AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD026011

Climate reconstruction using data assimilation of water isotope ratios from ice cores

Nathan J. Steiger[1], Eric J. Steig[2], Sylvia G. Dee[3], Gerard H. Roe[2], and Gregory J. Hakim[4]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA, [2]Department of Earth and Space Sciences, University of Washington, Seattle, Washington, USA, [3]Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, Rhode Island, USA, [4]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA

Key Points:
- Novel data assimilation-based paleoclimate reconstructions using water isotope-enabled climate models
- For several climate variables the spatial extent of reconstruction skill varies by region
- Some nonlocal, tropical climate information may also be reconstructable from ice core isotopes

Correspondence to:
N. J. Steiger,
nsteiger@ldeo.columbia.edu

Citation:
Steiger, N. J., E. J. Steig, S. G. Dee, G. H. Roe, and G. J. Hakim (2017), Climate reconstruction using data assimilation of water isotope ratios from ice cores, *J. Geophys. Res. Atmospheres*, *122*, 1545–1568, doi:10.1002/2016JD026011.

Received 27 SEP 2016
Accepted 11 JAN 2017
Accepted article online 14 JAN 2017
Published online 7 FEB 2017

Abstract Water isotope data from ice cores, particularly $\delta^{18}O$, have long been used in paleoclimatology. Although $\delta^{18}O$ has been primarily interpreted as a proxy for local air temperature, isotope-enabled climate models have established that there are many nonlocal and nontemperature-related climatic influences on isotopic signals at coring locations. Moreover, recent observational studies have linked ice core isotopes to nonlocal patterns of climate variability, particularly to midlatitude atmospheric circulation patterns and to variations in tropical climate. Therefore, paleoclimate reconstructions may better utilize ice core isotope proxies by combining them with isotope-enabled climate models. Here we employ a data assimilation-based technique that fuses isotopic proxy information with the dynamical constraints of climate models. Through several idealized and real proxy experiments we assess the spatial and temporal extent to which isotope records can reconstruct surface temperature, 500 hPa geopotential height, and precipitation. We find local reconstruction skill to be most robust across the reconstructions, particularly for temperature and geopotential height, as well as limited nonlocal skill in the tropics. 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 nonlocal climate information may also be inferred from ice cores. However, the spatial range of this information is nonuniform and depends on skillful modeling of the proxy data within the reconstruction process.

1. Introduction

Water isotope data from ice cores have long been used as a paleoclimate proxy, particularly the oxygen isotopic composition of precipitation deposited in layered ice core archives. $\delta^{18}O$, a measure of the ratio of the oxygen isotopes oxygen-18 and oxygen-16, has been used primarily as a proxy for local temperature in the mid to high-latitudes and for local precipitation in the tropics [e.g., *Dansgaard*, 1964; *Jouzel et al.*, 1997]. In the past few decades, isotope-enabled climate models have allowed for a richer understanding of the climate processes that produce the isotopic signals in precipitation [e.g., *Risi et al.*, 2012]. Such modeling-based studies complicate the simple temperature-isotope interpretation by identifying the many nonlocal influences on isotopic signals at ice coring locations. Recent observational studies have also linked ice cores to nonlocal patterns of climate variability, particularly to midlatitude atmospheric circulation patterns and to variations in tropical climate [e.g., *Schneider and Noone*, 2007; *Schneider and Steig*, 2008]. And although the temperature-isotope relationship is well established for the study of past climate on long time scales (e.g., millennial time scales), on annual and decadal timescales nonlocal climate influences can become quite important [*Masson-Delmotte et al.*, 2008; *Sturm et al.*, 2010]. The full spatial extent over which ice core isotope records can robustly inform past climate on these shorter timescales has yet to be fully explored.

Here we explore the question: What can ice core isotopes tell us about annual and decadal climate? Specifically, we identify the spatial range over which ice core isotope records can be used to robustly reconstruct annual and decadal climate. To do this, we employ a data assimilation (DA)-based reconstruction technique that optimally combines proxy information with the dynamical constraints of climate models [*Steiger et al.*, 2014; *Hakim et al.*, 2016]. Here we produce spatial field reconstructions using a combination of 38 ice cores [*Emile-Geay et al.*, 2015] and two different isotope-enabled climate models to simulate isotope ratios

in precipitation. We include modeling of the firn compaction and water vapor diffusion process that affect the isotope signal ultimately recorded in the ice cores [*Herron and Langway*, 1980; *Johnsen et al.*, 2000; *Dee et al.*, 2015]. In contrast to this approach, most ice core-based reconstructions of past climate are based on converting single isotope records into temperature through an estimated linear relationship [e.g., *Jouzel et al.*, 1997] or are based on averaging a collection of records to arrive at a regional reconstruction [e.g., *Steig et al.*, 2013]. This study is the first to model ice core isotopes explicitly in DA-based reconstructions using real proxy data. Previous DA reconstructions that have incorporated ice core proxies have estimated the isotopes through a linear-univariate fit with local temperature [e.g., *Goosse et al.*, 2012; *Hakim et al.*, 2016].

We perform a series of pseudoproxy and real proxy experiments to assess the spatial and temporal extent to which ice core isotopes can be used to reconstruct past climate. Pseudoproxy experiments are synthetic, controlled tests of a reconstruction technique that are performed within a climate model world (see *Smerdon* [2012], for a review). Because the experiments are contained within this world, the answer is known perfectly, unlike the real historical climate. Real-world reconstructions include many uncontrolled and sometimes unknown factors in both the proxy and observational climate data; these factors can make it unclear where the deficiencies are in a reconstruction technique. Importantly, pseudoproxy experiments can be used to estimate the upper bound on the reconstruction skill one could expect from a similar real proxy reconstruction. Therefore, in this study we reconstruct the climate variables of surface temperature, geopotential height at 500 hPa, and precipitation in both pseudo and real proxy experiments, where the pseudoproxy experiments provide this upper bound estimate while the real proxy experiments assess what is currently knowable in a real, state-of-the-art reconstruction. After presenting the results of both sets of experiments and assessing them at annual to decadal timescales, we discuss the reasons for the fundamental differences between reconstruction skill of different variables and the key results of the pseudoproxy and real proxy reconstructions. These experimental results will be particularly useful for guiding interpretations of ice core isotope-based reconstructions over the past two millennia.

2. Pseudoproxy Experiments

2.1. Experimental Framework

We employ a data assimilation technique that optimally combines observations (in this context, proxy data) with climate model states. The 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 may contain any climate model variables of interest. 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, chapter 5] are given by

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

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

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

and cov() represents a covariance expectation. $\mathbf{x}_b$ is the prior (or "background") estimate of the state vector and $\mathbf{x}_a$ is the posterior (or "analysis") state vector. Observations (or proxies) are contained in vector $\mathbf{y}$. 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; in the present work $\mathcal{H}(\mathbf{x}_b)$ is the estimate of ice core isotopes at the ice core locations based on output from the climate model. Matrix $\mathbf{K}$, the Kalman gain, weights $\mathbf{y} - \mathcal{H}(\mathbf{x}_b)$ and transforms it into state space. Matrix $\mathbf{R}$ is the error covariance matrix for the observations. The DA update process involves iteratively computing equation (1) to arrive at the posterior, $\mathbf{x}_a$, for each year of the reconstruction. In any given year of the reconstruction $\mathbf{x}_a$ is the probability distribution of states that are consistent with the proxy observations and errors, as well as the prior distribution; therefore, it is the full probability distribution of $\mathbf{x}_a$ that represents the probabilistic reconstruction of the climate state. In the figures we show the mean of $\mathbf{x}_a$, with uncertainty estimates derived from the posterior distribution. Additionally, we present a variety of performance measures of the climate reconstructions, including one that takes account of the full posterior distribution for both the time series and the space-time reconstructions (see the discussion of the "continuous ranked probability skill score" in section 2.2). For more mathematical details of the reconstruction methodology and the precise calculation procedure, see the Appendix of *Steiger et al.* [2014].

In all of the experiments shown here (both pseudo and real reconstructions) we use a "no-cycling" or "off-line" DA approach, where the prior ensembles are drawn from existing climate model simulations. Here the ensemble members are individual years instead of independent model simulations, as employed in "cycling" or "online" data assimilation. Because of how the prior ensemble is constructed, it will not contain year-specific forcings or boundary conditions nor does the method propagate information forward in time. This off-line approach has large computational benefits over an online approach where one must integrate an ensemble of climate model simulations forward in time after each DA update step. Indeed, for the paleoclimate reconstruction problem, it is almost always impractical to cycle an ensemble of tens to hundreds of Coupled Model Intercomparison Project Phase 5-class climate model simulations (as used here) for a paleoclimate reconstruction. Moreover, in the off-line case, one may use hundreds to thousands of ensemble members from multiple models and high-resolution simulations, reducing the potential for model bias and sampling error. It is also appropriate 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 skillfully forecast the climate state a year into the future, then no useful information is gained by cycling the model. *Matsikaris et al.* [2015] recently compared online and off-line approaches to paleoclimate DA with a fully coupled earth system model and found no improvement with the online method, suggesting that the model was unable to provide useful atmospheric information at analysis times. However, an online DA approach may be advantageous for ocean-focused DA, given the longer climatic timescales involved [*Goosse*, 2016; *Steiger and Hakim*, 2016].

The climate model simulations employed in these experiments for the prior, $\mathbf{x}_b$, are the isotope-enabled atmospheric models ECHAM5-wiso [*Werner et al.*, 2011] and iCAM5 (described in J. Nusbaumer et al., submitted manuscript, 2016). We ran ECHAM5-wiso at the spectral truncation T106 (~1° resolution) using monthly historical sea ice and sea surface temperatures from the Hadley Centre sea ice and sea surface temperature (SST) data set (HadISST) [*Rayner et al.*, 2003] as boundary conditions for the years 1871 through 2011. We also use a simulation of iCAM5 that was run at ~2° resolution and was also forced with monthly historical sea surface temperatures for the years 1850–2014, from a merged data set of the HadISST and version 2 of the National Oceanic and Atmospheric Administration optimum interpolation SST analysis data products [*Hurrell et al.*, 2008] (though only the years of overlap with the ECHAM5-wiso simulation were used). Both model simulations assume that the oxygen isotope ratios of ocean water are constant in time but vary in space according to the annual mean observation-based product of *LeGrande and Schmidt* [2006]. In section A1, we evaluate specific features of both simulations that are relevant to our reconstruction experiments.

These models provide the annual states for our prior ensembles as well as the climate and isotopic fields for the pseudoproxies (see below). We reconstruct 2 m air temperature, geopotential height at 500 hPa, and precipitation, all at the native model resolutions. The pseudoproxy reconstructions were performed over the years 1871–2011 of the simulations by reconstructing the annual climate states 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). Each year here is a calendar year, January to December; we also tested the seasonality of the prior averaging by alternatively defining a year as April to the next March, and we found very similar results (not shown) and therefore retained the usual annual definition. Though we note that for particular locations, such as for the Andes, the calendar year may not fully account for the local hydrological year and may limit reconstructions of seasonal El Niño–Southern Oscillation (ENSO) dynamics [*Hardy et al.*, 2003; *Vimeux et al.*, 2009; *Hurley et al.*, 2016]. Each reconstruction was repeated 100 times in a Monte Carlo fashion by random sampling 75% of the pseudoproxy network for each reconstruction iteration. Each of these iterations comes with a posterior ensemble, and uncertainty estimates are derived from the spread in all the posterior ensembles of the Monte Carlo iterations. We note that because our ensemble size is rather large, with 129, 130, or 131 members depending on the particular segment of the reconstruction, we did not employ covariance localization, which is a common (though ad hoc) method of reducing the effects of sampling error in small ensembles.

Pseudoproxy experiments, usually designed to test different methodological approaches to climate reconstruction, generally employ relatively simple pseudoproxies [*Smerdon*, 2012]. 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, observation

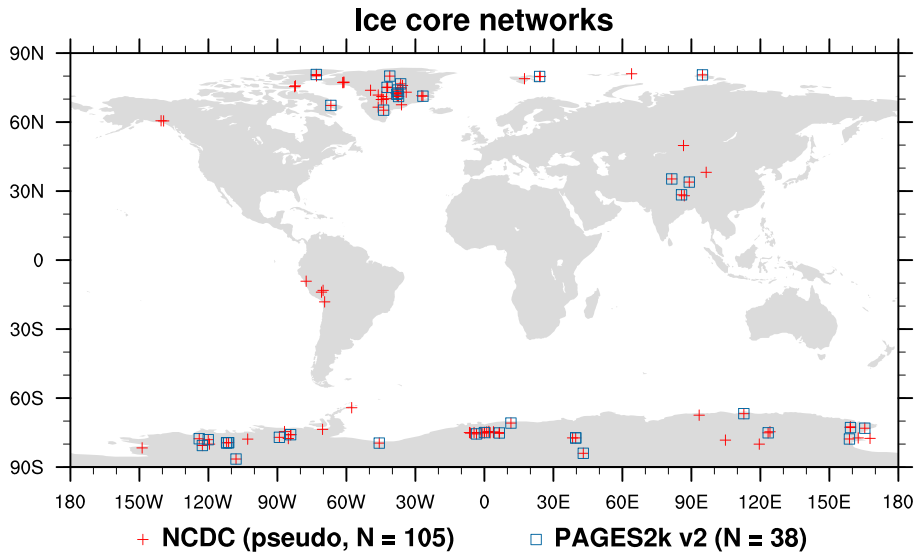

Figure 1. Pseudo and real proxy networks used in this study, with the number of proxy sites indicated for each network. Note that some site locations are shown here with approximate coordinates following what is entered in the proxy databases and also note that many proxy sites are very near one another or are colocated.

models, or forward models—to create our pseudoproxies (see *Evans et al.* [2013], for a review of PSMs). PSMs are physically based models that transform the simulated climate signal (e.g., temperature and precipitation) to a synthetic proxy time series. Here the pseudoproxy ice cores were modeled using the water isotope fields of ECHAM5-wiso/iCAM5 together with the PSM discussed in *Dee et al.* [2015]. Besides including water isotopes, ECHAM5-wiso and iCAM5 account for sublimation, melting, and wind erosion of snow on the land surface. The pseudoproxy network for these experiments is the "NCDC" network shown in Figure 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 $\delta^{18}O$); this large ice core network was chosen as a best case network coverage scenario.

The ice core PSM includes three submodels, 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.* [2015], 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 convolving the original isotope signal with a Gaussian kernel characterized by a diffusion length σ (see *Johnsen et al.* [2000], *Küttel et al.* [2012], and *Dee et al.* [2015], for mathematical details), and the observation model accounts for analytical uncertainty by sampling an age model. We note that for simplicity, and because we are interested in estimating an upper limit on reconstruction skill, in this study we do not account for age uncertainties, which would generally act to degrade the skill of the reconstructed variables. While the combination of the climate model output and the PSM produce single isotopic values for a given year (like real ice cores), this value represents the many short-time scale processes included in the simulations, such as daily weather events and seasonal snow melt or sublimation. 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 adjust the diffusion length, σ, in the ice core PSM. Diffusion in the firn can blur well-defined layering of isotope ratios, potentially limiting the ability to reconstruct annual climate variables [*Johnsen*, 1977; *Cuffey and Steig*, 1998; *Johnsen et al.*, 2000; *Küttel et al.*, 2012]. The extent to which diffusion affects an ice core isotope record is principally a function of site temperature and snow accumulation rate. We estimate the uncertainty associated with diffusion by directly setting σ in two end-member experiments and comparing the results with an unmodified control σ. For the low-diffusion

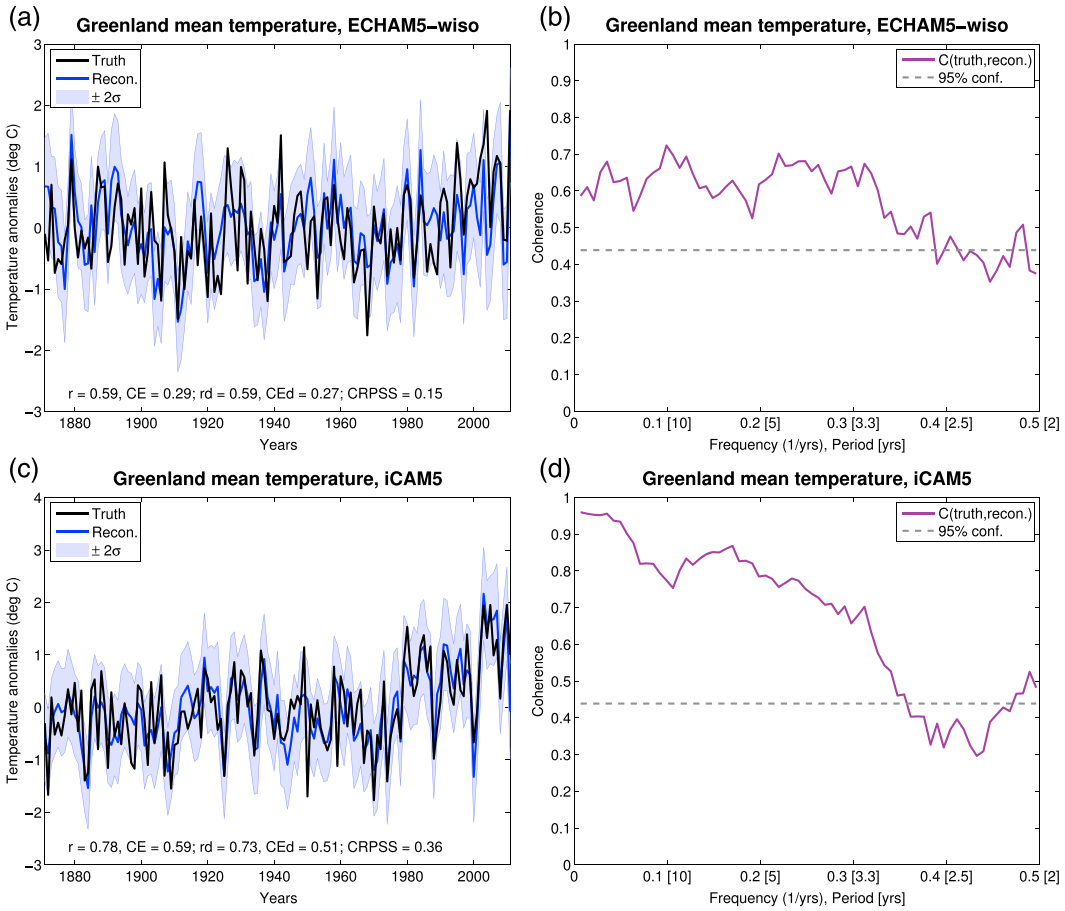

Figure 2. (a and c) Greenland mean temperature pseudoproxy reconstruction. Skill metrics are indicated on the lower left as correlation (*r*), coefficient of efficiency (CE), and both of these again after the true and reconstructed time series have been detrended (indicated by "d"); additionally the mean continuous ranked probability skill score (CRPSS) assesses both the reconstruction mean and uncertainty relative to the uninformed prior baseline. The uncertainty estimates are based on the spread in all the posteriors of the Monte Carlo iterations. (b and d) Coherence of the reconstruction mean with the true times series, along with a 95% confidence level.

experiment we set σ to one half its control value and for the high-diffusion experiment we set σ to 2 times its control value. The control values of σ are obtained following *Johnsen et al.* [2000], using a snow density profile calculated with a Herron-Langway densification model [*Herron and Langway*, 1980]. The Herron-Langway model uses inputs of temperature and snow accumulation that are taken from the climate model simulations and bias corrected to elevation. The sensitivity of the reconstruction results to modifying the diffusion length are shown in section A2.

2.2. Reconstruction Results

Let us first consider the local temperature reconstruction skill for the pseudoproxy experiments, focusing on Greenland. Figures 2a and 2c show the reconstructed Greenland mean temperature and the "true" Greenland mean temperature (from both general circulation model (GCM) simulations), with uncertainty estimates based on both the posterior ensembles and the Monte Carlo iterations. Skill for the reconstruction mean is shown at the bottom of the figure panels: correlation (*r*), coefficient of efficiency (CE) [*Nash and Sutcliffe*, 1970], and again both of these after the true and reconstructed time series have been detrended. Additionally, the figures indicate the mean continuous ranked probability skill score (CRPSS) of the posterior distributions of the reconstructions; this skill score is based on the continuous ranked probability score (CRPS), which is a "strictly proper" scoring rule that accounts for the skill of the entire posterior 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 skillful for

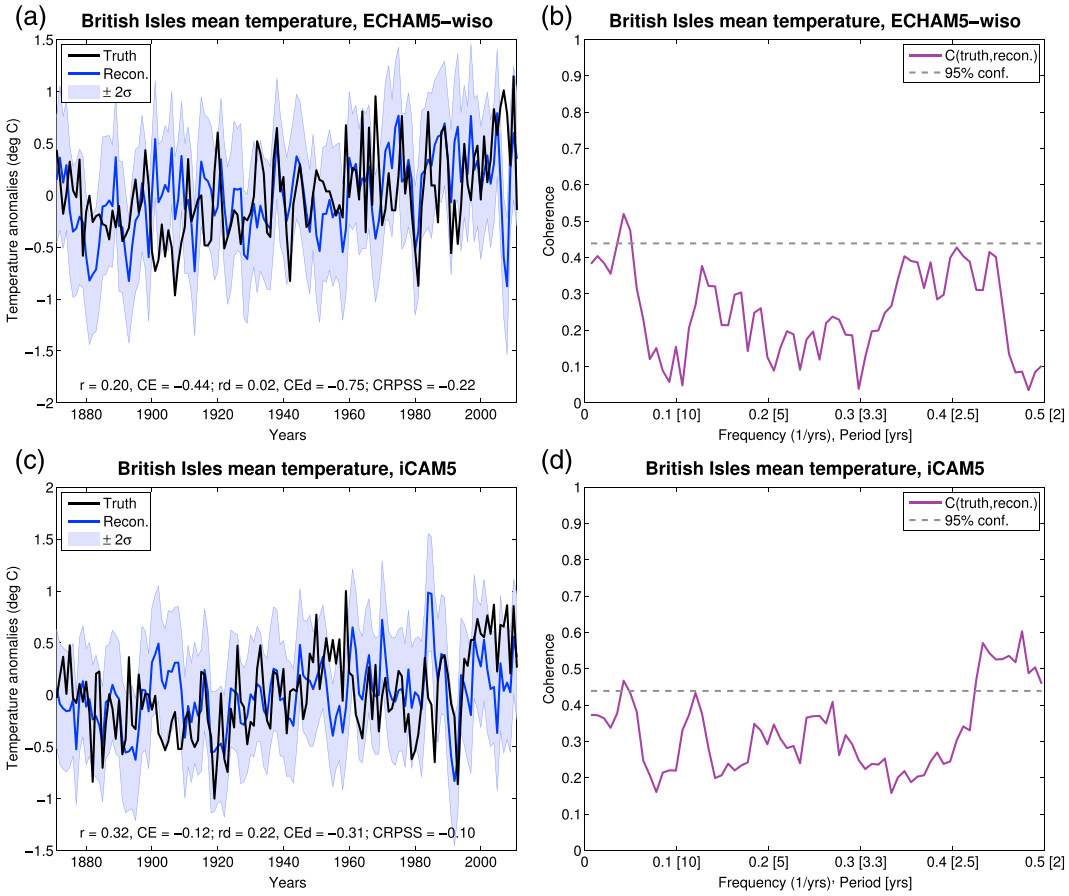

Figure 3. (a and c) British Isles mean temperature pseudoproxy reconstruction (cf. Figure 2). (b and d) Coherence of the reconstruction mean with the true times series.

this metric than the uninformed prior. CRPSS is calculated for each year of the reconstruction, and the "mean" CRPSS shown in all the figures indicates the time mean of the CRPSS values. We note that the implementation of CRPS used here (and elsewhere [*Barboza et al.*, 2014; *Werner and Tingley*, 2015]) does not account for errors in the verification data sets; this assumption is not an issue for the pseudoproxy results because the truth is known exactly, but for the real proxy results it should be kept in mind that there can be large uncertainties in some historical observation-based products. Figures 2b and 2d shows the cross-spectral coherence of the mean reconstructions with the true time series (computed using a multitaper method). Coherence is similar to correlation as a function of frequency, and in this figure it shows relatively constant or improved coherence at lower frequency or longer time scales. We note that the coherence is here computed without first detrending the time series since we are partly interested in how well the trends are reconstructed; therefore, depending on the particular time series, there can be some fragility of the coherence skill at longer than decadal time scales.

The North Atlantic Oscillation (NAO) is an atmospheric circulation pattern that extends over both Greenland and Europe and is the dominant teleconnection pattern in the North Atlantic-European region [*Hurrell et al.*, 2013]. It is therefore reasonable to expect that Greenland ice cores may contain important information about European climate. Furthermore, it has been argued that particular accumulation and $\delta^{18}O$ records in Greenland correlate with the NAO [*Appenzeller et al.*, 1998; *Rogers et al.*, 1998] and Greenland ice core $\delta^{18}O$ records have even been used to reconstruct the NAO [*Vinther et al.*, 2003]; though we note that different NAO reconstructions based on multiple proxies have found very different results from one another [e.g., *Schmutz et al.*, 2000; *Trouet et al.*, 2009]. As a convenient point of comparison with the Greenland mean reconstructions shown in Figure 2, we additionally reconstruct the British Isles (Britain and Ireland) mean temperature (Figure 3). Unlike the Greenland mean temperature reconstructions in Figure 2, this regional mean

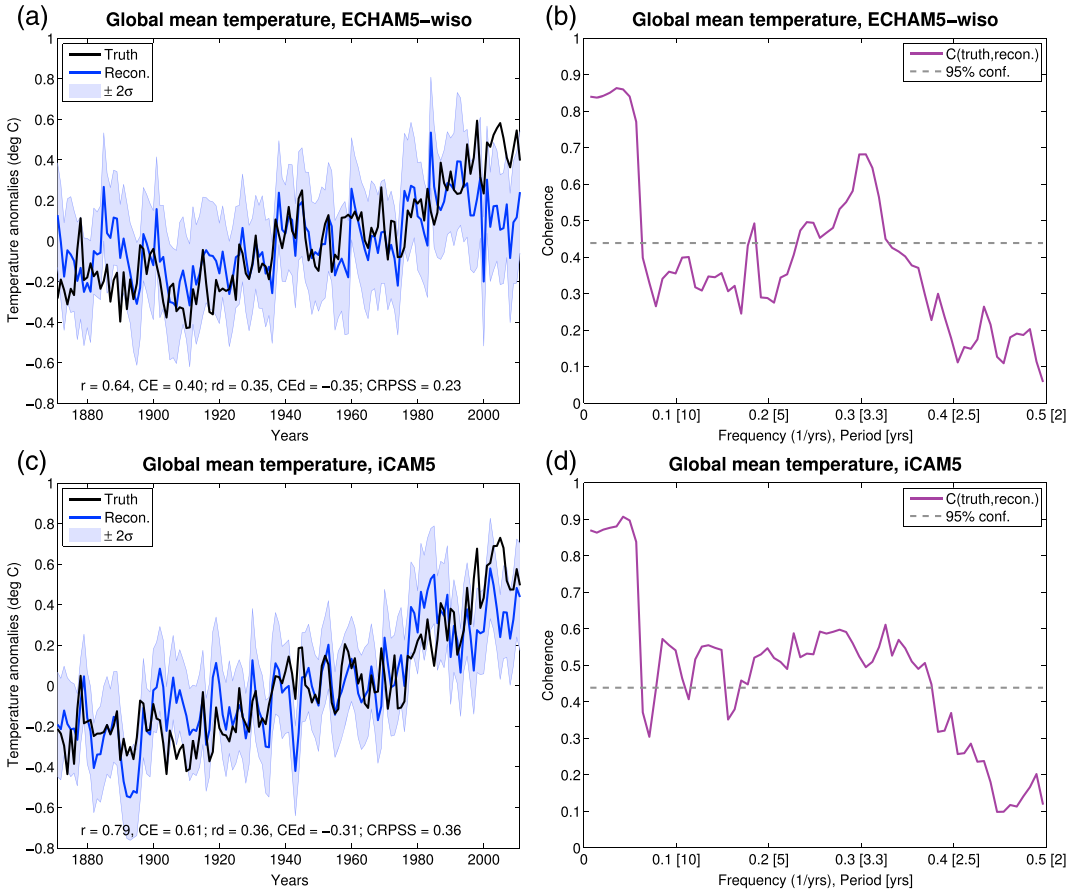

Figure 4. (a and c) Global mean temperature pseudoproxy reconstruction (cf. Figure 2). (b and d) Coherence of the reconstruction mean with the true times series.

reconstruction shows essentially no skill, even at longer timescales, (Figures 3b and 3d); later in this section we explore the reason for this result.

Because we are interested in exploring the full spatial extent to which ice core isotopes can inform past climate, we also reconstruct the global mean temperature (Figure 4); though we do not have a strong a priori reason for expecting that ice core isotope records in polar regions will skillfully reconstruct the annual mean global temperature. We note that the drop in skill for Figures 4a and 4c during the last decade for both models is likely a result of the experimental construction where the prior for that segment (years 1841–2000) is not sufficiently representative of the last decade's warmth. This situation is somewhat unique to this particular reconstruction scenario: for reconstructions over the last two millennia, the global climate changes are small relative to those driven by postindustrial global warming and using a range of last millennium simulations for the prior will very likely provide a sufficiently representative prior [e.g., *Otto-Bliesner et al.*, 2016]. By comparing the regular skill scores with the skill of the detrended reconstruction, we can see that much of the annual r and CE skill comes from the trend. Also, the positive mean CRPSS over the entire time frame indicates a probabilistic reconstruction that is better (for this metric) than an uninformed prior. Also, the spectral coherence is near or below the 95% confidence level at most frequencies, and if the time series are first detrended before computing the coherence, we do not see the large decadal-scale coherence increase at the low-frequency ends of Figures 4b and 4d. We therefore conclude that while ice core isotopes can potentially be used to reconstruct a centennial global mean trend, in principle they do not provide enough high signal-to-noise information to reconstruct global mean temperature on annual and decadal time scales.

Next, we consider the spatial skill of the pseudoproxy reconstructions for temperature, geopotential height at 500 hPa (Z500), and total precipitation. Figure 5 shows the skill of the 2 m air temperature reconstructions, for both GCMs. Each of the spatial skill metrics are computed either with the posterior mean (*r* and CE) or the full posterior distribution (CRPSS) of the grid point temperature series over the entire reconstruction

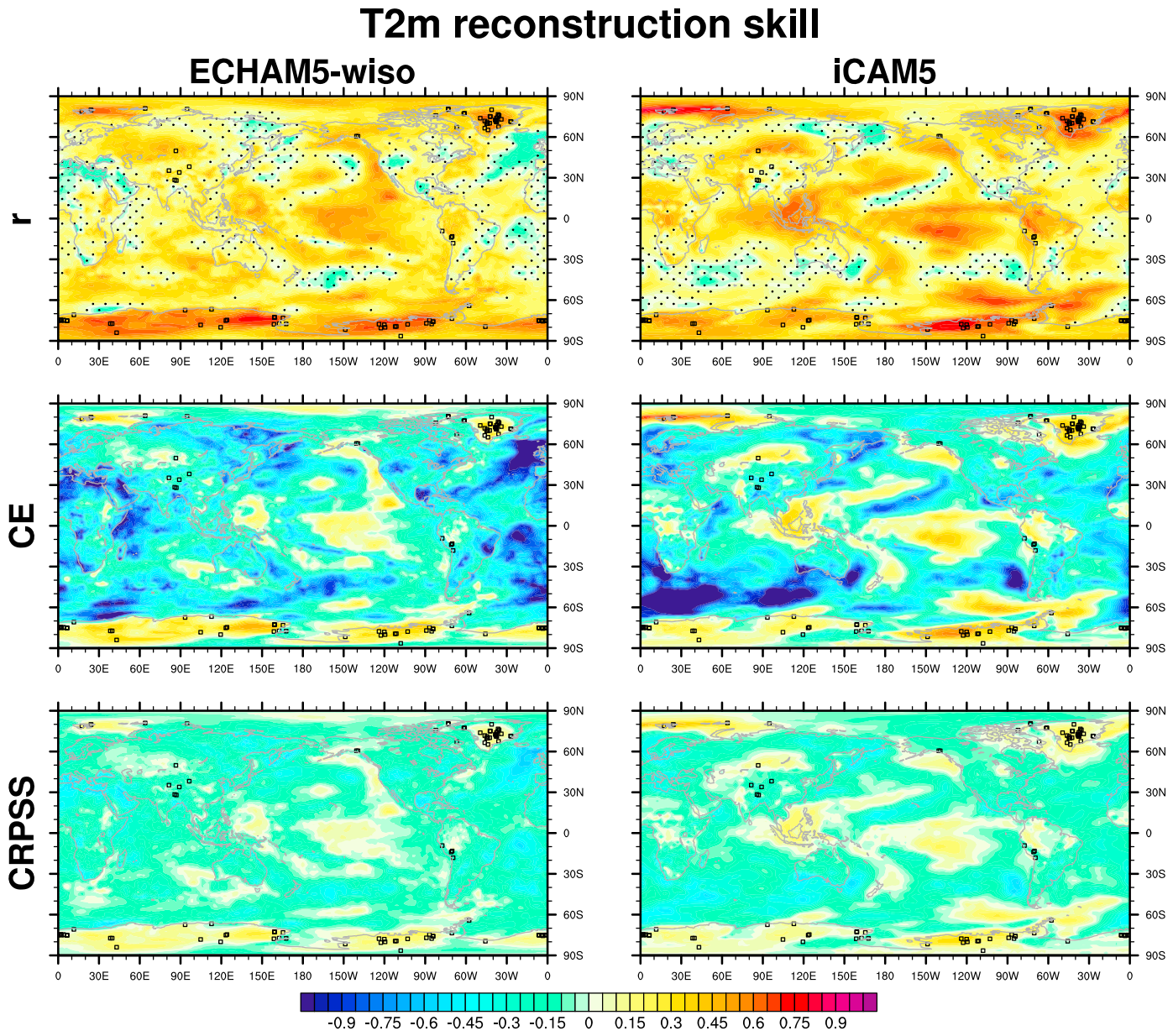

Figure 5. Spatial 2 m air temperature pseudoproxy reconstruction skill with (top row) correlation (*r*), (middle row) coefficient of efficiency (CE), and (bottom row) continuous ranked probability skill score (CRPSS); both CE and CRPSS are only shown down to −1. All skill metrics are computed for the length of the reconstruction, 1871–2011, and after detrending all of the reconstructed and true spatial time series. (left column) Reconstruction skill for ECHAM5-wiso; (right column) skill for iCAM5. For the correlation plots, stippling indicates correlations with *p* values greater than 0.05. The locations of the pseudo ice cores are indicated by the open squares.

after detrending the underlying grid point time series (removing the linear trend). Similar to the regional mean temperature reconstructions in Figures 2 and 3, there is clear skill near the pseudo ice cores, especially in Greenland and Antarctica, but not consistently in other locations. The possible exceptions to this include the Southern Ocean and the central Pacific. If *r* and CE are computed at 10 year averages (not shown), the ice cores tend not to add any new areas of high skill but simply amplify skill in a given location. Corresponding annual spatial results for Z500 shown in Figure 6 have similar features, with highest skill nearest the proxies. The pattern of positive skill across the tropics in both T2m and Z500 may directly reflect an ENSO teleconnection: the Rossby wave train-like pattern emanating poleward from the central Pacific in Figure 6 (most noticeable in iCAM) is very similar to the known tropospheric response to ENSO variability [e.g., *Trenberth et al.*, 1998]. Such a pattern has been linked to Andean ice core $\delta^{18}O$ [*Vuille et al.*, 2003].

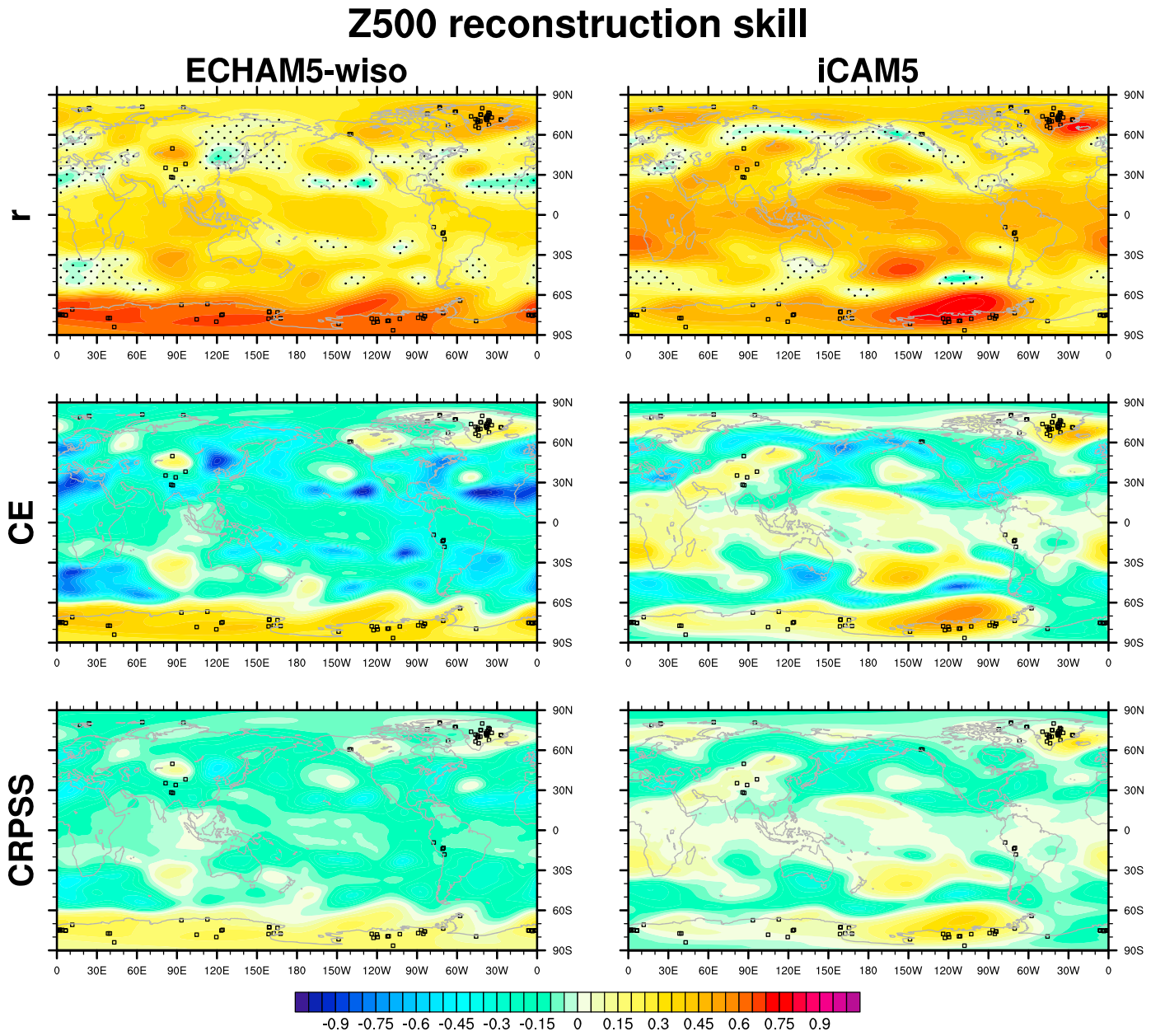

Figure 6. Geopotential height at 500 hPa (Z500) pseudoproxy reconstruction skill (cf. Figure 5).

Figure 7 shows the spatial skill for precipitation. Unlike temperature and Z500, for precipitation the skill is either nonexistent or localized to sites of high annual snow accumulation where the isotopic signal is less diffused, such as coastal and West Antarctica. However, as an exception, there is apparently skill for precipitation in the central Pacific and over the Amazon basin where T2m and Z500 skill are also good; this likely reflects the influence of ENSO on ice cores in the polar regions [*Schneider and Steig*, 2008; *Steig et al.*, 2013] and the Andes [*Hoffmann et al.*, 2003; *Samuels-Crow et al.*, 2014].

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 8 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 equation (2), we note that larger covariance values between the prior and the prior-estimated observations

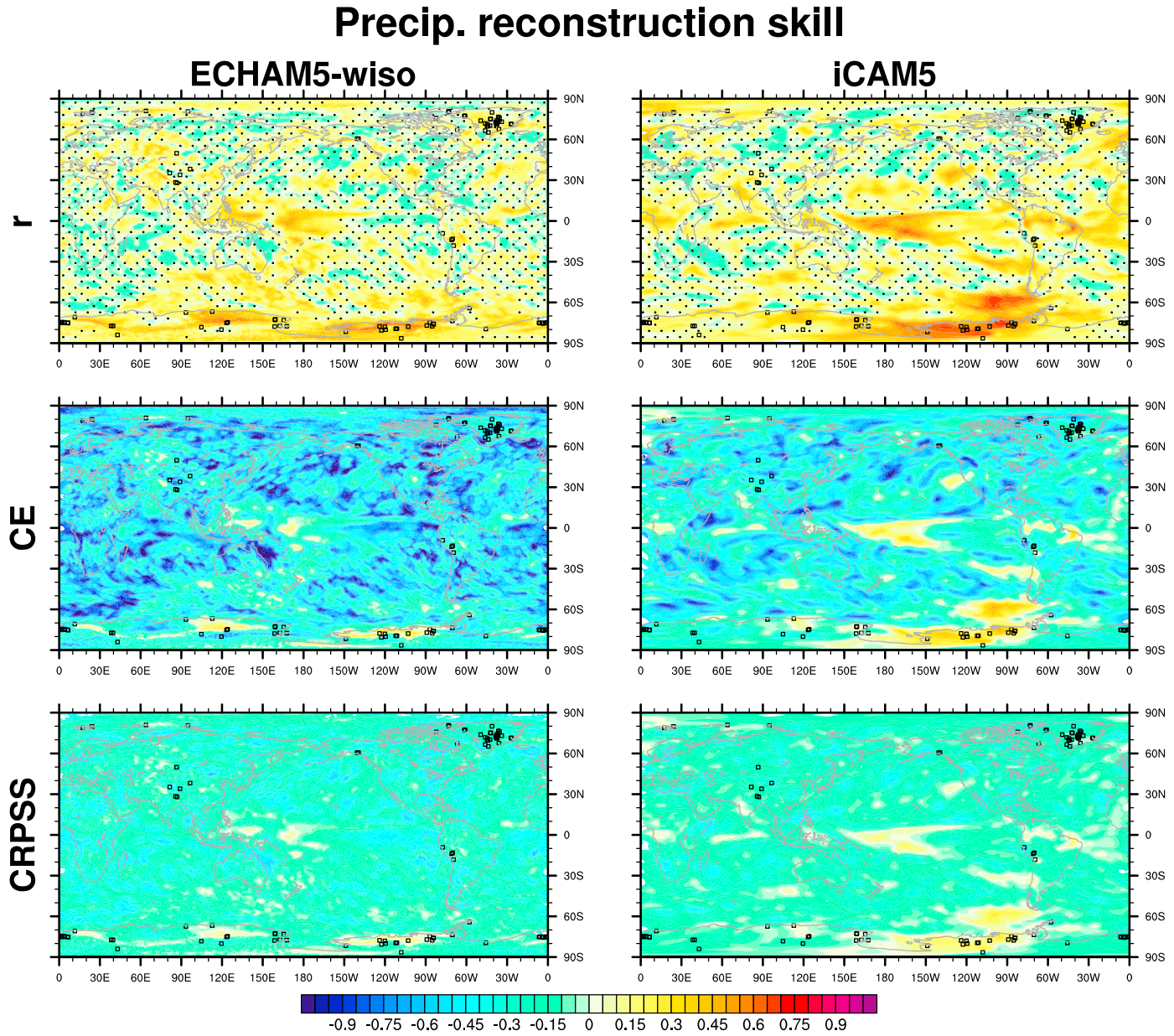

Figure 7. Precipitation pseudoproxy reconstruction skill (cf. Figure 5).

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 isotope record at one location has very little to say about precipitation on an annual time scale at other spatial locations. We note that snow accumulation could be a better proxy of precipitation than isotopes, though reconstructions using accumulation would still be affected by the reduced correlation length scale of precipitation.

Because the amount of diffusion within the firn has the potential to 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 in section A2; in summary, we do not find that diffusion uncertainty in the ice core PSM affects the spatial reconstruction skill results.

We also performed a set of experiments to determine if the skill in the central Pacific is dependent on any particular regional group of pseudo ice cores. In each reconstruction experiment we removed a different regional network, viz., Greenland, West Antarctica, and the Andes, and compared the spatial skill of temperature and Z500 in the tropics (not shown). We found that no single location was responsible for the tropical skill: removal

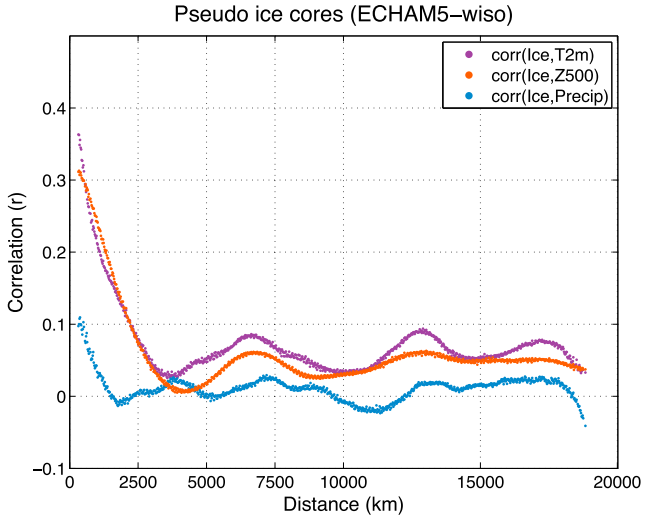

Figure 8. Correlation as a function of distance between the PSM-derived pseudo ice cores and the spatial reconstructed variables. For each pseudo ice core, detrended correlations are computed for each grid point time series of the spatial fields, then these correlations are binned by distance, and finally averaged over all pseudo ice cores.

of the Andes, West Antarctic, or Greenland proxies did not lead to clear reductions in tropical skill (though local to the removed regional network there was significant loss of skill, with much of the lost skill for the Andes happening over the Amazon basin). This result suggests that many of the pseudo ice cores in this network contain some amount of useful, nonlocal information that they can contribute in the absence of other ice cores.

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 initial information about nonlocal climate? Or would perfect instrumental observations of climate at the ice core locations be able to reconstruct more nonlocal climate? To test this question in the pseudoproxy framework, we generated temperature-only pseudoproxies at the ice core locations with a signal-to-noise ratio of 5 (typical annual proxies have a local signal-to-noise ratio of about 0.3 [*Wang et al.*, 2014]), which would be characteristic of thermometer measurements, and then we performed

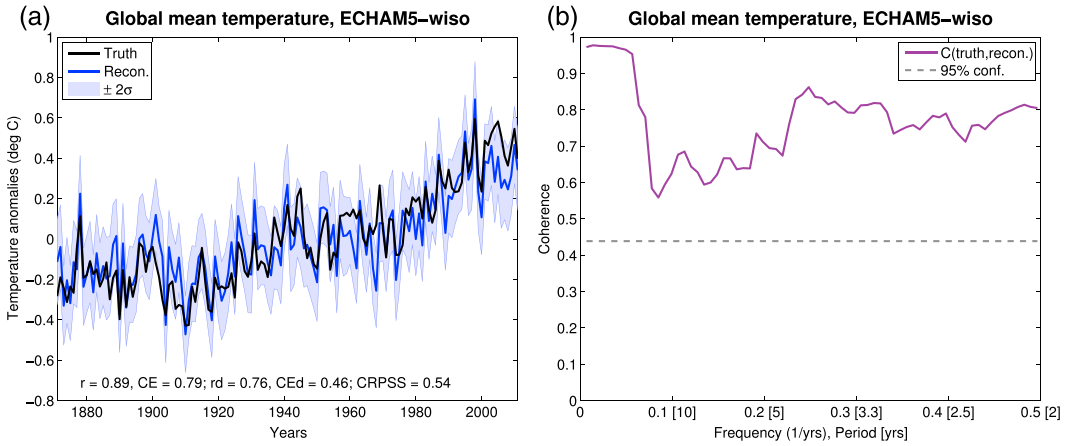

Figure 9. (a) Global mean temperature reconstruction based on purely temperature pseudoproxies at all ice core locations. Skill metrics are indicated on the lower left as correlation (*r*), coefficient of efficiency (CE), and both of these again after the true and reconstructed time series have been detrended (indicated by "d"); additionally, the continuous ranked probability skill score (CRPSS) assesses both the mean and uncertainty relative to the uninformed prior baseline. (b) Coherence of the reconstruction mean with the true times series, along with a 95% confidence level.

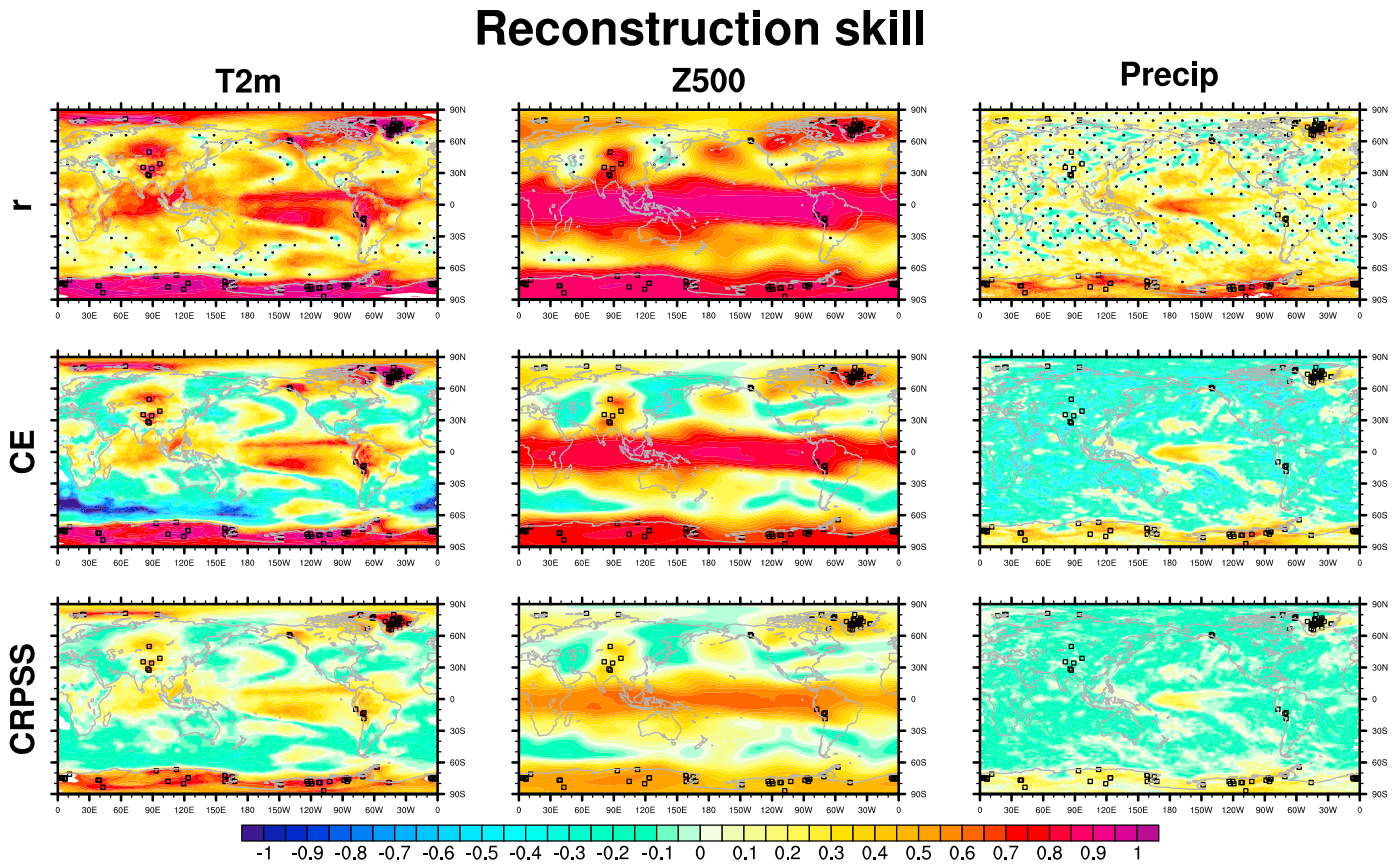

Figure 10. Reconstruction skill for each climate variable (left to right columns) based on purely temperature pseudoproxies at all ice core locations. The skill metrics are (top row) correlation (*r*), (middle row) coefficient of efficiency (CE), and (bottom row) continuous ranked probability skill score (CRPSS); both CE and CRPSS are only shown down to −1. The skill metrics are computed after detrending the underlying true and reconstruction data. For the correlation plots, stippling indicates correlations with *p* values greater than 0.05. The locations of the temperature pseudoproxies are indicated by the open squares.

reconstructions with ECHAM5-wiso leaving the rest of the reconstruction process the same. Figure 9 shows the global mean temperature reconstruction (analogous to Figure 4), and Figure 10 shows the spatial skill of the reconstructions for temperature, Z500, and precipitation. These two figures show that given sufficiently accurate observations (or sufficiently many noisy observations in the same locations), one could reconstruct global mean temperature and even consistent tropical climate and local precipitation with high skill. This implies that processes inherent to the proxy system itself blur the desired climate signal for such reconstructions. Interestingly, these results also show that even direct temperature observations in Greenland provide insufficient information to skillfully reconstruct European climate on annual timescales (near zero or negative CE and CRPSS over nearly all of Europe, Figure 10).

3. Real Proxy Experiments

3.1. Experimental Framework

For the real reconstructions (reconstructions using real ice core isotope data) we employ a similar framework as that in the pseudoproxy reconstructions, with the following differences. We used 38 annual ice core $\delta^{18}O$ proxies from the latest version of the PAGES2k database [*Emile-Geay et al.*, 2015], with the locations indicated by "PAGES2k v2" in Figure 1 and their temporal availability shown in Figure 11. Because the proxy data is composed of only ice cores, we did not standardize the raw isotopic measurements at any point in the reconstructions, though the time mean was removed from both the proxies and the model estimate of the proxies, $\mathcal{H}(\mathbf{x}_b)$. We estimated analytical errors in the isotope measurements as 0.1/‰. As part of the Monte Carlo reconstruction process, we performed six different types of reconstructions over the period 1900–2000; each of the experiments paired a climate model simulation, ECHAM5-wiso or iCAM5, with one of three choices

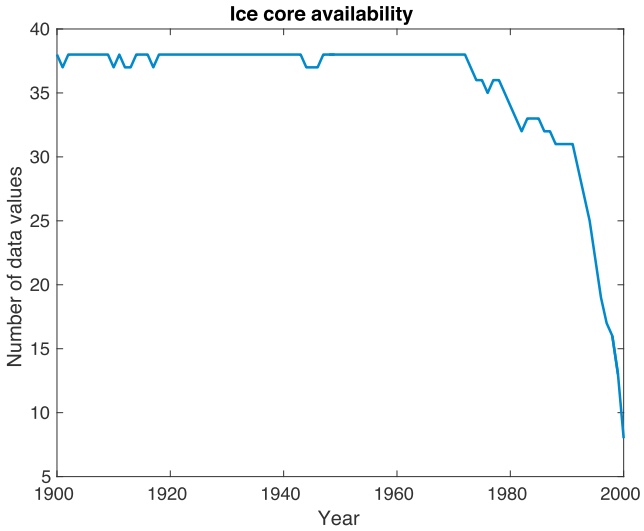

Figure 11. Ice core data values available in the PAGES2k-V2 data set for each year of the reconstruction. Thirty-eight ice cores are used in the reconstruction, but any given year may contain missing data or may not cover the entire 1900–2000 year reconstruction interval.

of diffusion in the PSM: normal, half diffusion, and no diffusion. We did not use more than normal diffusion because the proxy system model tends to be too diffusive for many low-accumulation locations [*Dee et al.*, 2015]. We repeated the reconstructions 100 times with all six combination pairs of GCM and PSM, sampling 75% of the proxy network for each iteration. These 600 individual reconstructions, each with an ensemble of size 140 (140 annual states drawn from the historical simulations), were then pooled together to create a "grand" posterior ensemble; this is similar to the pseudoproxy experiments except that all the reconstruction variations are pooled together into this grand posterior ensemble. The figures use both the mean of this ensemble and error estimates based on the grand ensemble spread. Note that the reconstructions overlap in time with the ECHAM5-wiso and iCAM5 simulations that were run from 1871 to 2011 using historical SST boundary conditions. However, in the off-line reconstruction process the only time axis information that the prior ensembles are aware of come from the proxies (unlike some other DA-based reconstructions that are additionally given year-specific forcing information [e.g., *Matsikaris et al.*, 2015]). Ideally, one would use simulations for the prior that did not have temporal overlap with the reconstruction, though these were the only extant simulations available that meet the criteria of being isotope enabled, high resolution, and at least centennial in length.

3.2. Reconstruction Results

Analogous to the pseudoproxy results, Figure 12 shows the mean temperature reconstructions for Greenland, the British Isles, and the globe, with Berkeley Earth [*Rohde et al.*, 2013] as verification; we chose Berkeley Earth for temperature verification because of its high spatial resolution and more extensive spatial coverage, particularly in the Arctic, as compared to HadCRUT4.3, for instance (though the skill results are not sensitive to the choice of instrumental verification data where they overlap spatially). We note that neither product can be considered reliable in the Antarctic due to the lack of instrumental data during the reconstruction period [*Nicolas and Bromwich*, 2014], and thus, we have chosen to focus on the reconstruction over Greenland. The results in Figure 12 share several features with the pseudoproxy experiments, yet also differ in important ways, such as the overall reduced levels of skill. However, in the 5–10 year coherence band two real proxy reconstructions reach similar levels as the upper bound pseudoproxy estimates, cf. the Greenland mean temperature in Figure 2b with Figure 12b and the global mean temperature in Figures 4b and 4d with Figure 12f. Similar to the pseudoproxy experiments, Greenland mean temperature is reconstructed better than for the British Isles or for the global mean at both annual and longer timescales.

Figure 13 (left column) 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

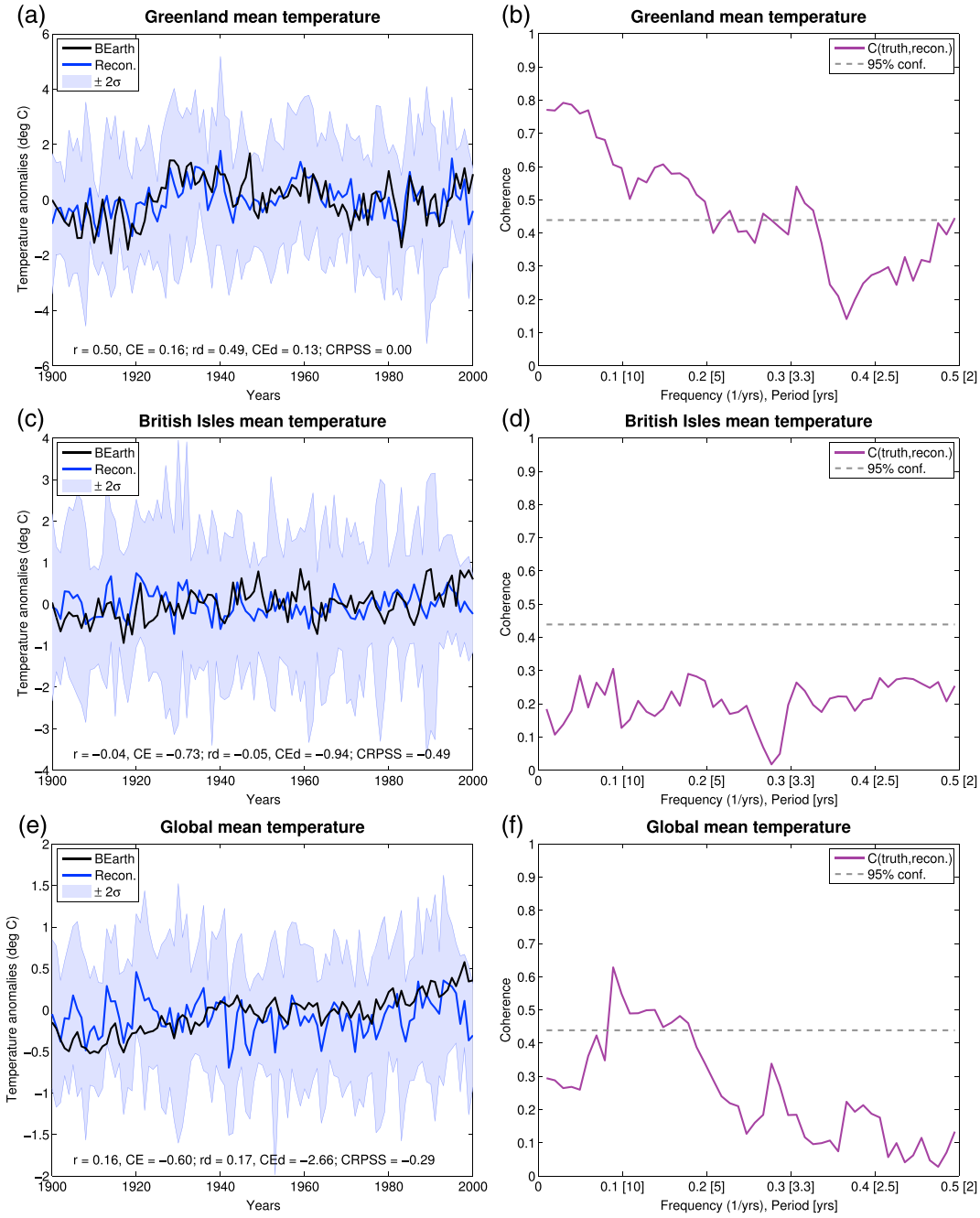

Figure 12. Temperature reconstructions using real ice core $\delta^{18}O$ proxy data and coherence of the reconstruction mean with the Berkeley Earth time series (BEarth). (a and b) Greenland mean temperature reconstruction, (c and d) British Isles mean temperature reconstruction, and (e and f) global mean temperature reconstruction.

results shown in Figure 5, 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 instrumental observations are rare in Antarctica prior to the late 1950s, whereas over Greenland such data extend back into the 1800s. This means that we have fewer and less-reliable data in Antarctica with which to validate the PSMs and to verify the reconstructions. As in the pseudoproxy experiments, there is a small amount of skill in the tropical Pacific. Figure 13 (middle and right columns) 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

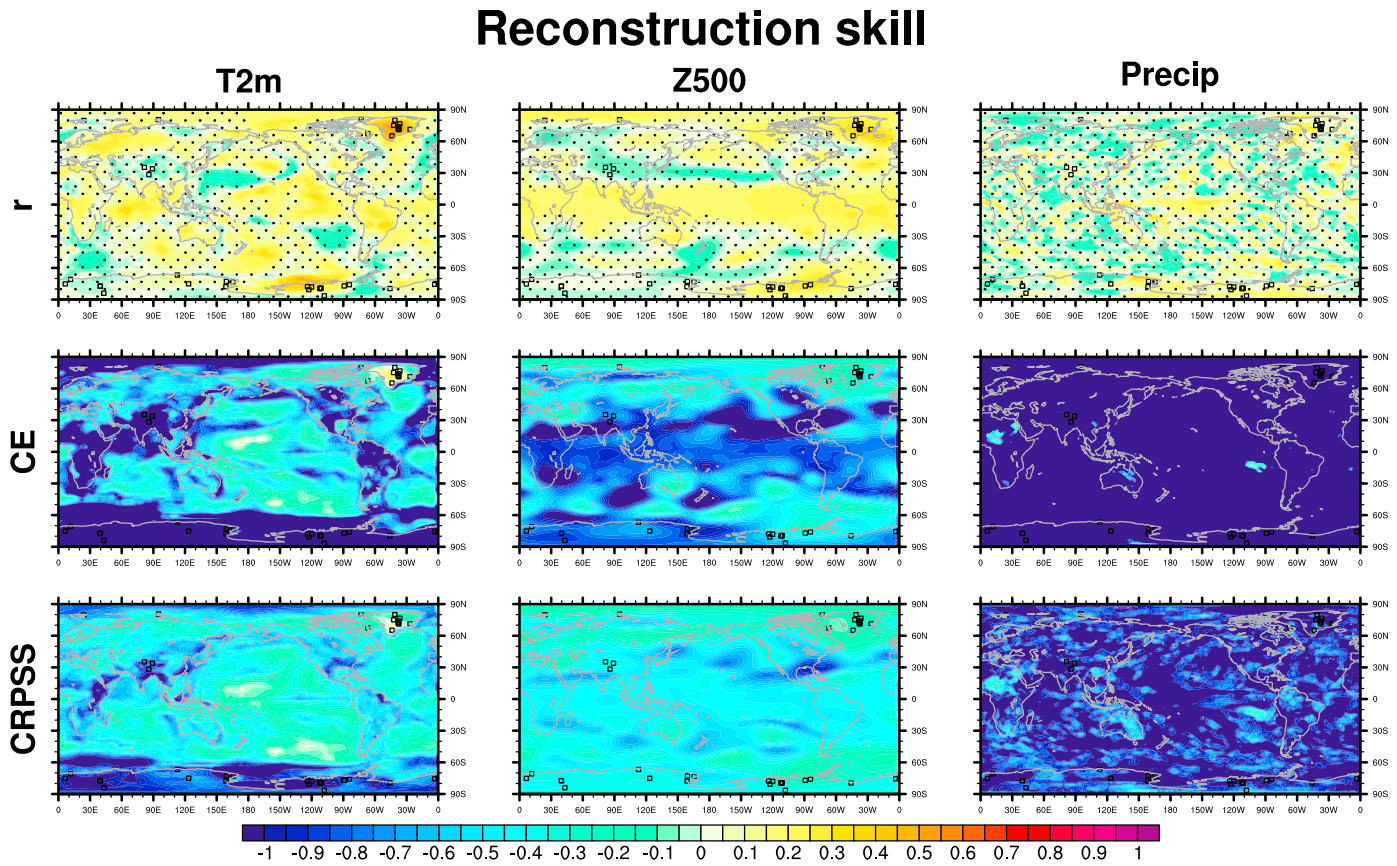

Figure 13. Spatial reconstruction using real proxies and PSM-based estimates of the proxies. (left to right columns) Different reconstructed variables. The skill metrics are (top row) correlation (*r*), (middle row) coefficient of efficiency (CE), and (bottom row) continuous ranked probability skill score (CRPSS); both CE and CRPSS are only shown down to −1. The skill metrics are computed after detrending the underlying true and reconstruction data. For the correlation plots, stippling indicates correlations with *p* values greater than 0.05. The locations of the temperature pseudoproxies are indicated by the open squares.

the pseudoproxy results (local skill as well as a band of skill in the tropics), though the skill is generally quite low. Precipitation as shown in the third column of Figure 13 appears to not be robustly reconstructed anywhere. The key results from these real reconstructions are that they show local skill to be most robust together with the possibilities of nonlocal reconstruction skill in the tropics (in agreement with the pseudoproxy experiments shown in section 2). It is not surprising to find reduced skill in the real proxy reconstructions relative to the pseudoproxy reconstructions (cf. Figures 13 and 5–7), and additional steps could be taken in an attempt to improve the real reconstructions, such as bias correcting the GCM output variables, optimizing the PSM parameters, increasing the size of the prior ensembles, and increasing the resolution of the GCM simulations, though these are all beyond the scope of this study.

Additionally, we performed experiments employing a "linear statistical PSM" as a point of reference for the reconstructions already shown in Figures 12 and 13. This linear statistical PSM is a simple linear fit of the $\delta^{18}O$ to local temperature, as used in a recent paleoclimate reanalysis [*Hakim et al.*, 2016]. Fitting ice core data to local temperature as part of a reconstruction has been standard practice for several decades [*Dansgaard*, 1964; *Jouzel et al.*, 1997]; thus, linear statistical modeling of ice core isotopes is important from a historical perspective in addition to its practical advantages of being much less labor intensive to implement. The specific method used to produce these linear statistical real proxy reconstructions is discussed in section A3. The results of these reconstructions are shown in Figures 14 and 15. Focusing on a comparison between the time series reconstructions of both types of reconstructions, Figures 12 and 14, we see that both are of comparable quality, with better skill metric values (*r*, CE, and CRPSS) for the linear statistical approach, while the physics-based PSM reconstructions tend to have better or comparable coherence (e.g., comparing Greenland mean temperature coherence in Figures 12b and 14b). The spatial skill maps also show similar local skill. These results are in agreement with the pseudoproxy-only reconstruction results of *Dee et al.* [2016], who found

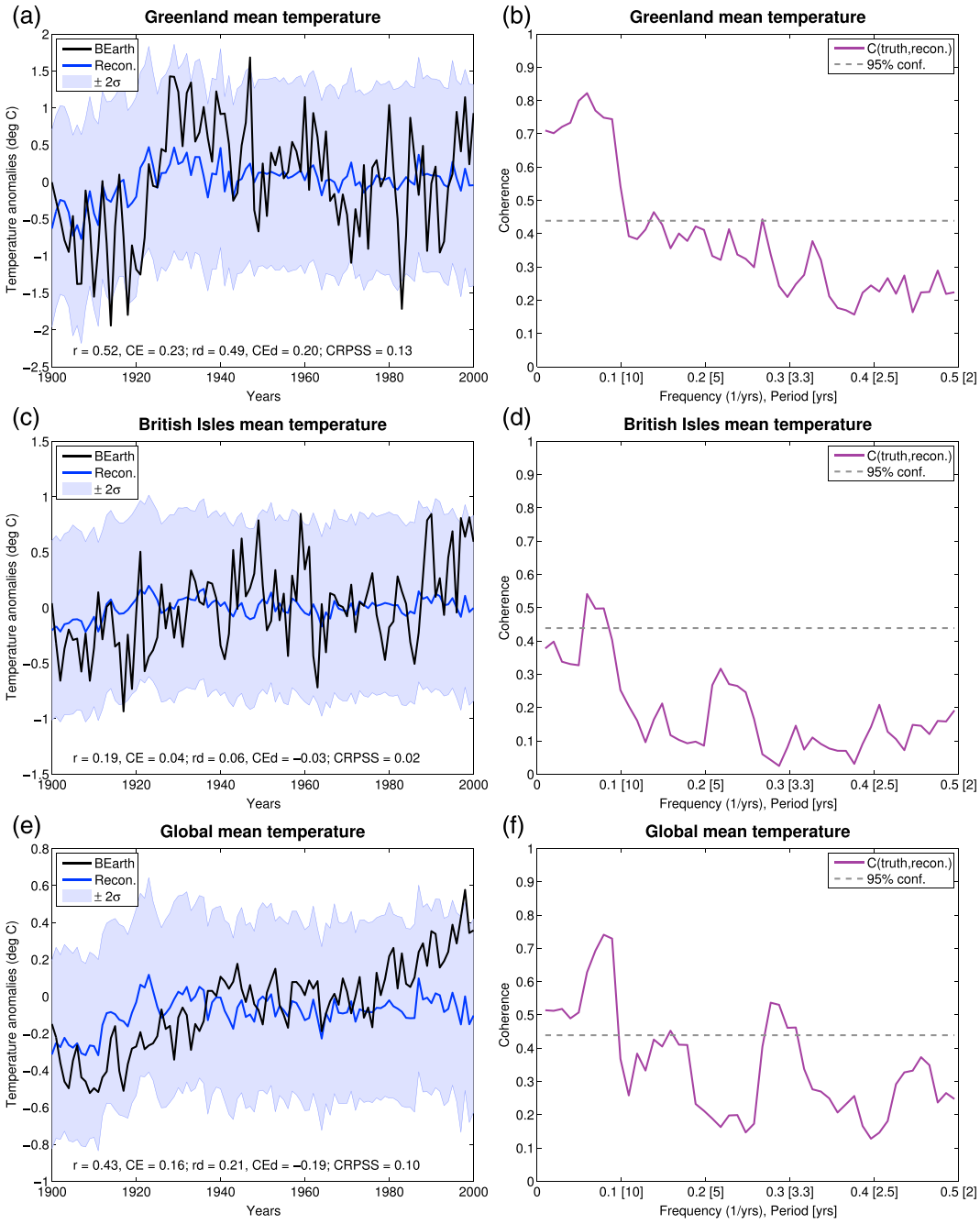

Figure 14. Greenland, British Isles, and global mean temperature real proxy reconstructions using the linear statistical PSM. (a, c, and e) The reconstructions in blue with $\pm 2\sigma$ error estimates and (b, d, and f) the coherence of the mean reconstruction (dark blue line) with the observational data set Berkeley Earth (BEarth; black line).

comparable skill between a linear statistical approach and a similar physics-based approach to that used here of isotope-enabled climate simulations in conjunction with ice core diffusion modeling. In making the comparisons between Figures 12 and 13 and Figures 14 and 15 it is critical to keep in mind that the physics-based PSMs used here have not been optimized or had their inputs bias corrected, while a linear optimization and bias correction are naturally part of the linear statistical approach. Therefore, the comparison between reconstruction techniques shown here likely underestimates the skill of the physics-based PSMs, which have the potential to capture more complex climatic and nonclimatic dependencies than linear statistical methods; as PSMs are refined and improved, it will be important to continually evaluate the relative performance of different methods for establishing climate-proxy relationships.

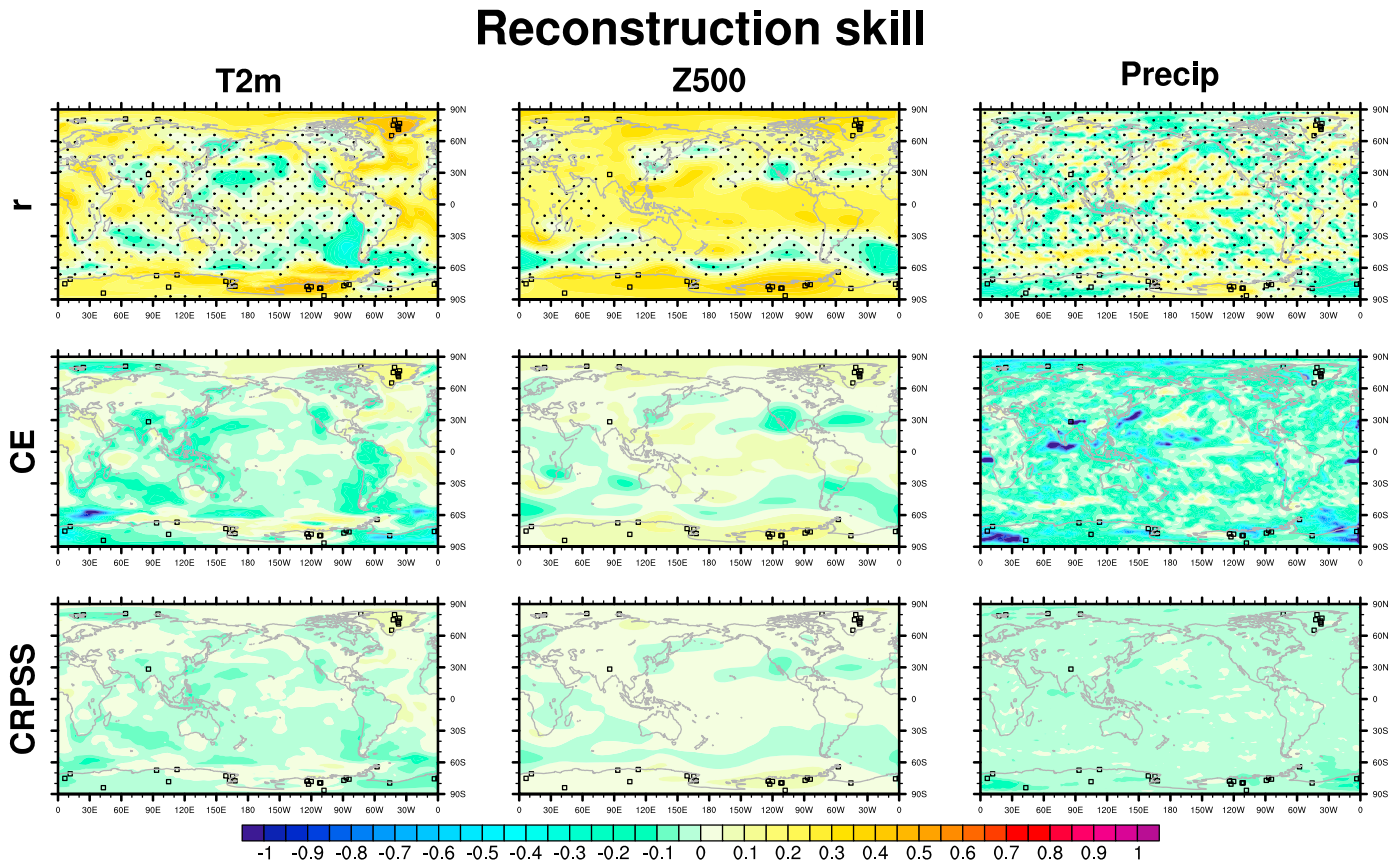

Figure 15. Spatial reconstruction skill using the linear statistical PSM and real proxies.

4. Conclusions

Through a series of pseudo and real ice core isotope reconstructions we have estimated the spatial and temporal extent of how ice core isotopes can inform on 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). Furthermore, the pseudoproxy experiments suggest that ice cores can in principle be used to reconstruct nonlocal variables, particularly in the tropics. Though this nonlocal skill is spatially variable and even if ice core locations have a known seasonal link to other regions, this does not guarantee skillful ice core isotope-based reconstructions of these other linked regions: as a test case, all sets of experiments show that information in Greenland ice core isotopes (in addition to other global ice cores) is uninformative for climate reconstructions over the British Isles and Europe; while both Greenland and Europe lie within the influence of modes of variability such as 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 reconstruction variables of surface temperature, geopotential height at 500 hPa, 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 core isotopes over most of the globe on annual timescales.

The real ice core proxy reconstructions presented in the second half of this paper represent the first ice core-based reconstructions to include PSMs together with data assimilation in the reconstruction process. These reconstructions compare most favorably with the idealized pseudoproxy experiments and a linear statistical real proxy reconstruction technique at the 5–10 year timescale. While this is only a first foray

into a novel approach, we suggest that with future refinements, this reconstruction approach could provide improved ice core-based reconstructions.

Appendix A: Additional Experiments and Figures

A1. Climate Model Evaluation

Figures A1–A4 provide a general evaluation of the ECHAM5-wiso and iCAM5 simulations against observations. Figure A1 compares the annual mean $\delta^{18}O$ of precipitation in a global observational data set, the global network of isotopes in precipitation (GNIP), with the simulated $\delta^{18}O$ of precipitation. From this figure, we see that for both models the simulated water isotope ratios are in broad 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 $\delta^{18}O$ 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 A2 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 A3 shows the mean accumulation of the model simulations together with the ERA-Interim reanalysis product, where data were taken from the nearest grid cells to the ice core locations. We note that both models correctly identify low-, moderate-, and high-accumulation sites, with the absolute value of the relative errors for most locations less than 1 (where relative error is defined as $\delta x = x_0/x_t - 1$, with x_0 as the modeled value and x_t as the observed value according to ERA-Interim). However, in Figure A3b we see that there is a general tendency to have too much

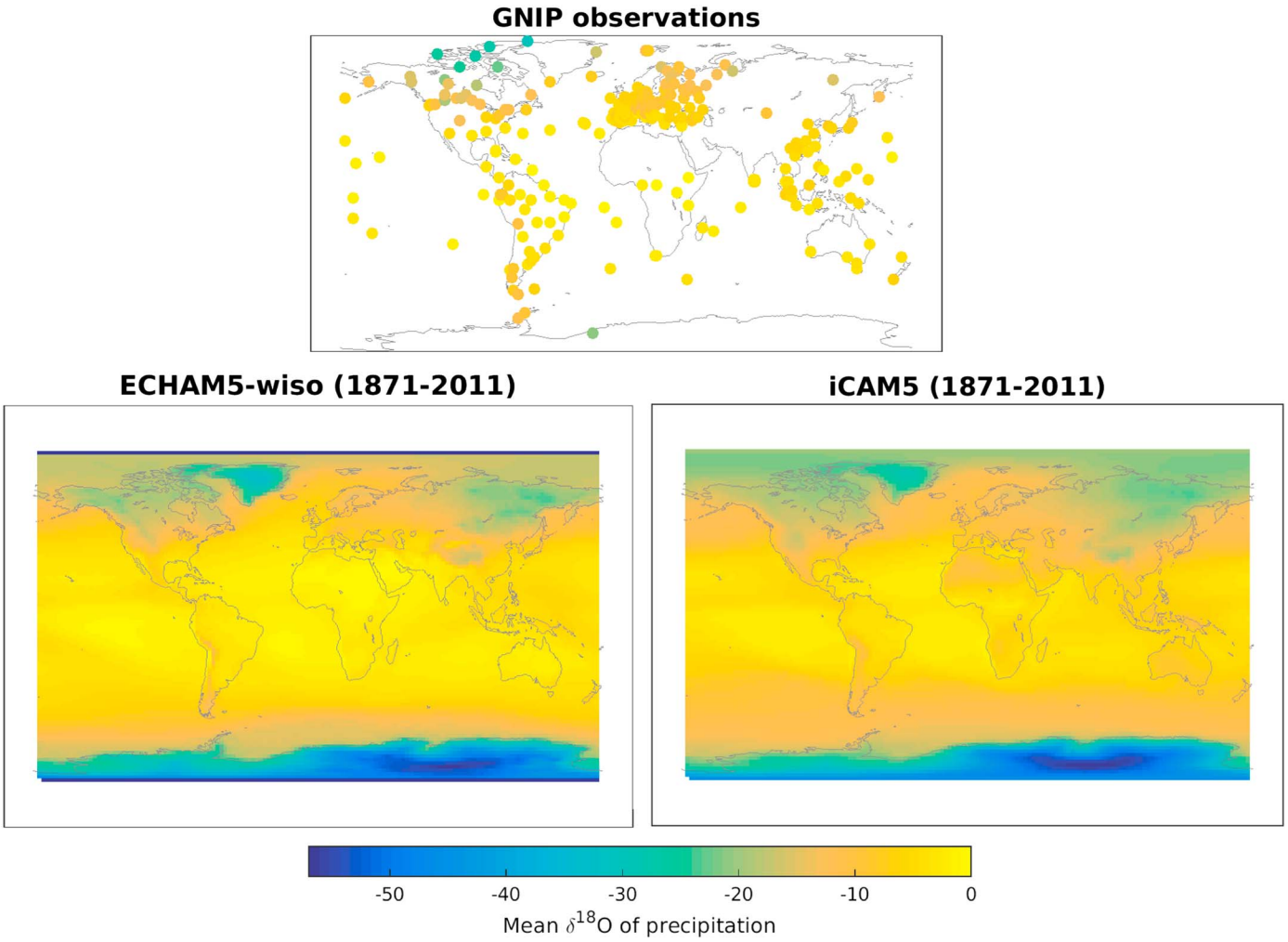

Figure A1. Global network of isotopes in precipitation (GNIP) and modeled $\delta^{18}O$ of precipitation in the two models employed in this study. The isotope values are the time mean over the available observations and the model simulation years of 1871–2011.

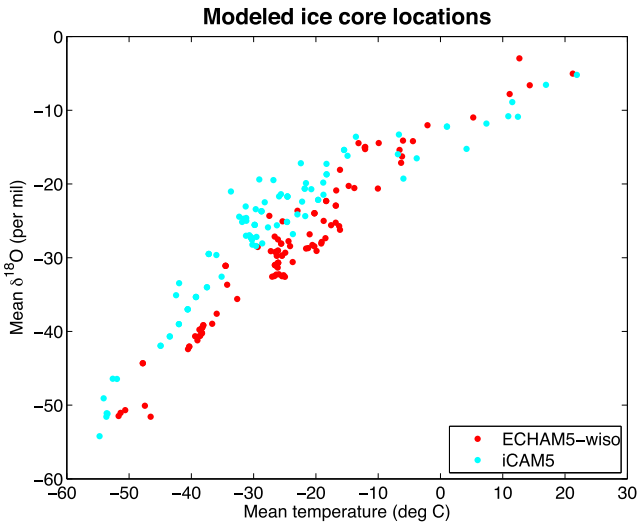

Figure A2. The relationship between annual mean $\delta^{18}O$ of precipitation and annual mean temperature at all ice core locations for both climate models.

accumulation at ice core sites. Figure A4 shows the spatial distribution of these relative errors. Neither of the models show spatially coherent patterns of bias, with the exception of ECHAM5-wiso generally simulating too much accumulation across East Antarctica.

A2. Diffusion Experiments

As mentioned in the text, the amount of diffusion within the firn has the potential to impact the isotopic signal within our modeled ice cores. Thus, we also performed two end-member diffusion experiments and compare these with the control or normal diffusion experiments. Figures A5–A7 summarize the spatial reconstruction skill for 2 m temperature, Z500, and precipitation for a control, $1/2\times\sigma$ and $2\times\sigma$ diffusion experiments, also for both GCMs. In these figures we can see that the largest differences are between the GCMs, with relatively small

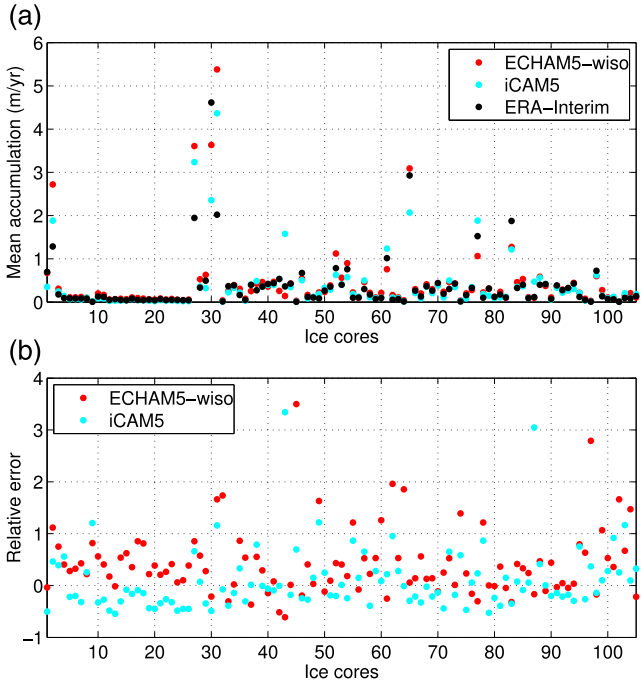

Figure A3. (top) Modeled mean annual accumulation of total precipitation at all ice core locations for the two models used in this study compared against the ERA-Interim reanalysis. (bottom) Relative error for ECHAM5-wiso and iCAM5 simulations, calculated against ERA-Interim.

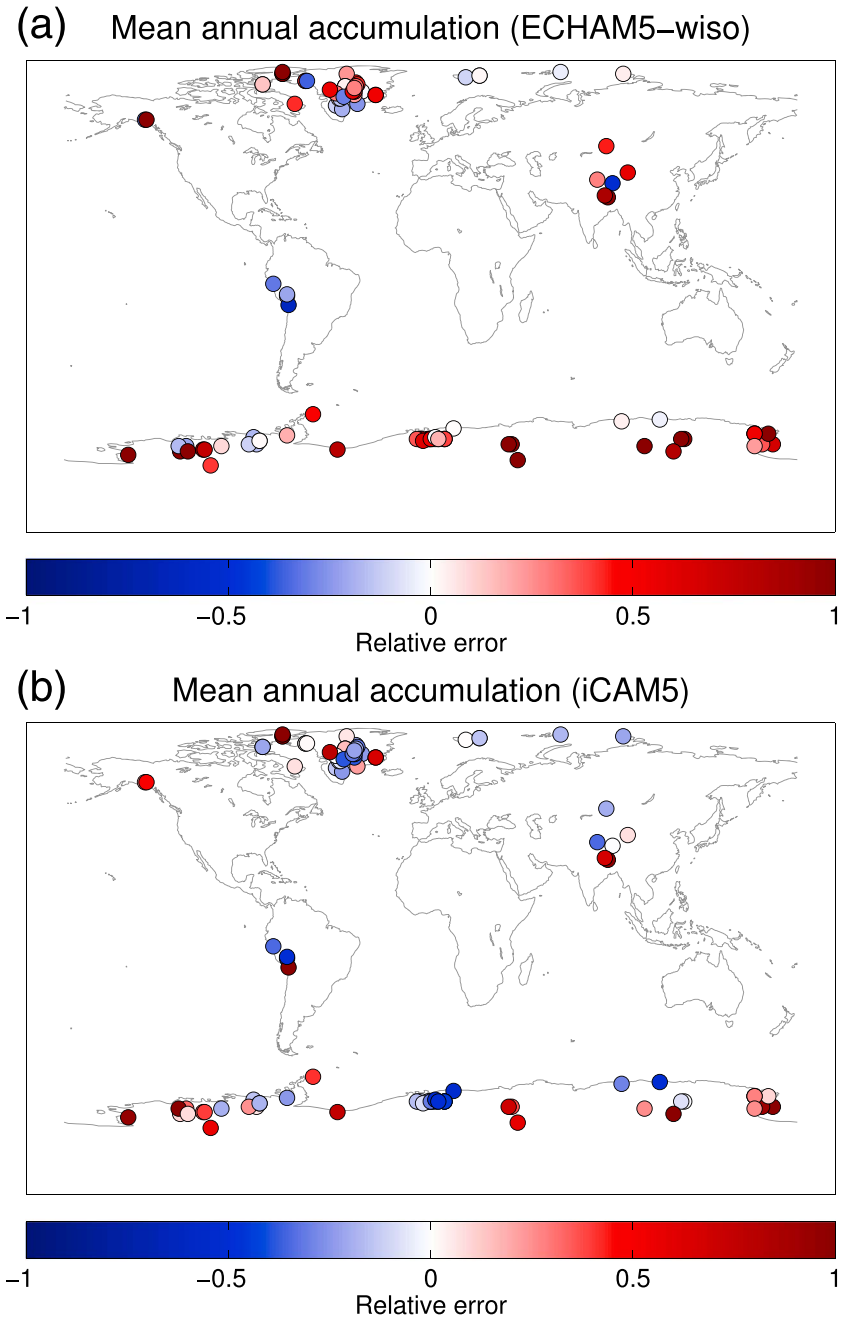

Figure A4. The relative error of GCM-modeled total precipitation accumulation at the pseudoproxy locations (same information as in Figure A3 but shown spatially). Reds indicate too much accumulation, while blues indicate too little accumulation relative to ERA-Interim.

changes in the distributions for different diffusion parameters; decreasing or reducing the normal diffusion by a factor of 2 does not change the nature of the reconstruction skill over these 141 year reconstructions.

A3. Linear Statistical Reconstruction Method

Additional real proxy experiments were performed using a simple linear statistical model, as in *Hakim et al.* [2016], rather than a physically based PSM. 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

$$p_i = \alpha_i + \beta_i T_i + \epsilon_i \,, \tag{A1}$$

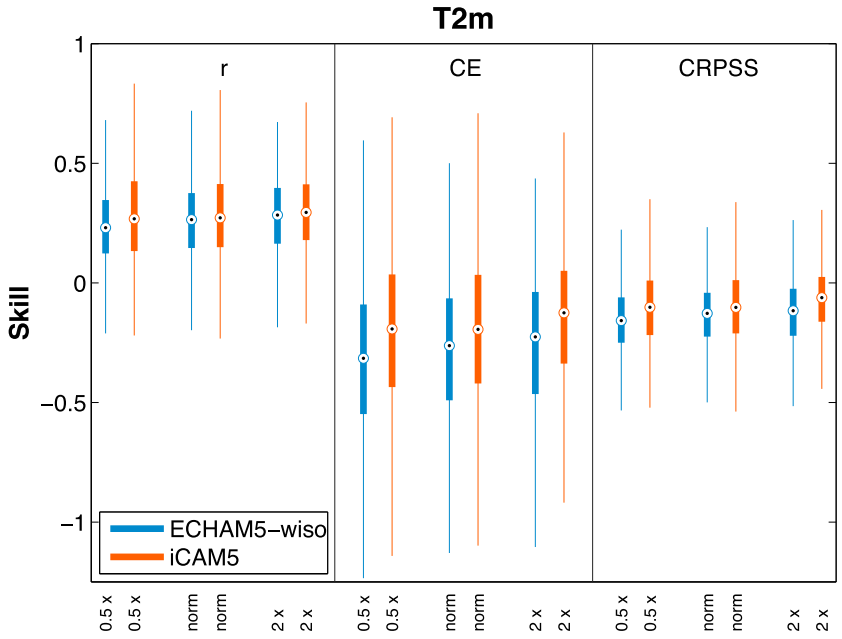

Figure A5. Box plot summaries of the distribution of spatial temperature pseudoproxy reconstruction skill as a function of GCM and ice core diffusion. The skill of six reconstructions, by model and diffusion type, is shown for the metrics correlation (*r*), coefficient of efficiency (CE), and continuous ranked probability skill score (CRPSS). The three ice core diffusion types are half (0.5 x), normal (norm), and double (2 x). The box plot whiskers extend to ±1.5 the interquartile range and outliers have been omitted for clarity.

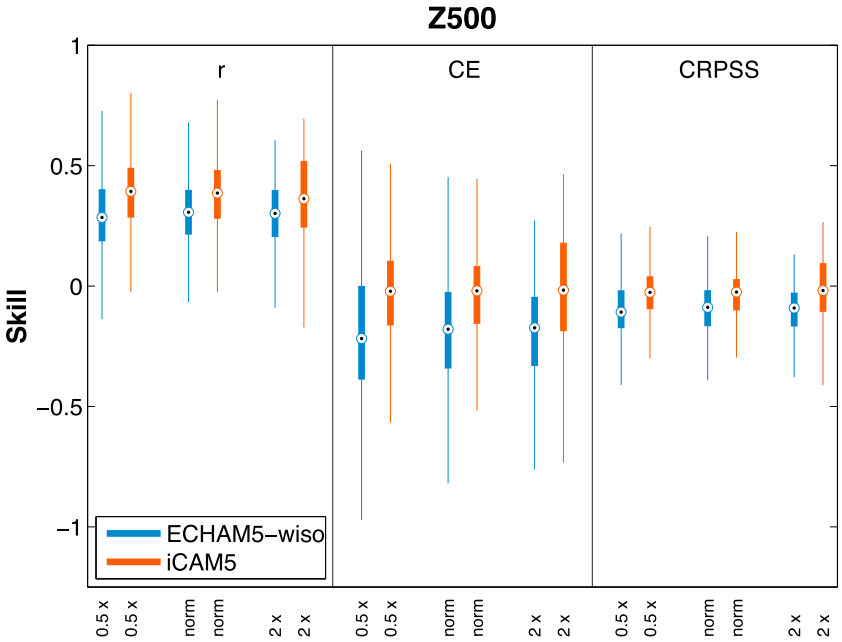

Figure A6. Box plot summaries of spatial Z500 pseudoproxy reconstruction skill as a function of GCM and ice core diffusion, as in Figure A5.

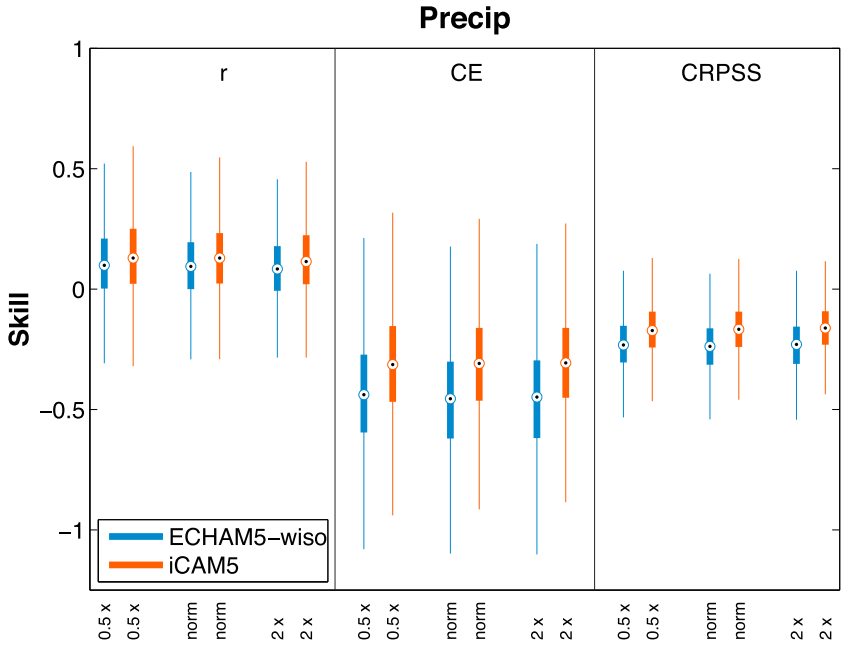

Figure A7. Box plot summaries of spatial precipitation pseudoproxy reconstruction skill as a function of GCM and ice core diffusion, as in Figure A5.

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

$$y_{ei} = \alpha_i + \beta_i \mathcal{T}_i \,, \tag{A2}$$

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 nonlocal climate variables, rather than embedding some of that nonlocal information within the PSM as done in the other ice core-based reconstructions presented here.

As part of the Monte Carlo reconstruction process, we performed a total of 400 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]. Four sets of 100 reconstructions were performed with a unique model calibration data pair, randomly sampling 75% of the proxy network for each iteration, as with the other experiments. We also used the same GCM simulations from ECHAM5-wiso and iCAM5 to construct the priors. The same 38 real $\delta^{18}O$ measurements were used here (Figure 1). Because the proxy data are 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 all the other reconstructions.

Acknowledgments
We would like to thank David Battisti and David Noone for helpful discussions and Jesse Nusbaumer for running the iCAM5 simulation. This research was supported by the following awards: National Science Foundation grants 1304263, 1503281, and 1504267 as well as the National Oceanic and Atmospheric Administration grant NA14OAR4310176. The authors declare no conflicting financial interests. The ice core isotope data are available from NOAA paleoclimatology: https://www.ncdc.noaa.gov/data-access/paleoclimatology-data or upon request from nsteiger@ldeo.columbia.edu. The PSM modeling code is available at https://github.com/sylvia-dee/PRYSM. The modeling output of the ECHAM5-wiso simulation is available upon request from nsteiger@ldeo.columbia.edu, while the modeling output of the iCAM5 simulation is available upon request from jesse.nusbaumer@nasa.gov.

References

Appenzeller, C., T. Stocker, and M. Anklin (1998), North Atlantic oscillation dynamics recorded in Greenland ice cores, *Science*, *282*(5388), 446–449.

Barboza, L., B. Li, M. P. Tingley, and F. G. Viens (2014), Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models, *Ann. Appl. Stat.*, *8*(4), 1966–2001, doi:10.1214/14-AOAS785.

Cuffey, K. M., and E. J. Steig (1998), Isotopic diffusion in polar firn: Implications for interpretation of seasonal climate parameters in ice-core records, with emphasis on central Greenland, *J. Glaciol.*, *44*(147), 273–284.

Dansgaard, W. (1964), Stable isotopes in precipitation, *Tellus*, *16*(4), 436–468, doi:10.1111/j.2153-3490.1964.tb00181.x.
Dee, S., J. Emile-Geay, M. N. Evans, A. Allam, E. J. Steig, and D. Thompson (2015), PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, *J. Adv. Model. Earth Syst.*, *7*(3), 1220–1247, doi:10.1002/2015MS000447.
Dee, S. G., N. J. Steiger, J. Emile-Geay, and G. J. Hakim (2016), On the utility of proxy system models for estimating climate states over the common era, *J. Adv. Model. Earth Syst.*, *8*, 1164–1179, doi:10.1002/2016MS000677.
Emile-Geay, J., N. P. McKay, J. Wang, D. Kaufman, and P. Consortium (2015), A semantic database of temperature proxies covering the Common Era, in *Proceedings of the Fifth International Workshop on Climate Informatics*, pp. 1–2, The Natl. Cent. for Atmos. Res. [Available at https://www2.cisl.ucar.edu/sites/default/files/38-%20Emile-Geay.pdf.]
Evans, M. N., S. Tolwinski-Ward, D. Thompson, and K. J. Anchukaitis (2013), Applications of proxy system modeling in high resolution paleoclimatology, *Quat. Sci. Rev.*, *76*, 16–28.
Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, *102*(477), 359–378.
Goosse, H. (2016), Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries, *Clim. Dyn.*, 1–19, doi:10.1007/s00382-016-3154-z.
Goosse, H., E. Crespin, S. Dubinkina, M.-F. Loutre, M. E. Mann, H. Renssen, Y. Sallaz-Damaz, and D. Shindell (2012), The role of forcing and internal dynamics in explaining the medieval climate anomaly, *Clim. Dyn.*, *39*(12), 2847–2866.
Hakim, G. J., J. Emile-Geay, E. J. Steig, D. Noone, D. M.Anderson, R. Tardif, N. Steiger, and W. A. Perkins (2016), The last millennium climate reanalysis project: Framework and first results, *J. Geophys. Res. Atmos.*, *121*, 6745–6764, doi:10.1002/2016JD024751.
Hardy, D., M. Vuille, and R. S. Bradley (2003), Variability of snow accumulation and isotopic composition on Nevado Sajama, Bolivia, *J. Geophys. Res.*, *108*(D22), 4693, doi:10.1029/2003JD003623.
Herron, M. M., and C. C. Langway (1980), Firn densification: An empirical model, *J. Glaciol.*, *25*, 373–385.
Hoffmann, G., et al. (2003), Coherent isotope history of Andean ice cores over the last century, *Geophys. Res. Lett.*, *30*(4), 1179, doi:10.1029/2002GL014870.
Hurley, J., M. Vuille, and D. Hardy (2016), Forward modeling of δ18O in Andean ice cores, *Geophys. Res. Lett.*, *43*(15), 8178–8188, doi:10.1002/2016GL070150.
Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski (2008), A new sea surface temperature and sea ice boundary dataset for the community atmosphere model, *J. Clim.*, *21*(19), 5145–5153, doi:10.1175/2008JCLI2292.1.
Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (2013), An overview of the North Atlantic Oscillation, in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, pp. 1–35, AGU, Washington, D. C., doi:10.1029/134GM01.
Johnsen, S. (1977), Stable isotope homogenization of polar firn and ice, in *Proceedings of Symposium on Isotopes and Impurities in Snow and Ice*, pp. 210–219, I.U.G.G. XVI, General Assembly, Washington, D. C.
Johnsen, S. J., H. B. Clausen, K. M. Cuffey, G. Hoffmann, J. Schwander, and T. Creyts (2000), Diffusion of stable isotopes in polar firn and ice: The isotope effect in firn diffusion, *paper presented at International Symposium on Physics of Ice Core Records*, pp. 121–140, Hokkaido University, Sapporo, Jpn., 14–17 Sept.
Jouzel, J., et al. (1997), Validity of the temperature reconstruction from water isotopes in ice cores, *J. Geophys. Res.*, *102*, 26–471.
Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge Univ. Press, Cambridge, U. K.
Küttel, M., E. J. Steig, Q. Ding, A. J. Monaghan, and D. S. Battisti (2012), Seasonal climate information preserved in west Antarctic ice core water isotopes: Relationships to temperature, large-scale circulation, and sea ice, *Clim. Dyn.*, *39*(7-8), 1841–1857.
LeGrande, A. N., and G. A. Schmidt (2006), Global gridded data set of the oxygen isotopic composition in seawater, *Geophys. Res. Lett.*, *33*, L12604, doi:10.1029/2006GL026011.
Masson-Delmotte, V., et al. (2008), A review of Antarctic surface snow isotopic composition: Observations, atmospheric circulation, and isotopic modeling, *J. Clim.*, *21*(13), 3359–3387, doi:10.1175/2007JCLI2139.1.
Matsikaris, A., M. Widmann, and J. H. Jungclaus (2015), On-line and off-line data assimilation in palaeoclimatology: A case study, *Clim. Past*, *11*, 81–93.
Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.
Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models. Part 1: A discussion of principles, *J. Hydrol.*, *10*, 282–290.
Nicolas, J. P., and D. H. Bromwich (2014), New reconstruction of Antarctic near-surface temperatures: Multidecadal trends and reliability of global reanalyses, *J. Clim.*, *27*(21), 8070–8093, doi:10.1175/JCLI-D-13-00733.1.
Otto-Bliesner, B. L., E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai, and G. Strand (2016), Climate variability and change since 850 CE: An ensemble approach with the community earth system model, *Bull. Am. Meteorol. Soc.*, *97*(5), 735–754, doi:10.1175/BAMS-D-14-00233.1.
Poli, P., et al. (2013), The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), ERA Rep. Ser., 59 pp., ECMWF, Shinfield Park, Reading, U. K.
Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.
Risi, C., et al. (2012), Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopologues: 1. Comparison between models and observations, *J. Geophys. Res.*, *117*, D05303, doi:10.1029/2011JD016621.
Rogers, J. C., J. F. Bolzan, and V. A. Pohjola (1998), Atmospheric circulation variability associated with shallow-core seasonal isotopic extremes near Summit, Greenland, *J. Geophys. Res.*, *103*(D10), 11,205–11,219.
Rohde, R., R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickham, and S. Mosher (2013), Berkeley Earth temperature averaging process, *Geoinfor. Geostat.*, *1*(2), 1–13.
Samuels-Crow, K. E., J. Galewsky, D. R. Hardy, Z. D. Sharp, J. Worden, and C. Braun (2014), Upwind convective influences on the isotopic composition of atmospheric water vapor over the tropical Andes, *J. Geophys. Res. Atmos.*, *119*, 7051–7063, doi:10.1002/2014JD021487.
Schmutz, C., J. Luterbacher, D. Gyalistras, E. Xoplaki, and H. Wanner (2000), Can we trust proxy-based NAO index reconstructions?, *Geophys. Res. Lett.*, *27*(8), 1135–1138.
Schneider, D. P., and D. C. Noone (2007), Spatial covariance of water isotope records in a global network of ice cores spanning twentieth-century climate change, *J. Geophys. Res.*, *112*, D18105, doi:10.1029/2007JD008652.
Schneider, D. P., and E. J. Steig (2008), Ice cores record significant 1940s Antarctic warmth related to tropical climate variability, *Proc. Natl. Acad. Sci.*, *105*(34), 12,154–12,158, doi:10.1073/pnas.0803627105.
Schneider, T. (2001), Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *J. Clim.*, *14*(5), 853–871.

Smerdon, J. E. (2012), Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments, *WIRs Clim. Change*, *3*(1), 63–77.

Steig, E. J., et al. (2013), Recent climate and ice-sheet changes in West Antarctica compared with the past 2,000 years, *Nat. Geosci.*, *6*(5), 372–375.

Steiger, N., and G. Hakim (2016), Multi-timescale data assimilation for atmosphere-ocean state estimates, *Clim. Past*, *12*(6), 1375–1388, doi:10.5194/cp-12-1375-2016.

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

Sturm, C., Q. Zhang, and D. Noone (2010), An introduction to stable water isotopes in climate models: Benefits of forward proxy modelling for paleoclimatology, *Clim. Past*, *6*(1), 115–129.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski (1998), Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, *J. Geophys. Res.*, *103*(C7), 14,291–14,324, doi:10.1029/97JC01444.

Trouet, V., J. Esper, N. E. Graham, A. Baker, J. D. Scourse, and D. C. Frank (2009), Persistent positive North Atlantic oscillation mode dominated the medieval climate anomaly, *Science*, *324*(5923), 78–80, doi:10.1126/science.1166349.

Vimeux, F., P. Ginot, M. Schwikowski, M. Vuille, G. Hoffmann, L. G. Thompson, and U. Schotterer (2009), Climate variability during the last 1000 years inferred from Andean ice cores: A review of methodology and recent results, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, *281*(3), 229–241, doi:10.1016/j.palaeo.2008.03.054.

Vinther, B. M., S. J. Johnsen, K. K. Andersen, H. B. Clausen, and A. W. Hansen (2003), NAO signal recorded in the stable isotopes of Greenland ice cores, *Geophys. Res. Lett.*, *30*(7), 1387, doi:10.1029/2002GL016193.

Vuille, M., R. S. Bradley, R. Healy, M. Werner, D. R. Hardy, L. G. Thompson, and F. Keimig (2003), Modeling $\delta^{18}O$ in precipitation over the tropical Americas: 2. Simulation of the stable isotope signal in Andean ice cores, *J. Geophys. Res.*, *108*(D6), 4175, doi:10.1029/2001JD002039.

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

Werner, J. P., and M. P. Tingley (2015), Technical note: Probabilistically constraining proxy age-depth models within a Bayesian hierarchical reconstruction model, *Clim. Past*, *11*(3), 533–545, doi:10.5194/cp-11-533-2015.

Werner, M., P. M. Langebroek, T. Carlsen, M. Herold, and G. Lohmann (2011), Stable water isotopes in the ECHAM5 general circulation model: Toward high-resolution isotope modeling on a global scale, *J. Geophys. Res.*, *116*, D15109, doi:10.1029/2011JD015681.