Global climate reconstruction across time and space using data assimilation

Nathan Steiger

A dissertation submitted in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Greg Hakim, Chair

David Battisti

Gerard Roe

Eric Steig

Program Authorized to Offer Degree:
Department of Atmospheric Sciences

University of Washington

Abstract

Global climate reconstruction across time and space using data assimilation

Nathan Steiger

Chair of the Supervisory Committee:
Professor Greg Hakim
Department of Atmospheric Sciences

Paleoclimate proxy data span seasonal to millennial time scales, and Earth's climate system has both high- and low-frequency components. Yet it is currently unclear how best to incorporate multiple time scales of proxy data into a single reconstruction framework and to also capture both high- and low-frequency components of reconstructed variables. In the first half of this dissertation we present a data assimilation algorithm that can explicitly incorporate proxy data at arbitrary time scales. Through a series of pseudoproxy experiments, we find that atmosphere–ocean states are most skilfully reconstructed by incorporating proxies across multiple time scales compared to using proxies at short (annual) or long ($\sim$ decadal) time scales alone. Additionally, reconstructions that incorporate long time-scale 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 insensitive to the choice of climate model, 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 solely on atmospheric surface temperature proxies.

Water isotope data from ice cores, particularly δ^{18}O, has long been used as a paleoclimate

proxy. In the past decade or so isotope-enabled climate models have allowed for a more rich understanding of the climate processes that produce the isotopic signals in precipitation. Such modeling-based studies tend to complicate simple temperature-isotope interpretations by pointing to the many non-local influences on isotopic signals at coring locations. Recent observational studies have also linked ice cores to non-local patterns of climate variability, particularly to mid-latitude atmospheric circulation patterns and to variations in tropical climate. However, the full spatial and temporal extent to which ice cores can robustly inform past climate is unknown. In the second half of this dissertation we estimate a realistic upper-bound on what ice cores can tell us about climate at annual and decadal time scales in a paleoclimate reconstruction context. We employ a similar data assimilation-based reconstruction technique that optimally combines isotopic proxy information with the dynamical constraints of climate models. Through several pseudo and real proxy experiments we assess the spatial and temporal extent to which ice cores can reconstruct the key variables of surface temperature, geopotential height at 500 hPa, and precipitation. We find local reconstruction skill to be the most robust across the reconstructions, particularly for temperature and geopotential heights. Non-local skill is also found for these variables in certain locations. For precipitation we find virtually no reconstruction skill outside of coastal and West Antarctica. These results are in agreement with long-held views that isotopes in ice cores have clear value as local climate proxies, particularly for temperature and atmospheric circulation. These results also show that in principle non-local climate information may also be reliably deduced from ice cores, though the spatial range of this information depends on how proxies are modelled within the reconstruction process, is non-uniform, and may not extend into plausible nearby locations: in particular, Greenland ice cores, by their nature, appear to be relatively uninformative for Europe and the British Isles at annual and decadal time scales.

TABLE OF CONTENTS

Page

Chapter 1: Multi-time scale data assimilation for atmosphere–ocean state estimates 1
1.1 Introduction . . . 1
1.2 Assimilation technique . . . 3
1.3 Experimental framework . . . 7
1.4 Reconstruction results . . . 16
1.5 Conclusions . . . 24

Chapter 2: What can ice cores tell us about past climate? . . . 26
2.1 Introduction . . . 26
2.2 Pseudoproxy experiments . . . 27
2.3 Real proxy experiments . . . 48
2.4 Conclusions . . . 56

Bibliography . . . 57

Appendix A: Supplementary figures . . . 63

Chapter 1

MULTI-TIME SCALE DATA ASSIMILATION FOR ATMOSPHERE–OCEAN STATE ESTIMATES

1.1 Introduction

Paleoclimate proxies range across widely different time scales. 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 have short records and are mostly limited to the past two millennia, whereas some low resolution proxies can reach back across the Cenozoic [e.g., Zachos et al., 2008]. In addition to the many time scales of proxies, the climate system itself varies across a large range of time scales: 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 time scales, captured by both proxies and climate models, as possible.

Few paleoclimate reconstruction methods have been created that specifically incorporate multiple proxy time scales. 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 easily constructed to calibrate in both low and high frequency domains. Nevertheless, a few methods have been modified for such purposes [e.g., Mann et al., 2005]. Additionally, this is not entirely a methodological problem but partly a temporal sampling issue: given that instrumental temperature data only span the past $\sim$ 150 years, low frequency reconstruction techniques

have few degrees of freedom on which to be calibrated and validated if the time scale is longer than about a decade. Only Li et al. [2010] (using a Bayesian hierarchical model approach) and Hanhijrvi et al. [2013] (using an approach based on pairwise comparisons) present reconstruction methods that can specifically incorporate proxies at any time resolution without linear interpolation of coarse proxies or coarsening high resolution proxies to some uniform time scale. These methods have thus far been used only for time series reconstructions, not space-time reconstructions.

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 that incorporates proxies across any arbitrary range of time scales. DA-based reconstructions have so far used only a single uniform time scale [e.g., Goosse et al., 2012] or have performed separate reconstructions at different uniform time scales [Mathiot et al., 2013]. 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. 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]. But besides this kind of benefit, 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 time scales.

1.2 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

$$\mathbf{x}_a \;=\; \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}_b)]\,, \tag{1.1}$$

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

$$\mathbf{K} \;=\; cov(\mathbf{x}_b, \mathcal{H}(\mathbf{x}_b))[cov(\mathcal{H}(\mathbf{x}_b), \mathcal{H}(\mathbf{x}_b)) + \mathbf{R}]^{-1}\,, \tag{1.2}$$

with *cov* representing a covariance expectation. The prior (or "background") estimate of the state vector is $\mathbf{x}_b$ and $\mathbf{x}_a$ is the posterior (or "analysis") state vector. Observations (or proxies) are contained in vector $\mathbf{y}$. The true value of the observations are estimated by the prior through $\mathcal{H}(\mathbf{x}_b)$, which is, in general, a nonlinear vector-valued observation operator that maps $\mathbf{x}_b$ 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 $\mathbf{y} - \mathcal{H}(\mathbf{x}_b)$, called the innovation, and transforms it into state space. Matrix $\mathbf{R}$ is the error covariance matrix for the observations. The DA update process involves computing Eqs. (1.1) and (1.2) 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 numerator of Eq. (1.2), we can interpret $\mathbf{K}$ 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 $cov(\mathbf{x}_b, \mathcal{H}(\mathbf{x}_b))$ 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 time scales.

For the update calculations we employ an ensemble square-root Kalman filter, 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 time scales by leveraging the serial observation processing approach to the Kalman filter [Houtekamer and Mitchell, 2001]. The following general algorithm allows one to assimilate any collection of observation or proxy data, including irregular time averages:

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 time scale:

   (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 $\mathbf{x}_b = \overline{\mathbf{x}}_b + \mathbf{x}'_b$, such that the time-average of $\overline{\mathbf{x}}_b$ matches the time scale of y.

   (b) Estimate the observation via the forward model, or "proxy system model" [Evans et al., 2013] $\mathcal{H}(\overline{\mathbf{x}}_b)$, and update (compute Eqs. (1.1) and (1.2)) the time averaged ensemble(s) $\overline{\mathbf{x}}_b \xrightarrow{DA} \overline{\mathbf{x}}_a$, with $\overline{\mathbf{x}}_a$ as the posterior (or "analysis") time-mean ensemble(s).

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

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

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. Prior ensembles can be composed of annually averaged climate states that have been randomly drawn from a long climate simulation and initially assigned to each year of a reconstruction, Fig. 1.1. 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. 1.1. 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 all of the experiments shown here we use a "no-cycling" or "off-line" DA approach, where the prior ensembles are drawn from existing climate model simulations. This approach has vast computational benefits over a "cycling" or "on-line" 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 off-line case one may use hundreds to thousands of ensemble

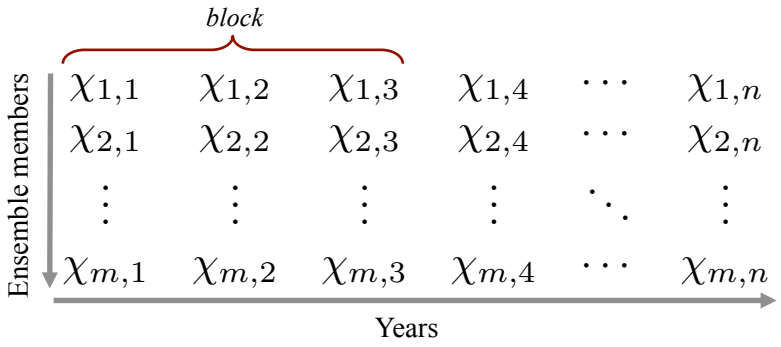

Figure 1.1: Schematic of the general reconstruction method using an off-line approach. Prior ensembles of m state vectors, χ, are assigned to each of the n years. To retain some temporal coherency, the rows are composed of time-coherent blocks drawn from a climate model simulation (arbitrarily illustrated here as a 3 year block, or 3 consecutive annual states). The method updates prior ensembles for specific years corresponding to annual proxy data points, while for long time-scale proxies prior ensembles are computed by time averaging across the rows corresponding to the years of a proxy data point.

members from multiple models and simulations, reducing the potential for model bias and sampling error. It is also advantageous to use an off-line approach when the predictability time limit of the model is shorter than the time scale 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 on-line and off-line approaches to paleoclimate DA with a fully-coupled earth system model and found no improvement with the on-line 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 on-line approach is by cycling on the shortest time scale (e.g., annual or seasonal) and updating longer time scales at the end of the appropriate interval without cycling.

If the reconstructions use the off-line approach together with multiple time scales, tem-

poral consistency of the priors will need to be maintained in order to have coherent long-time-scale covariance relationships. Thus portions of these priors can be randomly drawn in blocks of consecutive years from the employed climate model simulation, see Fig. 1.1. 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 time scale) 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. 1.1 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.

Also note that for the sake of simplicity in the above example and throughout the paper, we are assuming that an irregular, long-time-scale 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 ($\mathcal{H}(\overline{\mathbf{x}}_b)$ (in 1.1) and (1.2)); however, the algorithm described above is general and covers the case when such models are available.

1.3 Experimental framework

1.3.1 Models and variable characterizations

For the experiments presented here, we are interested in (1) how the reconstruction methodology proposed in Sect. 1.2 performs in both the atmosphere and ocean, (2) how the differing time scales of the atmosphere and ocean may be leveraged in the reconstruction process, and (3) how these results vary with two different models having very different spectral characteristics in their coupled-climate systems. To this end we choose two long pre-industrial 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 and the other from CCSM4. We also choose two illustrative reconstruction variables, global mean 2-m air temperature and the Atlantic meridional overturning circulation (AMOC). Figures

1.2 and 1.3 characterize the global mean temperature and an AMOC index for each simulation (defined here as the maximum value of the overturning streamfunction in the North Atlantic between 25°N and 70°N and between depths of 500 m and 2000 m), respectively. Note that even though these are only time series variables, the DA framework proposed here can trivially reconstruct spatial variables as well [Steiger et al., 2014]. From Figs. 1.2 and 1.3 we see that despite both being control simulations, these two models display very 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 observations, Eq. (1.2). For an experiment where the pseudoproxies, $\mathbf{y}$, are composed of atmospheric surface temperature (and thus what $\mathcal{H}(\mathbf{x}_b)$ estimates), we would need to know the covariance between these and the state variables of global mean temperature and the AMOC index, contained in $\mathbf{x}_b$. A simple assessment of this is shown in Fig. 1.4 with panels (a) and (b) showing box plot summaries of the correlations between the global mean temperature time series and all the surface temperature time series at every grid point for both climate simulations; these correlations are then computed over a range of averaging times. Panels (c) and (d) in Fig. 1.4 are the same calculation but for the AMOC index instead of the global mean temperature. (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 1.4 indicates that there is increased covariance information (or more locations with higher correlations) between surface temperature and the reconstruction variables at longer time scales. 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 are also reduced, potentially spuriously inflating the correlations. A test of statistical significance accounting for the reduced degrees of freedom is, however, not 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.

1.3.2 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. 1.5(a); 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. 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 SNR $\equiv \sqrt{var(X)/var(N)}$, and where X is a grid-point temperature series drawn from the true state and 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. 1.5(b). 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 modelled after Fig. 3 from Wang et al. [2014].

Also in contrast to nearly all pseudoproxy experiments, we use pseudoproxies at two different time scales for each model. Importantly for the comparison of results, we use the same SNR distribution for both time scales 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 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. (1.2). By drawing from the same SNR distribution for all pseudoproxy time scales we are also assuming that the distribution is an appropriate characterization of the error in long time scale 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 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 (e.g., pseudoproxies from the CCSM4 simulation are used to reconstruct the CCSM4 simulation, not the GFDL-CM3 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 novel algorithm, we therefore have not considered this additional layer of complexity, which can only be fully assessed within a study 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.

1.3.3 Pseudoproxy experiments

The primary results of this paper are presented with 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 time-scale 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. 1.3(a) and (c).

A choice of 20 years for GFDL-CM3 would effectively undersample the AMOC variability and so we have chosen a long time scale of 5 years for GFDL-CM3. Unfortunately, a long time scale of 5 years for CCSM4 shows little difference in the results over the annual time scale reconstructions (not shown), as would be suggested by the small difference in correlation (covariance) between 1 and 5 years, Fig. 1.4(b).

Both the short only and long only reconstructions use 100 pseudoproxies randomly drawn from the network of 274 proxy locations shown in Fig. 1.5(a). For the mixed resolution reconstructions, 100 pseudoproxies are randomly drawn from the network for each time scale, giving a total of 200. This is an approximation of the real world setting where one usually has proxies at multiple time scales and would like to use all of them. Following the algorithm outlined in Sect. 1.2, for the multi-scale reconstructions, we assimilate the long time-scale pseudoproxies first, followed by the annual time-scale pseudoproxies; we also performed these reconstructions by swapping which time scale 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 time scales.

We have reconstructed the first 500 years of each simulation while randomly drawing the prior, of size 500, in 20 year continuous blocks from the entire length of the simulations (800 years for GFDL-CM3 and 1051 years for CCSM4); this uniform block length was chosen because it was the longest time scale 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. 1.1 and the discussion in Sect. 1.2). Because of the large prior ensemble size, we did not employ covariance localization, a common DA practice for controlling sampling error. Each of the 12 reconstructions are repeated 100 times in a Monte Carlo fashion where new SNR values are randomly drawn from the distribution, Fig. 1.5(b), for each pseudoproxy location and a new pseudoproxy network is also randomly drawn for each time scale from the original network. All the reconstruction figures show the mean of 100 of these Monte Carlo reconstruction iterations.

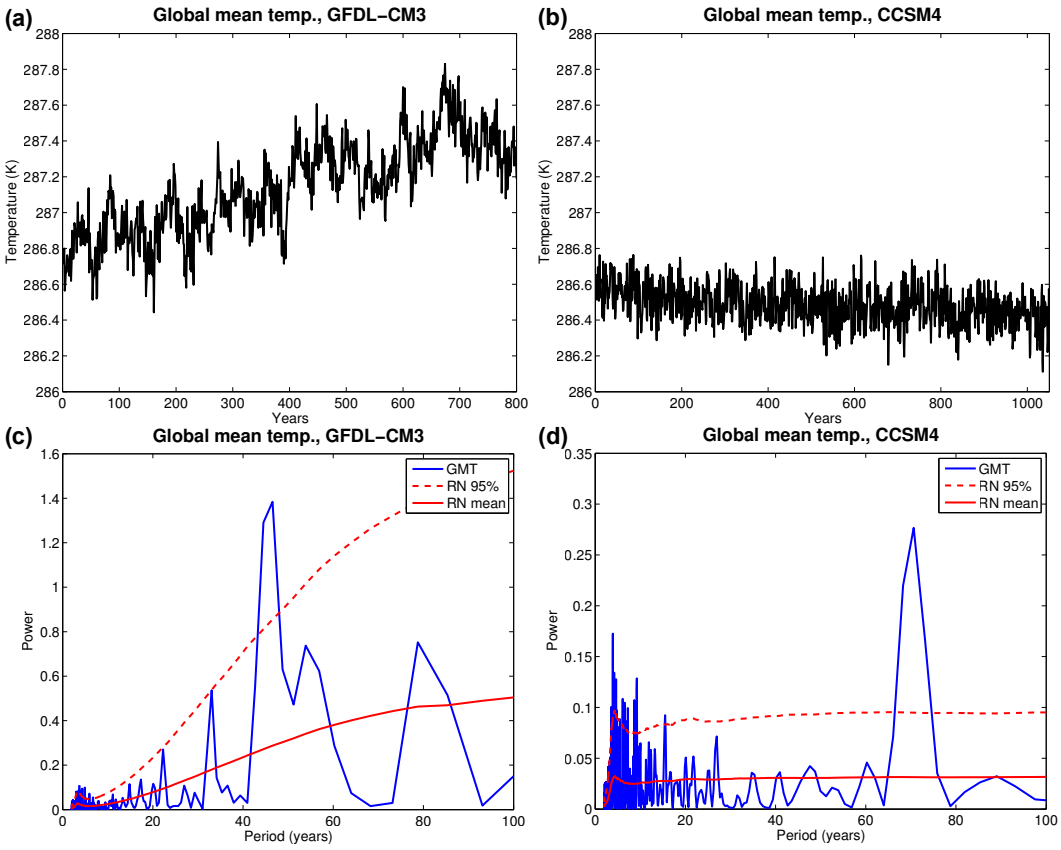

Figure 1.2: Characterization of the global mean 2-m air temperature variables used in this paper. Panels (a) and (b) show the global mean temperature time series for the pre-industrial control simulations of GFDL-CM3 and CCSM4, respectively. Panels (c) and (d) show their respective power spectra (GMT) with a best-fit red noise (RN) spectrum (computed as in Schneider and Neumaier [2001]) and an estimated 95% confidence interval.

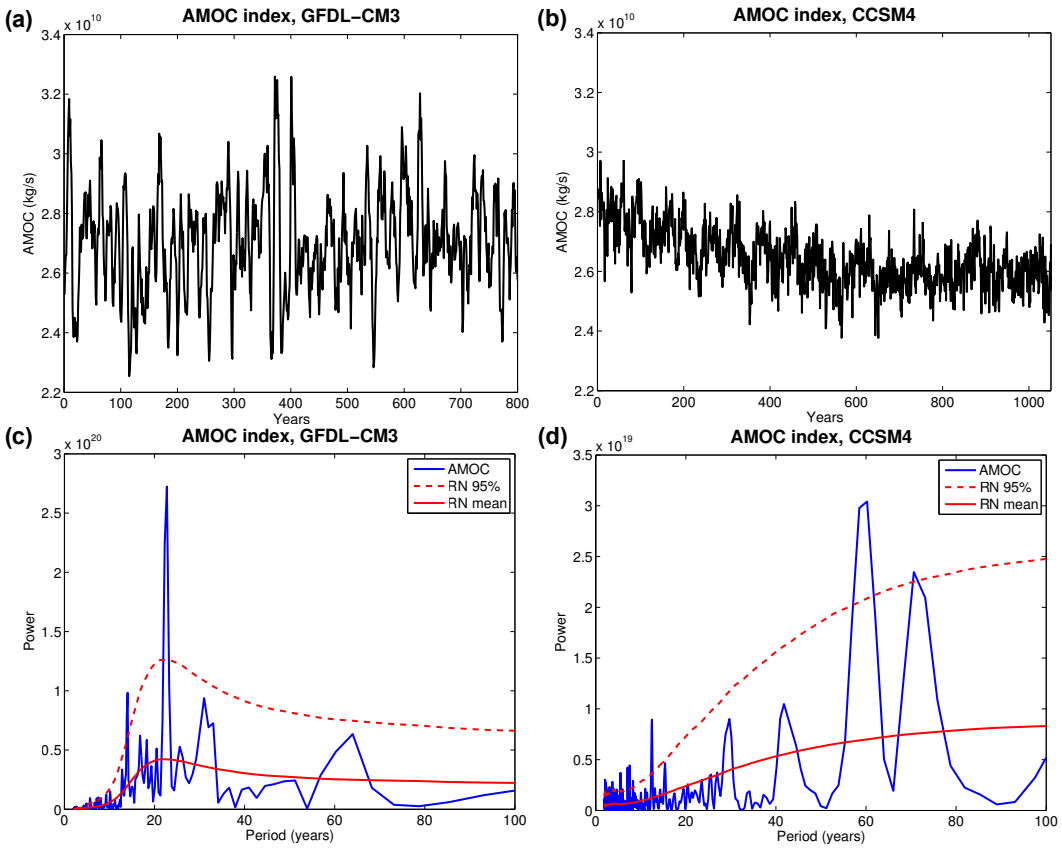

Figure 1.3: Characterization of the Atlantic meridional overturning circulation (AMOC) index variables used in this paper. Panels (a) and (b) show the AMOC index time series (defined in the text) for the pre-industrial control simulations of GFDL-CM3 and CCSM4, respectively. Panels (c) and (d) show their respective power spectra with a best-fit red noise (RN) spectrum (computed as in Schneider and Neumaier [2001]) and an estimated 95% confidence interval.

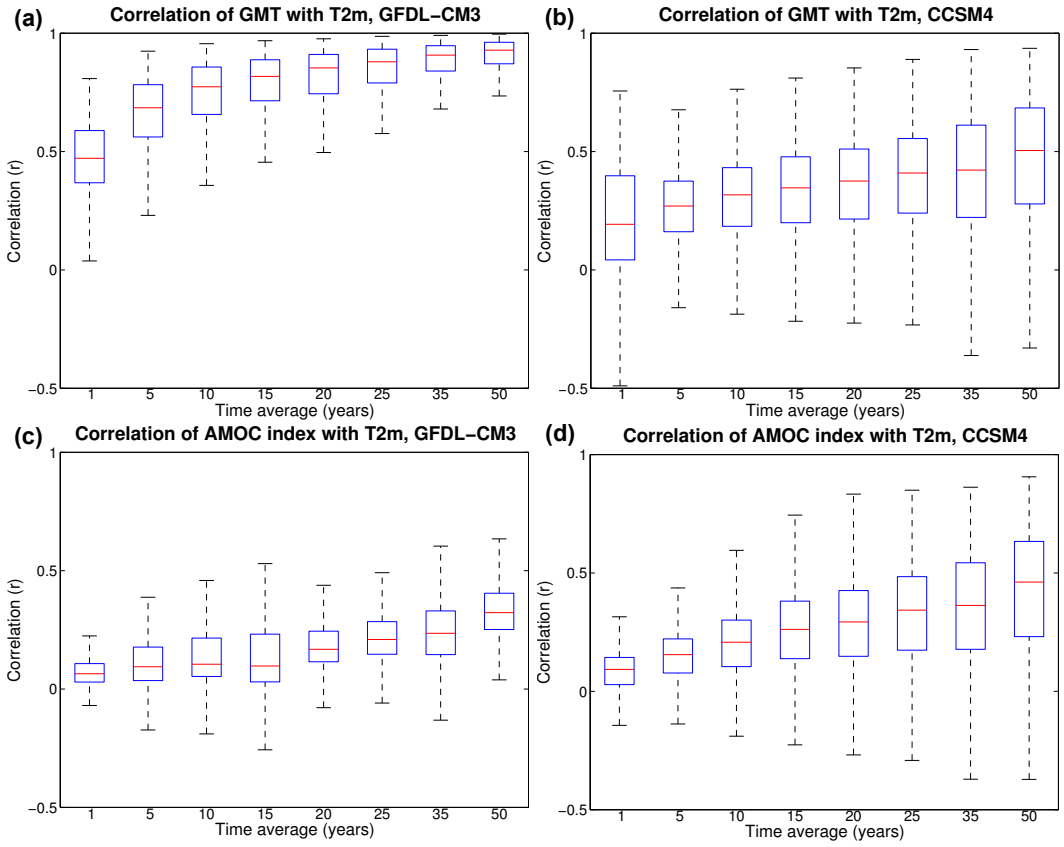

Figure 1.4: Correlation of global mean 2-m air temperature (GMT) with the spatial 2-m air surface temperatures (T2m), panels (a) and (b), and correlation of the Atlantic meridional overturning circulation (AMOC) index with the spatial 2-m air surface temperatures, panels (c) and (d), at a range of time scales. The correlations are computed for each spatial grid point at a given averaging time scale, with the spatial correlation information summarized with these box plots (outliers have been omitted for clarity).

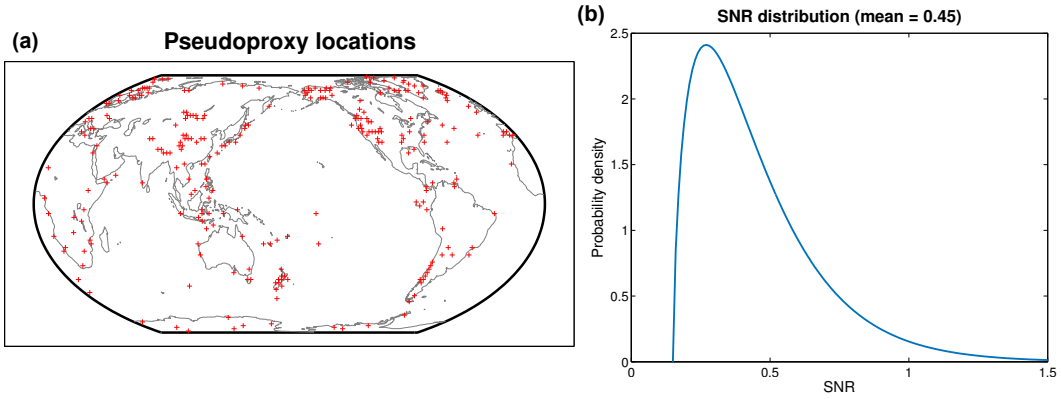

Figure 1.5: (a) Pseudoproxy locations used in this study (n = 274), drawn from the predominantly high resolution (annual) proxy collection of PAGES 2k Consortium [2013] and all the comparatively low resolution (decadal to centennial) proxy locations in Shakun et al. [2012] and Marcott et al. [2013]. (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.

1.4 Reconstruction results

The reconstructions of global mean temperature are shown in Fig. 1.6. The top panels of (a) and (b) show the reconstructions with the annual pseudoproxies, the middle panels show the reconstructions with the long time-scale proxies, and the bottom panels show the reconstructions for both time scales. Skill metrics, computed at annual resolution, are shown for each reconstruction: correlation (r) and coefficient of efficiency (ce), which for a data series comparison of length N is defined as

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

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 $-\infty < \text{ce} \leq 1$, where ce $= 1$ corresponds to a perfect match and ce < 0 indicates that the error variance is greater than the true time series variance. Comparing the reconstructions in Fig. 1.6 we see that the bottom panels with both time scales have the highest skill.

Note that the long-time scale reconstructions in the middle panels of Fig. 1.6(a) and (b) 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-time scale pseudoproxies are averages over a given 5 or 20 year period. As discussed in Sect. 1.2, this experimental design is only a single illustrative example of the general algorithm. The data from real proxies are not apportioned into specific time frames, but are scattered irregularly in time. Using many irregular proxies will act to smooth the long time-scale reconstructions. As long as the time scales can be estimated and an appropriate proxy system model is used, the algorithm of Sect. 1.2 can handle any real proxy data. While not dealt with explicitly here, real long time-scale (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 reconstructions many times and sampling from an ensemble of ages dates for a given proxy.

One assessment of skill as a function of time scale is to compute the cross spectra of the reconstructed time series with the true time series (Fig. 1.7). The cross spectra in this case reveals 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. 1.7(b) 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 time-scale reconstruction, 20+1, does better or just as well as the single time-scale reconstructions at both short and long periods. This same general result holds for Fig. 1.7(a), though it is more difficult to see because of the much larger power at longer periods in the GFDL-CM3 simulation.

Figure 1.8 shows three time scale reconstructions of the AMOC index for the two model simulations, similar to Fig. 1.6. 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 and ce). 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. Figure 1.9 shows the corresponding cross spectra for the reconstructions shown in Fig. 1.8. The most robust result from this figure 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. 1.4, where the correlation (covariance) information between the atmosphere and the ocean is particularly low at annual averages but improves at longer time

[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. 1.7 and 1.9 should not be expected to precisely match up with Figs. 1.2 and 1.3

averages.

We note that all the cross spectra of the reconstructions shown in Figs. 1.7 and 1.9 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 gray lines (not shown). Therefore the reduced power relative to the true state in Figs. 1.7 and 1.9 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. 1.6 and 1.8 with two time scales use twice as many pseudoproxies as the single time scale 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 time scale reconstructions used 200 randomly drawn pseudoproxies and the multi-scale reconstructions used 100 randomly drawn pseudoproxies each for the two time scales (the same as in the previous multi-scale reconstructions). Figure 1.10 is a characteristic example of the results of these additional tests. Figure 1.10(a) shows the reconstructions of the AMOC index with the CCSM4 model output and Fig. 1.10(b) shows the respective cross spectra. In (a), the skill is best for the multi-scale reconstructions and in (b) the cross spectra shows the same general result of improved low-frequency power for the time-averaged pseudoproxies. However, the cross spectra for the 20+1 reconstruction is not always closest to the true spectrum, suggesting that the number of pseudoproxies does play a role in improving the spectrum of the reconstructions. However, we note that the r and ce skill metrics for the single time scale reconstructions in Fig. 1.10(a) are hardly changed relative to those in Fig. 1.8(b) even though the number of observations are doubled.

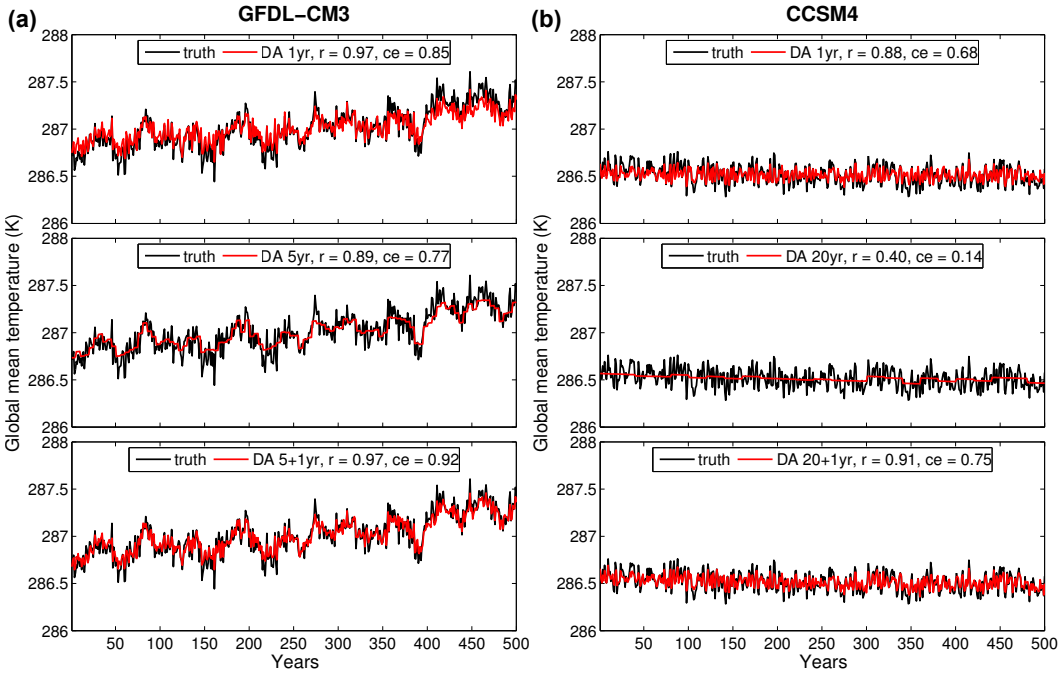

Figure 1.6: Global mean temperature reconstructions (mean of 100 Monte Carlo iterations) for the three types of experiments discussed in the text and for each climate model simulation. Black lines indicate the true time series while red lines indicate the reconstructed time series for only short time scale (annual) pseudoproxies, only long time scale (5 or 20 years) pseudoproxies, and both long and short time scale pseudoproxies. Skill metrics of the reconstructions, correlation (r) and coefficient of efficiency (ce), are shown at the top of each subpanel.

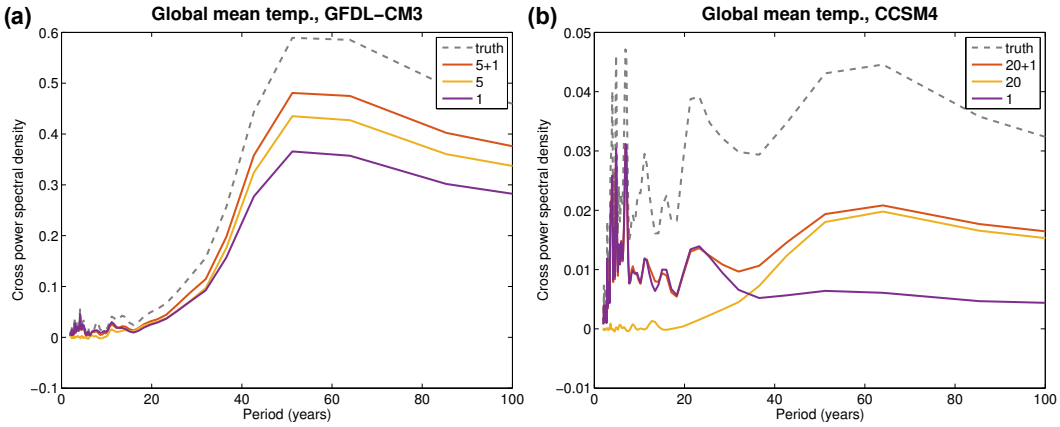

Figure 1.7: Cross spectra of the reconstructed global mean temperature time series with the true global mean temperature time series, for the reconstructions shown in Fig. 1.6. For reference, the dotted gray line indicates the cross spectra of the true time series with itself, or equivalently its own power spectrum.

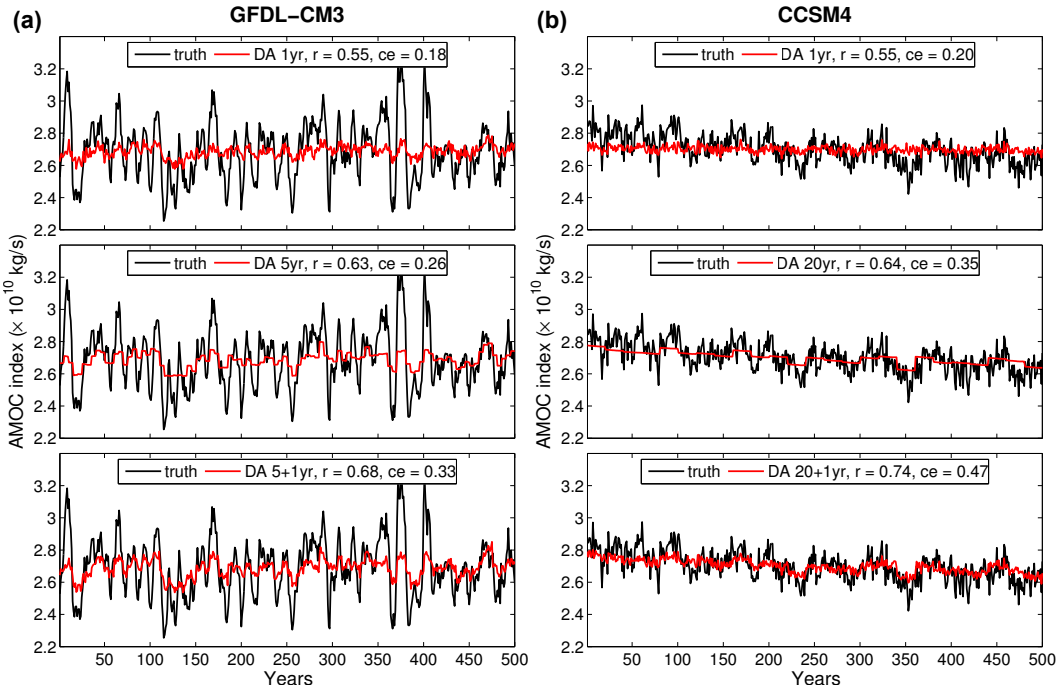

Figure 1.8: AMOC index reconstructions (mean of 100 Monte Carlo iterations) for the three types of experiments discussed in the text and for each climate model simulation. Black lines indicate the true time series while red lines indicate the reconstructed time series for only short time scale (annual) pseudoproxies, only long time scale (5 or 20 years) pseudoproxies, and both long and short time scale pseudoproxies. Skill metrics of the reconstructions, correlation (r) and coefficient of efficiency (ce), are shown at the top of each subpanel.

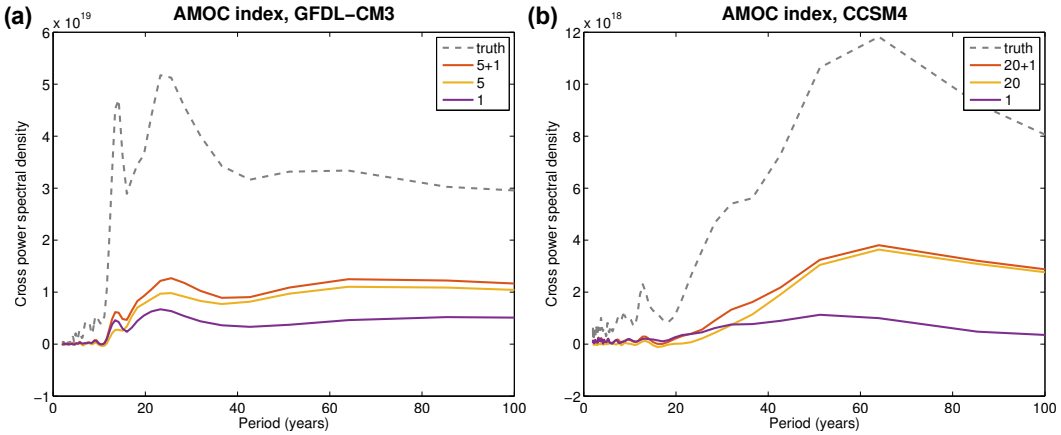

Figure 1.9: Cross spectra of the reconstructed AMOC index time series with the true AMOC index time series, for the reconstructions shown in Fig. 1.8. For reference, the dotted gray line indicates the cross spectra of the true time series with itself, or equivalently its own power spectrum.

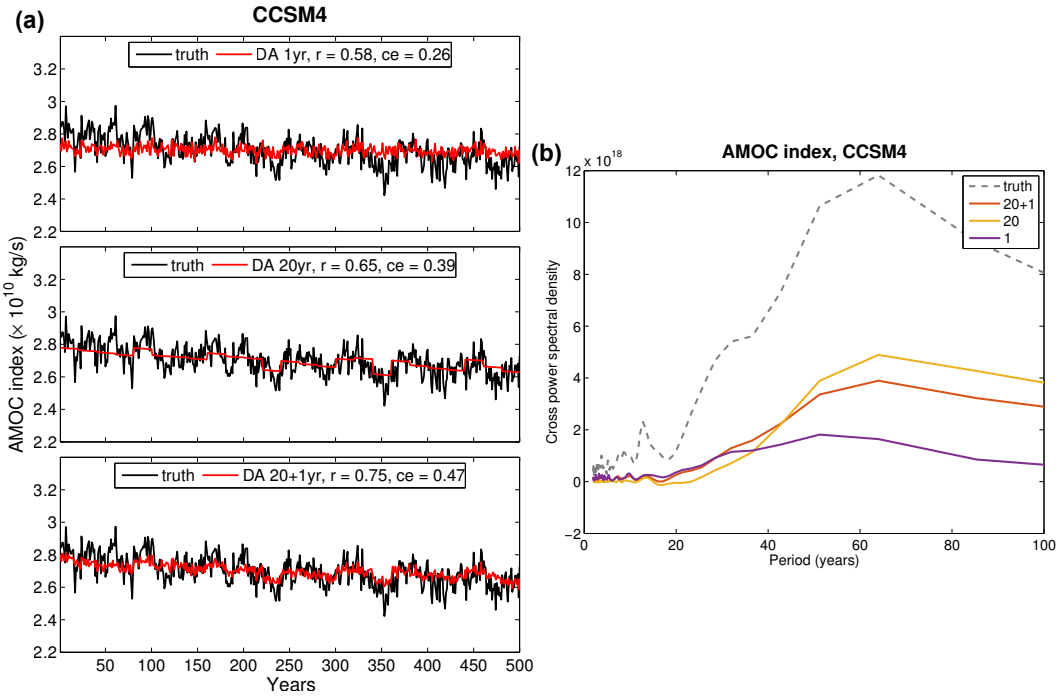

Figure 1.10: AMOC index reconstructions (mean of 100 Monte Carlo iterations) and corresponding cross-spectra similar to those shown in Figs. 1.8(b) and 1.9(b) but for the case where each experiment uses 200 pseudoproxies: the single time scale reconstructions use 200 pseudoproxies each, while the multi-time scale reconstructions use 100 pseudoproxies for the short time scale and 100 pseudoproxies for the long time scale.

1.5 Conclusions

This paper presents a data assimilation algorithm that can explicitly incorporate proxy data on arbitrary time scales. 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. This may be particularly useful given the many inherent time scales of the climate system, such as the fast time scales of the atmosphere and the slow time scales of the ocean. We performed three types of realistic atmosphere–ocean pseudoproxy reconstructions to assess the impact of using observations at multiple time scales: (1) short (annual) pseudoproxies only, (2) long ($\sim$ decadal) pseudoproxies only, and (3) both short and long time-averaged pseudoproxies. We found that for both global mean temperature and an index of the Atlantic meridional overturning circulation, the reconstructions that incorporated proxies across both short and long time scales were more skilful than reconstructions that used short or long time scales alone (Figs. 1.6 and 1.8). This result holds even when the number of pseudoproxies for the single time-scale reconstructions are doubled, Fig. 1.10(a). Multi-scale reconstructions would be expected to perform better than single-scale reconstructions because they can include information at multiple time scales and because the prior can be better conditioned as it's used from one time scale to the next.

We found that reconstructions incorporating long time scale pseudoproxies improve the low-frequency components of the reconstructions over reconstructions that only use annual time-scale pseudoproxies, Figs. 1.7, 1.9, and 1.10(b). 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. (1.2). 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 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. 1.3.2, we ensured that $\mathbf{R}$ from Eq. (1.2) remains fixed for both time scales.

These results indicate that data assimilation-based atmosphere–ocean state estimates may be improved by including proxies and climate states from multiple time scales. The general results outlined above appear to be insensitive to the choice of climate model simulation. These results also show, as suggested by Kurahashi-Nakamura et al. [2014], that given a representative prior ensemble, features of the Atlantic meridional overturning circulation may be reconstructed using observations of surface variables alone.

Chapter 2

WHAT CAN ICE CORES TELL US ABOUT PAST CLIMATE?

2.1 Introduction

Water isotope data from ice cores, particularly $\delta^{18}O$, has long been used as a paleoclimate proxy, particularly for local temperature in the mid to high-latitudes and for local precipitation in the tropics [Dansgaard, 1964, Jouzel et al., 1997]. In the past decade or so isotope-enabled climate models have allowed for a richer understanding of the climate processes that produce the isotopic signals in precipitation [e.g., Schmidt et al., 2007]. Such modeling-based studies highlight the complication of simple temperature-isotope interpretations by pointing to the many non-local influences on isotopic signals at coring locations. Recent observational studies have also linked ice cores to non-local patterns of climate variability, particularly to mid-latitude atmospheric circulation patterns and to variations in tropical climate [e.g., Schneider and Noone, 2007, Schneider and Steig, 2008]. However, the full spatial and temporal extent to which ice cores can robustly inform past climate has yet to be fully explored.

Here we estimate a realistic upper-bound on what ice cores can tell us about climate at annual and decadal time scales in a paleoclimate reconstruction context. We employ a data assimilation-based (DA) reconstruction technique that optimally combines proxy information with the dynamical constraints of climate models [Steiger et al., 2014]. Most previous ice-core-based reconstructions have been limited to converting single isotope records into temperature through some estimated linear relationship [e.g., Jouzel et al., 1997, Alley et al., 1997] or have averaged a collection of standardized records to arrive at a regional reconstruction [e.g., Steig et al., 2013]. For this study we produce spatial field reconstructions using both an isotope-enabled climate model and explicit modelling of the compaction and

diffusion process leading to ice core formation [Herron and Langway Jr, 1980, Johnsen et al., 2000]. To our knowledge, this represents the first time such models and processes have both been included for any spatial ice core-based pseudo or real proxy reconstruction. Through a series of pseudo and real proxy experiments we assess the spatial and temporal extent to which ice cores can be used to reconstruct the key variables of surface temperature, geopotential height at 500 hPa, and precipitation. After presenting the results of both sets of experiments, we discuss the reasons for the fundamental differences between reconstruction skill of different variables and the key differences between the pseudoproxy and real proxy reconstruction results.

2.2 Pseudoproxy experiments

2.2.1 Experimental framework

The climate model simulations employed in these experiments are the isotope-enabled atmospheric models ECHAM5-wiso [Werner et al., 2011] and iCESM (in development, Noone and Nusbaumer). We ran ECHAM5-wiso at T106 ($\sim$1 degree resolution) using historical sea ice and sea surface temperatures [Rayner et al., 2003] as boundary conditions for the years 1871 through 2011. We also obtained a simulation of iCESM that was run at $\sim$2 degree resolution and was also forced with historical sea surface temperatures for the years 1850–2014 (though only the years of overlap with the ECHAM5-wiso simulation were used). These models provided the annual states for our prior ensembles as well as the climate and isotopic fields for the pseudoproxies (see below). We chose to reconstruct 2 m air temperature, geopotential height at 500 hPa, and precipitation, at the native model resolutions. The reconstructions were performed over the years 1871–2011 of the simulations by reconstructing a decade at a time while using the remaining annual states for the prior ensemble (e.g., the years 1880–1889 were reconstructed using the years 1871–1879 and 1890–2011 as the prior ensemble). The same offline approach as used in Chapter 1 (without consideration of multiple time scales within the data assimilation process) is employed here. Each reconstruction was repeated

100 times in a Monte Carlo fashion by sampling 75% of the pseudoproxy network for each reconstruction iteration and uncertainty estimates were derived from both these iterations and the full analysis/posterior ensembles.

Pseudoproxy experiments, usually designed to test different methodological approaches to climate reconstruction, generally employ simple pseudoproxies. These are constructed by first choosing a climate model to provide the climate fields, then a realistic pseudoproxy network is chosen that mimics real world spatial proxy availability. Using this network, temperature time series are drawn from the model and realistic amounts of noise (usually white noise) are added to create the pseudoproxies. In contrast to this approach, here we employ proxy system models (PSMs)—also called process models or "forward" models—to create our pseudoproxies (see Evans et al. [2013] for a review). PSMs are physically-based models that transform the simulated climate signal (e.g., temperature, precipitation) to a synthetic proxy time series. Here, the pseudoproxy ice cores were modeled using the water isotope fields of ECHAM5-wiso together with the PSM discussed in [Dee et al., 2015a]. The pseudoproxy network for these experiments is the "NCDC" network shown in Figure 2.1, which is a network based on all the locations of ice core proxies in the National Climatic Data Center database (not necessarily just measurements of δ^{18}O); this large ice core network was chosen as a best-case network coverage scenario.

The ice core PSM includes three sub-models, each of which mimics a separate modification of the original input signal as it would occur in nature: a *sensor* model, which describes the physical and geochemical processes that respond to the climate signal, an *archive model*, which accounts for any processes that affect the encoding of the signal in the proxy medium, and an *observation* model, which involves proxy measurements and accounts for dating uncertainties and analytical errors in the final measurement made on the ice core data. Following Dee et al. [2015a], the sensor model calculates precipitation-weighted $\delta^{18}O_p$ and corrects for temperature and altitude bias of the modeled output; the archive model estimates compaction and diffusion down core by convoluting the original isotope signal with a Gaussian kernel characterized by a diffusion length σ (see Dee et al. [2015a] for mathematical details);

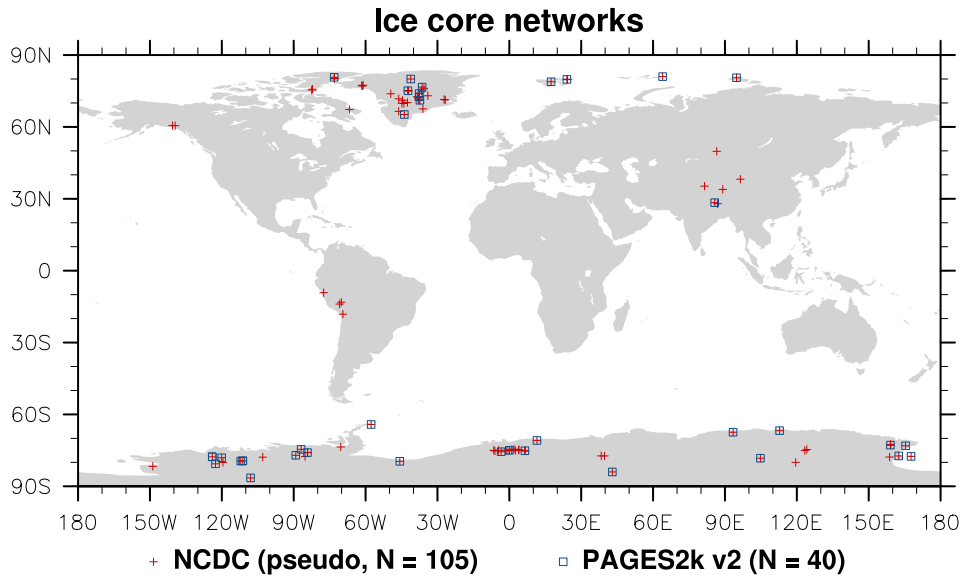

Figure 2.1: Pseudo and real proxy networks used in this study, with the number of proxy sites indicated for each network. Note that some site locations are shown here with approximate coordinates following what is entered in the proxy databases and also note that many proxy sites are very near one another or are co-located.

and the observation model accounts for analytical uncertainty by sampling an age model. We note that for simplicity in this study we do not account for age uncertainties, as would normally be included in the observation model, which would only act to degrade the skill of the reconstructed variables. In the data assimilation process, the prior estimates of the observations, $\mathcal{H}(\mathbf{x}_b)$, are the PSM-derived pseudoproxies for the years of the prior ensemble; this is a "perfect model" framework that would be equivalent in a real reconstruction to knowing precisely the physical processes that led to the formation of all the ice cores.

As a test of the robustness of the pseudoproxy results, we also perform two end-member reconstructions where we adjust the diffusion length, σ, in the ice core PSM. Diffusion in the firn can blur well-defined layering of isotope ratios, limiting the reconstructability of

annual temperature at a given site [Johnsen, 1977, Cuffey and Steig, 1998, Johnsen et al., 2000, Küttel et al., 2012]. Diffusion is principally controlled by average site temperature and accumulation rate, though its actual calculation involves some empirical approximations and simplifying assumptions. We therefore wanted to approximate this uncertainty by manually setting σ in two end member experiments and comparing the results with an unmodified control σ. For the low-diffusion experiment we set σ to 1/2 its control value and for the high-diffusion experiment we set σ to 2 times its control value. The sensitivity of the results to modifying the diffusion length are discussed in the following section and shown in the appendix.

Figures 2.2–2.4 provide a general verification of the ECHAM5-wiso and iCESM simulations against observations. Figure 2.2 compares the annual mean δ^{18}O of precipitation in a global observational dataset, the global network of isotopes in precipitation (GNIP), with the simulated δ^{18}O of precipitation. From this figure we see that both model simulations are in agreement with the observations in both their spatial patterns and magnitudes. An additional first-order test of the simulations is the modeled relationship between the mean δ^{18}O of precipitation and mean temperature at ice core locations; as shown in Jouzel et al. [1997], for example, this relationship is known from observations to be approximately linear. Figure 2.3 shows that for both model simulations this indeed holds. The remaining key component of the ice cores is the accumulation rates of precipitation at site locations. Figure 2.4 shows the mean accumulation of the model simulations together with the ERA-Interim reanalysis product. We note that both models accurately estimate low, moderate, and high accumulation sites, with the absolute value of the relative errors for most locations less than 1. However, from the bottom panel of Fig. 2.4 we see that there is a general tendency to have too much accumulation at ice core sites. Figures A.1 and A.2 show the spatial distribution of these relative errors. Considering both models, there are not many spatially-coherent patterns of bias, with the exception of ECHAM5-wiso generally simulating too much accumulation across East Antarctica.

2.2.2 Reconstruction results

Since we seek to understand both the spatial and temporal limits to what ice cores can tell us about past climate, we will look first to regional mean temperatures in Greenland and the British Isles (Britain and Ireland), followed by global mean temperature, and lastly the spatial skill of temperature, geopotential height at 500 hPa (Z500), and total precipitation; for each of these we will also consider how skill varies at longer time scales. The British Isles were chosen as a reconstruction target because of their proximity to Greenland ice cores and the a priori expectation that the British Isles and Europe would be climatically connected, via the North Atlantic oscillation for instance.

Figure 2.5 shows the reconstructed Greenland mean temperature and the true Greenland mean temperature, with uncertainty estimates based on the "grand ensemble" of both the analysis/posterior ensembles and the Monte Carlo iterations. Skill for the reconstruction mean is shown at the bottom of the left figure panel: correlation (r), coefficient of efficiency (CE), and both of these after the true and reconstructed time series have been detrended. Additionally, the figures indicate the continuous ranked probability skill score (CRPSS) of the ensemble of the reconstructions; the continuous ranked probability score (CRPS) is a "strictly proper" scoring rule that accounts for the skill of the entire reconstruction distribution (see Gneiting and Raftery [2007] for details). The skill score version, CRPSS, is the reconstructed CRPS computed with respect to the CRPS of a reference distribution, $\mathrm{CRPSS} = 1 - \mathrm{CRPS}_{\mathrm{rec}}/\mathrm{CRPS}_{\mathrm{ref}}$, which for this paper is the initial, uninformed prior; positive CRPSS indicates that the reconstructed distribution is more skilful than the uninformed prior. The right panel of Fig. 2.5 shows the coherence of the reconstruction mean with the true time series. Coherence is similar to correlation as a function of frequency (or period as shown here), and in this figure it shows that at both annual and longer time scales, the skill of the reconstruction remains relatively constant. Figure 2.6 shows the mean temperature reconstruction for the British Isles along with its skill and coherence. Unlike the Greenland mean temperature reconstructions in Fig. 2.5, this regional mean reconstruction shows no

skill, even at longer time scales. Figure 2.7 shows the same information but for the global mean temperature. In Fig. 2.7 note that by comparing the regular skill scores with the skill of the detrended reconstruction, we can see that nearly all of the annual r and CE skill comes from the trend. The coherence is consistent with noise except possibly at multi-decadal time scales, though because this reconstruction is only 141 years in length, this result should be interpreted with caution; in an additional 500 year reconstruction experiment we performed using a last-millennium simulation of the model SPEEDY-IER [Dee et al., 2015b] (forced by CCSM4 last-millenium sea surface temperatures), we do not see any long time-scale coherence for the global mean (not shown).

Next let us consider the spatial skill of the reconstructions. Figure 2.8 shows the skill of the 2 m air temperature reconstructions for each of the metrics, which are computed either with the mean (r and CE) or the distribution (CRPSS) of the grid point temperature series over the entire reconstruction. Similar to the regional mean temperature reconstructions in Figs. 2.5 and 2.6, there is clear skill nearby the pseudo ice cores, especially in Greenland and Antarctica, but not consistently in other locations. The possible exceptions to this include the Southern ocean off the coast of East Antarctica and the central Pacific. If r and CE are computed at 10 year averages (Fig. A.3) the ice cores tend not to add any new areas of high skill, but simply amplify skill or lack thereof in a given location. Corresponding results for Z500 shown in Fig. 2.9 have similar features, with highest skill nearest the proxies. At ten year averages (Fig. A.4) the tropical skill does tend to increase, however, the variability in Z500 in the tropics is low, especially at 10 year averages. Figure 2.10 shows the spatial skill for precipitation. Unlike temperature and Z500, only a few locations in coastal Antarctica have skill, even at 10 year averages (Fig. A.5).

The fundamental differences between the precipitation results and both temperature and Z500 can be explained by the fact that the covariance length scales for precipitation are much reduced compared to those for temperature and Z500. Figure 2.11 shows the correlation as a function of distance between the PSM-derived pseudo ice cores and the reconstructed variables. (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.) Recalling the discussion of the DA equations in Chapter 1, we note that larger covariance values between the prior and the prior-estimated observations end up weighting the innovation more heavily; thus larger covariances allow the information in the observations (or proxies) to be spread more widely. Evidently, an ice core at one location on an annual time scale tells one very little about precipitation at other spatial locations.

Because the amount of diffusion within the firn has the potential to significantly impact the isotopic signal within our modeled ice cores, we also performed two end-member diffusion experiments and compare these with the control or normal diffusion experiments. The results seen in surface temperature reconstructions are typical: Figs. A.6 and A.7 show the $1/2\sigma$ and 2σ versions of the ECHAM5-wiso reconstructions and Figs. A.8–A.10 show the control, $1/2$ and 2σ versions of the iCESM reconstructions. As one moves from low to normal to high diffusion, the skill tends to decrease locally and slightly increase non-locally. This is due to the fact that the smoothing effect of diffusion tends to remove some of the local "noise" of the precipitation-based isotopic signal. This smoothed data then covaries more strongly with non-local climate signals, Fig. A.11. With these longer covariance length scales the ice cores get a slightly increased capacity to inform non-local temperatures points.

One outstanding question related to these experiments is whether or not the primarily-local skill is driven by climate statistics at ice core sites or by the ice cores themselves: in the process of formation, do ice cores degrade much of their climate information? Or would even perfect observations of climate at the ice core locations only be able to reconstruct local climate? To test this question in the pseudoproxy framework, we generated temperature pseudoproxies at the ice core locations with a signal-to-noise ratio of 5, which would be characteristic of thermometer measurements (see Chapter 1 for discussion of signal-to-noise ratios), and then we performed reconstructions with ECHAM5-wiso leaving the rest of the reconstruction process the same. Figure 2.12 shows the global mean temperature reconstruction (analogous to Fig. 2.7) and Figs. 2.13–2.15 show the spatial skill of the reconstructions for temperature, Z500, and precipitation. These figures show that given sufficiently accurate

observations, one could reconstruct global mean temperature and even some tropical climate and local precipitation with high skill, though clearly local skill would be even more robust in this instance.

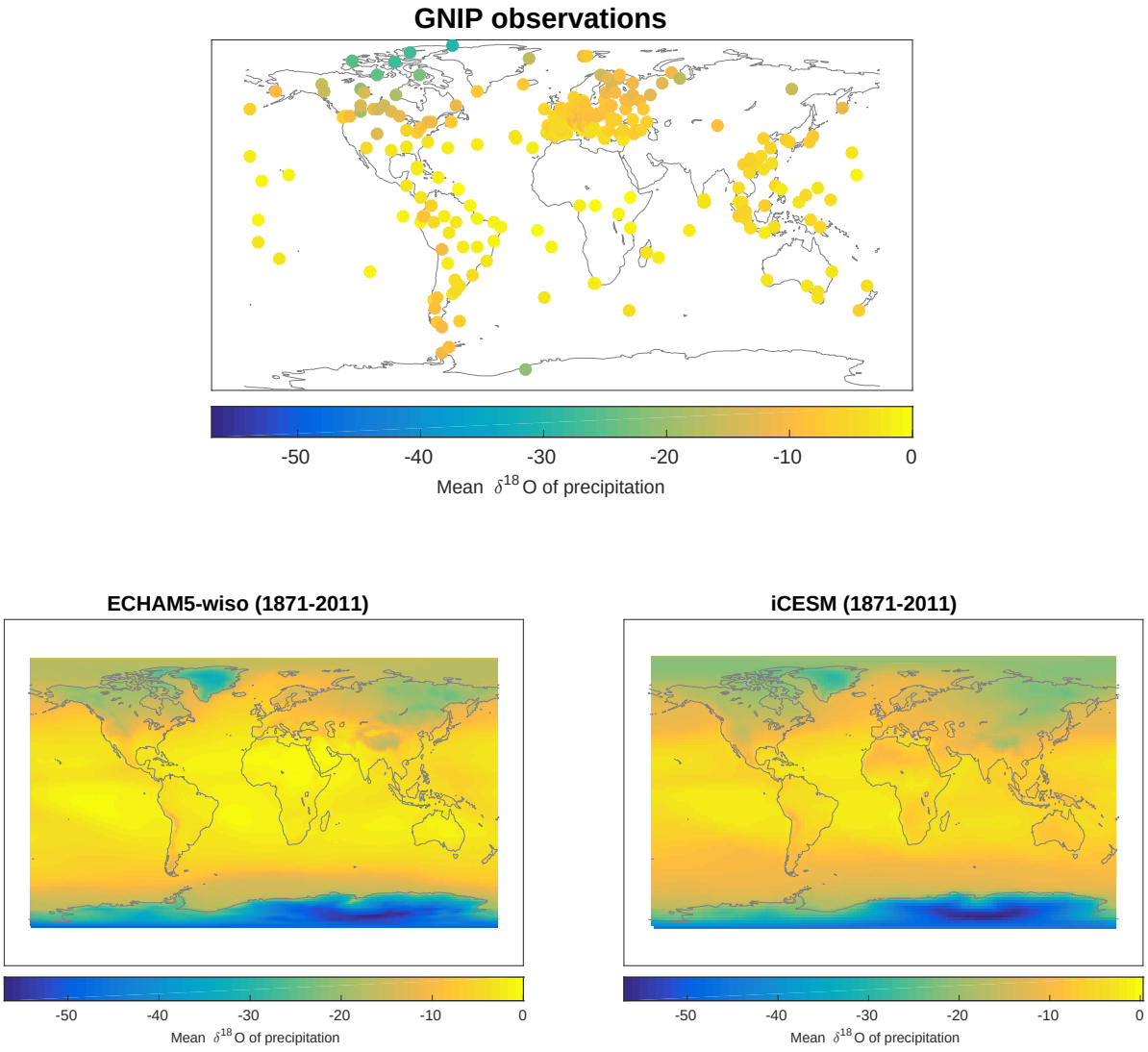

Figure 2.2: Global network of isotopes in precipitation (GNIP) and modeled δ^{18}O of precipitation in the two models employed in this study.

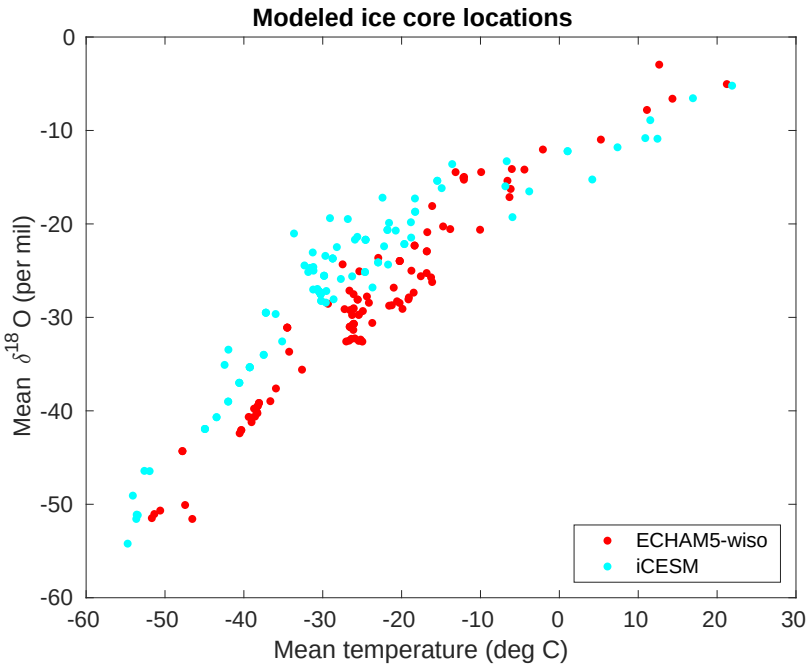

Figure 2.3: The relationship between annual mean δ^{18}O of precipitation and annual mean temperature at all ice core locations for both climate models.

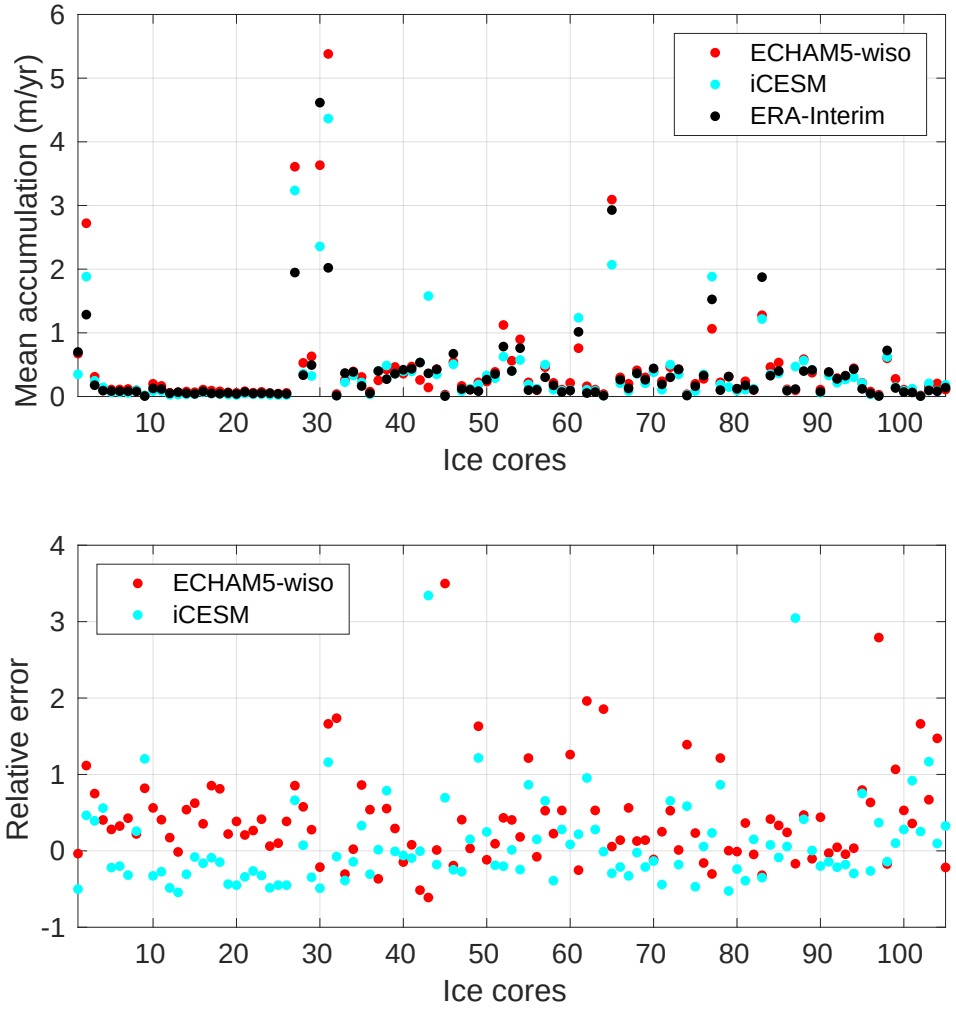

Figure 2.4: **Top:** Modeled mean annual accumulation of total precipitation at all ice core locations for the two models used in this study compared against the ERA-Interim renalaysis. **Bottom:** Relative error for ECHAM5-wiso and iCESM simulations, calculated against ERA-Interim.

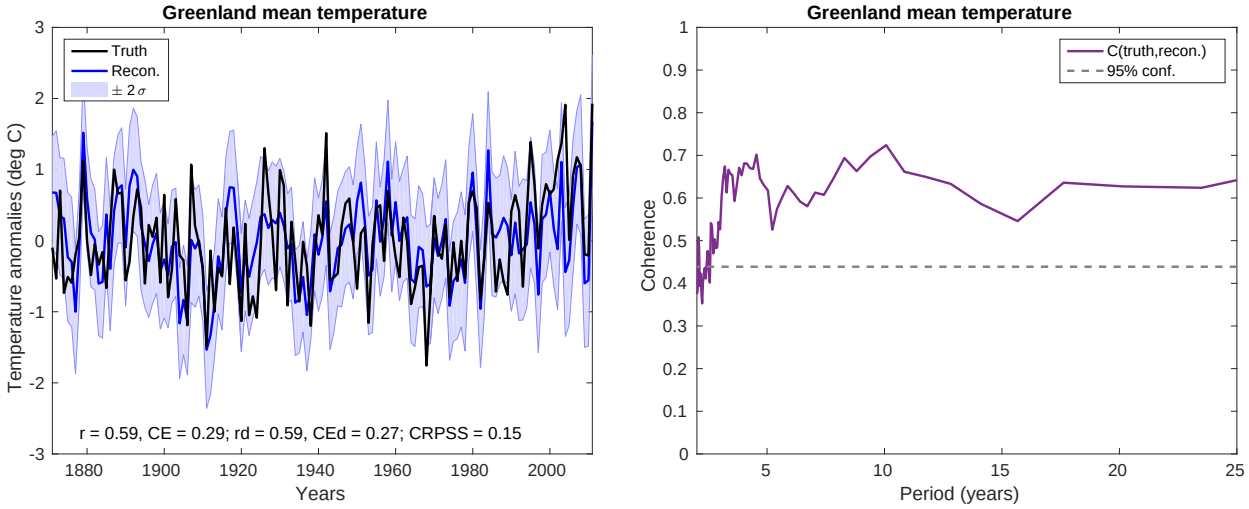

Figure 2.5: **Left:** Greenland mean temperature reconstruction. Skill metrics are indicated on the lower left as correlation (r), coefficient of efficiency (CE), and both of these again after the true and reconstructed time series have been detrended (indicated by "d"); additionally the continuous ranked probability skill score (CRPSS) assesses both the mean and uncertainty relative to the uninformed prior baseline. **Right:** Coherence of the reconstruction mean with the true times series, along with a 95% confidence level.

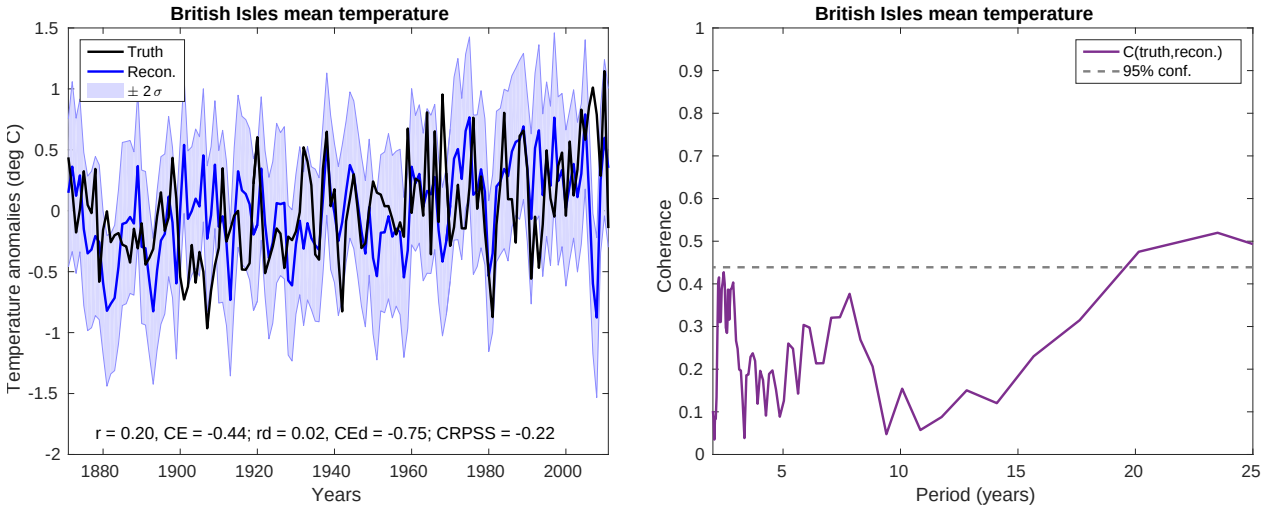

Figure 2.6: **Left:** British Isles mean temperature reconstruction (cf. Fig. 2.5). **Right:** Coherence of the reconstruction mean with the true times series.

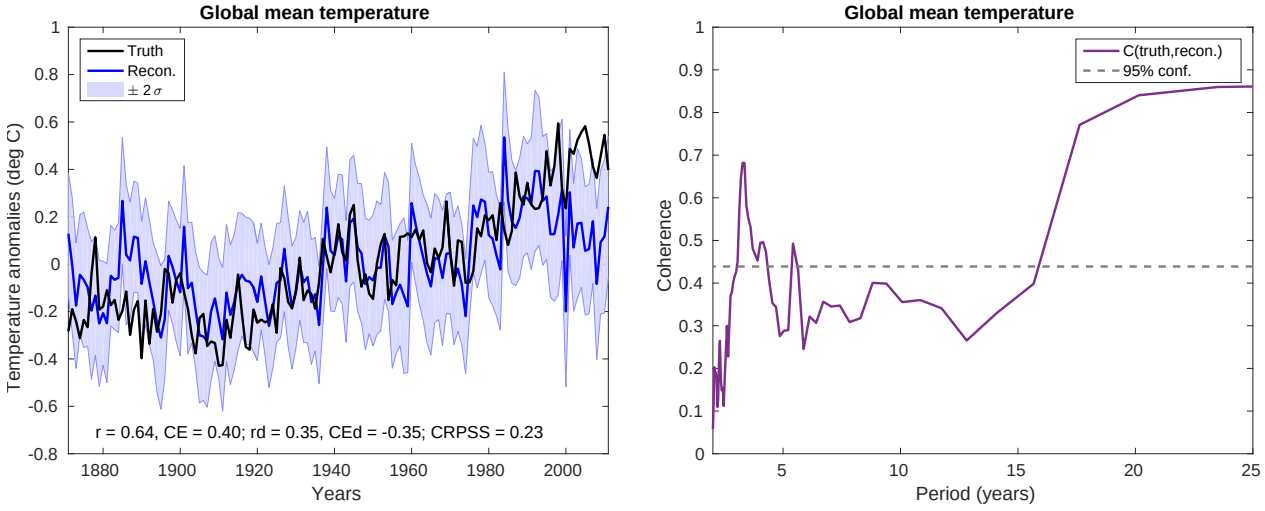

Figure 2.7: **Left:** Global mean temperature reconstruction (cf. Fig. 2.5). **Right:** Coherence of the reconstruction mean with the true times series.

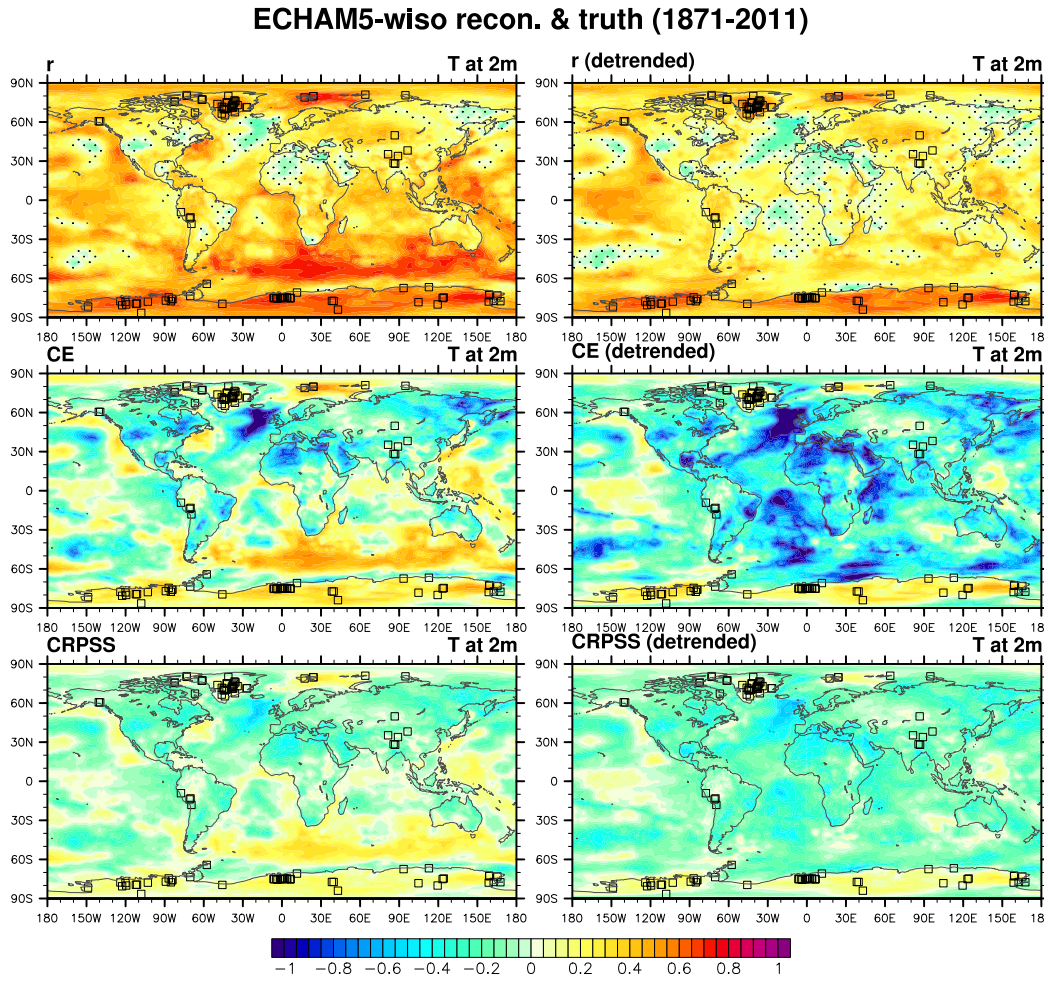

Figure 2.8: Spatial 2 m air temperature reconstruction skill with correlation (r) on top, coefficient of efficiency (CE) in the middle, and continuous ranked probability skill score (CRPSS) on bottom; both CE and CRPSS are only shown down to -1. The left column shows the raw reconstruction skill while the right column skill metrics are computed after detrending the underlying true and reconstruction data. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the pseudo ice cores are indicated by the open squares.

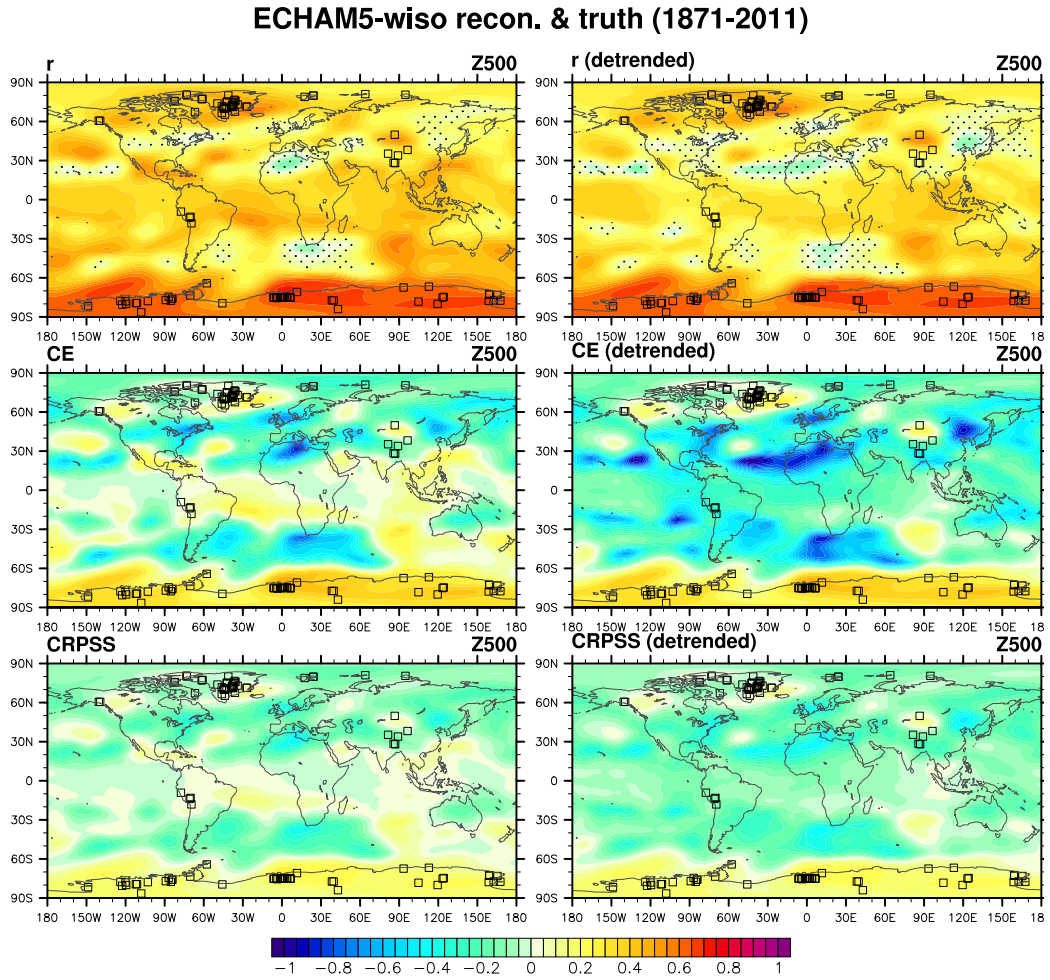

Figure 2.9: Geopotential height at 500 hPa (Z500) reconstruction skill (cf. Fig. 2.8).

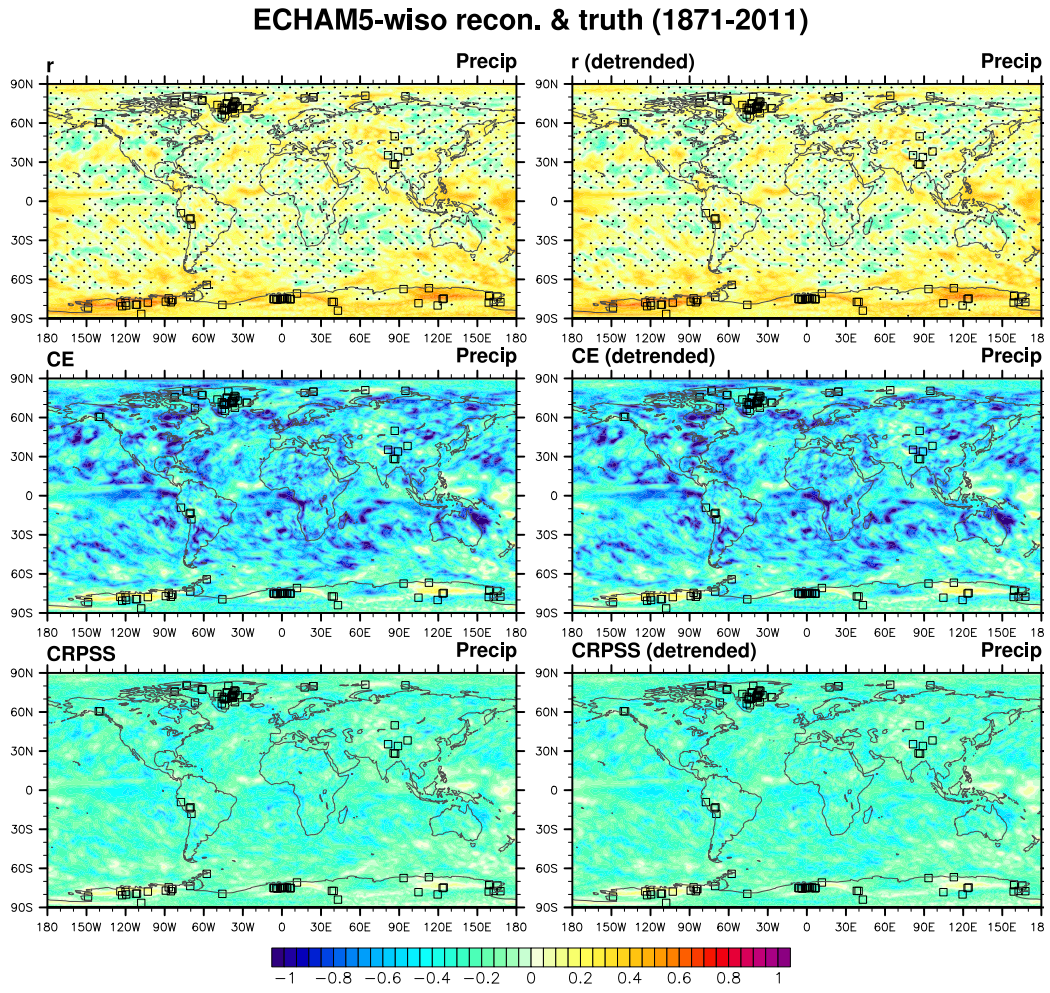

Figure 2.10: Precipitation reconstruction skill (cf. Fig. 2.8).

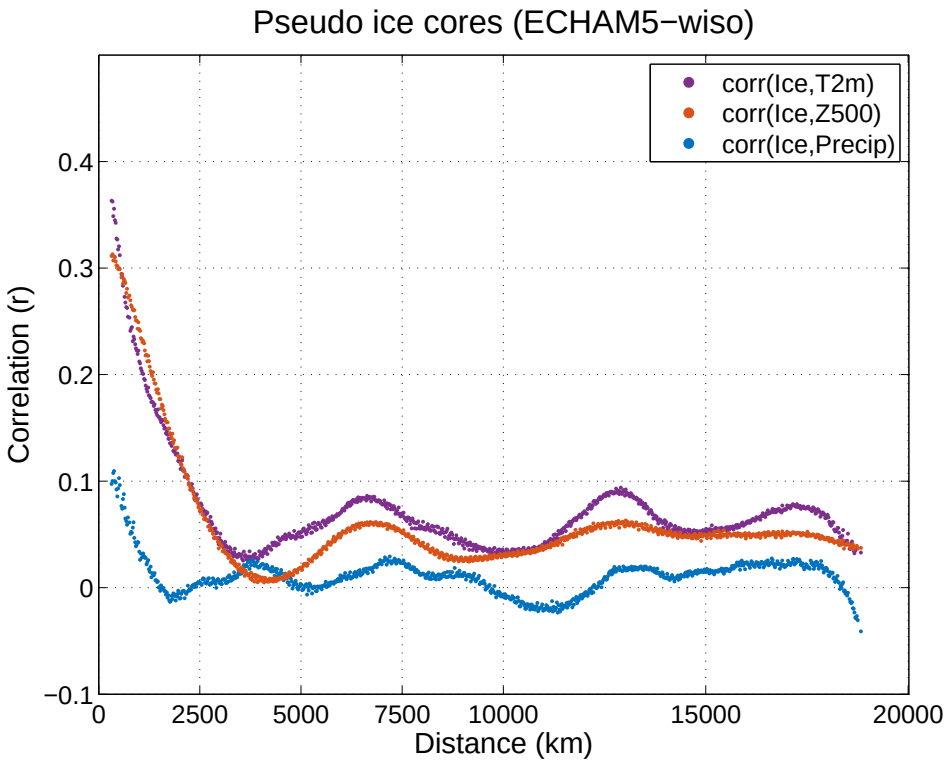

Figure 2.11: Correlation as a function of distance between the PSM-derived pseudo ice cores and the spatial reconstructed variables, averaged over all ice core locations.

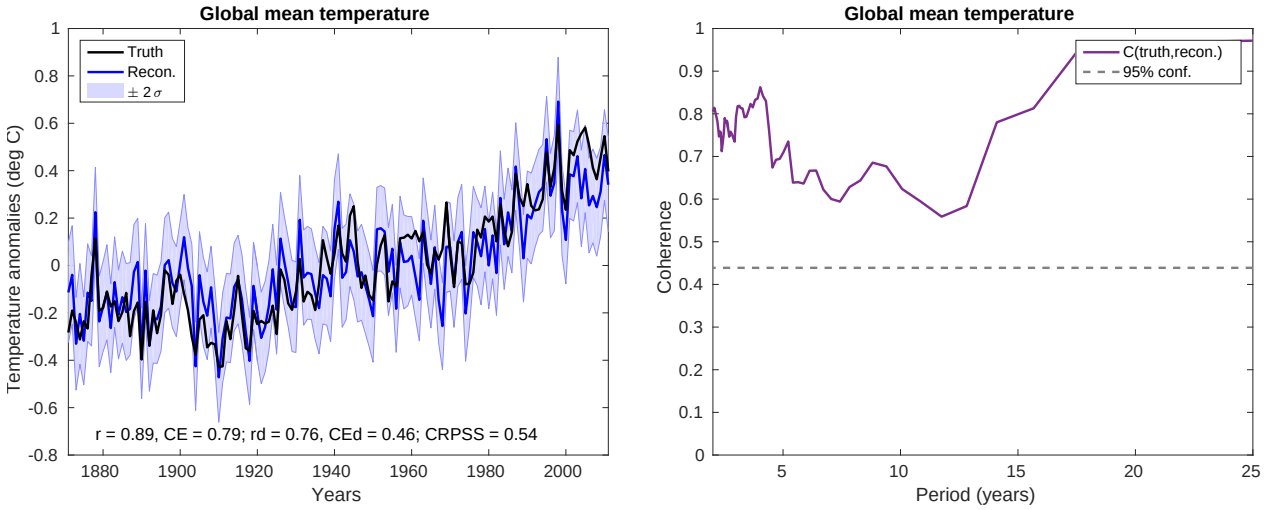

Figure 2.12: **Left:** Global mean temperature reconstruction for purely temperature pseudo-proxies at all ice core locations. Skill metrics are indicated on the lower left as correlation (r), coefficient of efficiency (CE), and both of these again after the true and reconstructed time series have been detrended (indicated by "d"); additionally the continuous ranked probability skill score (CRPSS) assesses both the mean and uncertainty relative to the uninformed prior baseline. **Right:** Coherence of the reconstruction mean with the true times series, along with a 95% confidence level.

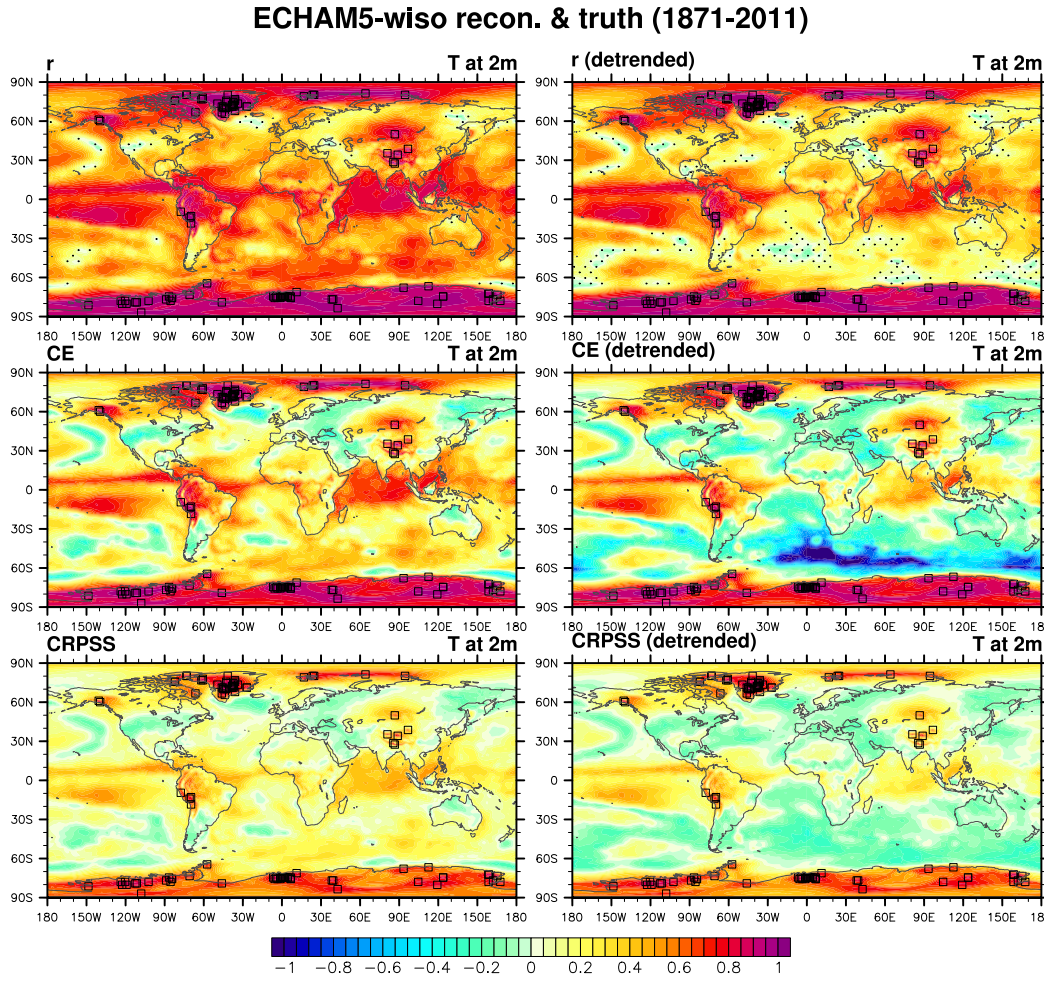

Figure 2.13: Spatial 2 m air temperature reconstruction skill for purely temperature pseudo-proxies at all ice core locations. The metrics are correlation (r) on top, coefficient of efficiency (CE) in the middle, and continuous ranked probability skill score (CRPSS) on bottom; both CE and CRPSS are only shown down to -1. The left column shows the raw reconstruction skill while the right column skill metrics are computed after detrending the underlying true and reconstruction data. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the temperature pseudoproxies are indicated by the open squares.

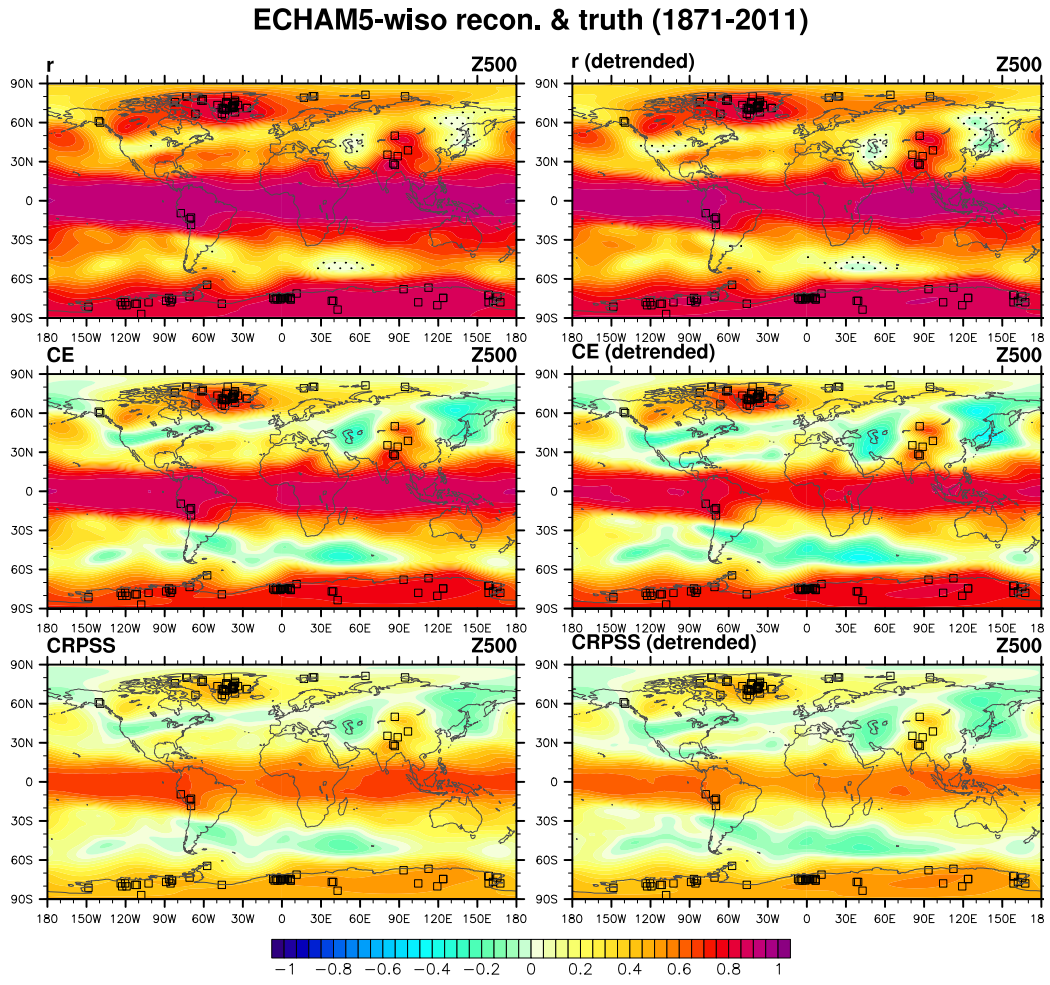

Figure 2.14: Geopotential height at 500 hPa (Z500) reconstruction skill for purely temperature pseudoproxies at all ice core locations (cf. Fig. 2.13).

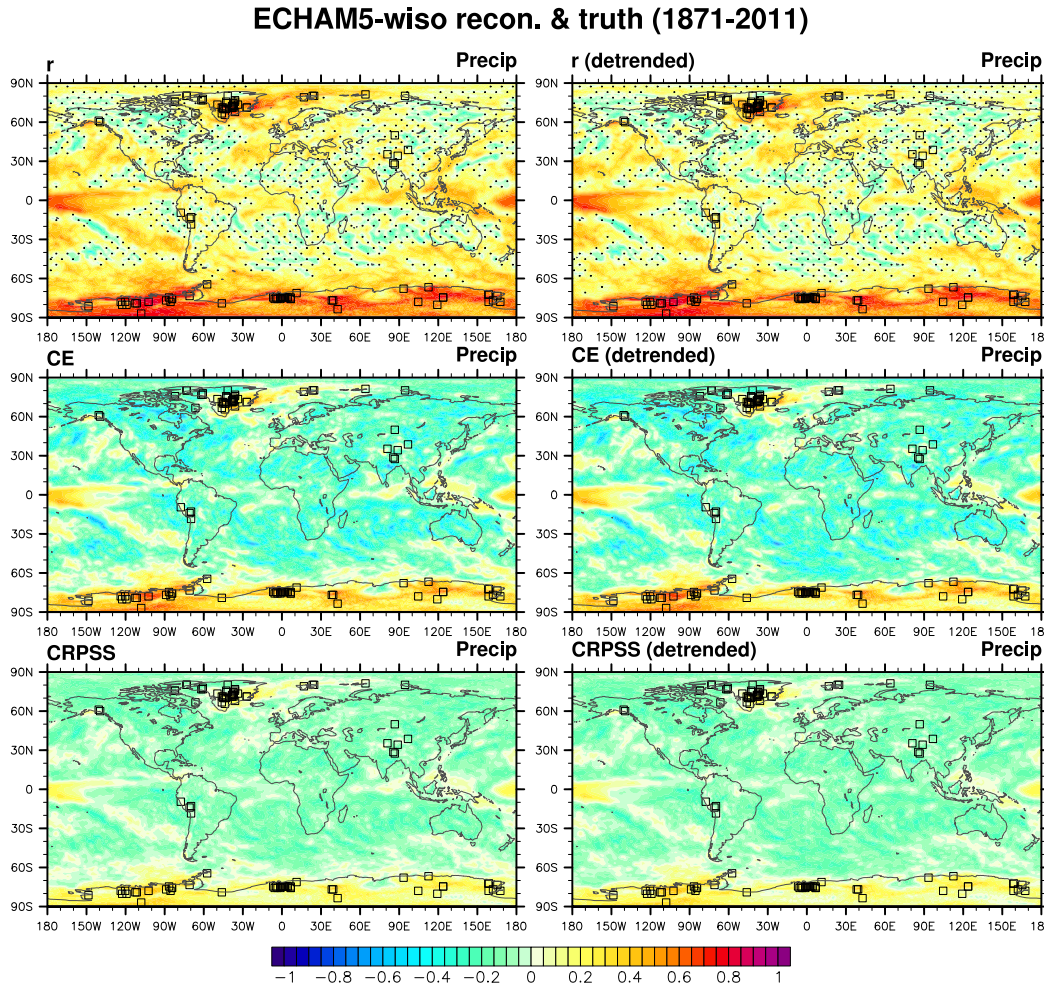

Figure 2.15: Precipitation reconstruction skill for purely temperature pseudoproxies at all ice core locations (cf. Fig. 2.13).

2.3 Real proxy experiments

2.3.1 Experimental framework

For the real reconstructions we employ a slightly different reconstruction framework, though the DA method remains the same. For the prior ensemble, we employ multiple climate model simulations intended to represent a variety of models and model simulations. Specifically, we use the same ECHAM5-wiso simulation as in the pseudoproxy reconstructions, along with the last millennium or "past 1000" simulations of the CCSM4 and MPI-ESM-P global climate models that were part of the coupled model intercomparison project phase 5 (CMIP5) experiments; these models were driven by similar historical climate forcings (such as volcanic aerosols, land use changes, and solar variations) from the period 850–1850 C.E. Additionally, we used the "historical" simulation, forced by observed atmospheric composition changes, of the GFDL-CM3 global climate model. Thus we employed two last millennium simulations and two historical simulations, though the forcings for the ECHAM5-wiso simulation were not included explicitly, but were implicit somewhat in using observed sea surface temperatures as boundary conditions.

The PSM employed for these real reconstructions also differs from that presented in the pseudoproxy section. Instead of a physically-based PSM, we employ a simple linear statistical model to estimate the proxies from the prior ensemble. This statistical model is derived from a linear regression between a given annual proxy time series, p_i, and a given local instrumental annual mean temperature series, T_i, as

$$p_i = \alpha_i + \beta_i T_i + \epsilon_i \,, \tag{2.1}$$

which is calculated over a period of temporal overlap between the two. The prior estimate of the proxies, $H(\mathbf{x}_b) \equiv \mathbf{y}_e$, is then found for each proxy by using the calibrated parameters α_i and β_i in

$$y_{ei} = \alpha_i + \beta_i \mathcal{T}_i \,, \tag{2.2}$$

where $\mathcal{T}_i$ are the grid point temperature values nearest to the proxy location in each prior

ensemble member. The residuals for each proxy, ϵ_i, are then used as the mean squared error in the DA update equations via $\overline{\epsilon_i^2} = R_i$. Note that this statistical model only considers local temperature and relies solely on the DA covariance relationships to inform non-local climate variables, rather than embedding some of that non-local information within the PSM as we did previously with the pseudo ice cores.

As part of the Monte Carlo reconstruction process, we performed a total of 200 reconstructions over the period 1900–2000. For temperature calibration data sets, we used HadCRUT4.3 [Morice et al., 2012] and Berkeley Earth [Rohde et al., 2013]. Eight sets of 25 reconstructions were performed with a unique model–calibration-data pair, randomly sampling 75% of the proxy network for each iteration, as with the pseudoproxy experiments. The 40 annual proxies for these reconstructions are from the latest version of the PAGES2k database (in prep.) and include mostly δ^{18}O measurements but also a few other isotopic variables, with the locations indicated by "PAGES2k v2" in Fig. 2.1. Because the proxy data is composed of only ice cores, we did not standardize the raw isotopic measurements at any point in the reconstructions. The calibrations were computed over the period 1880-2014, though individual locations had various proportions of missing data for either the proxy and/or the temperature calibration data (particularly there is very little temperature data over Antarctica prior to the late 1950s). For the reconstructions, we drew a random prior of 140 annual states from each climate model simulation (the prior size was limited to 140 by the extent of the historical simulations). We also tested the seasonality of the prior averaging and the proxy–temperature statistical PSMs by alternatively defining a year as April to the next March; we found very similar results (not shown) and therefore retained the usual annual definition of January to December in all the results shown here, which is consistent with the previous pseudoproxy reconstructions.

2.3.2 Reconstruction results

Analogous to the pseudoproxy results, Figs. 2.16–2.18 show the mean temperature reconstructions for Greenland, the British Isles, and the globe, with Berkeley Earth as verification;

we chose Berkeley Earth for temperature verification because of its high spatial resolution and more extensive spatial coverage, particularly in the polar regions, as compared to HadCRUT4.3. These results share several features with the pseudoproxy experiments, yet also differ in important ways. Like the pseudoproxy experiments, Greenland mean temperature is reconstructed better than for the British Isles or for the global mean at both annual and longer time scales. For the global mean in Fig. 2.18, detrended skill is significantly worse than the original time series, with r = 0.44 and CE = 0.15 while their detrended counterparts are $r_d = 0.23$ and $CE_d = -0.04$. One clear difference between the pseudoproxy and real reconstructions is that the real temperature reconstructions tend to have too little variance, particularly for Greenland and the British Isles, Figs. 2.16 and 2.17. This is likely the result of poor regression fits and thus loss of variance information for the statistical PSMs: the median correlation of all the ice core and local temperature series used in the fits is only 0.15 (with a maximum of 0.43).

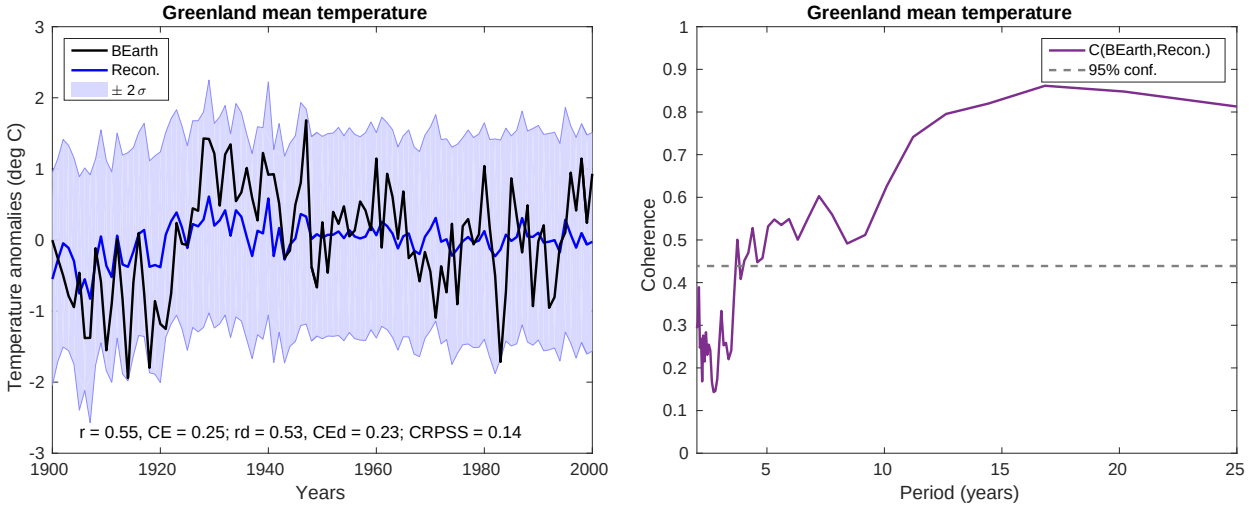

Figure 2.16: **Left:** Greenland mean temperature reconstruction. Skill metrics are computed with respect to Berkeley Earth (BEarth). **Right:** Coherence of the reconstruction mean with the Berkeley Earth times series, along with a 95% confidence level.

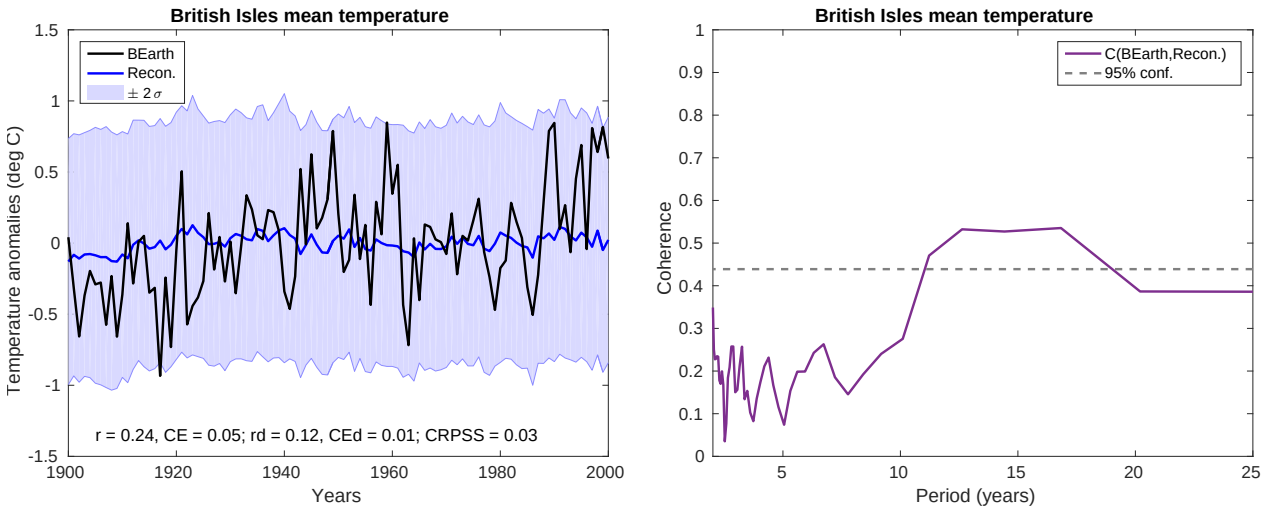

Figure 2.17: **Left:** British Isles mean temperature reconstruction. **Right:** Coherence of the reconstruction mean with the Berkeley Earth times series.

Figure 2.19 shows the annual, detrended spatial skill of the temperature reconstructions as verified against Berkeley Earth (infilled with RegEM iridge [Schneider, 2001]). Like the pseudoproxy reconstruction results shown in Fig. 2.8, the local skill over Greenland is high. The CE skill is not correspondingly high in Antarctica and we note that this result is confounded by the fact that observational data (and hence Berkeley Earth and HadCRUT data) is rare in Antarctica prior to the late 1950s whereas data over Greenland extend back into the 1800s. This means that we have less data in Antarctica to calibrate the PSMs and to verify the reconstructions. Another important difference between the real and pseudoproxy reconstructions is the lack of skill in the tropical Pacific. This difference is most likely due to the nature of the local-statistical PSMs that do not embed non-local information like the pseudoproxy PSMs do. Because of the atmospheric teleconnections between the tropical Pacific and much of the globe, we would expect more sophisticated PSMs to pick up some tropical reconstruction skill. Another difference of note is the large areas of negative CE skill in the pseudoproxy reconstructions that are not present in the real reconstructions. Large

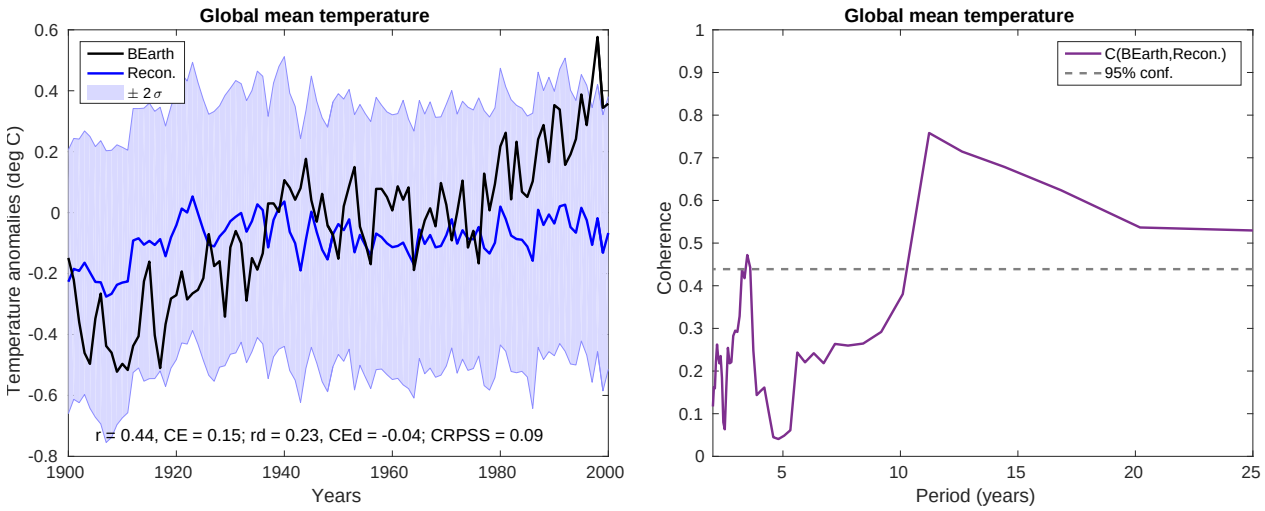

Figure 2.18: **Left:** Global mean temperature reconstruction. **Right:** Coherence of the reconstruction mean with the Berkeley Earth times series.

negative CE can arise from having bias and/or too much variance relative to the true state. Since the results shown in these figures are detrended first, they contain the same mean, so the large negative CE values in the pseudoproxy results are because of too much variance in the reconstruction mean. As seen in Figs. 2.16–2.18, the real reconstructions seem to suffer from the opposite problem of too little variance in many locations, which does not show up as large negative CE.

Figures 2.20 and 2.21 show the detrended spatial skill for Z500 and precipitation as verified against the European Centre For Medium-Range Weather Forecasts reanalysis product ERA-20C, which has data available starting in the year 1900 [Poli et al., 2013]. The spatial skill for Z500 shares similar patterns with the pseudoproxy results, though the skill is quite low with r values around 0.3 locally near ice core locations and CE values that are essentially zero. Precipitation as shown in Fig. 2.21 appears to not be robustly reconstructed anywhere, with higher r values in West Antarctica counter-balanced by low CE skill there.

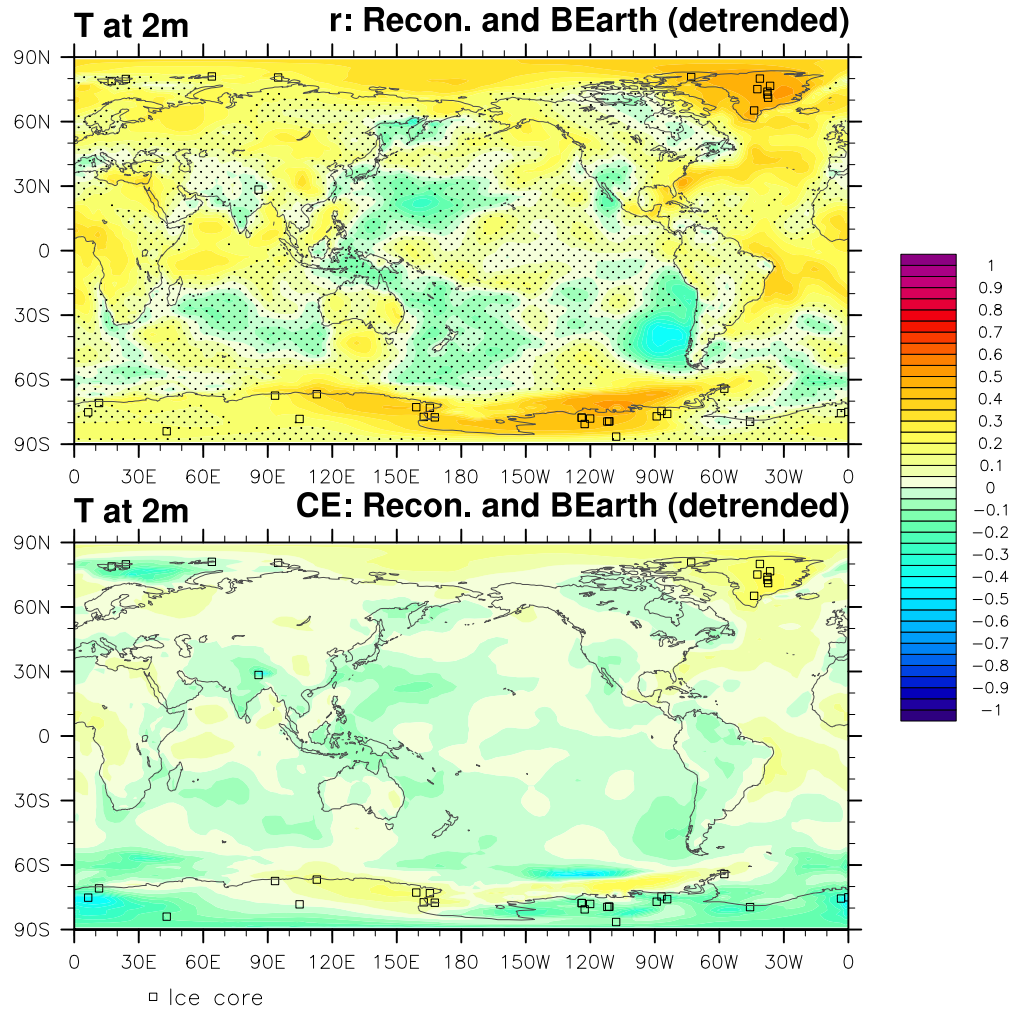

Figure 2.19: Spatial 2 m temperature reconstruction skill with correlation (r) on top and coefficient of efficiency (CE) on bottom, which is only shown down to -1. Both skill metrics are computed after detrending the underlying true and mean reconstruction data. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the pseudo ice cores are indicated by the open squares.

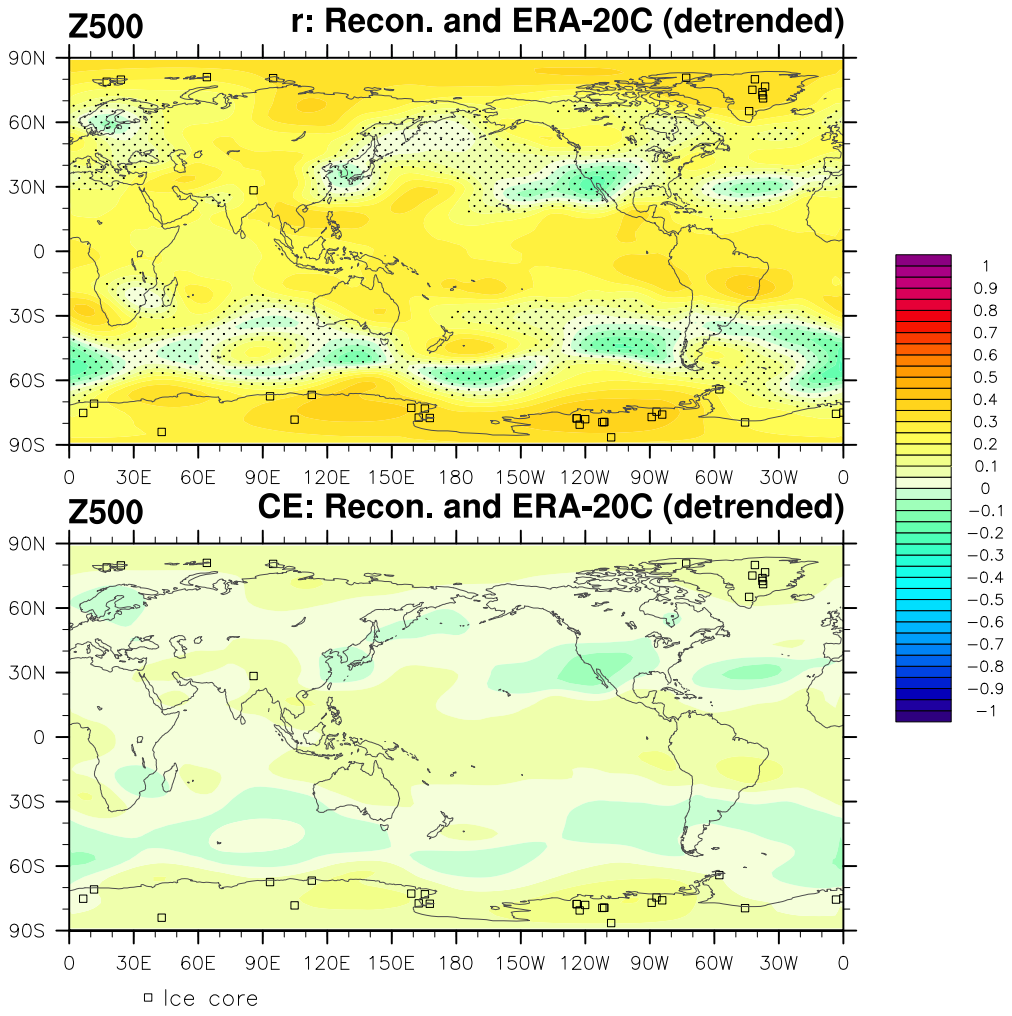

Figure 2.20: Geopotential height at 500 hPa (Z500) reconstruction skill (cf. Fig. 2.19).

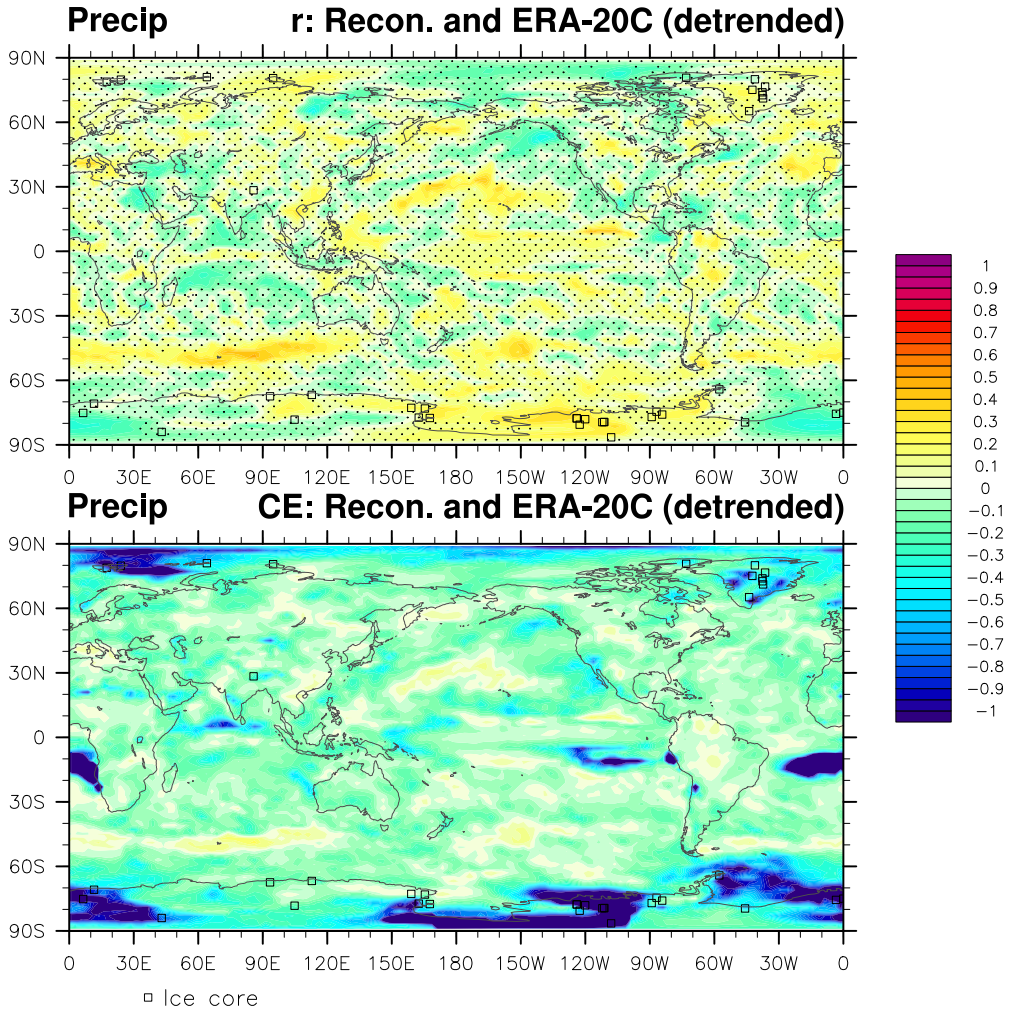

Figure 2.21: Precipitation reconstruction skill (cf. Fig. 2.19).

2.4 Conclusions

Through a series of pseudo and real ice core reconstructions we have estimated the spatial and temporal extent of how ice cores can inform past climate. Confirming much previous evidence (e.g. [Jouzel et al., 1997]), we find that ice cores can most robustly reconstruct local temperature and measures of atmospheric circulation (geopotential height). As the pseudoproxy experiments and recent observational studies suggest, ice cores can in principle be used to reconstruct non-local variables, particularly in the tropics. However, this result appears to depend on the sophistication and fidelity of the proxy system models employed. Additionally, the non-local skill is non-uniform and may not extend into plausible nearby locations: as a regional test case, all sets of experiments show that information in Greenland ice cores (in addition to other global ice cores) is uninformative for climate in the British Isles and Europe. Therefore despite seasonal linkages between Greenland and Europe via teleconnections like the NAO, it appears unlikely that Greenland ice cores can robustly inform European temperatures or atmospheric circulation at annual and decadal time scales.

We also found fundamental differences in skill between the key reconstruction variables of surface temperature, geopotential height at 500 hPa (Z500), and precipitation, with precipitation skill being clearly less robust. These differences are determined by the covariance relationships between the prior estimate of the proxies and the reconstructed variables (as used by the data assimilation method): the covariance length scale for precipitation is simply too short for that variable to be reconstructed by ice cores over much of the globe.

An important extension of the work presented here is to use the more realistic PSMs from the pseudoproxy experiments for the real proxy experiments. This is part of ongoing work and would be the first paleoclimate reconstruction to use isotope-enabled climate models. Based on the pseudoproxy work here we expect real reconstruction skill to improve, particularly non-locally for temperature and Z500.

BIBLIOGRAPHY

Richard B Alley, Paul A Mayewski, Todd Sowers, Minze Stuiver, Kendrick C Taylor, and Peter U Clark. Holocene climatic instability: A prominent, widespread event 8200 yr ago. *Geology*, 25(6):483–486, 1997.

R. S. Bradley. *Paleoclimatology*. Academic Press, Oxford, UK, 3rd edition, 2014.

Kurt M Cuffey and Eric J Steig. Isotopic diffusion in polar firn: implications for interpretation of seasonal climate parameters in ice-core records, with emphasis on central greenland. *Journal of Glaciology*, 44(147):273–284, 1998.

W. Dansgaard. Stable isotopes in precipitation. *Tellus*, 16(4):436–468, 1964. ISSN 2153-3490. doi: 10.1111/j.2153-3490.1964.tb00181.x. URL `http://dx.doi.org/10.1111/j.2153-3490.1964.tb00181.x`.

S. Dee, J. Emile-Geay, M. N. Evans, A. Allam, E. J. Steig, and D.M. Thompson. Prysm: An open-source framework for proxy system modeling, with applications to oxygen-isotope systems. *Journal of Advances in Modeling Earth Systems*, 7(3):1220–1247, 2015a. ISSN 1942-2466. doi: 10.1002/2015MS000447. URL `http://dx.doi.org/10.1002/2015MS000447`.

S. Dee, D. Noone, N. Buenning, J. Emile-Geay, and Y. Zhou. Speedy-ier: A fast atmospheric gcm with water isotope physics. *Journal of Geophysical Research: Atmospheres*, 120(1):73–91, 2015b. ISSN 2169-8996. doi: 10.1002/2014JD022194. URL `http://dx.doi.org/10.1002/2014JD022194`. 2014JD022194.

Sbastien Dirren and Gregory J. Hakim. Toward the assimilation of time-averaged observations. *Geophys. Res. Lett.*, 32(4), 2005. ISSN 1944-8007. doi: 10.1029/2004GL021444.

M.N. Evans, S.E. Tolwinski-Ward, D.M. Thompson, and K.J. Anchukaitis. Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, 76:16 – 28, 2013. ISSN 0277-3791. doi: http://dx.doi.org/10.1016/j.quascirev.2013.05.024. URL `http://www.sciencedirect.com/science/article/pii/S0277379113002011`.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Hugues Goosse, Elisabeth Crespin, Svetlana Dubinkina, Marie-France Loutre, Michael E Mann, Hans Renssen, Yoann Sallaz-Damaz, and Drew Shindell. The role of forcing and internal dynamics in explaining the medieval climate anomaly. *Climate dynamics*, 39(12): 2847–2866, 2012.

Sami Hanhijrvi, MartinP. Tingley, and Atte Korhola. Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2,000 years. *Clim. Dyn.*, 41 (7-8):2039–2060, 2013. ISSN 0930-7575. doi: 10.1007/s00382-013-1701-4.

Michael M Herron and Chester C Langway Jr. Firn densification: an empirical model. *Journal of Glaciology*, 25:373–385, 1980.

Peter L Houtekamer and Herschel L Mitchell. A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001.

Helga S. Huntley and Gregory J. Hakim. Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Clim. Dyn.*, 35(6):995–1009, 2010. ISSN 0930-7575. doi: 10.1007/s00382-009-0714-5.

Sigfús J Johnsen, Henrik B Clausen, Kurt M Cuffey, Georg Hoffmann, and Timothy T Creyts. Diffusion of stable isotopes in polar firn and ice: the isotope effect in firn diffusion. *Physics of ice core records*, pages 121–140, 2000.

SJ Johnsen. Stable isotope homogenization of polar firn and ice. *Isotopes and impurities in snow and ice*, pages 210–219, 1977.

J Jouzel, RB Alley, KM Cuffey, W Dansgaard, P Grootes, G Hoffmann, SJ Johnsen, RD Koster, D Peel, CA Shuman, et al. Validity of the temperature reconstruction from water isotopes in ice cores. *Journal of Geophysical Research*, 102:26–471, 1997.

Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge, Cambridge, UK, 2003.

T. Kurahashi-Nakamura, M. Losch, and A. Paul. Can sparse proxy data constrain the strength of the Atlantic meridional overturning circulation? *Geosci. Model Dev.*, 7(1): 419–432, 2014. doi: 10.5194/gmd-7-419-2014.

Marcel Küttel, Eric J Steig, Qinghua Ding, Andrew J Monaghan, and David S Battisti. Seasonal climate information preserved in west antarctic ice core water isotopes: Relationships to temperature, large-scale circulation, and sea ice. *Climate dynamics*, 39(7-8): 1841–1857, 2012.

Bo Li, Douglas W Nychka, and Caspar M Ammann. The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association*, 105(491):883–895, 2010.

Michael E Mann, Scott Rutherford, Eugene Wahl, and Caspar Ammann. Testing the fidelity of methods used in proxy-based reconstructions of past climate. *Journal of Climate*, 18 (20):4097–4107, 2005.

Michael E. Mann, Zhihua Zhang, Malcolm K. Hughes, Raymond S. Bradley, Sonya K. Miller, Scott Rutherford, and Fenbiao Ni. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci.*, 105 (36):13252–13257, 2008. doi: 10.1073/pnas.0805721105.

Shaun A. Marcott, Jeremy D. Shakun, Peter U. Clark, and Alan C. Mix. A reconstruction of regional and global temperature for the past 11,300 years. *Science*, 339(6124):1198–1201, 2013. doi: 10.1126/science.1228026.

Pierre Mathiot, Hugues Goosse, Xavier Crosta, Barbara Stenni, Martina Braida, Hans Renssen, Cédric J Van Meerbeeck, Valérie Masson-Delmotte, Aurélien Mairesse, and Svetlana Dubinkina. Using data assimilation to investigate the causes of southern hemisphere high latitude cooling from 10 to 8 ka bp. *Climate of the Past*, 9(2):887–901, 2013.

Anastasios Matsikaris, Martin Widmann, and Johann H Jungclaus. On-line and off-line data assimilation in palaeoclimatology: a case study. *Climate of the Past*, 11:81–93, 2015.

Colin P. Morice, John J. Kennedy, Nick A. Rayner, and Phil D. Jones. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres*, 117(D8):n/a–n/a, 2012. ISSN 2156-2202. doi: 10.1029/2011JD017187. URL `http://dx.doi.org/10.1029/2011JD017187`. D08101.

PAGES 2k Consortium. Continental-scale temperature variability during the past two millennia. *Nature Geosci.*, 6(5):339–346, 2013. doi: 10.1038/ngeo1797.

P. Poli, H. Hersbach, D. Tan, D. Dee, J.-N. Thpaut, A. Simmons, C. Peubey, P. Laloyaux, T. Komori, P. Berrisford, R. Dragani, Y. Trmolet, E. Holm, M. Bonavita, L. Isaksen, and M. Fisher. The data assimilation system and initial performance evaluation of the ecmwf pilot reanalysis of the 20th-century assimilating surface observations only (era-20c). *ERA Report Series*, (14), 2013.

N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14):n/a–n/a, 2003. ISSN 2156-2202. doi: 10.1029/2002JD002670. URL `http://dx.doi.org/10.1029/2002JD002670`. 4407.

Robert Rohde, Richard Muller, Robert Jacobsen, Saul Perlmutter, Arthur Rosenfeld,

Jonathan Wurtele, J Curry, Charlotte Wickham, and S Mosher. Berkeley earth temperature averaging process. *Geoinfor. Geostat.: An Overview*, 1(2):1–13, 2013.

Gavin A. Schmidt, Allegra N. LeGrande, and Georg Hoffmann. Water isotope expressions of intrinsic and forced variability in a coupled ocean-atmosphere model. *Journal of Geophysical Research: Atmospheres*, 112(D10):n/a–n/a, 2007. ISSN 2156-2202. doi: 10.1029/2006JD007781. URL `http://dx.doi.org/10.1029/2006JD007781`. D10103.

David P. Schneider and David C. Noone. Spatial covariance of water isotope records in a global network of ice cores spanning twentieth-century climate change. *Journal of Geophysical Research: Atmospheres*, 112(D18):n/a–n/a, 2007. ISSN 2156-2202. doi: 10.1029/2007JD008652. URL `http://dx.doi.org/10.1029/2007JD008652`. D18105.

David P. Schneider and Eric J. Steig. Ice cores record significant 1940s antarctic warmth related to tropical climate variability. *Proceedings of the National Academy of Sciences*, 105(34):12154–12158, 2008. doi: 10.1073/pnas.0803627105. URL `http://www.pnas.org/content/105/34/12154.abstract`.

Tapio Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14(5):853–871, 2001.

Tapio Schneider and Arnold Neumaier. Algorithm 808: Arfita matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, 27(1):58–65, 2001.

Jeremy D Shakun, Peter U Clark, Feng He, Shaun A Marcott, Alan C Mix, Zhengyu Liu, Bette Otto-Bliesner, Andreas Schmittner, and Edouard Bard. Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, 484(7392): 49–54, 2012.

Jason E. Smerdon. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs Clim. Change*, 3(1):63–77, 2012. ISSN 1757-7799. doi: 10.1002/wcc.149.

Eric J Steig, Qinghua Ding, James WC White, Marcel Küttel, Summer B Rupper, Thomas A Neumann, Peter D Neff, Ailie JE Gallant, Paul A Mayewski, Kendrick C Taylor, et al. Recent climate and ice-sheet changes in west antarctica compared with the past 2,000 years. *Nature Geoscience*, 6(5):372–375, 2013.

N. J. Steiger, G. J. Hakim, E. J. Steig, D. S. Battisti, and G. H. Roe. Assimilation of time-averaged pseudoproxies for climate reconstruction. *J. Climate*, 27(1):426–441, 2014. doi: 10.1175/JCLI-D-12-00693.1.

Robert Tardif, GregoryJ. Hakim, and Chris Snyder. Coupled atmosphere-ocean data assimilation experiments with a low-order model and CMIP5 model data. *Clim. Dyn.*, 2014. doi: 10.1007/s00382-014-2390-3.

J Wang, J Emile-Geay, Jason E Smerdon, D Guillot, and B Rajaratnam. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate of the Past*, 10(1):1–19, 2014.

Martin Werner, Petra M. Langebroek, Tim Carlsen, Marcus Herold, and Gerrit Lohmann. Stable water isotopes in the echam5 general circulation model: Toward high-resolution isotope modeling on a global scale. *Journal of Geophysical Research: Atmospheres*, 116 (D15):n/a–n/a, 2011. ISSN 2156-2202. doi: 10.1029/2011JD015681. URL `http://dx.doi.org/10.1029/2011JD015681`. D15109.

James C Zachos, Gerald R Dickens, and Richard E Zeebe. An early cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451(7176):279–283, 2008.

Appendix A

SUPPLEMENTARY FIGURES

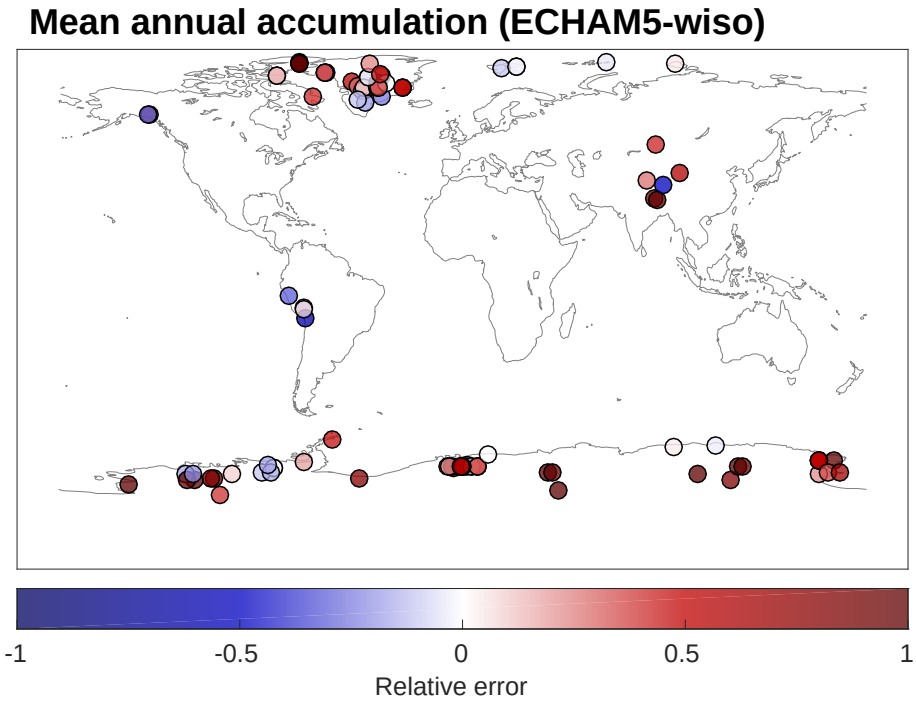

Figure A.1: The relative error of ECHAM5-wiso-modeled total precipitation accumulation at the pseudoproxy locations. Reds indicate too much accumulation while blues indicate too little accumulation relative to ERA-Interim.

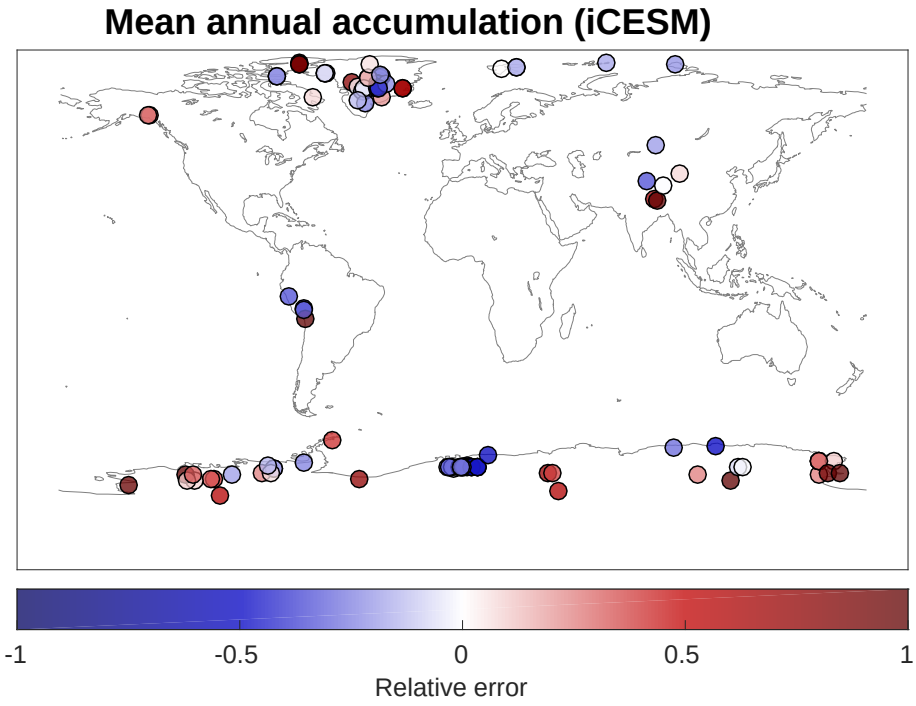

Figure A.2: The relative error of iCESM-modeled total precipitation accumulation at the pseudoproxy locations. Reds indicate too much accumulation while blues indicate too little accumulation relative to ERA-Interim.

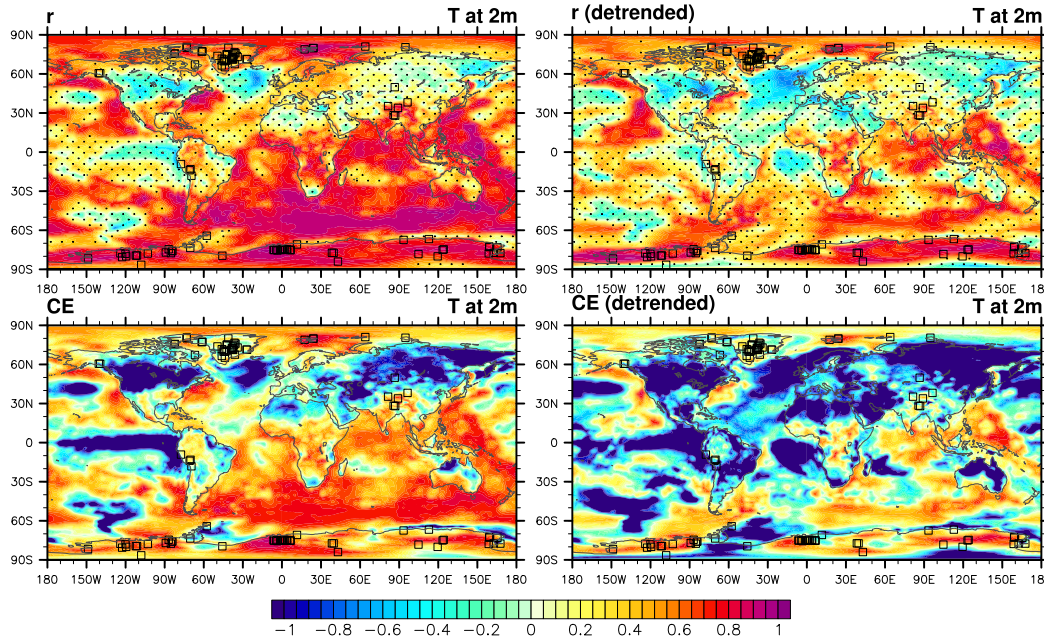

Figure A.3: Spatial 2 m air temperature reconstruction skill, at 10 year averages, with correlation (r) on top and coefficient of efficiency (CE) on bottom, which is only shown down to -1. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the pseudo ice cores are indicated by the open squares.

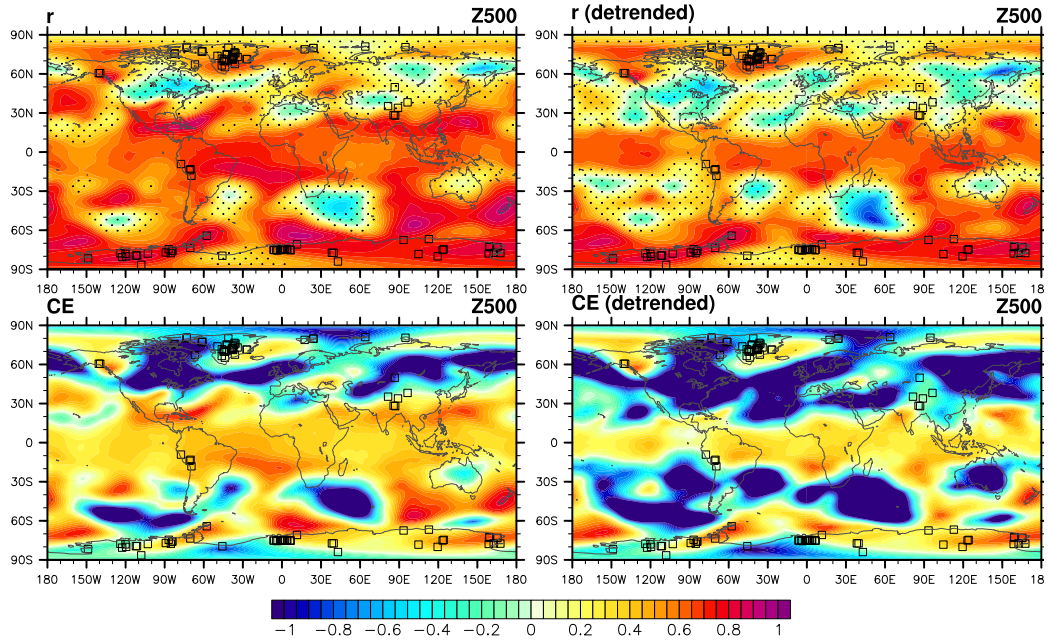

Figure A.4: Geopotential height at 500 hPa (Z500) reconstruction skill, at 10 year averages, with correlation (r) on top and coefficient of efficiency (CE) on bottom, which is only shown down to -1. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the pseudo ice cores are indicated by the open squares.

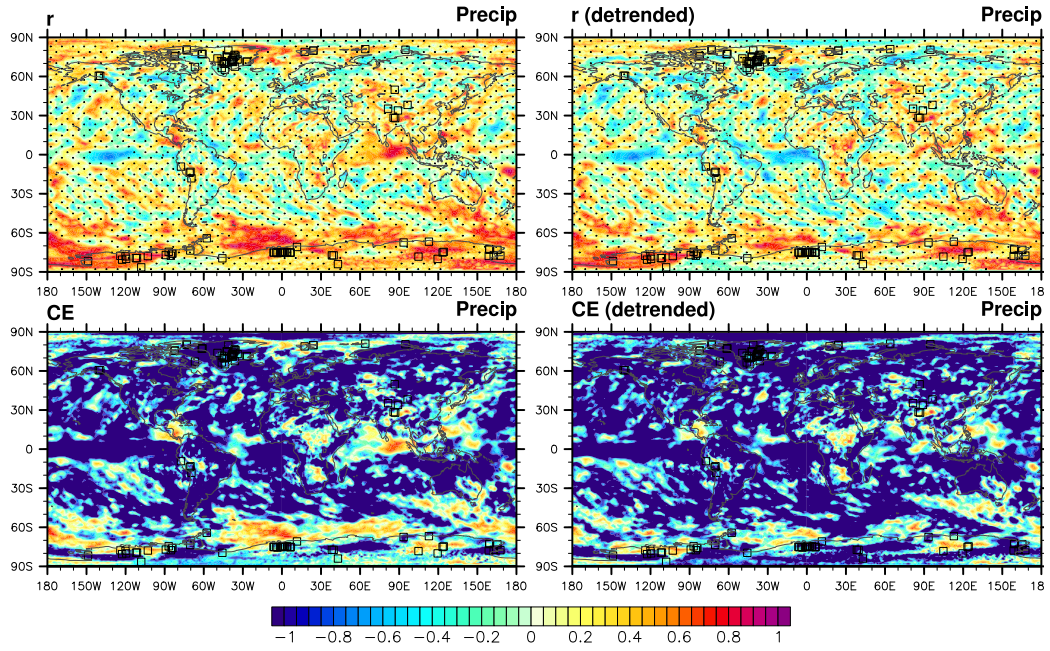

Figure A.5: Precipitation reconstruction skill, at 10 year averages, with correlation (r) on top and coefficient of efficiency (CE) on bottom, which is only shown down to -1. For the correlation plot, stippling indicates correlations that are not significant at the 95% level. The locations of the pseudo ice cores are indicated by the open squares.

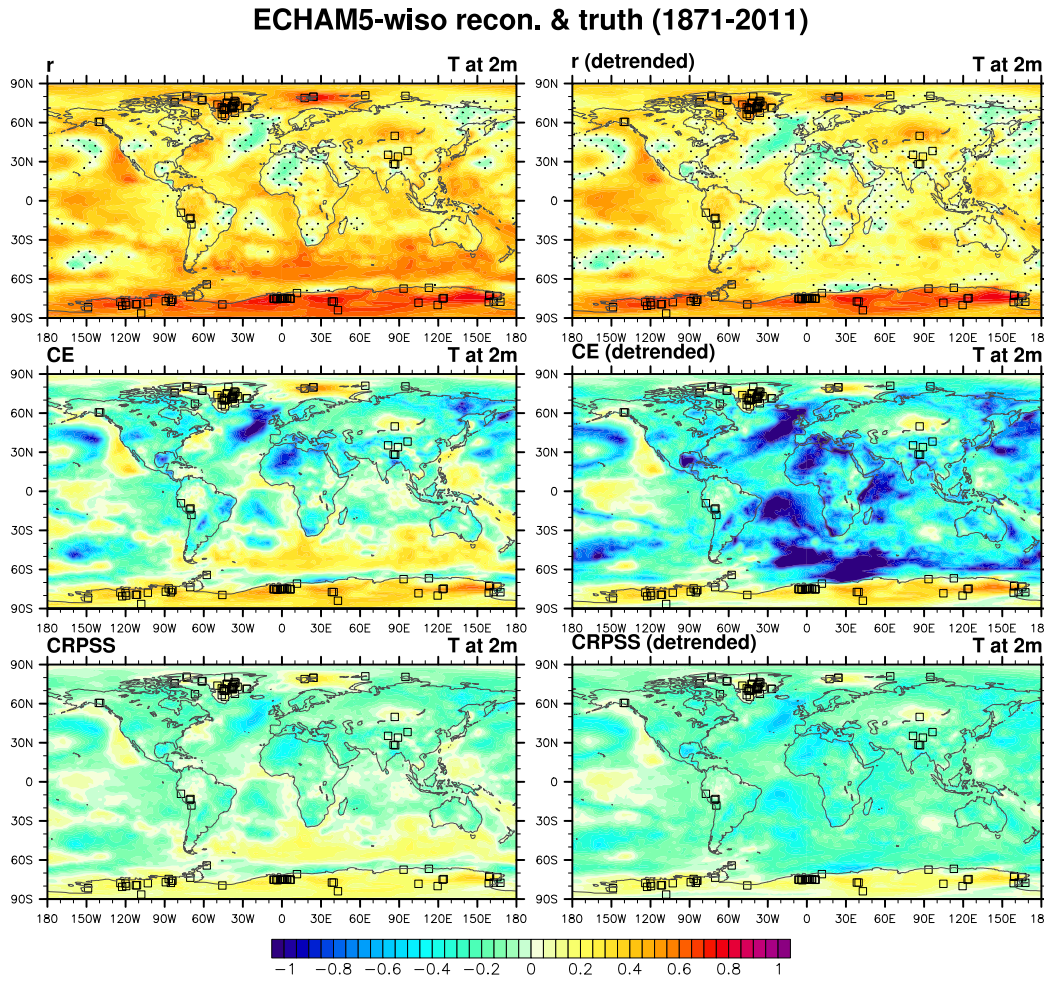

Figure A.6: Reconstruction skill for 2 m temperature when the ice core diffusion is reduced by 1/2.

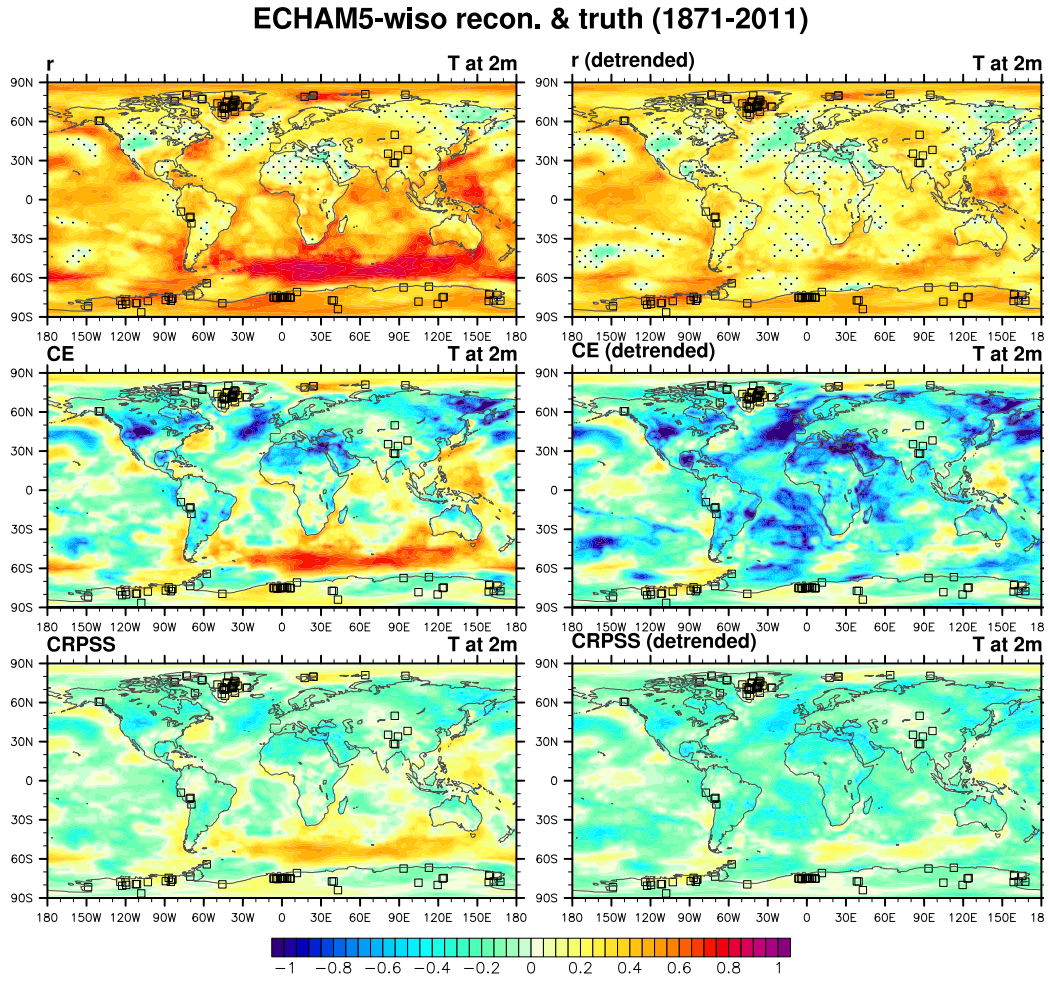

Figure A.7: Reconstruction skill for 2 m temperature when the ice core diffusion is increased by 2.

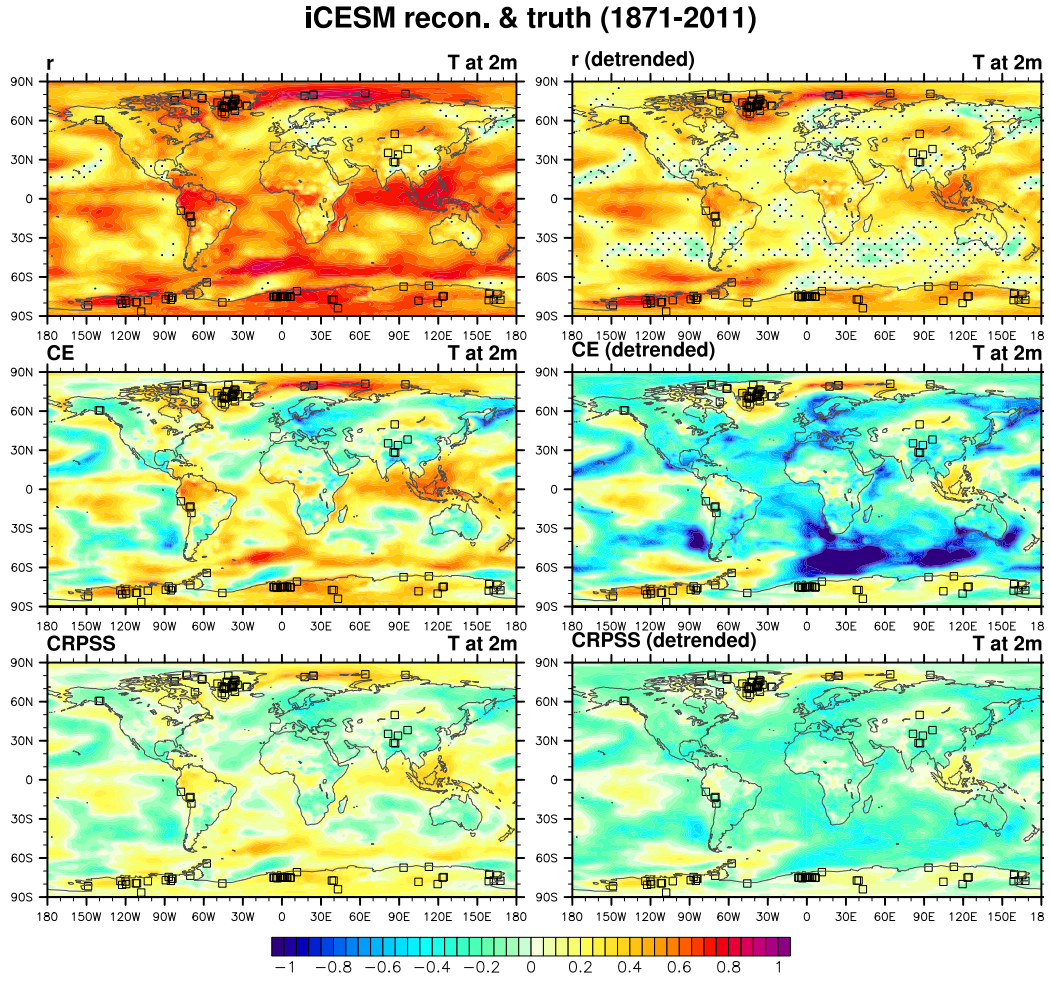

Figure A.8: Reconstruction skill for 2 m temperature except using iCESM instead of ECHAM5-wiso (see Fig. 2.8).

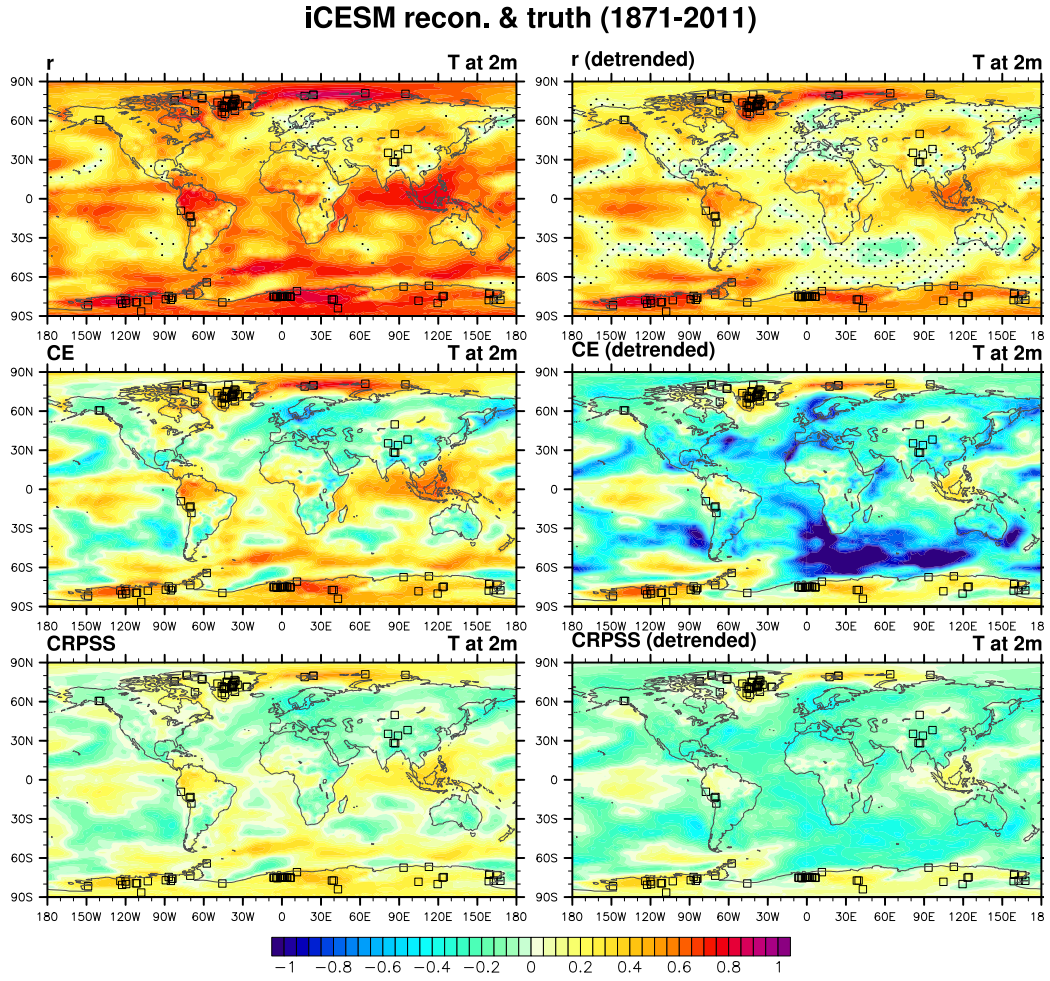

Figure A.9: Reconstruction skill for 2 m temperature (using iCESM) when the ice core diffusion is reduced by 1/2.

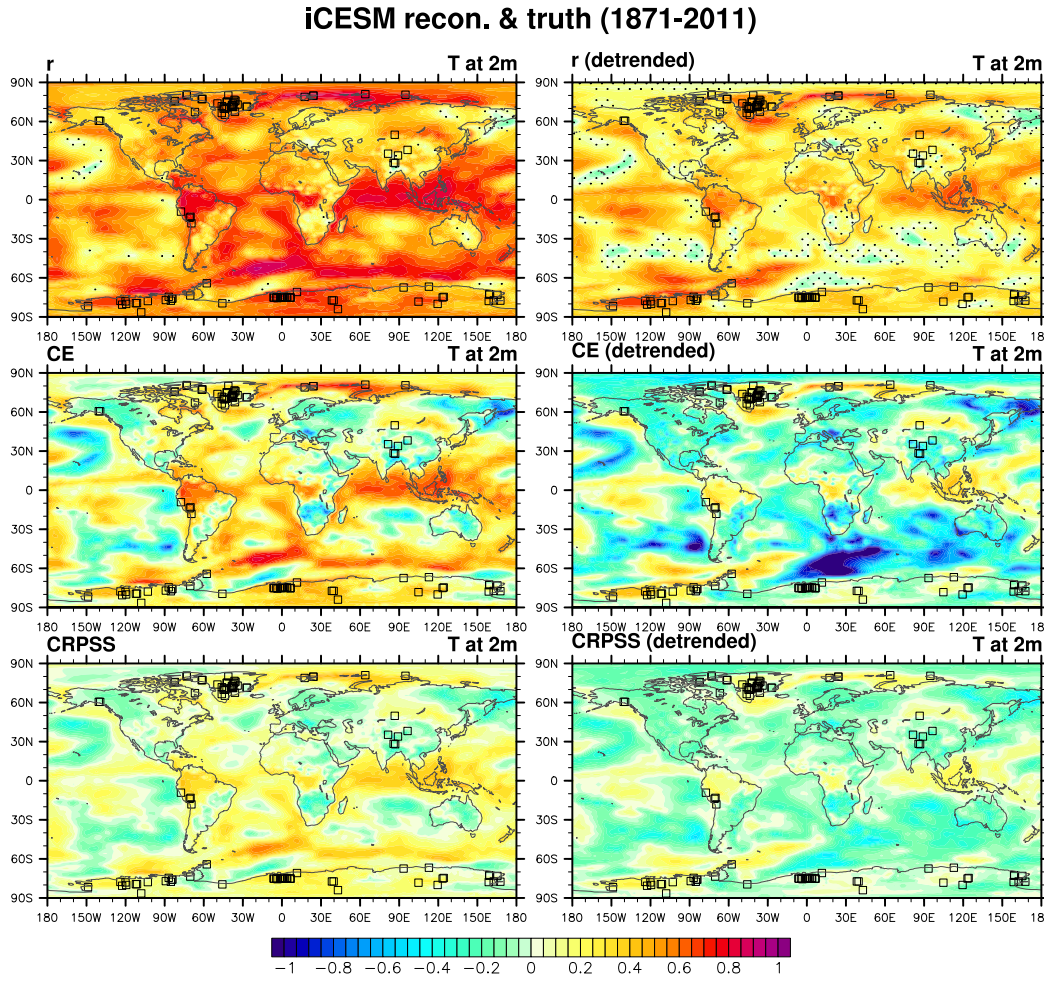

Figure A.10: Reconstruction skill for 2 m temperature (using iCESM) when the ice core diffusion is increased by 2.

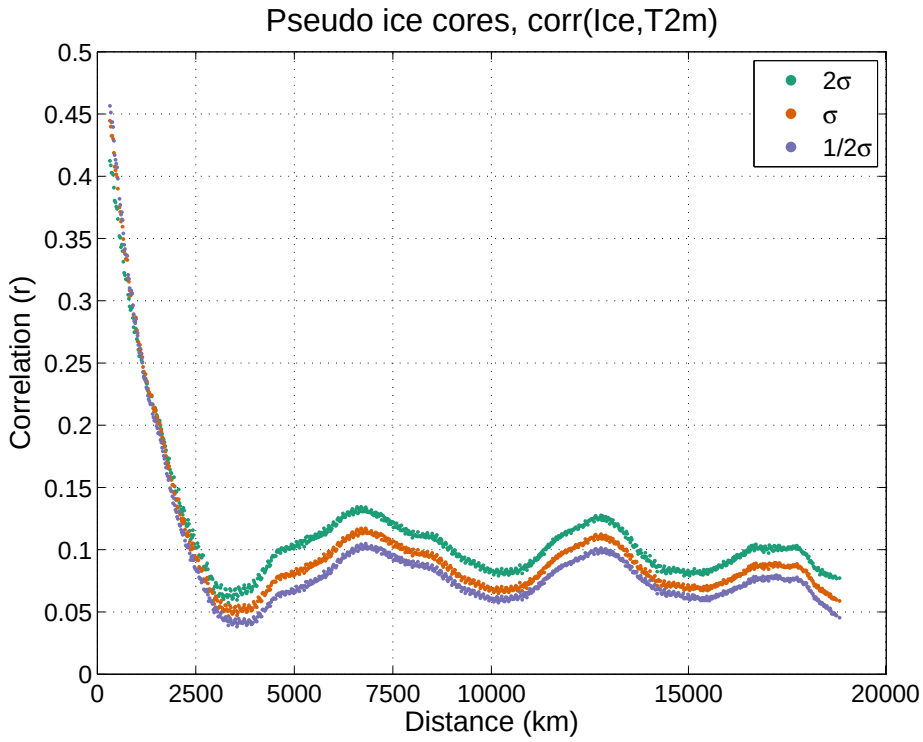

Figure A.11: Correlation as a function of distance between the PSM-derived pseudo ice cores and 2 m temperature, averaged over all ice core locations, for each version of diffusion.

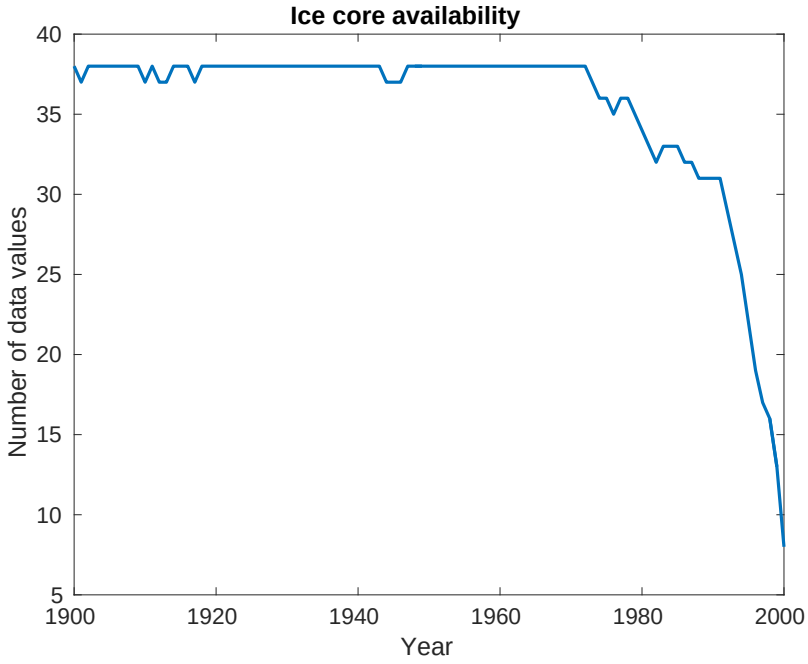

Figure A.12: Ice core data values available in the PAGES2k-V2 dataset for each year of the reconstruction. Forty ice cores are used in the reconstruction, but any given year may contain missing data or may not cover the entire 1900-2000 year reconstruction interval.