The DOI for this manuscript is doi: 10.1175/JCLI-D-12-00693.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Steiger, N., G. Hakim, E. Steig, D. Battisti, and G. Roe, 2013: Assimilation of Time-averaged Pseudoproxies for Climate Reconstruction. J. Climate. doi:10.1175/JCLI-D-12-00693.1, in press.

Assimilation of Time-averaged Pseudoproxies for Climate Reconstruction

Nathan J. Steiger[1] *, Gregory J. Hakim[1], Eric J. Steig[2]

David S. Battisti[1], and Gerard H. Roe[2]

[1] *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*

[2] *Department of Earth and Space Sciences, University of Washington, Seattle, Washington*

* *Corresponding author address:* Nathan Steiger, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195.

E-mail: nathanjs@atmos.washington.edu

ABSTRACT

We examine the efficacy of a novel ensemble data assimilation (DA) technique in climate field reconstruction (CFR) of surface temperature. We employ a minimalistic, computationally inexpensive DA technique that requires only a static ensemble of climatologically plausible states. We perform pseudoproxy experiments with both general circulation model (GCM) and 20th Century Reanalysis (20CR) data by reconstructing surface temperature fields from a sparse network of noisy pseudoproxies. We compare the DA approach to a conventional CFR approach based on Principal Component Analysis (PCA) for experiments on global domains. DA outperforms PCA in reconstructing global-mean temperature in all experiments, and is more consistent across experiments, with a range of time-series correlations of 0.69–0.94 compared to 0.19–0.87 for the PCA method. DA improvements are even more evident in spatial reconstruction skill, especially in sparsely sampled pseudoproxy regions and for 20CR experiments. We hypothesize that DA improves spatial reconstructions because it relies on coherent, spatially local temperature patterns, which remain robust even when glacial states are used to reconstruct non-glacial states and vice versa. These local relationships, as utilized by DA, appear to be more robust than the orthogonal patterns of variability utilized by PCA. Comparing results for GCM and 20CR data indicates that pseudoproxy experiments that rely solely on GCM data may give a false impression of reconstruction skill.

1. Introduction

Climate reconstructions seek to extract useful information from noisy and sparse paleoclimate proxy data. These reconstructions usually take one of two forms: broad indices, such as global mean surface temperature; and climate fields, such as spatial maps of surface temperature. While index reconstructions may yield large-scale information, climate field reconstructions (CFRs) offer important spatial details and regional information. Additionally, it is possible to compute global or hemispheric means from the reconstructed fields, though these can sometimes suffer loss of variance (see discussion in Mann et al. (2012)).

The best way to perform CFRs remains an open question, with no universally superior approach (Smerdon et al. 2011). One important way to examine CFR techniques is through pseudoproxy experiments (PPEs), which provide a synthetic, controlled testbed (see Smerdon (2012) for a review). Based on PPEs, large-scale indices have been shown to be skillfully recovered using most of the well-known CFR techniques (Smerdon et al. 2011; Jones et al. 2009), while skill in reconstructing the climate fields themselves has been much more variable (Smerdon et al. 2011).

In addressing the climate reconstruction problem, data assimilation (DA) has emerged as a potentially very useful CFR technique. DA provides a flexible framework for combining information from paleoclimate proxies with the dynamical constraints of a climate model. The majority of DA approaches utilized thus far can be roughly assigned to four categories: pattern nudging (von Storch et al. 2000), ensemble filters (Dirren and Hakim 2005; Huntley and Hakim 2010; Pendergrass et al. 2012; Bhend et al. 2012), forcing singular vectors (van der Schrier and Barkmeijer 2005), and the selection of ensemble members best matching proxy data (Goosse et al. 2006, 2010; Franke et al. 2011; Annan and Hargreaves 2012). Forcing singular vectors and the selection of ensemble members have been applied to real proxy data using Earth System models of intermediate complexity while pattern nudging has been used to prescribe atmospheric circulation anomalies that then give temperature anomalies consistent with proxy data (Widmann et al. 2010); each of these approaches give results that are consistent with spatially dense empirical knowledge over Europe (Widmann et al. 2010).

Ensemble DA provides a particularly compelling approach to paleoclimate reconstruction because it allows for spatially and temporally changing statistics that may use proxy data more effectively. However, exploiting temporally changing statistics requires forecast models with predictability limits longer than the timescale of the proxy data. Branstator et al. (2012) demonstrate that up to decadal persistence exists in the North Atlantic in several GCMs, yet the location of persistence varies widely by model; how ocean persistence translates into atmospheric predictability is also an open question. Moreover, simulating ensembles using climate models over hundreds, if not thousands, of years presents a tremendous computational cost. These realities motivate an "off-line" approach to DA, where background ensembles are constructed from existing climate model simulations (e.g., Huntley and Hakim 2010), without the need to cycle analyses forward in time with a climate model. Traditional "online" DA approaches, such as those used in operational weather forecasting, become feasible for climate reconstruction only when it has been demonstrated that forecast predictability issues have been overcome and when the reconstruction skill significantly improves upon a vastly cheaper off-line equivalent.

Off-line approaches have been advanced by Bhend et al. (2012) and Annan and Hargreaves (2012). Bhend et al. (2012) applied the time-average assimilation method of Dirren and Hakim (2005) and Huntley and Hakim (2010), based on an ensemble square root filter, while Annan and Hargreaves (2012) applied a degenerate particle filter approach, similar to Goosse et al. (2006, 2010). Both methods reconstruct a "true" model simulation selected out of their ensemble of model simulations, all of which were given identical forcings; additionally, the Bhend et al. (2012) simulations were given identical boundary conditions. Both methods show positive reconstruction skill, particularly for near-surface temperature over land in the Northern Hemisphere. Annan and Hargreaves (2012) note, however, that their ensemble tended to "collapse" (a dramatic loss of ensemble variance) even for a very large ensemble size, a known limitation of the particle filter approach (Snyder et al. 2008). They also discuss that they obtain little to no forecast skill by using the analysis as the initial conditions to generate the following year's background estimate.

The off-line approach and experiments reported here differ from previous DA-based climate reconstruction papers in the following ways: (1) We use a novel time-averaged algorithm that reconstructs the global-mean temperature separately from the temperature field. This allows the global-mean surface temperature to be unaffected by covariance localization, effectively permitting, rather than suppressing, spatially remote covariance relationships with the global mean. This algorithm also has the effect of decreasing variance loss in reconstructions of the global mean (a common problem with CFR approaches). (2) We use the same background ensemble (or prior) for every reconstruction year; the background ensemble is drawn from part of a single climate model simulation or reanalysis data, where ensemble members are individual years instead of independent model simulations, as is typically done in DA schemes and as used by Bhend et al. (2012). This approach allows for more flexibility in the sense that it does not require multiple model simulations to generate large ensembles, though it could be extended to include many model simulations over many time periods or even a collection of different models. Because of how the background ensemble is constructed, it will not contain year-specific boundary condition and forcing information (which act to constrain ensemble variance), nor does it allow for the forward propogation of information in time. (3) We compare our results directly with a standard CFR approach based on Principal Component Analysis (PCA). This PCA approach uses an optimized regression technique known as "truncated total least squares" (TTLS), which has been shown to be robust in a pseudoproxy framework (Mann et al. 2007). (4) We provide analyses (i.e, reconstructions) of only surface temperature so as to directly compare the DA and PCA approaches. In principle DA can provide analyses of the full system state, which constitutes all model variables at all levels and grid cells, but this is not required in the off-line approach. Consequently, this minimalistic DA approach is computationally inexpensive and can be extended to other fields and variables. (5) We also perform DA and PCA pseudoproxy reconstructions with 20th Century Reanalysis (20CR) (Compo et al. 2011) and a Last Glacial Maximum climate model simulation, which tests the robustness of the algorithms and of pseudoproxy experiments in general.

In Sections 2 and 3 we review the DA and PCA techniques and the details of our methodology. Section 4 gives results for global PPEs using data from the 20th Century Reanalysis (20CR) project and from the CCSM 4.0 model (CCSM4). Robustness tests in section 4 include using PPE results for reconstructions of pre-industrial climate given LGM data for the background ensemble (for DA) and for the calibration period (for PCA), as well as tests of differently chosen time periods and red-noise pseudoproxies. In Section 5 we draw conclusions and discuss the benefits of DA in addition to discussing the issue of data choice in PPEs (GCM vs. Reanalysis).

2. Mathematical Background

a. PCA-Based Reconstruction

Here we outline the chief features of the PCA-based reconstruction technique used for comparison with DA. We follow the essential aspects of the method outlined in Mann et al. (1998), except that the Truncated Total Least Squares (TTLS) method is used for the regression of PCs with proxies, described below[1]. We take as given a field of climate data (in our case annual-mean surface temperature) over a calibration period, which we denote $\mathbf{T}_c$, and also proxy data over the calibration and reconstruction periods, denoted as $\mathbf{T}_{pc}$ and $\mathbf{T}_{pr}$ respectively. $\mathbf{T}_c$ is an $m \times n$ matrix where m is the spatial domain and n the time domain, $\mathbf{T}_{pc}$ is an $n \times q$ matrix where q is the number of proxies, and $\mathbf{T}_{pr}$ is an $r \times q$ matrix where r is the number of reconstruction years. We remove the time mean[2] at each grid point of $\mathbf{T}_c$,

[1] We used T. Schneider's implemenation available at `http://www.gps.caltech.edu/~tapio/software.html`, with the default truncation parameter, which we found to give the best results.

[2] In our analysis we do not standardize $\mathbf{T}_c$ so that we can more easily compare the results with our DA approach. We tested the effects of standardization on the PCA-based approach, and in the pseudoproxy

which we then denote as $\mathbf{T}'_c$, and area-weight $\mathbf{T}'_c$ by $\sqrt{\cos(\text{latitude})}$ yielding $\widetilde{\mathbf{T}}'_c$. A singular value decomposition of $\widetilde{\mathbf{T}}'_c$ gives

$$\widetilde{\mathbf{T}}'_c = \mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_c^T \tag{1}$$

where $\mathbf{U}_c$ are the EOFs, $\mathbf{\Sigma}_c$ are the singular values (SVs), $\mathbf{V}_c$ are the PCs, and $\mathbf{V}_c^T$ denotes the transpose of $\mathbf{V}_c$. Preisendorfer's Rule N (as discussed in Wilks 2006, p. 485) is used to determine the number, p, of significant PCs to retain. The following regression equation is solved using TTLS

$$\mathbf{T}_{pc} = \mathbf{V}_c \beta \tag{2}$$

for matrix β, which consists of $p \times 1$ coefficient vectors found for each of the q proxies. During the reconstruction period we solve the regression equation

$$\mathbf{T}_{pr} = \mathbf{V}_r \beta \tag{3}$$

for $\mathbf{V}_r$ (using TTLS) which is an $r \times q$ matrix of the reconstructed PCs. The reconstructed climate field $\widetilde{\mathbf{T}}'_r$ is then found via

$$\widetilde{\mathbf{T}}'_r = \mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_r^T , \tag{4}$$

where $\mathbf{\Sigma}_c$ and $\mathbf{U}_c$ are assumed to remain constant through both the calibration and reconstruction periods.

As discussed in Jones et al. (2009), several of the most prominent CFR techniques share equations (1) and (4) as key steps in their reconstruction processes. In Section 4 we discuss some of the potential pitfalls inherent in assuming that the EOFs and SVs remain constant in time.

b. DA-Based Reconstruction

Here we briefly review the background mathematics of our DA approach to CFR, and leave the details to the Appendix. We also compare the mathematics of DA with the PCA-

experiments no differences of consequence were found.

based method discussed in Section 2a. Data assimilation typically handles observations (or "pseudoproxies" in this paper) by either filtering, which proceeds sequentially at discrete times, or smoothing, which proceeds over time intervals. The paleoclimate reconstruction problem, however, tends to blur this distinction due to the integrated nature of many proxies, and the treatment of time-averaged observations in DA has been discussed in Dirren and Hakim (2005), Huntley and Hakim (2010), and Pendergrass et al. (2012). In either filtering or smoothing, an essential element of DA is the notion of a background, or prior, estimate of the observations. In weather forecasting, the prior comes from a short-term forecast based on an earlier analysis, but this need not always be the case. In a climate context, the prior could be a climate forecast based on a reconstructed state at an earlier time, which if the simulation interval is long enough, amounts to using randomly selected states from the model climate. DA applies weights to the two estimates of the true value of the state, the observations and the prior estimate, to arrive at a posterior, or analysis state. Assuming Gaussian-distributed errors, the classical solution is given by the "update equation" for the Kalman filter (Kalnay 2003):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}_b)]\,, \tag{5}$$

where $\mathbf{x}_b$ is the prior ("background") estimate of the state vector and $\mathbf{x}_a$ is the posterior ("analysis") state vector. Observations (pseudoproxies) 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. The difference between the observations and the prior estimate of the observations, $\mathbf{y} - \mathcal{H}(\mathbf{x}_b)$, is called the innovation. The innovation represents the new information in the observations not known already from the prior. Matrix $\mathbf{K}$, the Kalman gain, weights the innovation and transforms the innovation into state space,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\,, \tag{6}$$

where $\mathbf{B}$ is the error covariance matrix for the prior and $\mathbf{R}$ is the error covariance matrix for the observations. Matrix $\mathbf{H}$ represents a linearization of $\mathcal{H}$ about the prior estimate. For the off-line approach used here, both $\mathbf{B}$ and $\mathbf{R}$ are constant, though in general they may be time-dependent. Since $\mathbf{B} = \langle \mathbf{x}_b \mathbf{x}_b^{\mathrm{T}} \rangle$, where angle brackets denote an expectation, we note that $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ can be written as $\langle \mathbf{x}_b (\mathbf{H}\mathbf{x}_b)^{\mathrm{T}} \rangle$ and $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ can be written as $\langle \mathbf{H}\mathbf{x}_b (\mathbf{H}\mathbf{x}_b)^{\mathrm{T}} \rangle$, and

$$\mathbf{K} = cov(\mathbf{x}_b, \mathbf{H}\mathbf{x}_b)[cov(\mathbf{H}\mathbf{x}_b, \mathbf{H}\mathbf{x}_b) + \mathbf{R}]^{-1} \tag{7}$$

where *cov* represents a covariance expectation. Thus, the numerator of $\mathbf{K}$ "spreads" the information contained in observations through the covariance between the prior and the prior-estimated observations. Comparing (6) and (7) also reveals that $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ represents the error covariance matrix of the prior-estimated observations, which is directly comparable to $\mathbf{R}$. From (5) and (7) we see that the change in the posterior over the prior, $\mathbf{x}_a - \mathbf{x}_b$, is determined by linear regression of the prior on the innovation. New information in the observations is spread from the observation locations to the state variables through the covariance between these quantities. For high-dimensional problems such as weather and climate estimation, the prior error covariance is typically known only through an ensemble estimate, which is subject to sampling error.

c. *Comparison of DA- and PCA-Based Reconstructions*

A superficial comparison of the DA method to the PCA method described previously suggests that they are closely related, since both represent linear-regression solutions to the estimation problem. An essential difference between the methods concerns the use of the prior in the DA method: the innovation is the independent variable for the DA method, whereas for the PCA method the observations or proxies are the independent variable. As a result, in the present context where we consider a "calibration" period, the calibration data is used differently by the two methods. For the DA method, it is assumed that errors in the prior and the observations are uncorrelated, so that the covariance between the prior estimate and the innovation is given by

$$\langle \mathbf{x}_b(\mathbf{y} - \mathbf{H}\mathbf{x}_b)^{\mathrm{T}} \rangle \;=\; \mathbf{B}\mathbf{H}^{\mathrm{T}}. \tag{8}$$

Therefore, in the DA reconstruction method, the observational or proxy data during the calibration period plays no "training" role in the calculation: DA does not use $\mathbf{T}_{pc}$. Errors in the observations contribute "noise" to the calculation through the known error-covariance matrix $\mathbf{R}$. For the PCA method, the observational data during the calibration period is crucial, providing the relationship between the dependent variables, the PCs, and the observations; errors in the observations do not explicitly enter the calculation. The PCA truncation of PCs adds an additional approximation since it affects the relationship between locations and the observations. For situations where temperatures at a location covaries strongly with a proxy observation, but happens to fall on a node of all retained principal components, the PCA method yields a zero reconstruction. We emphasize that a difficulty with the DA method concerns the operator $\mathbf{H}$, which may not be well known for some proxies.

3. Methods

a. Data Sources and Treatment

In this study we use surface temperature data from "The Twentieth Century Reanalysis Project" (20CR; Compo et al. (2011)).[3] We also use surface temperature output from the "Last Millennium run" (covering 850-1850), the "Last Millennium Extension" simulation (covering 1850-2005), a "Last Glacial Maximum" (LGM) simulation, and a pre-industrial control simulation all from the CCSM 4.0 model (CCSM4).[4] Both "Millennium" CCSM4 data sets are from forced runs. Note that we only use the surface temperature data from these existing simulations, and not the *models* that produced the data.

[3]Data provided by NOAA and available at `http://www.esrl.noaa.gov/psd/`

[4]Available at `http://www.earthsystemgrid.org/`

For both the DA-based and PCA-based reconstructions we utilize the full resolution of the 20CR and CCSM4 datasets and do not interpolate the data onto coarser grids as has been done in some other pseudoproxy experiments. For the PCA-based reconstruction, we do not detrend the data since detrending is known to significantly reduce variance in the data set and adversely affect reconstruction skill (Wahl et al. 2006). Global-mean temperature is computed by area-weighting.

b. Pseudoproxy network and proxy noise

We choose pseudoproxy locations based the collation of 1209 proxies published in Mann et al. (2008); the number of proxies as a function of time rapidly decreases in time from this value. For global reconstructions shown in Section 4, we select locations for pseudoproxies where there are continuous records dating back to at least 1300. This choice is somewhat arbitrary but does not significantly affect the results we discuss in this paper. We select this network for several reasons: (1) The full network of 1209 proxy locations greatly over-represents global proxy network density over time periods longer than a few hundred years. (2) Reconstructions with real proxy data must screen proxy records for quality assurance purposes which diminishes the total number actually used (e.g., Mann et al. 2008). (3) While more sparse than the full proxy network, our choice of network still maintains global coverage and the general geographical features of the full network. (4) This reconstruction interval starts near the beginning of the so-called European "Little Ice Age," a possibly significant climatic feature.

We construct two types of pseudoproxies by adding either white or red noise to the annual-mean temperature time-series at the locations discussed in the previous paragraph. Proxy locations are interpolated onto model grid points and we remove duplicates where closely spaced proxies interpolate onto the same grid point. Because the 20CR and CCSM4 data sets have different resolutions (which we retain) they differ in some proxy locations after interpolation: for the global reconstructions, 20CR has 78 pseudoproxies, while CCSM4 has

88. These differences do not substantially change the geographical coverage of the proxy network.

To construct the white noise pseudoproxies, we add to the annual-mean grid point temperature series Gaussian white noise with a signal-to-noise ratio (SNR) of 0.5, where SNR is defined as

$$SNR = \sqrt{\frac{var(X)}{var(N)}} \tag{9}$$

where X is the grid-point temperature series and N is the additive noise series, and var is the variance. SNR values of 0.5 are considered to be consistent with real proxy noise levels (Smerdon 2012) and so we use this value throughout. Red noise with a given SNR is defined by

$$N_r(i) = aN_r(i-1) + s_n\epsilon(i)\sqrt{1-a^2} \tag{10}$$

where N_r is a red noise time series with index i, a is the lag-one autocorrelation, $s_n = \sqrt{var(N)}$ is the desired standard deviation of the noise, and ϵ is a random number drawn from a standardized normal distribution. Similar to the white noise pseudoproxies, those with red noise are constructed by adding red noise to annual-mean grid point temperature series. For a typical multiproxy network, Mann et al. (2007) estimate a mean autocorrelation of $a = 0.32$ to be a conservative (i.e., "redder" than in reality) value. We use this autocorrelation value in our red-noise pseudoproxy tests (see Table 3). For both the DA and PCA approaches, we compute $var(X)$ from the calibration period data. Bootstrap error estimates are derived by performing each reconstruction 30 times for both DA and PCA. For each reconstruction we generate different random noise signals which are added to the grid-point temperature series to create the pseudoproxies. Every reconstruction figure shows the mean of the 30 reconstructions and 1 standard deviation about this mean for the figures showing global mean temperature.

c. *DA Implementation*

For the DA-based approach, we solve the state "update equation" (5) for an analysis ensemble based upon a background ensemble, pseudoproxies, ensemble estimates of the observations, and error estimates for the background ensemble and the observations. The procedure, as detailed in the Appendix, follows Huntley and Hakim (2010), but with the important generalization that the global-mean temperature is solved separately from the spatial fields, which allows covariance localization to be applied only to the spatially varying part of the field.

We begin with a background ensemble that is identical to the data given to PCA during the calibration period: the annually averaged global surface temperature fields $\mathbf{T}_c$ described in Section 2a. These fields are derived from part of a single model simulation or reanalysis data set, where ensemble members are the annually averaged surface temperature fields over the chosen calibration period (such as over the years 1880-1980). This background ensemble is the same for each year of the reconstruction. This approach differs from most online DA approaches that use the previous time's analysis ensemble as the background ensemble for the current time. We note that in general background ensembles may be drawn from any collection of reasonable states and need not be composed of an ensemble of model simulations; in Bayesian terminology this can be referred to as a "non-informative prior" that is constrained to climatologically plausible states. This approach allows for more flexibility in the sense that it does not require multiple model runs to generate large ensembles, though it could be trivially extended to include many model runs over many time periods or even a collection of different models. Because of how the background ensemble is constructed, it does not contain year-specific boundary condition and forcing information, but does contain the spatial covariance relationships among fields associated with forcing variability. We also note that even though the background ensemble for each reconstruction is composed of consecutive years of some model run, the ensemble members are linearly independent for all reconstructions shown in this paper.

For the DA approach, the observations or pseudoproxies are identical to those given the PCA technique, $\mathbf{T}_{pr}$; they are the white or red noise-added time series at the pseudoproxy locations during the reconstruction time period. Ensemble estimates of the proxies and background error estimates are derived directly from the background ensemble. Observation error estimates are derived through the signal-to-noise equation (9) using an assumed signal-to-noise ratio and data during the calibration period (see Appendix for details).

Assimilation is performed one year at a time by serially processing the observations one at a time (a standard technique based on Houtekamer and Mitchell (2001) and discussed in Whitaker and Hamill (2002) and Tippett et al. (2003)), yielding an annual-mean, ensemble-mean analysis, which is the climate field reconstruction for that year, as well as an estimate of the ensemble-mean, annual-mean, global-mean surface temperature; the analysis ensemble mean state is analogous to $\widetilde{\mathbf{T}}_r'$ in (4) in the PCA method. The off-line nature of the DA approach means that a climate model is not needed to integrate from analyses to future times, which results in a tremendous computational cost savings. We provide analyses of only surface temperature so that the comparison between DA- and PCA-based methods is direct. In principle, DA can provide analyses for up to the full system state, which constitutes all model variables at all levels and grid cells.

4. Reconstructions

a. Results

In this section we focus on four global surface temperature reconstructions that we compare with the actual GCM/Reanalysis output during the reconstruction period. The first is a millennial-scale reconstruction using CCSM4 model output (which includes estimates of solar and volcanic forcing), with a calibration period from 1881-1980 and a reconstruction period from 1300-1880. The second and third reconstructions are centennial-scale, with calibrations over 1956-2005 and reconstructions over 1871-1955. The second reconstruction uses data from 20CR and the third uses data from CCSM4. The reason for this smaller time-frame is because the 20CR data only extends to 1871. The fourth reconstruction uses a 100 year CCSM4 Last Glacial Maximum simulation for the calibration period and 100 years of a CCSM4 pre-industrial control simulation for the reconstruction period; this reconstruction seeks to test the sensitivity of the results when the calibration and reconstruction climate differ significantly. Sensitivity to the white-noise pseudoproxy approximation and chosen time period is addressed by another set of experiments that use red noise and different time periods for calibration and reconstruction.

Figs. 1 and 2 show the reconstruction skill for the CCSM4 data for the period 1300-1880. For global-mean temperature, DA slightly outperforms PCA, with a time-series correlation of 0.92 compared to 0.87, respectively. Improvement of DA over PCA is more evident in spatial reconstruction skill as measured by the reconstruction–truth time-series correlation at each point (Fig. 2a,b) and by the "coefficient of efficiency" (CE) metric (Fig. 2c,d). The CE metric for a data series comparison of length N is defined by (Nash and Sutcliffe 1970)

$$\mathrm{CE} = 1 - \frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}\,, \tag{11}$$

where x is the "true" time series, $\overline{x}$ is the true time series mean, and $\hat{x}$ is the reconstructed time series. CE has the range $-\infty < \mathrm{CE} \leq 1$, where $\mathrm{CE} = 1$ corresponds to a perfect match and $\mathrm{CE} < 0$ indicates that the error variance is greater than the true time series variance (in all CE figures we show only the range $-1 \leq \mathrm{CE} \leq 1$). The DA approach reconstructs temperature with higher correlations in Asia, Greenland, and Europe as well as around lone pseudoproxies, such as those in the Southern Hemisphere, near New Zealand, Tasmania, Chile, and South Africa (Fig. 2a,b). The CE maps show positive skill for DA throughout most of the Northern Hemisphere while PCA has positive skill mainly around the dense North American pseudoproxy network (Fig. 2c,d).

The results in Figs. 1 and 2 are generally consistent with reconstructions we performed using other CMIP5 GCM data sets. For example, reconstructions based on data from the NASA GISS and MPI-ESM climate models over millennial time scales yield results roughly similar to those shown in Figs. 1 and 2 (not shown). We present results with CCSM4 for brevity and because the DA reconstruction showed similar skill across models while the PCA-based approach performed best with CCSM4; hence, the differences in skill between the DA and PCA reconstructions shown in Figs. 1 and 2 represent a rough lower bound on the differences between the DA and PCA reconstructions in the CIMP5 models we tested.

The second reconstruction uses 20CR and has a calibration (background ensemble) period for PCA (DA) of 1956–2005 and a reconstruction period of 1871–1955. The global-mean time series is reconstructed with a correlation of 0.69 for DA as compared to 0.19 for PCA (Fig. 3). Fig. 4 shows that for 20CR reconstructions, the DA method also has much higher skill in reconstructing regional temperature compared to the PCA method. Fig. 4 also shows that both DA and PCA are able to skillfully reconstruct temperatures over North America, where the proxy network is most dense, while only DA has high skill around most of the remaining pseudoproxies. Interestingly, in comparing Fig. 1 and 3, we see that neither PCA or DA is able to reproduce the global-mean temperature in the 20CR data as well as for the CCSM4 data.

As a check against our choice of proxy network, we performed a reconstruction for each method using 20CR where we increased the number of pseudoproxies to 278, corresponding to a network from the Mann et al. (2008) proxy collation that would extend back to the year 1600. We find the same general results for PCA as shown in Figs. 3 and 4: slightly improved, yet still low correlation with the global-mean temperature ($r = 0.49$) and areas of higher correlation ($r > 0.35$) and positive CE values only in the densest pseudoproxy networks in Europe and North America (not shown). For DA however, spatial r and CE values in most locations improved and the reconstructed global-mean temperature correlation increases to $r = 0.78$ (not shown).

The third reconstruction is over the same time frame as the second except we use CCSM4 data (Figs. 5 and 6). Comparing Fig. 3 with Fig. 5 for global-mean temperature shows that both methods are sensitive to the data source (i.e., GCM vs. reanalysis data)[5]. The source of the difference between the reconstructions with 20CR and CCSM4 could be due to several effects that will be discussed in the next section, but one clear difference is that 20CR is constrained by observations whereas CCSM4 is not. Comparing the spatial skill of both methods in Figs. 4 and 6 reveals that DA again outperforms PCA and that the PCA results are more dataset dependent than those for DA.

The fourth reconstruction seeks to test the approach in a situation with no trend in the underlying data (no global warming signal) and very different training and target climates for the reconstruction. Here we take as our DA background ensemble and the PCA calibration data a 100 year CCSM4 run of the Last Glacial Maximum (LGM) and reconstruct 100 years of a CCSM4 pre-industrial control run. Pseudoproxy locations are the same as in the previous CCSM4 reconstructions. We note that this is not intended as a realistic climate reconstruction scenario (e.g., the calibration/reconstruction periods are reversed from a typical setting and the proxy network is not consistent with proxy availability during the LGM), but rather a markedly different scenario intended to explore the robustness and range of applicability of the reconstruction techniques. Figs. 7 and 8 show the global-mean temperature reconstructions and the spatial-performance maps, respectively. These results show that the DA reconstructions give robust results, consistent with previously shown reconstructions, despite the radically different calibration and reconstruction states. The PCA results are less robust and show a global mean temperature reconstruction that has much reduced variance compared with the true variance.

Fig. 9 summarizes the spatial maps of r and CE in box-and-whisker plots. The distributions of the DA reconstructions are statistically significant improvements over the PCA reconstructions (via t-tests at the 95% level), with the largest improvement in the case of

[5]Note that the global-mean trends in these portions of the 20CR and CCSM4 data sets are slightly different.

the 20CR reconstruction shown in Fig. 4. Table 1 summarizes the mean and median values of each spatial map.

As a check against our choice of time frames, we perform reconstructions similar to the four previously shown, but with different or approximately "reversed" calibration–reconstruction periods while keeping everything else the same (see Table 2). As a counterpoint to the first reconstruction, we choose a calibration period of 1781-1880, to avoid calibration with a global warming signal, and reconstruct from 1300-1780. In juxtaposition to the remaining three reconstructions, we reverse the calibration and reconstruction periods while adjusting two of them to keep each period the same size as the original; for example, the second reconstruction shown in Figs. 3 and 4 uses a 50 year calibration period from 1956–2005 and an 85 year reconstruction period from 1871–1955, while the "reversed" reconstruction uses a 50 year calibration period from 1871–1920 and an 85 year reconstruction period from 1921–2005. A comparison of Table 2 with Table 1 reveals generally consistent results for both DA and PCA methods: DA always improves upon PCA and usually by similar magnitudes as those shown in Figs. 1–8 and Table 1. We also perform the same reconstructions as shown in Figs. 1–8, except with red noise pseudoproxies and find similar results compared to the white noise pseudoproxies (cf. Table 3 and Table 1), though PCA tends to increase the global mean correlation and tends to decrease the spatial mean CE values in some red noise reconstructions.

b. *Discussion and Analysis*

Many of the most common CFR methods rely on the assumption of constant EOFs and singular values (SVs) throughout the reconstruction and calibration periods, as in (4), and discussed in Jones et al. (2009). Investigating the 20CR and CCSM4 data sets, we find that for both the 20CR and CCSM4 data, the surface temperature EOFs and SVs change over time: the EOFs and SVs of the calibration period are different from the reconstruction period (Fig. 10, EOFs not shown). The 20CR data also has a broader SV spectrum compared to CCSM4 during the calibration period in that more EOFs are required to explain the same amount of variance. The variance explained, Λ, is related to the SVs (or "amplitude explained"), Σ, by the relationship $\Lambda = \Sigma^2/n$, where n is the size of the sampling dimension, in our case time. Given Λ from Σ, the cumulative variance explained is determined. Fig. 10 shows that 20CR has a shallower SV spectrum compared to CCSM4 during the calibration period, so that a given amount of variability is spread over a larger number of patterns in 20CR.

We now speculate on the reasons for consistent spatial skill in DA relative to the less consistent spatial skill of PCA. The discussion of the PCA and DA techniques in Section 2 suggests that, through $\mathbf{K}$, DA depends on local spatial correlations remaining consistent through time; this contrasts with PCA which relies upon stationary EOFs and SVs as well as consistent proxy-PC relationships through time. As discussed in the previous paragraph and shown in Fig. 10, the EOFs and SVs change in time. We consider it likely that several factors lead to PCA's poor spatial reconstruction in Fig. 4b, including the fact that the SV spectrum of 20CR is flatter in the calibration period than for CCSM4. It may also be that nature (at least as reflected in 20CR) has less spatially coherent variability than the climate model, helping to explain (i) a modest reduction in the skill of the reconstructed 20CR temperature compared to that of the reconstructed GCM temperature using the DA method, and (ii) the poor skill of the 20CR temperature reconstruction (locally and in the global average) using the PCA method. Given the potentially changing nature of the basis upon which PCA is founded, we argue that the local grid-point correlations exploited by the DA technique may offer a more reliable basis for reconstructions, particularly for spatial reconstructions. We also emphasize that in light of the fact that pseudoproxy experiments to date have almost exclusively relied on GCM data, our results suggest that these experiments may give a false impression of reconstruction skill.

5. Conclusions

The main purpose of this paper was to evaluate a data assimilation (DA) approach for climate field reconstructions (CFRs), and to compare the results with a standard approach based on principal component analysis (PCA). Using several pseudoproxy experiments (PPEs), we have shown that DA consistently outperforms PCA in reconstructions of both the global-mean temperature and regional patterns, although differences are especially evident in the spatial fidelity of the reconstructions. Relative to the PCA method, the DA method improves GCM temperature reconstructions around isolated pseudoproxies and in several sparsely sampled regions; DA also has much higher correlations and coefficient of efficiency values in most geographical regions when reconstructing 20CR temperatures.

DA does not involve any form of PCA and is thus able to avoid several assumptions inherent in many PCA-based CFR techniques: that empirical orthogonal functions (EOFs) and singular values remain roughly constant through time; that principal components (PCs) are well correlated with proxy time series trough time; and that standard selection criteria can consistently be applied across reconstruction scenarios. We attribute the consistency of the DA spatial reconstructions to the fact that DA relies on local temperature correlations, which are more robust to the assumption of stationarity than are EOFs. Moreover, we conclude that these spatial relationships are insensitive to details in the choice of background ensemble, as demonstrated by the high skill of the reconstructions of a pre-industrial simulation using background ensemble data from a simulation of the Last Glacial Maximum.

The results of this paper show that a novel off-line DA technique provides both robust spatial reconstructions in addition to global means. The approach is straightforward to extend to real proxy data and can easily handle practical challenges in the climate reconstruction problem such as missing values, time averaged proxies, and error estimates. Additionally, our experiments show that reanalysis data appears to differ from model simulated data in ways that impact the skill of reconstruction techniques. This suggests that PPEs that rely solely on GCM data may give a false impression of reconstruction skill.

Acknowledgments.

We thank three anonymous referees for their comments on the manuscript, which improved the clarity of presentation. This research was supported by National Science Foundation Award 0902500, granted to the University of Washington. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office.

APPENDIX

DA Implementation

Our DA method and equations are defined in section *a*, followed by a description of the numerical algorithm in section *b*.

a. Data assimilation method and equations

State updates for the Kalman filter are determined by (5) and (6), which are approximated here by an ensemble square root technique applied to time averages (Dirren and Hakim 2005; Huntley and Hakim 2010). Here we extend this technique to handle the global-mean average separately from deviations from this average by augmenting the state vector $\mathbf{x}$ (here composed of annual mean surface temperatures drawn from a portion of a GCM or reanalysis run) with the global-mean; we denote the augmented vector by $\mathbf{z}$. As will be described further, this is done so that the global-mean surface temperature is not affected by covariance localization. Following Huntley and Hakim (2010), we can use $\mathbf{z}$ in the Kalman filter equations as long as the global mean and the deviations from this mean—the rest of the state vector—do not signifcantly co-vary.

Following Whitaker and Hamill (2002), the update equation is split into an ensemble mean update (denoted by an overbar) and an update of the perturbations from the ensemble mean (denoted by a prime):

$$\overline{\mathbf{z}}_a = \overline{\mathbf{z}}_b + \mathbf{K}(\mathbf{y} - \overline{\mathbf{y}}_e)\,, \tag{A1}$$

$$\mathbf{z}'_a = \mathbf{z}'_b - \widetilde{\mathbf{K}}\mathbf{y}'_e\,. \tag{A2}$$

The analysis and background ensemble-mean states, $\overline{\mathbf{z}}_a$ and $\overline{\mathbf{z}}_b$, are column vectors of dimension $m \times 1$; we include only annually averaged surface temperatures in $\mathbf{z}$, with the global-mean removed and placed at the end of the state vector, so that m in this particular instance is the number of grid points plus one. The analysis and background perturbations from the ensemble mean, $\mathbf{z}'_a$ and $\mathbf{z}'_b$, are of dimension $m \times n$ where n is the ensemble size. Observations (proxy data) are given in $\mathbf{y}$ as a $p \times 1$ vector where p is the number of observations, and $\mathbf{y}_e = \mathbf{H}\mathbf{x}_b$ are observation estimates from the prior; $\overline{\mathbf{y}}_e$ is the ensemble-mean value, of dimension $p \times 1$, and $\mathbf{y}'_e$ are deviates from the mean, of dimension $p \times n$.

We solve (A1) and (A2) by processing the observations serially, one at a time (Houtekamer and Mitchell 2001), for computational expedience. In this case, at a single grid point, $\mathbf{K}$ simplifies (cf. 7) to the scalar

$$K = \frac{cov(z'_b, y'_e)}{var(y'_e) + r} \tag{A3}$$

where the covariance and variance estimates apply over the ensemble, and r is the error variance for the observation. For our pseudoproxy experiments we determine r for each observation location through the signal-to-noise (SNR) equation, (9): after assuming a fixed value of SNR (here SNR = 0.5; see discussion in Section 3b), we compute $var(X)$ for each location during the calibration time period, and then solve for $r = var(N)$. In addition to the ensemble-mean update, the ensemble perturbations are updated by (A2), where

$$\widetilde{K} = \left(1 + \sqrt{\frac{r}{var(y'_e) + r}}\right)^{-1} K, \tag{A4}$$

and $var(y'_e)$ applies over the ensemble. The process repeats for each observation, with y_e determined each time from the updated ensemble.

Once $\overline{\mathbf{z}}_a$ and $\mathbf{z}'_a$ are computed, we compare the results of DA with the true climate fields by adding the global-mean value, the last entry in the column vector $\overline{\mathbf{z}}_a$, back into the rest of $\overline{\mathbf{z}}_a$ so that we recover $\overline{\mathbf{x}}_a$ (of dimension $m - 1$), which is the annually averaged surface temperatures at all grid points. The last entry of $\overline{\mathbf{z}}_a$ is the global-mean temperature reconstruction.

We note that in order to compare DA and PCA, we let $\mathbf{x}_b$ (from which we derive $\mathbf{z}_b$) be the annually averaged climate field temperatures during the calibration period, the same as $\mathbf{T}_c$ discussed in Section 2a; we do not use an ensemble of climate models to produce $\mathbf{x}_b$, but rather the annually averaged fields of surface temperatures from a single climate model simulation (or reanalysis) for the ensemble members. For the off-line approach presented here, $\mathbf{x}_b$ (and thereby $\mathbf{z}_b$) is numerically identical for each reconstruction year. Also, the observations or pseudoproxies $\mathbf{y}$ are the same noise-added pseudoproxy time series used for the PCA reconstructions, $\mathbf{T}_{pr}$ in (3).

To control spurious long-distance correlations due to sampling error, we use a localization function (Gaspari and Cohn 1999) applied to the gain, K, with a length scale of 12,000 km during the update step. We determine this localization length by finding a minimum in mean error variance and a "smooth" analysis field, so that no "edges" of the localization mask are discernible. For the reconstructions, the mean error variance is a smooth function of localization radius with a wide range of values (from about 4,000 km to about 16,000 km) that were very near (within $\sim 0.01°\mathrm{C}^2$) the minimum mean error variance. We do not apply localization to the global-mean value.

b. Algorithm Sketch

For each reconstruction year we perform the following steps:

i. Construct $\mathbf{x}_b$, then $\mathbf{z}_b$ from $\mathbf{x}_b$, and the annual pseudoproxy vector $\mathbf{y}$

ii. Find the error r from (9) for each pseudoproxy.

iii. Split $\mathbf{z}_b$ into an ensemble mean and perturbations from this mean:

$$\mathbf{z}_b = \overline{\mathbf{z}}_b + \mathbf{z}_b'$$

iv. For each pseudoproxy:

(a) Compute $y_e = \mathbf{H}\mathbf{x}_b$

(b) Split up y_e into an ensemble mean and perturbations from this mean

$$y_e = \overline{y}_e + y_e'$$

(c) Compute K from (A3).

(d) Apply the localization function, if desired, to K except for the last entry (the global-mean value)

(e) Compute $\widetilde{K}$ from (A4)

(f) At each grid point, update the analysis ensemble-mean and perturbations from this mean

$$\overline{z}_a = \overline{z}_b + K(y_i - \overline{y}_e)$$

$$z_a' = z_b' - \widetilde{K} y_e'$$

(g) Use $\overline{\mathbf{z}}_a$ and $\mathbf{z}_a'$ as $\overline{\mathbf{z}}_b$ and $\mathbf{z}_b'$ respectively for the next observation

v. The full analysis ensemble may be recovered through

$$\mathbf{z}_a = \overline{\mathbf{z}}_a + \mathbf{z}_a'$$

where the column vector $\overline{\mathbf{z}}_a$ is added to each column vector of $\mathbf{z}_a'$

vi. After each year's pseudoproxies have been assimilated, we add the last column entry of $\overline{\mathbf{z}}_a$ to the rest of $\overline{\mathbf{z}}_a$ to recover $\overline{\mathbf{x}}_a$, the reconstructed temperature field for that year. We also use the last column entry of $\overline{\mathbf{z}}_a$ as the reconstructed global-mean temperature for that year.

REFERENCES

Annan, J. and J. Hargreaves, 2012: Identification of climatic state with limited proxy data. *Clim. Past*, **8**, 1141–1151.

Bhend, J., J. Franke, D. Folini, M. Wild, and S. Brönnimann, 2012: An ensemble-based approach to climate reconstructions. *Clim. Past*, **8**, 963–976.

Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six aogcms. *J. Climate*, **25**, 1827–1846.

Compo, G. P., et al., 2011: The twentieth century reanalysis project. *Q. J. Roy. Meteorol. Soc.*, **137**, 1–28.

Dirren, S. and G. Hakim, 2005: Toward the assimilation of time-averaged observations. *Geophys. Res. Lett.*, **32**, L04 804.

Franke, J., J. Gonzalez-Rouco, D. Frank, and N. Graham, 2011: 200 years of european temperature variability: insights from and tests of the proxy surrogate reconstruction analog method. *Clim. Dyn.*, **37**, 133–150.

Gaspari, G. and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteor. Soc.*, **125**, 723–757.

Goosse, H., E. Crespin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J. Geophys. Res.*, **115**, D09 108.

Goosse, H., H. Renssen, A. Timmermann, R. Bradley, and M. Mann, 2006: Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Clim. Dyn.*, **27**, 165–184.

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

Huntley, H. and G. Hakim, 2010: Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dyn.*, **35**, 995–1009.

Jones, P., et al., 2009: High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene*, **19**, 3–49.

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

Mann, M., J. Fuentes, and S. Rutherford, 2012: Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nature Geosci.*, **5**, 202–205.

Mann, M., S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12 109.

Mann, M., Z. Zhang, M. Hughes, R. Bradley, S. Miller, S. Rutherford, and F. Ni, 2008: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci. USA*, **105**, 13 252 –13 257.

Mann, M. E., R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392**, 779–787.

Nash, J. and J. Sutcliffe, 1970: River flow forecasting through conceptual models part 1: A discussion of principles. *J. Hydrology*, **10**, 282–290.

Pendergrass, A., G. Hakim, D. Battisti, and G. Roe, 2012: Coupled airmixed layer temperature predictability for climate reconstruction. *J. Climate*, **25**, 459–472.

Smerdon, J., 2012: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs Clim. Change*, **3**, 63–77.

Smerdon, J., A. Kaplan, E. Zorita, J. Gonzlez-Rouco, and M. Evans, 2011: Spatial performance of four climate field reconstruction methods targeting the common era. *Geophys. Res. Lett.*, **38**, L11 705.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

van der Schrier, G. and J. Barkmeijer, 2005: Bjerknes' hypothesis on the coldness during ad 1790–1820 revisited. *Clim. Dyn.*, **24**, 355–371.

von Storch, H., U. Cubasch, J. Gonzalez-Rouco, J. Jones, R. Voss, M. Widmann, and E. Zorita, 2000: Combining paleoclimatic evidence and gcms by means of data assimilation through upscaling and nudging (datun). *11th Symposium on Global Change Studies*.

Wahl, E., D. Ritson, and C. Ammann, 2006: Comment on "reconstructing past climate from noisy data". *Science*, **312**, 529.

Whitaker, J. and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Widmann, M., H. Goosse, G. van der Schrier, R. Schnur, and J. Barkmeijer, 2010: Using data assimilation to study extratropical northern hemisphere climate over the last millennium. *Climate of the Past*, **6 (5)**, 627–644.

Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. Elsevier, London, UK.

List of Tables

1 Summary statistics for Figs. 1–8. The correlation of the reconstructed global mean temperature with the actual is r_{gmt}, shown at the top of Figs. 1, 3, 5, and 7. Both $\bar{\mathrm{r}}$ and $\overline{\mathrm{CE}}$ are the mean values of the spatial r and CE maps shown in Figs. 2, 4, 6, and 8. Both $\widetilde{r}$ and $\widetilde{CE}$ are the median values of the spatial r and CE maps and also correspond to those center values indicated in the box-and-whisker plots, Fig. 9. The CCSM4 data types refer to the runs Last Millennium (LM), Last Millennium Extension (LM Ext.), Last Glacial Maximum (LGM), and pre-industrial control (PI). 28

2 Summary statistics for reconstructions with different or reversed calibration and reconstruction periods, cf. Figs. 1–8 and Table 1. Variables and data types are the same as those defined in Table 1. 29

3 Summary statistics for reconstructions that are akin to those shown in Figs. 1–8, except with red noise pseudoproxies (as defined and discussed in Section 3b). Variables and data types are the same as those defined in Table 1. 30

Table 1. Summary statistics for Figs. 1–8. The correlation of the reconstructed global mean temperature with the actual is r_{gmt}, shown at the top of Figs. 1, 3, 5, and 7. Both $\bar{r}$ and $\overline{CE}$ are the mean values of the spatial r and CE maps shown in Figs. 2, 4, 6, and 8. Both $\widetilde{r}$ and $\widetilde{CE}$ are the median values of the spatial r and CE maps and also correspond to those center values indicated in the box-and-whisker plots, Fig. 9. The CCSM4 data types refer to the runs Last Millennium (LM), Last Millennium Extension (LM Ext.), Last Glacial Maximum (LGM), and pre-industrial control (PI).

Figs.	Method	Data Type	r_{gmt}	$\bar{r}$	$\widetilde{r}$	$\overline{CE}$	$\widetilde{CE}$
1 & 2	DA	CCSM4 LM	0.92	0.36	0.38	0.13	0.12
1 & 2	PCA	CCSM4 LM	0.87	0.26	0.27	-0.023	-0.028
3 & 4	DA	20CR	0.69	0.29	0.29	0.054	0.046
3 & 4	PCA	20CR	0.19	0.090	0.076	-0.46	-0.36
5 & 6	DA	CCSM4 LM Ext.	0.94	0.38	0.37	0.14	0.11
5 & 6	PCA	CCSM4 LM Ext.	0.71	0.26	0.26	-0.015	-0.024
7 & 8	DA	CCSM4 LGM & PI	0.85	0.27	0.30	0.091	0.070
7 & 8	PCA	CCSM4 LGM & PI	0.78	0.15	0.12	-0.094	-0.068

TABLE 2. Summary statistics for reconstructions with different or reversed calibration and reconstruction periods, cf. Figs. 1–8 and Table 1. Variables and data types are the same as those defined in Table 1.

Method	Data Type	Cal. (yrs)	Recon. (yrs)	r_{gmt}	$\bar{r}$	$\overline{CE}$
DA	CCSM4 LM	1781–1880	1300–1780	0.92	0.34	0.14
PCA	CCSM4 LM	1781–1880	1300–1780	0.85	0.26	-0.0039
DA	20CR	1871–1920	1921–2005	0.86	0.48	0.21
PCA	20CR	1871–1920	1921–2005	0.65	0.076	-0.21
DA	CCSM4 LM Ext.	1871–1920	1921–2005	0.92	0.51	0.25
PCA	CCSM4 LM Ext.	1871–1920	1921–2005	0.87	0.26	0.030
DA	CCSM4 LGM & PI	100 of PI	100 of LGM	0.69	0.25	0.065
PCA	CCSM4 LGM & PI	100 of PI	100 of LGM	0.57	0.13	-0.18

TABLE 3. Summary statistics for reconstructions that are akin to those shown in Figs. 1–8, except with red noise pseudoproxies (as defined and discussed in Section 3b). Variables and data types are the same as those defined in Table 1.

Method	Data Type	r_{gmt}	$\bar{r}$	$\overline{CE}$
DA	CCSM4 LM	0.91	0.36	0.13
PCA	CCSM4 LM	0.85	0.24	-0.26
DA	20CR	0.69	0.29	0.057
PCA	20CR	0.40	0.095	-0.75
DA	CCSM4 LM Ext.	0.92	0.38	0.14
PCA	CCSM4 LM Ext.	0.80	0.26	-0.043
DA	CCSM4 LGM & PI	0.84	0.27	0.092
PCA	CCSM4 LGM & PI	0.53	0.11	-0.22

List of Figures

1 Global-mean temperature anomaly reconstructions of the (a) DA and (b) PCA techniques using CCSM4. Solid black lines are the mean reconstruction out of 30, dash-dotted lines are the actual model mean temperature. Gray shading is 1 standard deviation of the reconstructions. The calibration period is 1881-1980 and the reconstruction period is 1300-1880. The correlation coefficient, r, is noted at the top of each figure along with the number of PCs used for the PCA-based reconstruction. The anomalies are shown with respect to the reconstruction mean. 33

2 Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 1 (calibration period: 1881-1980, reconstruction period: 1300-1880), for (a) DA and (b) PCA. These maps show the correlation and coefficient of efficiency between each grid point temperature series of the mean reconstruction (mean of 30) and each actual grid point temperature series. Empty black boxes are centered over pseudoproxy locations and stippling indicates correlations that are *not* significant at the 95% level. 34

3 Global-mean temperature anomaly reconstructions of the (a) DA and (b) PCA techniques using 20CR. The calibration period is 1956-2005 and the reconstruction period is 1871-1955. 35

4 Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 3 (calibration period: 1956-2005, reconstruction period: 1871-1955), for (a) DA and (b) PCA. For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level. The lower bound of CE values shown are cut off at CE $= -1$. 36

5 Global-mean temperature anomaly reconstructions using (a) DA and (b) PCA techniques with CCSM4 over the same calibration and reconstruction periods as in Figs. 3 and 4 (calibration period: 1956-2005, reconstruction period: 1871-1955). 37

6 Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 5 with CCSM4, (calibration period: 1956-2005, reconstruction period: 1871-1955). For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level. 38

7 Global-mean temperature anomaly reconstructions using (a) DA and (b) PCA techniques with 100 years of CCSM4 LGM data for the calibration period and 100 years of a CCSM4 pre-industrial control run for the reconstruction period. 39

8 Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 7. For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level. The lower bound of CE values shown are cut off at CE $= -1$. 40

9 Box-and-whisker plots (for clarity, outliers are not shown) of each of the spatial reconstruction figures, for (a) correlation coefficient (r) maps and (b) coefficient of efficiency (CE) maps. Labels refer to DA or PCA techniques and the figure number of the data that the box-and-whisker plots represent. All DA-PCA distribution pairs are statistically distinct according to t-tests for each DA-PCA comparison. 41

10 Cumulative variance explained (CVE) of the retained EOFs for 20CR and CCSM4 during both the (a) calibration and (b) reconstruction periods shown in Figs. 3–6 (calibration period: 1956-2005, reconstruction period: 1871-1955). With 20CR we retain 12 PCs and with CCSM4 we retain 10 PCs; the CVE values are normalized by the total variance explained. 42

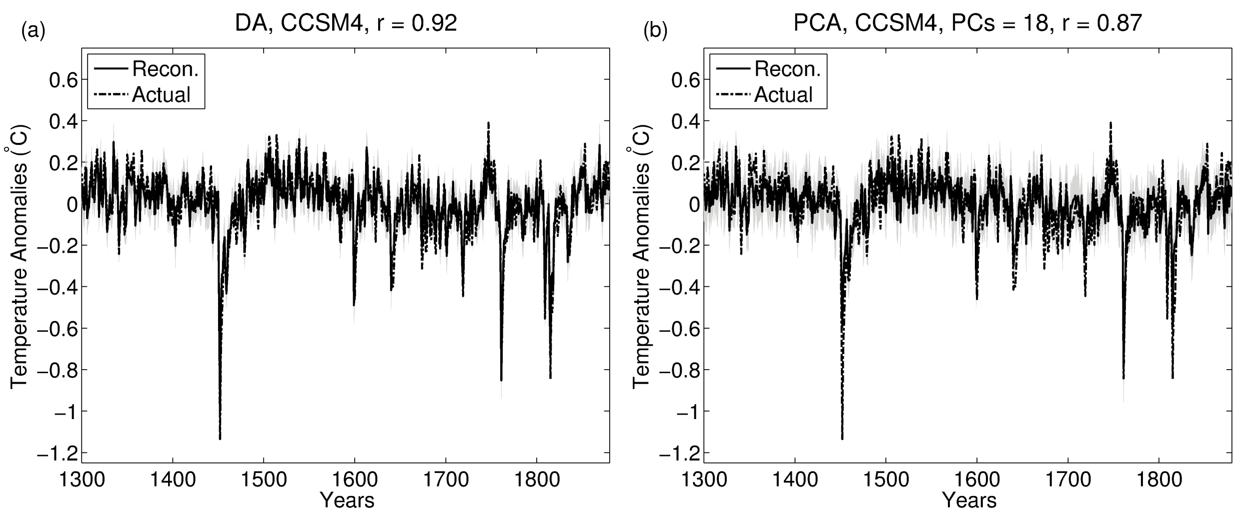

Fig. 1. Global-mean temperature anomaly reconstructions of the (a) DA and (b) PCA techniques using CCSM4. Solid black lines are the mean reconstruction out of 30, dash-dotted lines are the actual model mean temperature. Gray shading is 1 standard deviation of the reconstructions. The calibration period is 1881-1980 and the reconstruction period is 1300-1880. The correlation coefficient, r, is noted at the top of each figure along with the number of PCs used for the PCA-based reconstruction. The anomalies are shown with respect to the reconstruction mean.

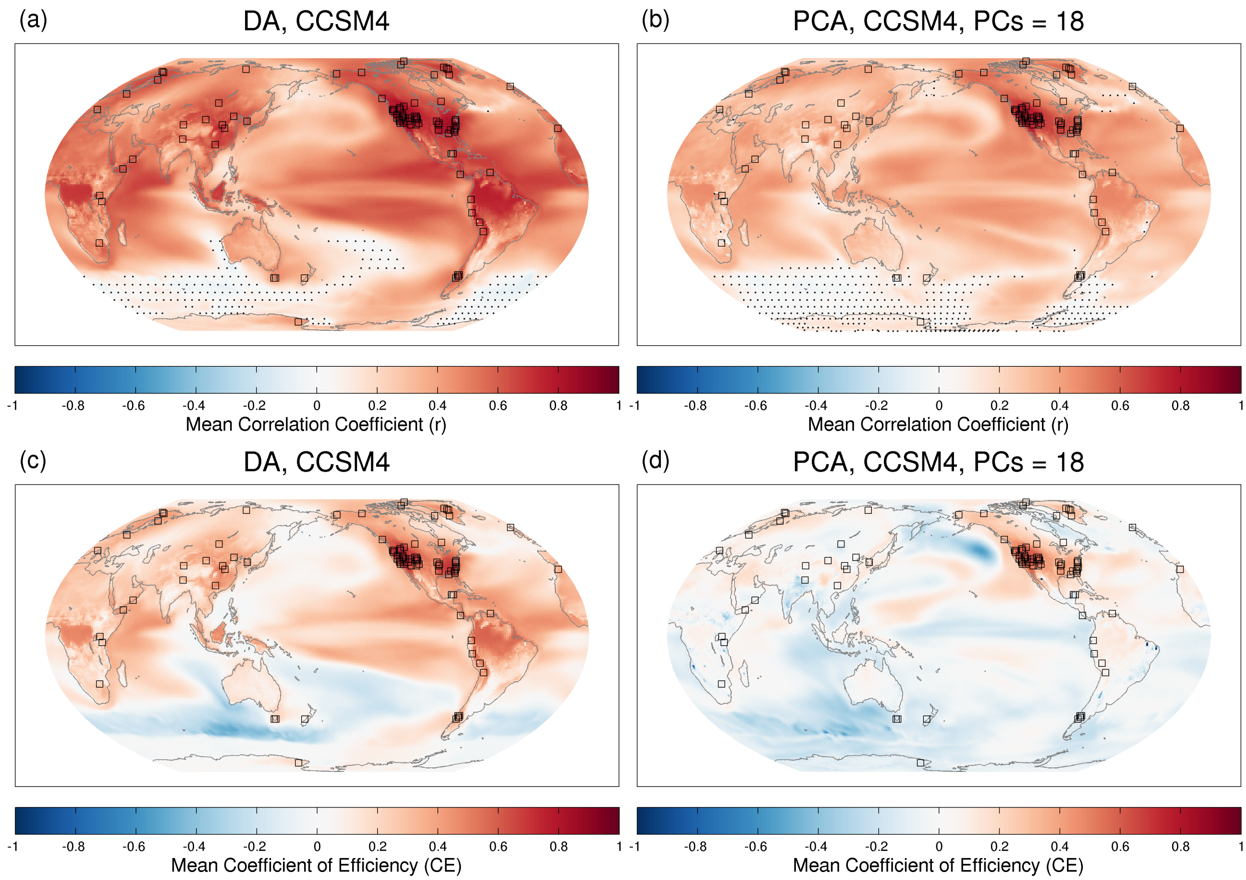

FIG. 2. Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 1 (calibration period: 1881-1980, reconstruction period: 1300-1880), for (a) DA and (b) PCA. These maps show the correlation and coefficient of efficiency between each grid point temperature series of the mean reconstruction (mean of 30) and each actual grid point temperature series. Empty black boxes are centered over pseudoproxy locations and stippling indicates correlations that are *not* significant at the 95% level.

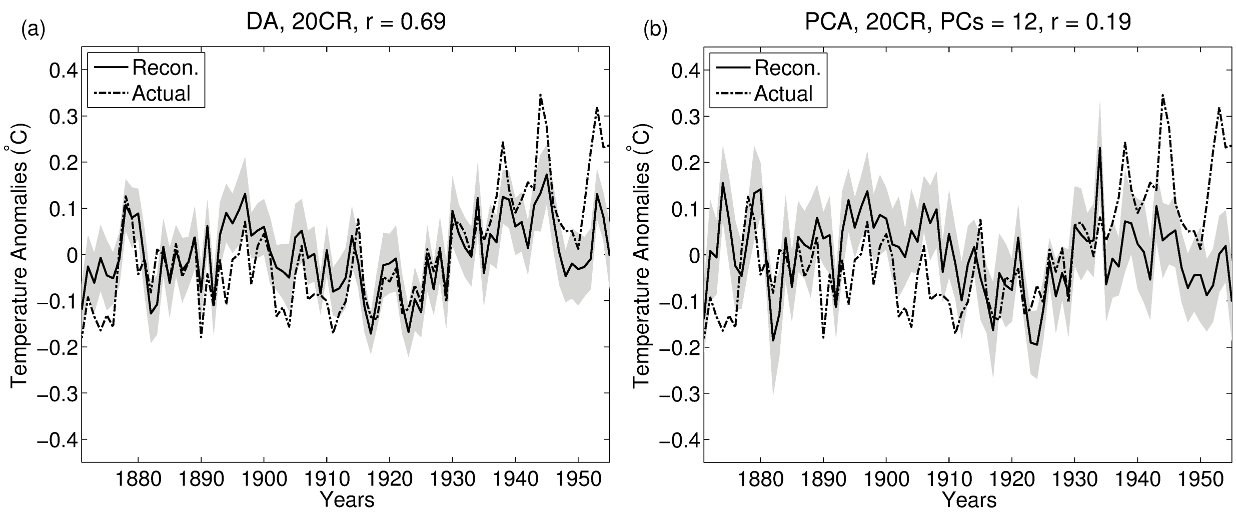

FIG. 3. Global-mean temperature anomaly reconstructions of the (a) DA and (b) PCA techniques using 20CR. The calibration period is 1956-2005 and the reconstruction period is 1871-1955.

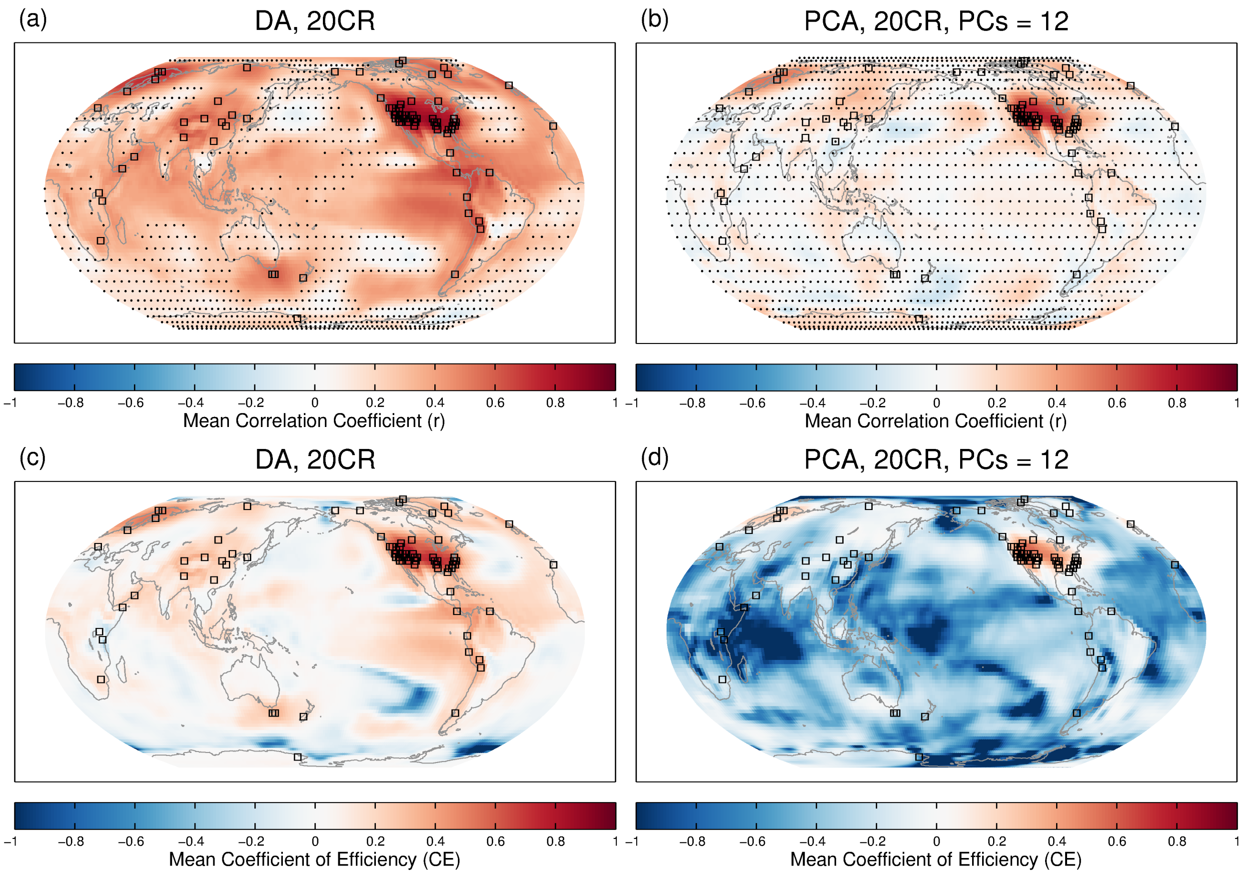

FIG. 4. Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 3 (calibration period: 1956-2005, reconstruction period: 1871-1955), for (a) DA and (b) PCA. For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level. The lower bound of CE values shown are cut off at CE $= -1$.

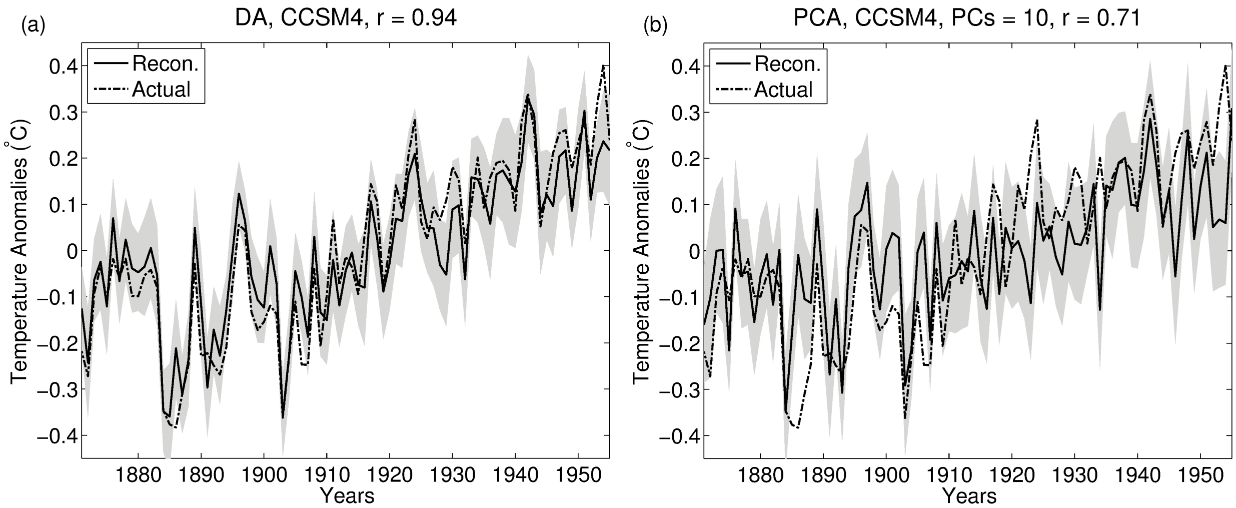

FIG. 5. Global-mean temperature anomaly reconstructions using (a) DA and (b) PCA techniques with CCSM4 over the same calibration and reconstruction periods as in Figs. 3 and 4 (calibration period: 1956-2005, reconstruction period: 1871-1955).

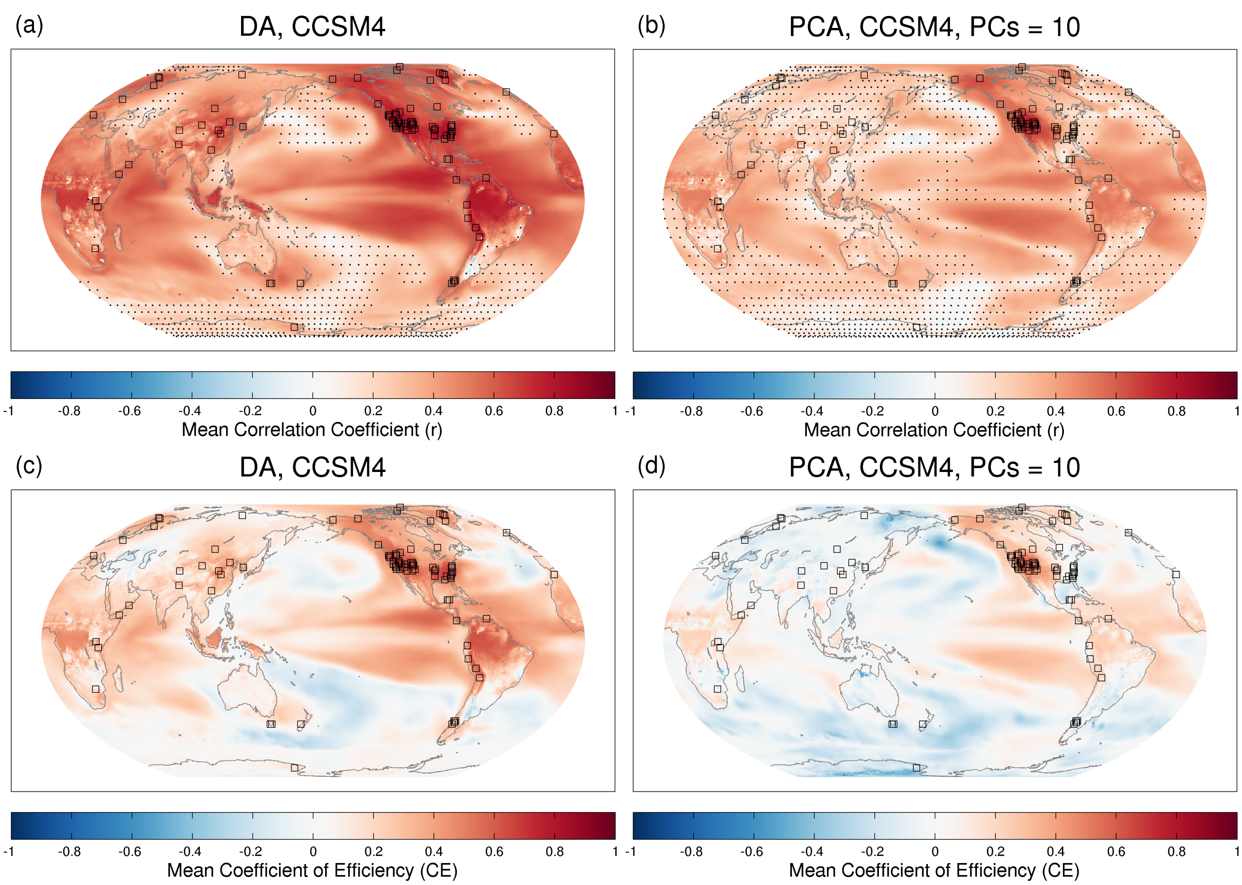

FIG. 6. Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 5 with CCSM4, (calibration period: 1956-2005, reconstruction period: 1871-1955). For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level.

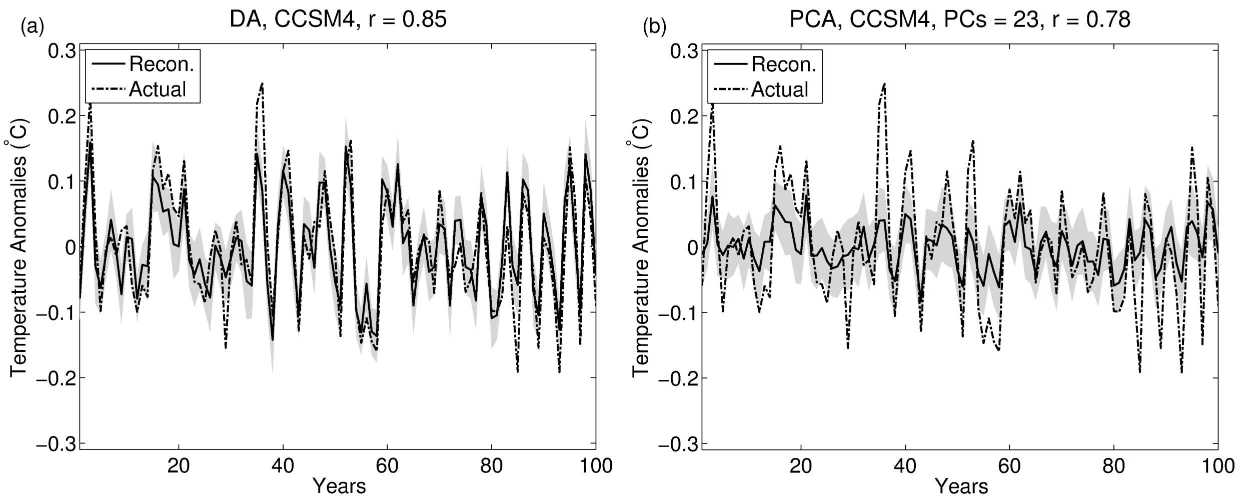

Fig. 7. Global-mean temperature anomaly reconstructions using (a) DA and (b) PCA techniques with 100 years of CCSM4 LGM data for the calibration period and 100 years of a CCSM4 pre-industrial control run for the reconstruction period.

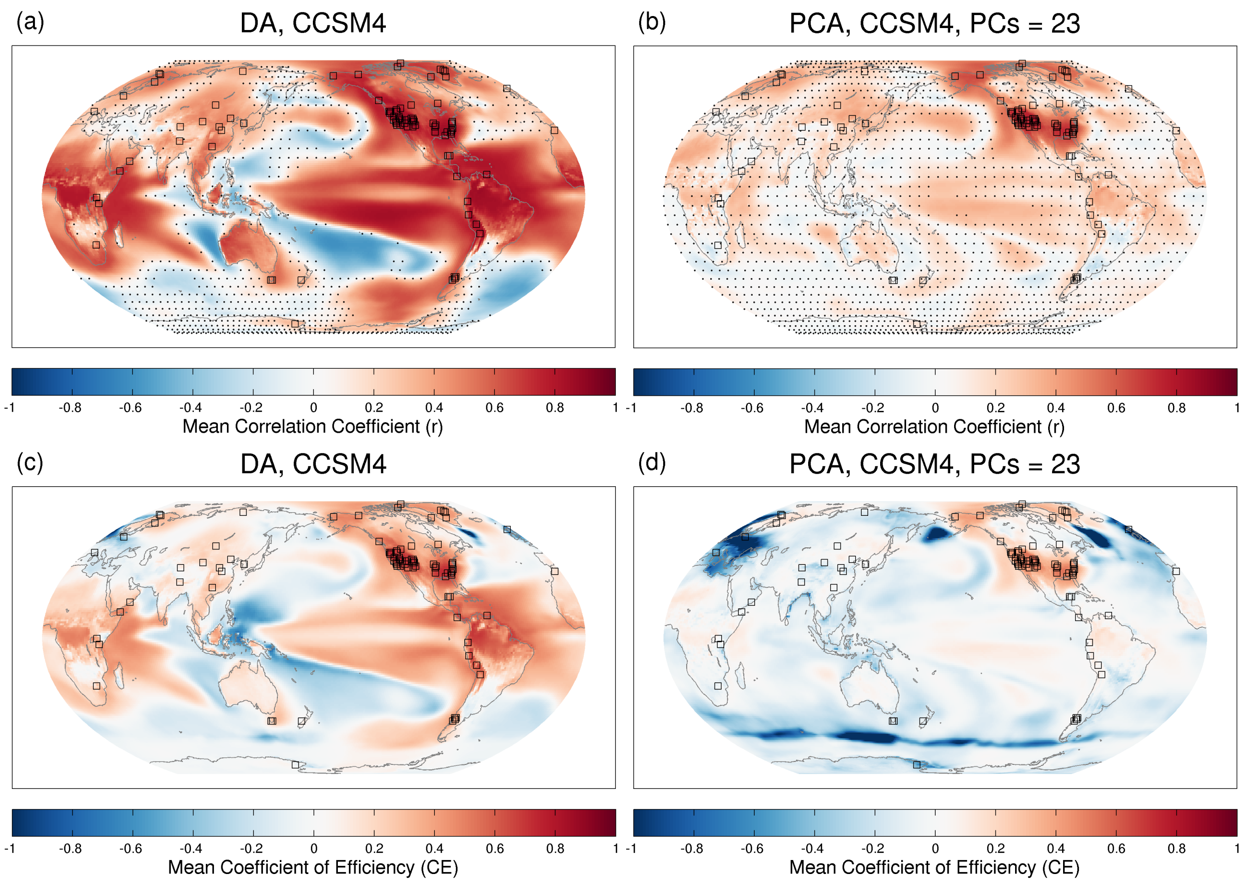

FIG. 8. Spatial maps of correlation coefficient, (a) and (b), as well as coefficient of efficiency, (c) and (d), corresponding to the reconstructions shown in Fig. 7. For (a) and (b), stippling indicates correlations that are *not* significant at the 95% level. The lower bound of CE values shown are cut off at CE $= -1$.

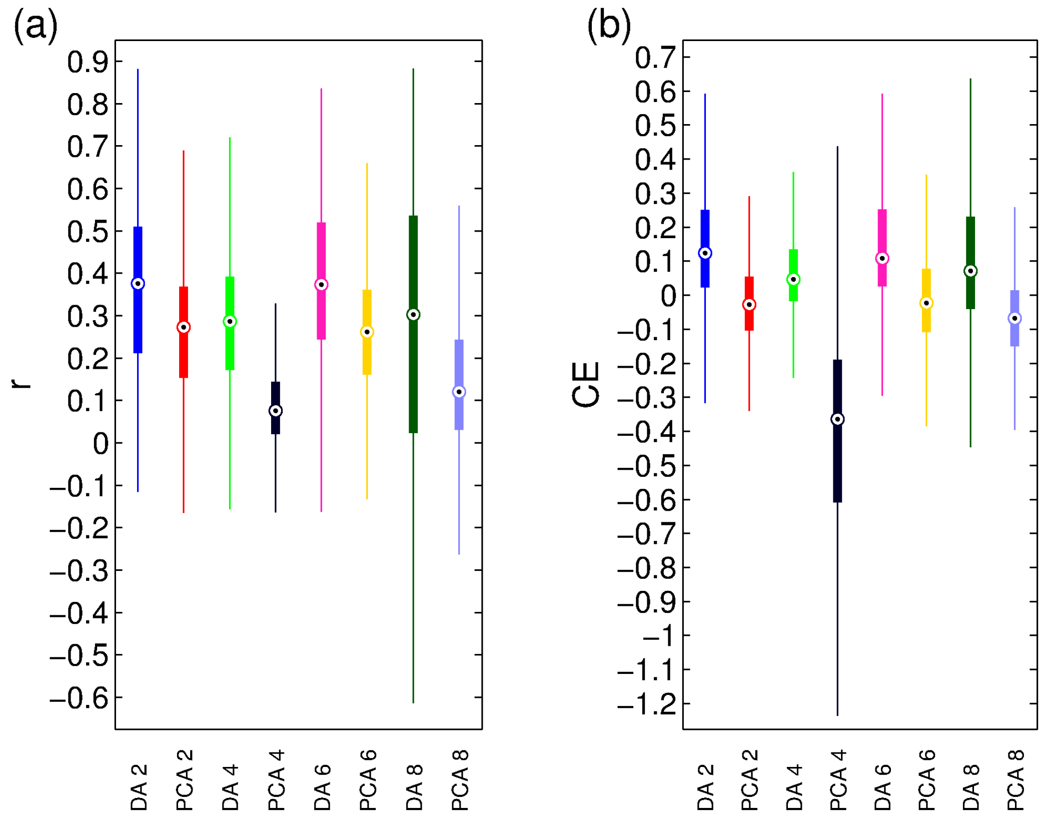

FIG. 9. Box-and-whisker plots (for clarity, outliers are not shown) of each of the spatial reconstruction figures, for (a) correlation coefficient (r) maps and (b) coefficient of efficiency (CE) maps. Labels refer to DA or PCA techniques and the figure number of the data that the box-and-whisker plots represent. All DA-PCA distribution pairs are statistically distinct according to t-tests for each DA-PCA comparison.

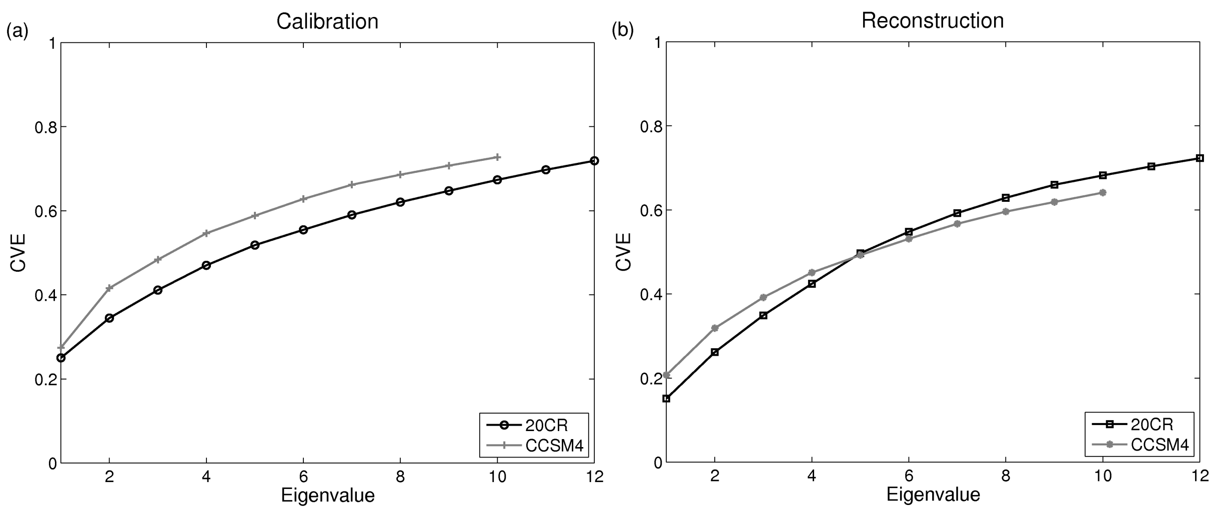

FIG. 10. Cumulative variance explained (CVE) of the retained EOFs for 20CR and CCSM4 during both the (a) calibration and (b) reconstruction periods shown in Figs. 3–6 (calibration period: 1956-2005, reconstruction period: 1871-1955). With 20CR we retain 12 PCs and with CCSM4 we retain 10 PCs; the CVE values are normalized by the total variance explained.