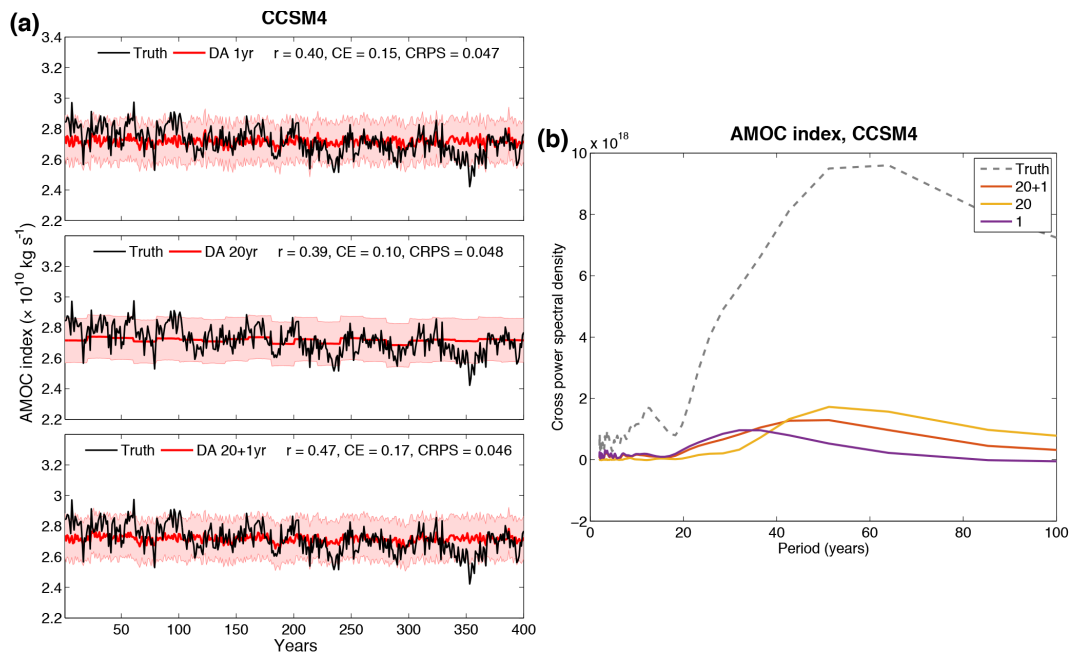

Figure 10. AMOC index reconstructions (mean of 100 Monte Carlo iterations, with error bars indicating $\pm 2\sigma$ of the iterations and analysis ensembles) and corresponding cross spectra similar to those shown in Figs. 8b, d, f and 9b but for the case where each experiment uses 200 pseudoproxies: the single-timescale reconstructions use 200 pseudoproxies each, while the multi-timescale reconstructions use 100 pseudoproxies for the short timescale and 100 pseudoproxies for the long timescale.

has a large covariance with the AMOC, the posterior will be more influenced by the observations. This result is not controlled by the noise added to the pseudoproxies because, as noted in Sect. 3.2, we ensured that **R** from Eq. (2) remains fixed for both timescales.

These results indicate that DA-based atmosphere–ocean state estimates may be improved by including proxies and climate states from multiple timescales. The general results outlined above are consistent across the employed climate

models. These results also show, as suggested by Kurahashi-Nakamura et al. (2014), that given a representative prior ensemble, features of the AMOC may be reconstructed using observations of surface variables. However, the reconstructions lack spectral power across all frequencies, which we attribute to the relatively small covariances between surface temperatures and the AMOC in the models we employed (Fig. 4c and d). Therefore observations of ocean quantities

such as salinity or indirect measures of ocean circulation may be better suited to reconstructing the AMOC.

6 Data availability

The GFDL-CM3 simulation is publicly available at http://nomads.gfdl.noaa.gov/.

The CCSM4 simulation is publicly available at https://www.earthsystemgrid.org/home.html.

References

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Weather Rev., 129, 2884–2903, 2001.

Annan, J. D., Lunt, D. J., Hargreaves, J. C., and Valdes, P. J.: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, Nonlin. Processes Geophys., 12, 363–371, doi:10.5194/npg-12-363-2005, 2005.

Bradley, R. S.: Paleoclimatology, Academic Press, Oxford, UK, 3rd edn., 2014.

Carro-Calvo, L., Salcedo-Sanz, S., and Luterbacher, J.: Neural computation in paleoclimatology: General methodology and a case study, Neurocomputing, 113, 262–268, 2013.

Dirren, S. and Hakim, G. J.: Toward the assimilation of time-averaged observations, Geophys. Res. Lett., 32, 4, doi:10.1029/2004GL021444, 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.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association, 102, 359–378, 2007.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly", Clim. Dynam., 39, 2847–2866, 2012.

Guiot, J. and Corona, C.: Growing season temperatures in Europe and climate forcings over the past 1400 years, PloS one, 5, e9972, doi:10.1371/journal.pone.0009972, 2010.

Hanhijärvi, S., Tingley, M., and Korhola, A.: Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2,000 years, Clim. Dynam., 41, 2039–2060, doi:10.1007/s00382-013-1701-4, 2013.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.

Huntley, H. S. and Hakim, G. J.: Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model, Clim. Dynam., 35, 995–1009, doi:10.1007/s00382-009-0714-5, 2010.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge, Cambridge, UK, 2003.

Kurahashi-Nakamura, T., Losch, M., and Paul, A.: Can sparse proxy data constrain the strength of the Atlantic meridional overturning circulation?, Geosci. Model Dev., 7, 419–432, doi:10.5194/gmd-7-419-2014, 2014.

Li, B., Nychka, D. W., and Ammann, C. M.: The value of multi-proxy reconstruction of past climate, J. Am. Stat. Assoc., 105, 883–895, 2010.

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Testing the fidelity of methods used in proxy-based reconstructions of past climate, J. Climate, 18, 4097–4107, 2005.

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, doi:10.1073/pnas.0805721105, 2008.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A Reconstruction of Regional and Global Temperature for the Past 11,300 Years, Science, 339, 1198–1201, doi:10.1126/science.1228026, 2013.

Mathiot, P., Goosse, H., Crosta, X., Stenni, B., Braida, M., Renssen, H., Van Meerbeeck, C. J., Masson-Delmotte, V., Mairesse, A., and Dubinkina, S.: Using data assimilation to investigate the causes of Southern Hemisphere high latitude cooling from 10 to 8 ka BP, Clim. Past, 9, 887–901, doi:10.5194/cp-9-887-2013, 2013.

Matsikaris, A., Widmann, M., and Jungclaus, J.: On-line and off-line data assimilation in palaeoclimatology: a case study, Clim. Past, 11, 81–93, doi:10.5194/cp-11-81-2015, 2015.

PAGES 2k Consortium: Continental-scale temperature variability during the past two millennia, Nat. Geosci., 6, 339–346, doi:10.1038/ngeo1797, 2013.

Schneider, T. and Neumaier, A.: Algorithm 808: ARfit – A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models, ACM Transactions on Mathematical Software (TOMS), 27, 58–65, 2001.

Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z., Otto-Bliesner, B., Schmittner, A., and Bard, E.: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation, Nature, 484, 49–54, 2012.

Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Clim. Change, 3, 63–77, doi:10.1002/wcc.149, 2012.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of time-averaged pseudoproxies for climate reconstruction, J. Climate, 27, 426–441, doi:10.1175/JCLI-D-12-00693.1, 2014.

Tardif, R., Hakim, G., and Snyder, C.: Coupled atmosphere-ocean data assimilation experiments with a low-order model and CMIP5 model data, Clim. Dynam., 43, 1631–1643, doi:10.1007/s00382-014-2390-3, 2014.

Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E., and Rajarat-
nam, B.: Evaluating climate field reconstruction techniques using
improved emulations of real-world conditions, Clim. Past, 10, 1–
19, doi:10.5194/cp-10-1-2014, 2014.

Zachos, J. C., Dickens, G. R., and Zeebe, R. E.: An early Cenozoic
perspective on greenhouse warming and carbon-cycle dynamics,
Nature, 451, 279–283, 2008.