AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2016MS000677

On the utility of proxy system models for estimating climate states over the common era

Sylvia G. Dee[1], Nathan J. Steiger[2], Julien Emile-Geay[3], and Gregory J. Hakim[4]

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

Key Points:

- A new modeling framework combines data assimilation (DA)-based approaches with proxy system modeling (PSM)
- The impacts of assuming a linear mapping between climate variables and proxy data in climate reconstructions are explored
- Structural model errors must be mitigated to effectively combine PSMs with DA

Correspondence to:

S. G. Dee,
sylvia_dee@brown.edu

Citation:

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.

Received 25 MAR 2016
Accepted 16 JUN 2016
Accepted article online 20 JUN 2016
Published online 10 AUG 2016

Abstract Paleoclimate data assimilation has recently emerged as a promising technique to estimate past climate states. Here we test two of the underlying assumptions of paleoclimate data assimilation as applied so far: (1) climate proxies can be modeled as linear, univariate recorders of temperature and (2) structural errors in GCMs can be neglected. To investigate these two points and related uncertainties, we perform a series of synthetic, paleoclimate data assimilation-based reconstructions where "pseudo" proxies are generated with physically based proxy system models (PSMs) for coral $\delta^{18}O$, tree ring width, and ice core $\delta^{18}O$ using two isotope-enabled atmospheric general circulation models. For (1), we find that linear-univariate models efficiently capture the GCM's climate in ice cores and corals and do not lead to large losses in reconstruction skill. However, this does not hold for tree ring width, especially in regions where the trees' response is dominated by moisture supply; we quantify how the breakdown of this assumption lowers reconstruction skill for each proxy class. For (2), we find that climate model biases can introduce errors that greatly reduce reconstruction skill, with or without perfect proxy system models. We explore possible strategies for mitigating structural modeling errors in GCMs and discuss implications for paleoclimate reanalyses.

1. Introduction

The instrumental climate record is relatively brief and spatially incomplete. Through the use of proxies for past climate, paleoclimate reconstructions can dramatically extend the record of historical climate variability. This extension puts present and future climate changes in context and allows for the study of climate phenomena over long time scales, potentially improving predictions of future global change. Climate field reconstructions (CFRs) seek to estimate space-time climate states based on noisy and sparse paleoclimate proxy data. Most CFR techniques employ an inverse approach, directly relating climate fields (e.g., precipitation, temperature) to proxy observations (e.g., $\delta^{18}O$ in ice cores or corals) via linear, statistical relationships [e.g., *Mann et al.*, 1998, 1999; *Luterbacher et al.*, 2004; *Moberg et al.*, 2005; *Mann et al.*, 2008, 2009; *Kaufman et al.*, 2009; *Barriopedro et al.*, 2011; *Tingley and Huybers*, 2013; *Emile-Geay et al.*, 2013; *Guillot et al.*, 2015; *Wang et al.*, 2015]. However, extracting robust climate information from paleo-observations via such relationships is nontrivial: data from proxy sites do not offer uniform spatial coverage, can differ sharply in quality, and the availability of such data dramatically decreases back in time [*Wang et al.*, 2014]. Further, proxy records are in general nonlinear, nonunique, and noisy transforms of the input climate [*Tingley et al.*, 2012] and many are also time-uncertain. Thus, inverse approaches result in important losses of information.

Data assimilation (DA) provides a complimentary alternative reconstruction approach (for an exposition of the advantages of DA in a paleoclimate context, see *Steiger et al.* [2014]). A key feature of DA-based reconstructions is that they blend dynamical information from climate models with proxy observations to estimate past climate. This blending considers error estimates in both the models and the proxies and provides an optimal estimate of past climates given the models and the proxy data. Importantly, DA differs from traditional climate field reconstruction methods in that it need not rely on linear, statistical relationships between proxies and climate.

In practice, however, most data assimilation approaches have employed strictly linear temperature-proxy models, whereby temperature and proxy signal are linearly related via calibrations over the instrumental period [e.g., *Goosse et al*., 2006, 2010; *Widmann et al*., 2010; *Steiger et al*., 2014]. One central aim of this paper is to scrutinize the implications of this assumption for climate reconstruction purposes. Like Bayesian hierarchical models used for climate reconstruction [e.g., *Tolwinski-Ward et al*., 2014], DA-based reconstructions can alternatively employ nonlinear, physically based proxy system models (PSMs) [*Evans et al*., 2013] that map between climate states or climate model variables and proxy observations. Using these PSMs allows for a more physically-grounded climate reconstruction. The second aim of this paper is to consider the errors in global climate models (GCMs) alone, and evaluate the effects of GCM structural biases on reconstructions with and without embedded PSMs.

We first investigate the utility of PSMs for use in paleoclimate data assimilation [*Steiger et al*., 2014; *Hakim et al*., 2016]. We explore this within a synthetic, pseudoproxy framework [see *Smerdon*, 2012, for a review] because it allows for a controlled test bed for both the PSMs and the reconstruction technique. Typically, pseudoproxies are generated by simply adding noise to temperature series, but here we generate pseudoproxies using several PSMs [*Dee et al*., 2015]. PSMs translate relevant dynamical and isotope variables to modeled proxy space for a direct comparison between GCM output and observations, thereby alleviating the need for a calibration. Thus far, only *Evans et al*. [2014] have used PSM-generated proxies in pseudoproxy reconstructions; using a tree ring width model only, the authors found that the PSM-based reconstructions differed substantially from a standard "temperature plus noise" pseudoproxy model. Our study builds on this work by including PSMs for oxygen isotopes in corals and ice cores in addition to tree ring width. Importantly, this study uses PSMs for both the production of the pseudoproxies and their estimation in the DA reconstruction process; this differs from *Evans et al*. [2014] where a PSM for tree ring width was only used for the production of the pseudoproxies. Through a series of experiments reconstructing both surface temperature and geopotential height at 500 hPa, the PSM-pseudoproxy framework serves as a test bed for investigating assumptions concerning our understanding of proxy-temperature relationships, and how errors in these assumptions trickle down to the reconstructed fields. We compare baseline PSM-based reconstructions using the full, nonlinear PSM-mapping (`FULL-PSM`) with reconstructions using statistical relationships that are simply linear in temperature (`LU-PSM`).

Second, we are led to consider the competing influence of structural modeling error [*Frigg et al*., 2014; *Bradley et al*., 2014], accounting for uncertainty in GCMs. That is, since the climate models (used to describe relationships within and among climate fields) are imperfect representations of the true climate, is it worse to use a simplified, linear representation of proxy-climate relationships or an uncertain representation of climate (which ultimately propagates forward through PSMs)? For this second question, we reconstruct a CMIP-class climate simulation using an Earth system model of intermediate complexity as an approximation for real-world reconstructions. We investigate whether the biases inherent in climate models prohibit the effective use of proxy system models in climate reconstructions.

This study is structured as follows: section 2 describes the paleoclimate state estimation framework, as well as our experimental design. Section 3 explores the impacts of the linearity assumption as well as the impacts of structural modeling error on reconstructed global temperature. Finally, broader implications and extensions of this work are discussed in section 4.

2. Experimental Framework

2.1. Data Assimilation Approach

The essential feature of DA is that it optimally combines observations (within this context, proxy data) with models. In the context of this study, a climate model provides an initial, or prior, state estimate that one can update in a Bayesian sense using the observations. This update also takes account of an estimate of the errors in both the observations and the prior. The prior contains the climate model variables of interest, and the updated prior, called the posterior, is the best estimate of the climate state given the observations and the error estimates. The update equations of DA [e.g., *Kalnay*, 2003] (applied in optimality for linear and Gaussian approximations) 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

Journal of Advances in Modeling Earth Systems 10.1002/2016MS000677

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

with $\langle \cdot, \cdot \rangle$ representing a covariance expectation. The prior (or "background") estimate of the state vector is $\mathbf{x}_b$ and $\mathbf{x}_a$ is the posterior (or "analysis") state vector. Observations (or proxies) are contained in vector $\mathbf{y}$. The true value of the observations is 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. Within a paleoclimate reconstruction context, this is the PSM. Matrix $\mathbf{K}$, the Kalman gain, optimally weighs $\mathbf{y}-\mathcal{H}(\mathbf{x}_b)$ (the innovation), and transforms it into state space. Matrix $\mathbf{R}$ is the error covariance matrix for the observations and is assumed to be diagonal. The DA update process involves computing equations (1) and (2) to arrive at the posterior state; within the context of climate reconstruction, the posterior state is the reconstructed state for a given time. Space-time reconstructions are obtained by iteratively estimating the posterior state for each year of the reconstruction.

The specific DA implementation used in this study follows closely that of *Steiger et al*. [2014]. Briefly, it uses an off-line approach, wherein the prior ensemble, $\mathbf{x}_b$, consists of annually averaged climate states drawn from a climate model simulation; the ensemble is not integrated forward in time because of the massive computational constraints involved and because online DA for paleoclimate reconstructions appears to provide little improvement in skill over off-line DA, at least for atmospheric variables [*Matsikaris et al*., 2015]. The most important difference from *Steiger et al*. [2014] is that we use PSMs for generating $\mathbf{y}$ and for computing $\mathcal{H}(\mathbf{x}_b)$ (in most cases). This provides a more realistic reconstruction framework than that typically employed in pseudoproxy experiments. We also note that the particular time-averaged DA method employed here is formally valid when $\mathcal{H}(\mathbf{x}_b)$ is linear [*Huntley and Hakim*, 2010]. Because the "full" versions of the tree ring, coral, and ice core PSMs used in this study are actually nonlinear, we are therefore making an approximation in using this DA method for the experiments that include full PSMs. Therefore, the results for these experiments may have reduced skill relative to what may be possible with a fully nonlinear DA approach.

2.2. Pseudoproxy Network

Our pseudoproxy network is broadly representative of the availability of annually resolved tree, coral and ice core records in the PAGES2k and Ocean2k databases [*PAGES2K Consortium*, 2013; *Tierney et al*., 2015]. We have added 32 coral record locations to the existing PAGES2k Phase 1 network, as this original network contained only 13 coral sites. The additional coral sites are based on published data, matching the coral network described in *Comboul et al*. [2014]. The network includes coral aragonite $\delta^{18}O$, tree ring width, and ice core $\delta^{18}O$ records only. We focus on these three proxy types alone due to richness of existing data, age

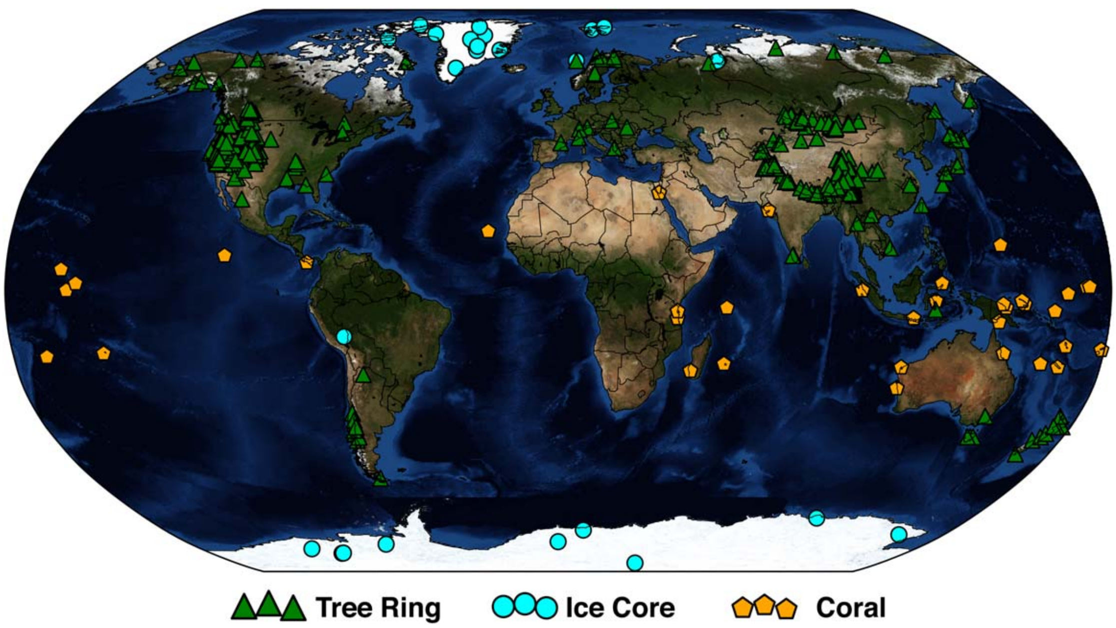

Figure 1. PRYSM Simulated Multi-Proxy System Model Network, PAGES2k Phase 1 + Tier1 Corals. Pseudoproxy network design: proxy locations and types for `FULL-PSM` and `LU-PSM`.

control, demonstrated sensitivity to climate variables of interest, and the existence of available PSMs for each proxy type.

Proxy data types and locations are shown in Figure 1. Most of the pseudoproxy network (467/544 total records) is comprised of tree ring width (TRW) locations. The resulting pseudoproxy network achieves fairly wide coverage; note however that most of the tree ring data are confined to Northern Hemisphere land masses and that extra-tropical oceans are not sampled at all.

2.3. Climate Model Simulations

To provide climate state vectors for the DA prior (see equation (1)), and to provide climate fields for the PSM-generated network, we primarily use the isotope-enabled GCM `SPEEDY-IER` [*Dee et al*., 2014]. `SPEEDY-IER` is an intermediate complexity atmospheric GCM (IC-AGCM), and constitutes an efficient option for long paleoclimate integrations. Despite its simplicity, the IC-AGCM computes climatic and water isotope fields that are comparable with higher-order AGCMs at a fraction of the cost. A simulation of `SPEEDY-IER` was forced with sea surface temperatures from the last-millennium simulation of the CCSM4 coupled model [*Landrum et al*., 2013], spanning the years 1000–2004 from that simulation. A second water isotope-enabled simulation forced with historical SST boundary conditions spanning 1871–2011 from `ECHAM5-wiso` was employed for the tests of structural biases. `ECHAM5-wiso` is a higher-order, higher resolution (run here at ~1° resolution) AGCM than `SPEEDY-IER`, and the computational expense restricted the simulation to a relatively shorter time interval.

We are limited to these two GCM simulations in this study: we require embedded water isotope physics to run the PSMs, and a large prior (e.g., a last-millennium simulation) for the DA. While a number of water isotope-enabled GCMs have been published (see Stable Water Isotope Inter-comparison Group projects SWING and SWING2, <http://www.giss.nasa.gov/staff/gschmidt/SWING2.html>), these GCMs are quite computationally expensive, and none have performed a publicly available last-millennium simulation with water isotopes.

2.4. PSM-Generated Pseudoproxies

We generate synthetic proxy fields via process (forward) models for each proxy type [see *Evans et al*., 2013 for a review], which transform the simulated climate signal (e.g., temperature and precipitation) to synthetic proxy observations. Each PSM includes three submodels, each of which mimics a separate transformation of the original input signal as it would occur in nature: a *sensor* model, which describes any physical, geochemical or biological processes altering the climate signal; an *archive model*, which accounts for any processes that affect the emplacement of the signal in the proxy medium; and an *observation* model, which accounts for dating uncertainties and

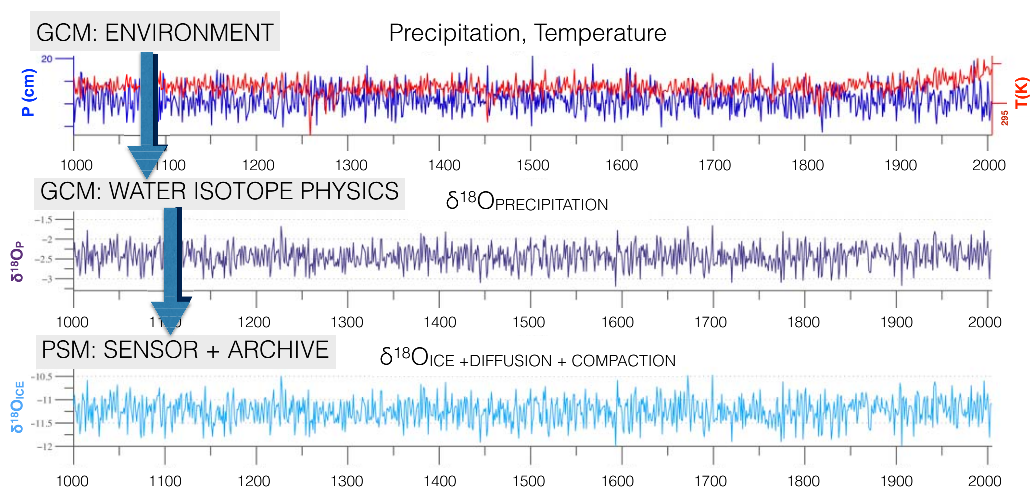

Figure 2. From climate signal to proxy: water isotope ratios and the ice core PSM. The GCM + PSM framework simulates the proxy signal from environmental inputs. The figure shows an example for water isotope ratios in ice core annual layers: modeled ice core $\delta^{18}O$ for Quelccaya, Peru. The signal starts as modeled T, P from the climate model simulation. The isotope-enabled GCM water isotope fields are used to drive the PSM's sensor and archive models which mimic diffusion and compaction processes in the core.

analytical errors in the measurements made on the archive [*Dee et al*., 2015]. Example output for a forward model of ice core $\delta^{18}O$, showing the breakdown and output for each submodel, is given in Figure 2.

Proxy data are lossy transforms of the original climate signal (i.e., processes inherent to the proxy system can filter out and subsequently result in considerable loss of the desired climate information) [*Tolwinski-Ward et al*., 2014]. The submodel (*sensor*, *archive*, and *observation*) framework of PSMs helps to quantify the information loss at each stage of the climate signal's evolution through the proxy system. For example, Figure 2 illustrates how the original climate signal (temperature and precipitation) and the water isotope fields are transformed by the ice core's *archive* model, which accounts for compaction and diffusion. Therefore, this proxy system modeling attempts to mimic the many climatic and/or biological processes that occur in real proxies at seasonal to interannual time scales. While the PSMs employed here take monthly climate inputs, they output single values in proxy units for a given calendar year, just as would be measured in most annually-resolved proxies.

Proxy system models are driven by monthly output from water isotope-enabled climate models for each of the locations in the above-described network. Each proxy type employs its own PSM. We used VS-Lite [*Tolwinski-Ward et al*., 2010] to generate tree ring width records for all of the tree proxy locations in the network using GCM temperature and precipitation fields; this model accounts for the seasonal dependence on tree-growth via simple parameterizations of temperature and moisture threshold responses at monthly scales. While a number of other variables have been used to reconstruct temperature from trees (e.g., maximum latewood density), we are not aware of a publicly available forward model for these additional variables, and so do not consider them here. The coral forward model follows the parameterization described in *Thompson et al*. [2011] and depends on both sea surface temperature and salinity anomalies; both the ice core and coral records were modeled using the water isotope-enabled model fields coupled with a synthesis of previously published models for water isotopes in high-resolution proxy data (PRYSM) [*Dee et al*., 2015]. The ice core model takes into account precipitation accumulation, local temperature, and incorporates dynamical information from the water isotope physics fields ($\delta^{18}O$ of precipitation) of `SPEEDY-IER` and `ECHAM5-wiso`. Seasonal controls on the simulated $\delta^{18}O$ of ice are accounted for via precipitation-weighting of the water isotopes stored in each annual layer of ice.

2.5. Reconstruction Experiments

The first of the questions we address is: how are reconstructions affected by assuming proxies are linear, univariate responders to temperature? For this question, we conduct a baseline experiment, called `FULL-PSM`, in which we reconstruct climate fields using PSM-derived pseudoproxies for both the proxies **y** and the prior estimate of the proxies $\mathcal{H}(\mathbf{x}_b)$ (equation 1); this would be equivalent in the real reconstruction problem to saying that we know the precise processes that lead to each proxy measurement, and these processes are fully represented by the PSM. Then we test how well a linear temperature-proxy approximation of $\mathcal{H}(\mathbf{x}_b)$, called here `LU-PSM` for "linear univariate," performs relative to the baseline experiment; this linear model is simply $\hat{y}=\beta_1+\beta_2*T+\epsilon$, where $\hat{y}$ is a vector of PSM-derived pseudoproxy values, β_i are coefficients, T is the local surface temperature, ϵ is a Gaussian white noise process with zero mean ($\epsilon\sim\mathcal{N}(0,\sigma^2)$) and the fit is established over the years of the prior. We use this linear model to substitute $\beta_1+\beta_2*T$ for $\mathcal{H}(\mathbf{x}_b)$ in equation (2). The error variance residuals from the linear fit are then used to define the diagonal elements of **R** (we assume off-diagonal elements are zero).

For these experiments, we use the `SPEEDY-IER` simulation and reconstruct the period 1251–1755. The prior consists of the 500 total years surrounding this period, 1000–1250 and 1756–2004. The reconstructed spatial variables are 2 m air surface temperature and 500 hPa geopotential height, or Z_{500}. These experiments were performed with all the pseudoproxies (ice cores, corals, and tree rings) and each proxy individually to see which proxies were affected most by the linearization. We repeated all of the reconstructions 100 times in a Monte Carlo fashion, randomly sampling 75% of the proxies on each iteration. We added different noise realizations to the proxies for each iteration to approximate small measurement errors (errors of 0.1 ‰ for the isotope-based pseudoproxies and 1% errors for tree ring width measurements); the specific values of these noise realizations were different for each iteration while the statistics of the realizations were the same. We note that the error estimates shown in the figures are based on a combination of this Monte Carlo process and the posterior ensemble spread.

The second question we address is: do the biases inherent in climate models strongly affect the use of proxy system models in climate reconstructions? To answer this, we use the simplified `SPEEDY-IER` as approximation to the more realistic `ECHAM5-wiso`. In other words, we claim that `SPEEDY-IER` is to `ECHAM5-wiso` what `ECHAM5-wiso` or any other CMIP5-class model is to nature (the levels of simplification are comparable).

We thus used the PSMs to generate both the proxies **y** and the prior estimate of the proxies $\mathcal{H}(\mathbf{x}_b)$, but the input fields for **y** come from the `ECHAM5-wiso` simulation while the input fields for $\mathcal{H}(\mathbf{x}_b)$ come from SPEEDY-IER. The prior for these experiments consists of 500 randomly drawn years from the `SPEEDY-IER` simulation and we reconstruct the entire 1871–2011 period of the `ECHAM5-wiso` simulation. The same Monte Carlo approach employed in the first set of reconstructions was used for this second experiment.

3. Results

3.1. Linear, Univariate PSMs

Reconstructions and associated skill values for both `FULL-PSM` and `LU-PSM` experiments (as described in section 2.5) are given in Figures 3, 4, and 5. Figure 3 shows the global mean temperature reconstructions for both experiments. We see that the two reconstructions are largely comparable, and suggest that the linear approximation performs similarly to the `FULL-PSM` case. As measures of skill we report the correlation coefficient (R) and coefficient of efficiency (CE), which is defined as *Nash and Sutcliffe* [1970]:

$$CE = 1 - \frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{n} (x_i - \overline{x_i})^2} \tag{3}$$

where x is the "true" time series, $\bar{x}$ is the mean of the true time series, and $\hat{x}$ is the reconstruction time series. CE is a scaled measure of the mean-squared error, and rewards adequate estimation of the phase,

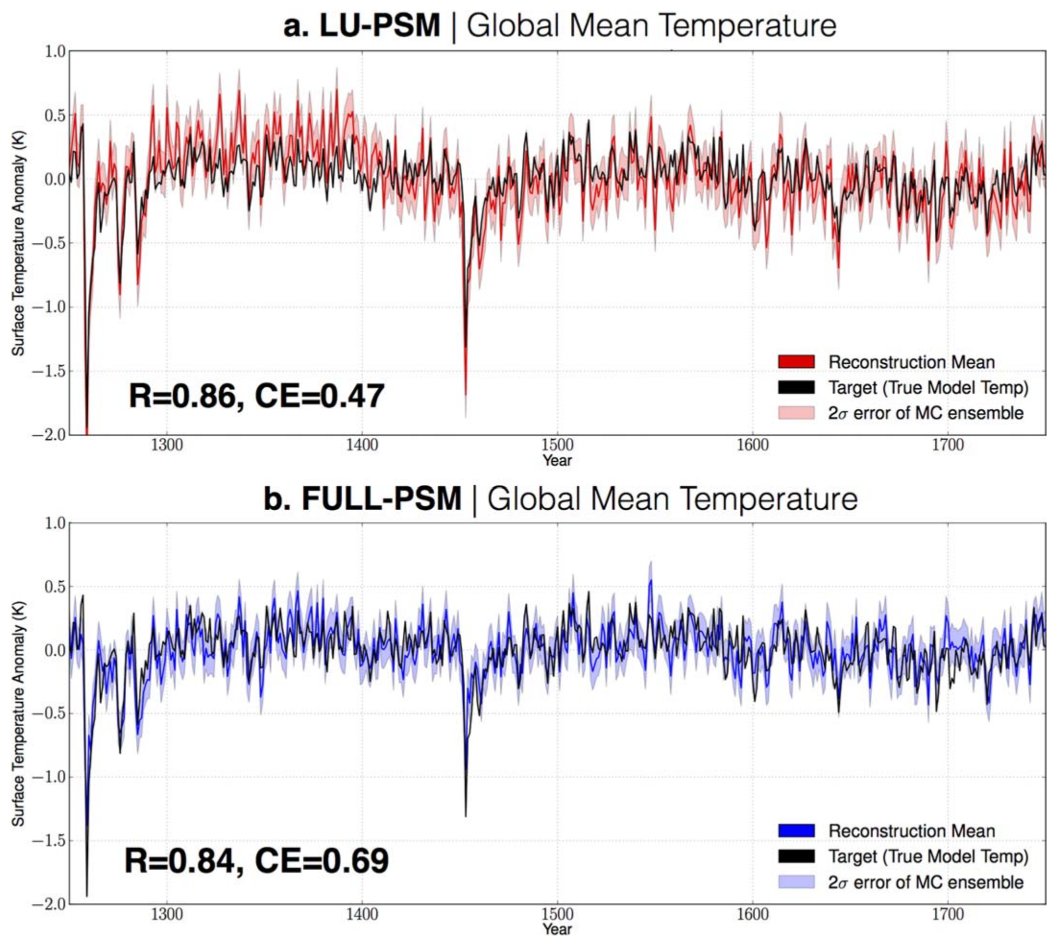

Figure 3. Global mean temperature reconstructions for `LU-PSM` and `FULL-PSM` experiments. (a) Global mean surface temperature reconstruction for linearized network `LU-PSM`. (b) Global mean surface temperature reconstruction for PSM-generated network `FULL-PSM`. In both figures, the red (Figure 3a)/blue (Figure 3b) line shows the reconstruction mean, and the black line is the true model temperature (the target field). Color shading indicates the $\pm 2\sigma$ range of the MC iterations. Skill scores (R, CE) are for the reconstructed global mean time series.

amplitude, and mean value of the field. The correlation coefficient, R, gives a measure of the covariability in terms of phasing between the reconstructed variables and the true variables. CE is considered the more stringent metric, and is limited to a value less than or equal to R^2. For reference, a CE of 1 indicates a perfect reconstruction, while positive CE values are generally indicative of a skillful reconstruction. For the global mean temperature reconstruction, R = 0.86, CE = 0.47 for the `LU-PSM` experiment, and R = 0.84, CE = 0.69 for the `FULL-PSM` experiment. Thus, marked improvements in reconstruction skill are observed as per the more stringent CE metric with the `FULL-PSM`.

Exploring the improvement in the CE score in more detail, Figure 4 compares the global skill maps for R (Figures 4a–4d) and CE (Figures 4e–4h) for both surface temperature and Z_{500}. This figure tells a similar story to the global mean temperature reconstructions in Figure 3, in that both `FULL-PSM` and `LU-PSM` perform similarly for all proxies together, though with notable exceptions in areas that contain high-density groupings of tree rings such as over North America. The `FULL-PSM` shows large improvements over `LU-PSM`; these results hold for both temperature and Z_{500}, though each have different spatial patterns of skill. The distributions of these skill metrics are summarized in Figure 5 via box plots, and confirm that for all cases, the `FULL-PSM` experiment does improve the overall spatial skill of the reconstruction.

We further investigate the spatial variability in CE scores between the two methods by computing difference maps of CE for the all-proxy network (`FULL-PSM` - `LU-PSM`), as well as difference maps for individual proxy experiments. Figure 6 shows the difference maps for reconstruction skill for the full proxy network and by individual proxy type for surface temperature and Z_{500}. This allows us to assess the importance of the FULL versus LU methodology in different proxy classes. Figure 6b highlights the fact that with the linear-univariate model, climate information can be lost when controls other than annual temperature are important.

When all proxy types are included (Figure 6a), the largest improvements using the `FULL-PSMs` occur in areas with high concentration of tree ring width records. This is a result of the fact that the tree ring width model, VS-lite, is poorly approximated by a linear fit with temperature; indeed, VS-lite is a nonlinear, nonstationary model that allows for both temperature and moisture to interact in their effects on tree ring growth. We note that the large difference in skill across the tropics in Figure 6b is not because the VS-lite-based pseudo tree rings are uniquely suited to reconstructing the tropics; temperature observations at annual time scales in the tree ring locations can provide satisfactory tropical skill [e.g., *Wang et al.*, 2014]. Rather, the particularly poor linear temperature approximation is the reason for the large *differences* in skill.

Figure 7 shows the range of the regression slopes (β_2 described in section 2.5) for the `LU-PSM` experiment by proxy type. The small range in the coral and ice core data demonstrates that compared to tree ring width, these proxies have very consistent linear relationships with temperature. For the ice core network, because the relationship between temperature and $\delta^{18}O_{ICE}$ is locally linear in `SPEEDY-IER` (as seen in observations of ice cores and local temperature) [*Jouzel et al.*, 1997], the `FULL-PSM` offers little improvement on a linear fit. However, this locally linear response to temperature may vary amongst isotope-enabled models, especially given the intermediate complexity of `SPEEDY-IER`.

Interestingly, the `FULL-PSM` for corals does add some skill to the reconstruction locally in the western Pacific, but actually causes a very slight reduction in skill across much of the tropics (Figure 6c). From the first covariance term of equation (2), we can interpret **K** as "spreading" the information contained in the observations through the covariance between the prior and the prior-estimated observations. This implies that, other things being equal, larger values of $cov(\mathbf{x}_b, \mathcal{H}(\mathbf{x}_b))$ will weight the innovation more heavily; thus this new information not contained in the prior has a bigger influence. Figure 8 shows the mean correlation length scale between the two different pseudoproxy coral types ($\mathcal{H}(\mathbf{x}_b)$) and the prior variables ($\mathbf{x}_b$). The mean correlation length scale is found by computing point correlation maps for the coral locations, binning these correlations by distance, and finally computing the mean of each bin. From Figure 8 we see that the `LU-PSM` estimates of the corals have consistently higher correlations with surface temperature and geopotential height. Thus the linear fit with temperature is effectively extracting a better climate signal from the perspective of the reconstruction methodology: these coral proxy estimates covary more strongly with 2 m air temperature and Z_{500}. Certainly some information about the corals is lost in this process, but this information turns out not to be as useful for the reconstruction.

To summarize and answer the question: *how are reconstructions affected by assuming proxies are linear, univariate responders to temperature?*, we find that for proxy systems that exhibit a generally linear response to

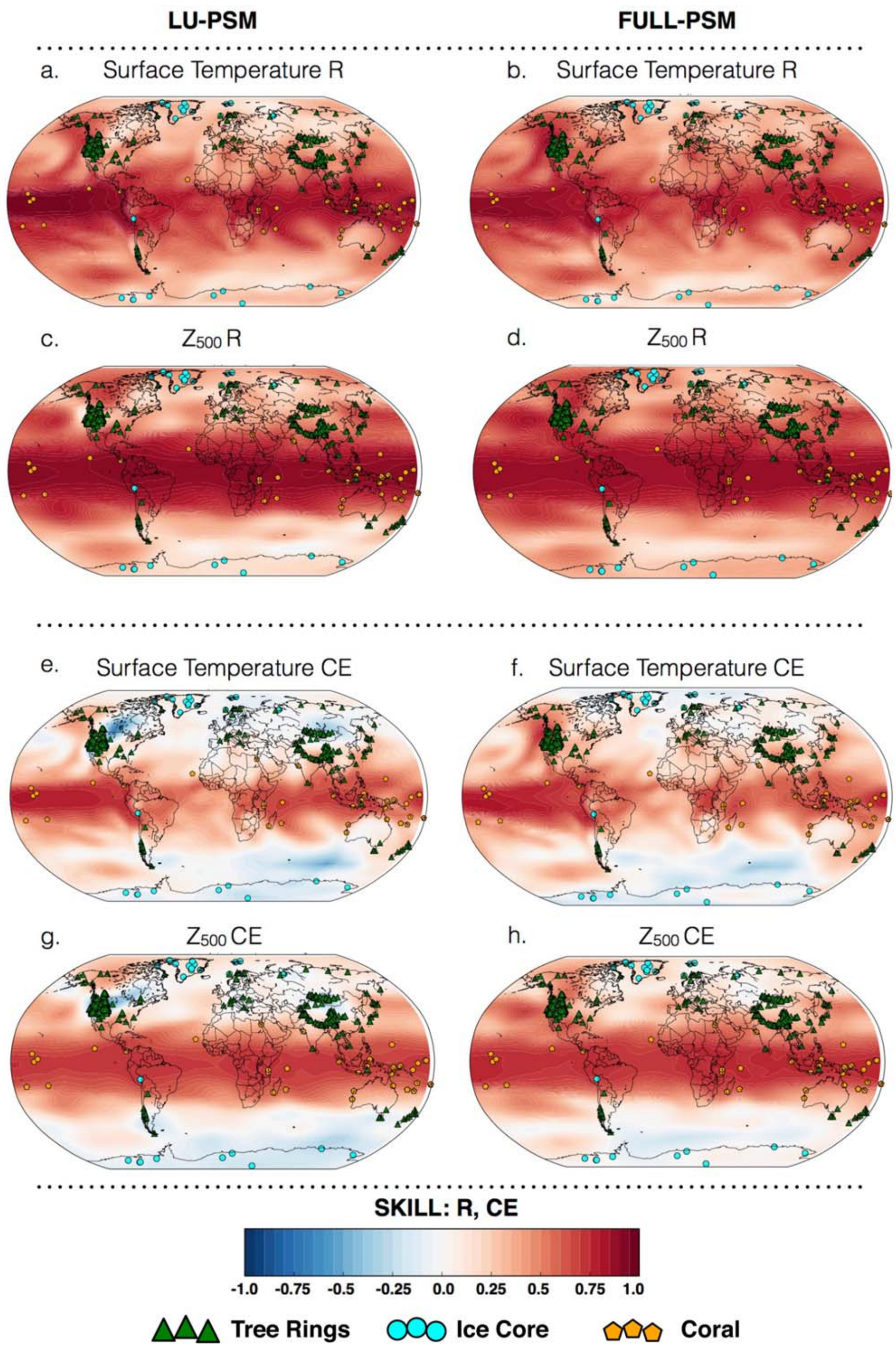

Figure 4. Spatial reconstruction skill for `LU-PSM` and `FULL-PSM` experiments. (a) R statistic, surface temperature, `LU-PSM`. (b) R statistic, `FULL-PSM`. (c) R statistic, Z_{500}, `LU-PSM`. (d) R statistic, Z_{500}, `FULL-PSM`. (e) CE statistic, surface temperature, `LU-PSM`. (f) CE statistic, `FULL-PSM`. (g) CE statistic, Z_{500}, LU-PSM. (h) CE statistic, Z_{500}, `FULL-PSM`. Proxy sites/types are superimposed on all maps.

temperature, the `LU-PSM` experiment performs just as well reconstructing past temperature as the `FULL-PSM` representation. However, we note that as discussed in section 2.1, the particular assumptions of the DA method used here may underestimate the potential skill of a `FULL-PSM`-like experiment that one could achieve with a fully nonlinear DA method (e.g., a particle filter). As expected, nonlinear, multivariate proxy system models offer improved reconstruction skill when the proxy response is clearly

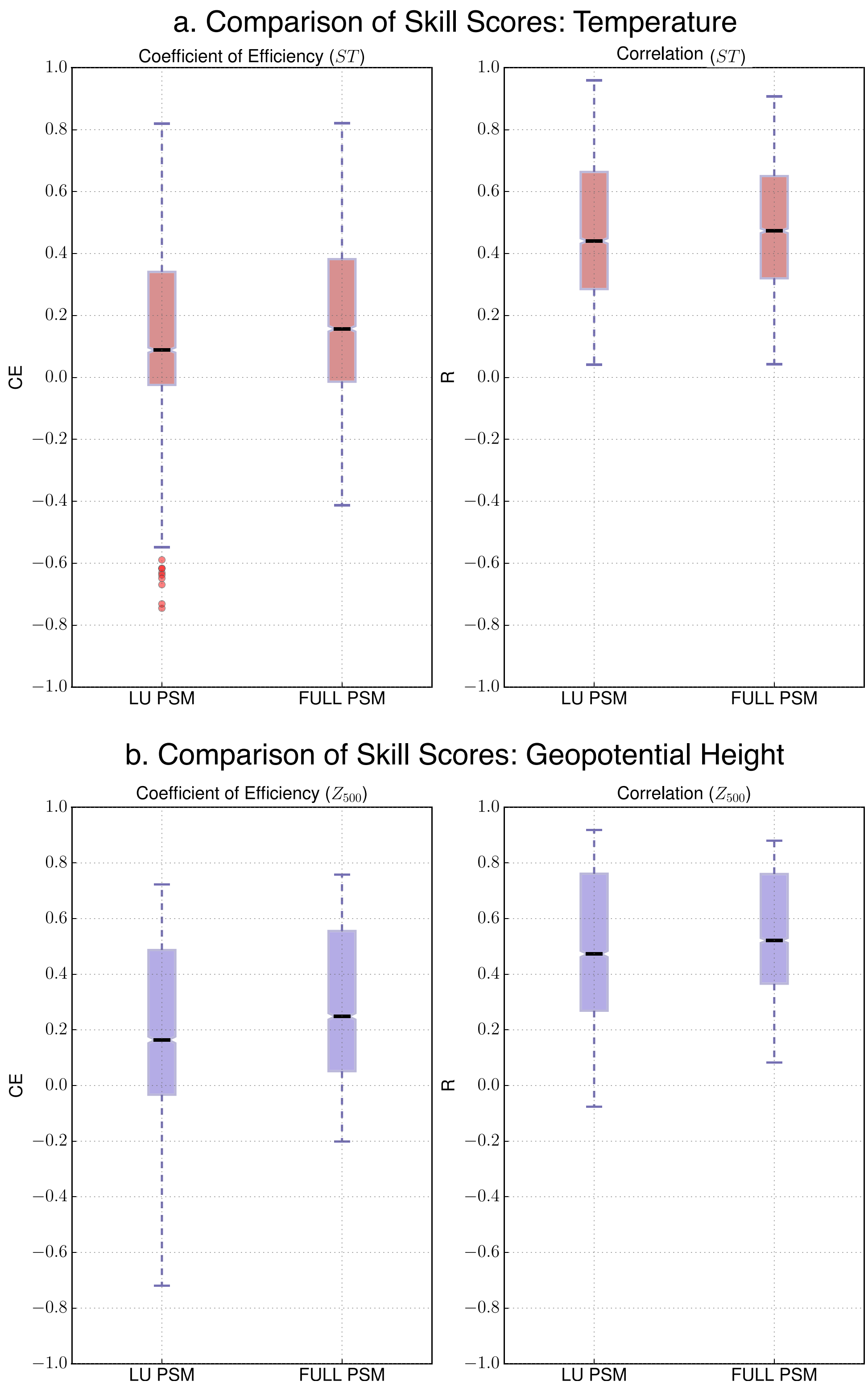

Figure 5. Box plot summary of the distribution of all spatial values for skill scores, `LU-PSM` versus `FULL-PSM`. Range of R and CE values at all model grid cells for (a) global mean surface temperature and (b) geopotential height at 500 hPa.

nonlinear and/or multivariate. In the case of tree ring width, which harbors dual sensitivity to temperature and moisture, as well as threshold effects in both, using a `FULL-PSM` representation yields a large improvement in reconstruction skill. Put differently, approximating temperature-proxy relationships with

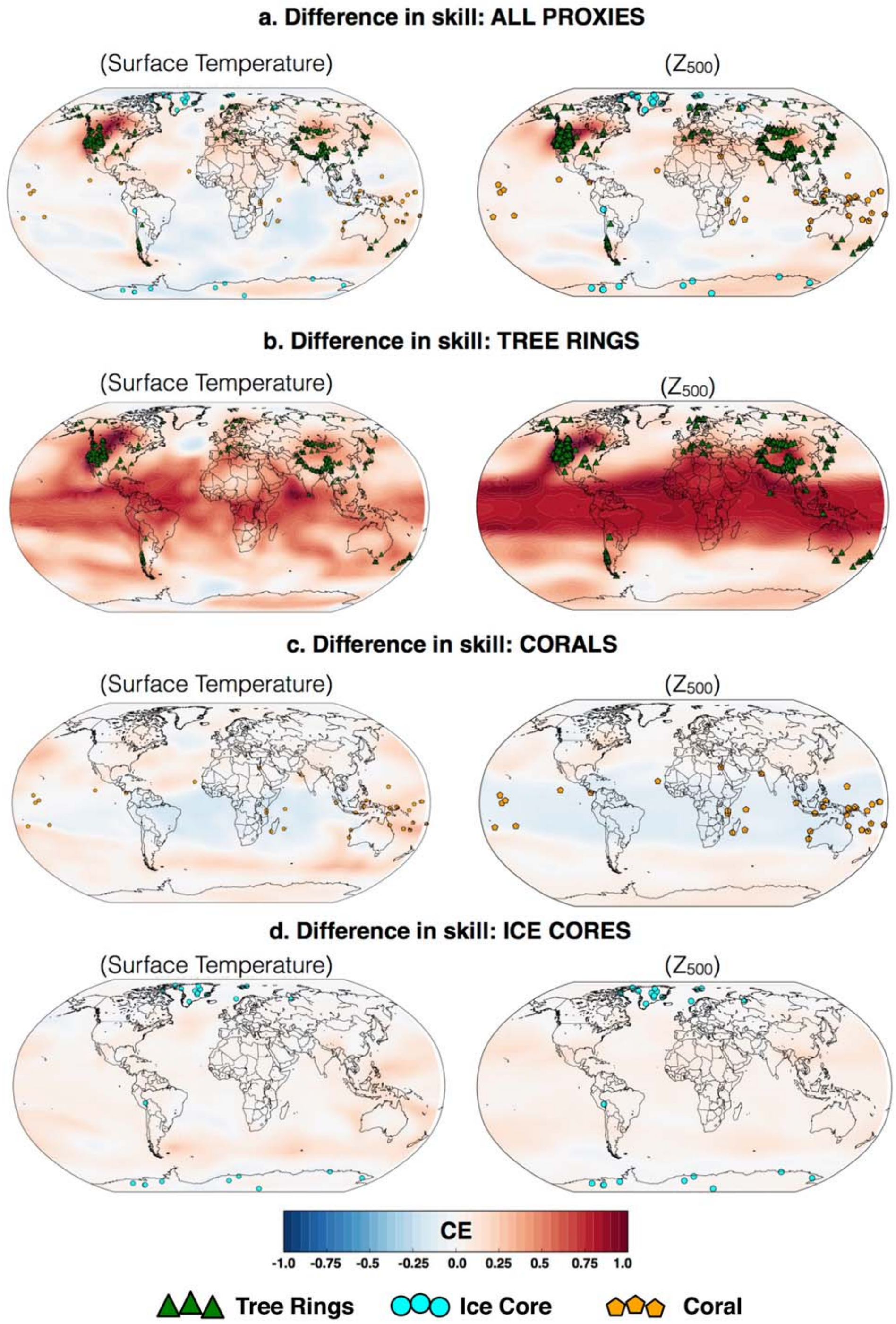

Figure 6. Reconstruction skill sensitivity to imposed linearity in proxy systems: difference maps (CE) for [FULL-PSM - LU-PSM]. Improvement in CE for (a) full network, (b) tree ring width network only, (c) coral network only, and (d) ice core network only. Proxy sites/types superimposed on all maps.

a linear, univariate model causes a considerable loss in the amount of retrievable climate information when this assumption is violated.

3.2. Structural Modeling Error

We now tackle our second motivating question and assess the impacts of structural modeling error (SME) in GCMs. Climate models do not provide a perfect representation of nature, and thus their use in the DA

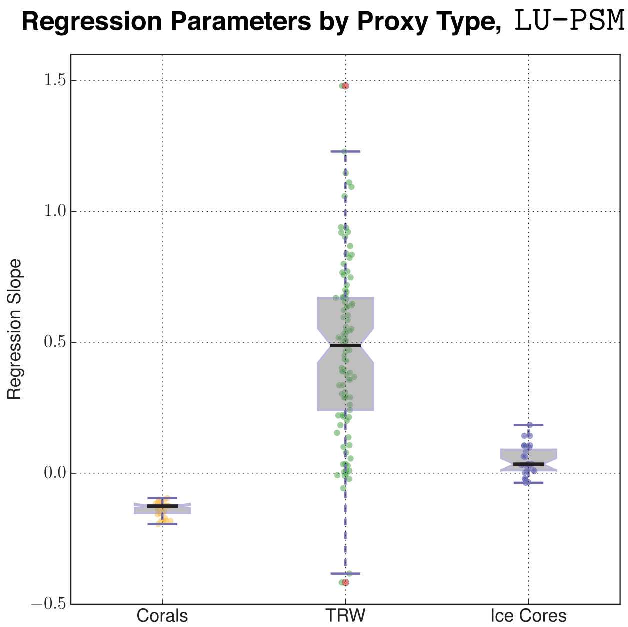

Figure 7. Regression parameters for the `LU-PSM` experiment. (left) Corals, (middle) TRW, and (right) ice cores. See description for regression parameters (β) in section 2.5.

framework as inputs for the PSMs introduces uncertainties in the reconstruction process. In this context, *structural model error* pertains to the differences between GCM(s) and nature (via both structural and parametric uncertainty), and not variability among different GCMs.

To explicitly test the impacts of GCM SME, we use a different GCM (`ECHAM5-wiso`) [*Werner et al.*, 2011] and assign it to represent the "truth" (nature), while `SPEEDY-IER` generates the background climate state that is used to reconstruct the true state from noisy pseudoproxies. Referring back to equation (1), `ECHAM5-wiso`-generated pseudoproxies are used as y, or the "real" proxy observations, and `SPEEDY-IER`-generated pseudoproxies are $\mathcal{H}(x_b)$, the "estimated" proxy. The experiment simulates the case where the prior is not drawn from the "true" distribution of climate state vectors, but instead an imperfect approximation of that distribution. The present work is the first to embed multiple PSMs in the DA framework, so we conducted this experiment to broaden awareness of uncertainties that may limit the usefulness of the method.

Figure 9 shows box plot summaries of the reconstructions from section 3.1 together with the SME experiment. From this comparison, we see that the imposition of SME to the `FULL-PSM` reconstruction causes a large skill reduction in both R and CE. These results suggest that if the DA prior is not drawn from the true distribution, SMEs carry forward through the PSMs and markedly reduce reconstruction skill. Therefore, structural biases housed in GCMs may prove an important limiting factor in using PSMs for climate reconstructions.

We thus performed two companion experiments to investigate how the impacts of SME may be mitigated. The first approach is to simply revert to the linear mapping between climate and proxy as done in parts of section 3.1. Using the `LU-PSM` method could prevent the introduction and subsequent propagation of

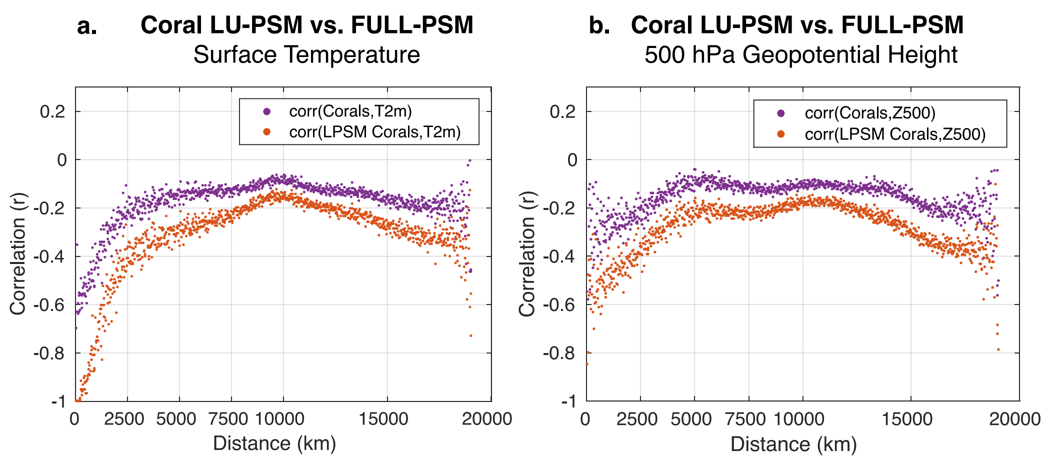

Figure 8. Correlation length scale. Mean correlation length scales for the `FULL-PSM` (purple) and `LU-PSM` (orange) corals and the reconstructed variables, (a) surface temperature and (b) geopotential height.

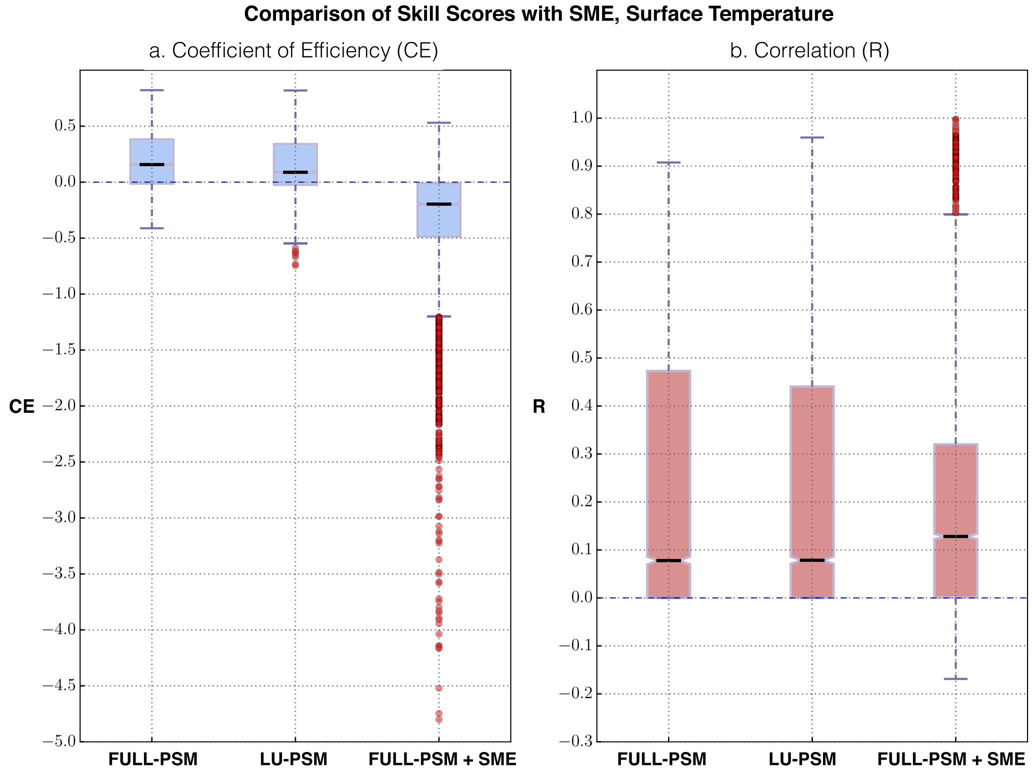

Figure 9. Reconstruction skill sensitivity to structural modeling error (SME). The two box plots compare the skill metrics ((left) CE and (right) R) for each of the experiments in this study: `FULL-PSM`, `LU-PSM`, and the `FULL-PSM` with imposed SME. Red dots indicate outliers beyond the [±1.5·IQR] range.

large structural errors via the PSMs because the proxy estimates, $\mathcal{H}(x_b)$, are calibrated to the true state. The second approach to mitigating SME is to apply a bias-correction to the outputs of `SPEEDY-IER` to more closely match `ECHAM5-wiso`; these outputs are the inputs to the PSMs, and so we would expect the reconstruction to improve if the PSM had comparable inputs to the true state. (The bias-corrections were made by correcting the mean and standard deviation of `SPEEDY-IER'` `s` monthly temperature, sea surface temperature, sea surface salinity, and water isotope fields at individual spatial points to match those of `ECHAM5-wiso`; for the precipitation fields, the bias-correction involved fitting a two-parameter gamma distribution to the monthly `ECHAM5-wiso` precipitation and applying that fit to `SPEEDY-IER`.)

Figure 10 shows the temperature reconstruction results of the two companion experiments together with the initial SME experiment. The figure contrasts the initial SME experiment (Figure 10a) with the two SME-mitigated experiments (Figures 10b–10g). Figures 10c, 10d, 10f, and 10g show the difference in skill (for both R and CE) between the SME-mitigating experiments and the initial SME experiment. Positive values (red) indicate that the SME-mitigating experiments improve the skill. It is evident that the temperature reconstructions are clearly improved by both SME mitigating strategies.

Our results applying a simple bias-correction to the `SPEEDY-IER` climate fields suggests that to mitigate the existence of SME while taking advantage of fully nonlinear/multivariate PSMs, one can draw upon the extensive literature on GCM bias-correction and potentially retain the benefits provided by the nonlinear PSMs. For example, recent studies have made significant gains using medical imaging techniques to modify erroneous GCM features towards observed structures [*Levy et al.*, 2013, 2014]. These and other corrective approaches could help mitigate SME, hence boosting the utility of PSMs within the DA framework. Statistical bias-corrections are not a panacea, however, and improvements to GCM resolution and physics should remain a priority.

Thus, to answer our second question, *do the biases inherent in climate models prohibit the effective use of proxy system models in climate reconstructions?*: structural model biases introduce uncertainties propagated by PSMs in our experimental framework, and these uncertainties systematically reduce reconstruction skill.

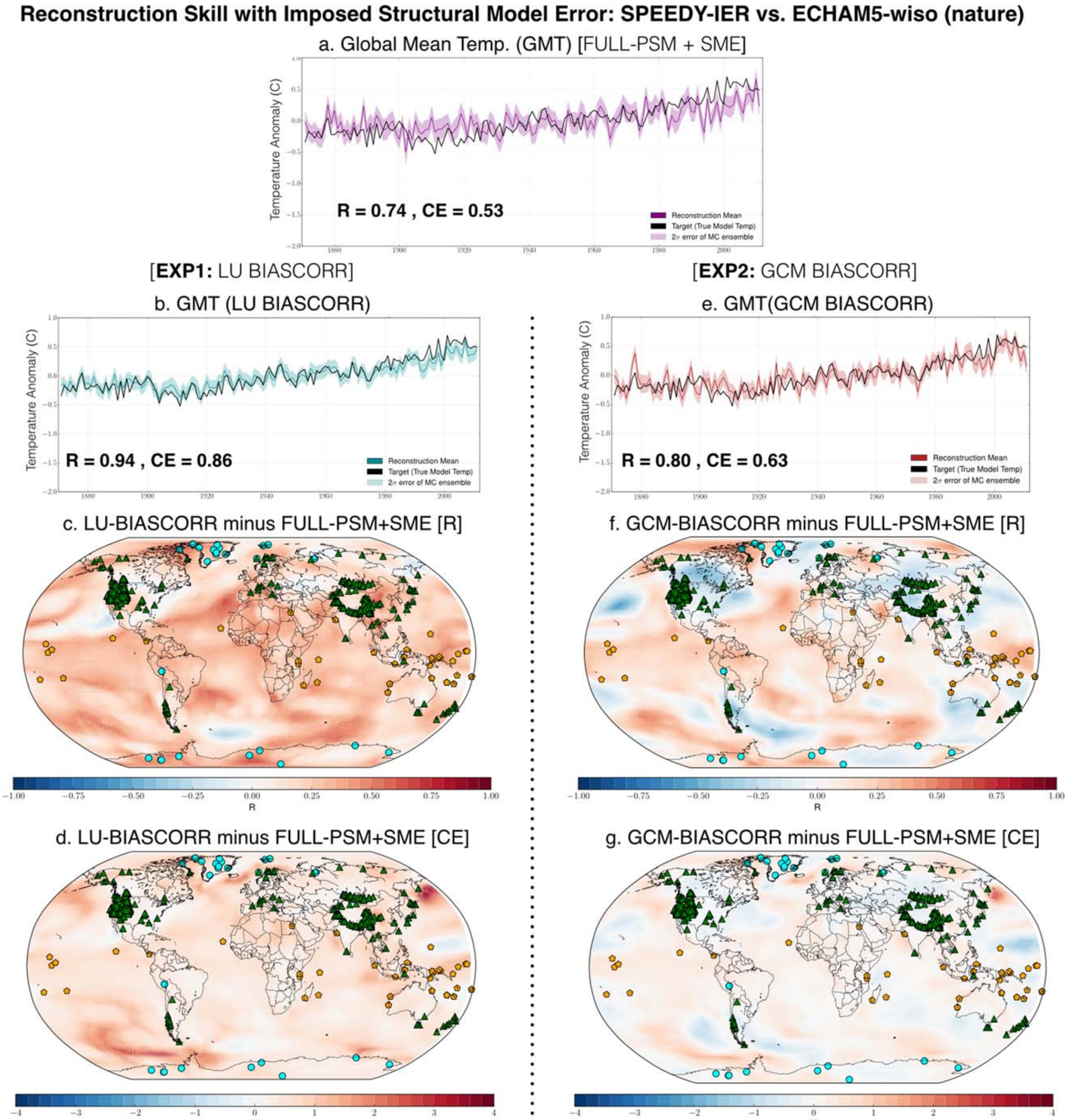

Figure 10. Reconstruction skill sensitivity to structural model errors: `ECHAM5-wiso` and `SPEEDY-IER`. (a) Global mean temperature reconstruction, initial experiment (FULL-PSM + SME), (b) global mean temperature reconstruction using Experiment 1: linear-univariate mapping as a bias-correction (LU-BIASCORR), (c) difference in R, LU-BIASCORR experiment minus FULL-PSM+SME experiment, (d) as in Figure 10c but for CE, showing improvement in reconstruction when LU-BIASCORR is used to offset the effects of SME. (e) Global mean temperature reconstruction using Experiment 2: GCM-bias-correction prior to use as GCM-inputs (GCM-BIASCORR), (f) difference in R, GCM-BIASCORR experiment minus FULL-PSM + SME experiment, (g) as in Figure 10f but for CE, showing improvement in reconstruction when GCM-BIASCORR is used to offset the effects of SME.

The reduction in skill due to SME is also larger than that induced by using linear-univariate approximations to the proxy system models (Figure 9). However, the reduction in skill due to SME alone can be mitigated using linear mapping where proxy data is calibrated against the true prior. Further, applying a bias-correction to the GCM fields can partially mitigate SME for application with the `FULL-PSM` implementation. In general, because model biases outweigh information loss due to proxy system simplification, a strategy for coping with SME is critical, even in the ideal case where PSMs perfectly represent the natural filtering of proxy systems.

4. Discussion

This study investigated the potential for embedding process-based models of proxy systems (PSMs) in a paleoclimate DA framework. We explored key assumptions surrounding the implementation of PSMs using DA to reconstruct climate over the past several centuries. In particular, our study scrutinized the impacts of two key assumptions of this framework.

First, we compared two experiments, one which used a linear, univariate representation (`LU-PSM`) and one which used more complex models (`FULL-PSM`) for proxy systems, to evaluate the impact on temperature reconstruction skill due to the assumption of a linear relationship between proxies and temperature. If a simple linear model to estimate all proxy data will suffice, it could be argued that building PSMs into our modeling framework is unnecessary. Indeed, we have shown that for some proxy types (coral and ice core δ^{18}O), assuming a linear, univariate mapping between climate and proxy is probably robust. However, as shown in section 3.1, a very accurate `FULL-PSM` will likely yield superior DA results compared to those based on empirical linear `LU-PSMs`. For tree ring width proxies in particular, we demonstrated that reconstruction skill using a nonlinear, multivariate model increases markedly when the proxies are sensitive to variables other than temperature. For tree rings (a major indicator of hydroclimate variability in western North America) and potentially other proxies not considered here, PSMs may improve reconstruction skill considerably when the proxy response is clearly nonlinear and/or multivariate.

Second, we explored the effects of structural modeling error. This test assumed that the DA prior is not drawn from the true distribution of climate states, acknowledging the fact that GCMs are an imperfect representation of nature. Because GCM structural errors propagate through the PSMs, these biases are found to adversely and systematically affect reconstruction skill. In fact, these reductions in skill exceed those incurred when assuming a simple linear-univariate model between proxy and temperature. This caveat suggests a need for caution when including PSMs in a climate reconstruction framework. As shown in section 3.2, at least two strategies are possible for mitigating the detrimental effects of model biases: (1) the use of a linear mapping between climate and proxy (calibrated to observations) and (2) employing a bias-correction to the GCM fields prior to their use in PSMs. In general, our results suggest that DA-based reconstructions rest more critically on the accuracy of GCMs than that of PSMs, and we call for a renewed focus on reducing GCM biases [*Flato et al*., 2013].

PSMs move us closer to higher-order, physically motivated representations of proxy systems; however, the nonlinear PSM networks designed in this study are only a first pass. In nature, many of these systems are highly complex and difficult to accurately model. As discussed, our results suggest that reconstruction skill improves for some proxy systems using nonlinear PSMs within DA, but these results may be dependent on the pseudoproxy experimental design. Factors such as GCM complexity and structural model errors, water isotope physics scheme, proxy network distribution (type and spatial coverage), or the structural design of each PSM all contribute to a modeling framework which may not be representative of nature. Thus, ongoing and future work towards real-proxy reconstructions must validate both the PSMs and the GCMs against observations.

This paper provides a stepping stone toward a fully operational paleoclimate reanalysis at annual scales, building upon the work of *Annan et al*. [2005], *Dirren and Hakim* [2005], *Goosse et al*. [2006], *Bhend et al*. [2012], *Steiger et al*. [2014], *Tardif et al*. [2014], and many others. By exploring some of the technical details of embedding nonlinear proxy system models in a data assimilation framework, we showed that these methods have the potential to enhance the utility of paleoclimate observations for constraining climate models. However, this hinges critically on a strategy for mitigating structural biases introduced by GCMs, and on the existence of well-specified PSMs. We leave a complete investigation of structural, dating, and parametric uncertainties in PSMs for future study.

Acknowledgments
This work was supported by grants NA14OAR431017[5,6] from the U.S. Department of Commerce, and NSF award AGS-1304263 made to the University of Washington. The authors thank David Noone and our two reviewers for their helpful suggestions and input. The modeling results used in this study are available upon request: sylvia_dee@brown.edu, nathanjs@uw.edu.

References

Annan, J., J. Hargreaves, N. Edwards, and R. Marsh (2005), Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter, *Ocean Modell.*, *8*(1), 135–154.

Barriopedro, D., E. M. Fischer, J. Luterbacher, R. M. Trigo, and R. García-Herrera (2011), The hot summer of 2010: Redrawing the temperature record map of Europe, *Science*, *332*(6026), 220–224, doi:10.1126/science.1201224.

Bhend, J., J. Franke, D. Folini, M. Wild, and S. Brönnimann (2012), An ensemble-based approach to climate reconstructions, *Clim. Past*, *8*(3), 963–976, doi:10.5194/cp-8-963-2012.

Bradley, S., R. Frigg, H. Du, and L. A. Smith (2014), Model error and ensemble forecasting: A cautionary tale, *Sci. Explanation Methodol. Sci.*, *1*, 58–66.

Comboul, M., J. Emile-Geay, M. Evans, N. Mirnateghi, K. M. Cobb, and D. M. Thompson (2014), A probabilistic model of chronological errors in layer-counted climate proxies: Applications to annually banded coral archives, *Clim. Past*, *10*(2), 825–841, doi:10.5194/cp-10-825-2014.

Dee, S., J. Emile-Geay, M. N. Evans, A. A. Allam, E. Steig, and D. N. Thompson (2015), PRYSM: An open-source framework for proxy system modeling, with applications to oxygen-isotope systems, *J. Adv. Model. Earth Syst.*, *7*, 1220–1247, doi:10.1002/2015MS000447.

Dee, S. G., D. C. Noone, N. Buenning, J. Emile-Geay, and Y. Zhou (2014), SPEEDY-IER: A fast atmospheric GCM with water isotope physics, *J. Geophys. Res. Atmos.*, *120*, 73–91, doi:10.1002/2014JD022194.

Dirren, S., and G. J. Hakim (2005), Toward the assimilation of time-averaged observations, *Geophys. Res. Lett.*, *32*, L04804, doi:10.1029/2004GL021444.

Emile-Geay, J., K. M. Cobb, M. E. Mann, and A. T. Wittenberg (2013), Estimating central equatorial pacific SST variability over the past millennium. Part I: Methodology and validation, *J. Clim.*, *26*(7), 2302–2328.

Evans, M., S. Tolwinski-Ward, D. Thompson, and K. Anchukaitis (2013), Applications of proxy system modeling in high resolution paleoclimatology, *Quat. Sci. Rev.*, *76*(0), 16–28, doi:10.1016/j.quascirev.2013.05.024.

Evans, M. N., J. E. Smerdon, A. Kaplan, S. E. Tolwinski-Ward, and J. F. Gonzalez-Rouco (2014), Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors, *Geophys. Res. Lett.*, *41*, 9127–9134, doi:10.1002/2014GL062063.

Flato, G., et al. (2013), Evaluation of climate models, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 741–866, Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9781107415324.020.

Frigg, R., S. Bradley, H. Du, and L. A. Smith (2014), Laplace's demon and the adventures of his apprentices, *Philos. Sci.*, *81*(1), 31–59.

Goosse, H., H. Renssen, A. Timmermann, R. S. Bradley, and M. E. 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, doi:10.1007/s00382-006-0128-6.

Goosse, H., E. Crespin, A. de Montety, M. 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*, D09108, doi:10.1029/2009JD012737.

Guillot, D., B. Rajaratnam, and J. Emile-Geay (2015), Statistical paleoclimate reconstructions via Markov random fields, *Ann. Appl. Stat.*, *9*, 324–352, doi:10.1214/14-AOAS794.

Hakim, G., J. Emile-Geay, E. Steig, D. Noone, D. Anderson, R. Tardif, N. Steiger, and W. Perkins (2016), The last millennium climate reanalysis project: Framework and first results, *J. Geophys. Res. Atmos.*, *121*, doi:10.1002/2016JD024751, in press.

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

Jouzel, J., et al. (1997), Validity of the temperature reconstruction from water isotopes in ice cores, *J. Geophys. Res.*, *102*(C12), 26,471–26,487.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge Univ. Press, Cambridge, U. K.

Kaufman, D. S., et al. (2009), Recent warming reverses long-term arctic cooling, *Science*, *325*(5945), 1236–1239, doi:10.1126/science.1173983.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng (2013), Last millennium climate and its variability in CCSM4, *J. Clim.*, *26*(4), 1085–1111, doi:10.1175/JCLI-D-11-00326.1.

Levy, A. A., W. Ingram, M. Jenkinson, C. Huntingford, F. Hugo Lambert, and M. Allen (2013), Can correcting feature location in simulated mean climate improve agreement on projected changes?, *Geophys. Res. Lett.*, *40*, 354–358, doi:10.1002/2012GL053964.

Levy, A. A., M. Jenkinson, W. Ingram, F. H. Lambert, C. Huntingford, and M. Allen (2014), Increasing the detectability of external influence on precipitation by correcting feature location in GCMS, *J. Geophys. Res. Atmos.*, *119*, 12,466–12,478, doi:10.1002/2014JD02235.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004), European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, *303*, 1499–1503, doi:10.1126/science.1093877.

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.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, *26*(6), 759–762.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. 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.*, *105*(36), 13,252–13,257, doi:10.1073/pnas.0805721105.

Mann, M. E., Z. Zhang, S. Rutherford, R. S. Bradley, M. K. Hughes, D. Shindell, C. Ammann, G. Faluvegi, and F. Ni (2009), Global signatures and dynamical origins of the little ice age and medieval climate anomaly, *Science*, *326*(5957), 1256–1260, doi:10.1126/science.1177303.

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.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable northern hemisphere temperatures reconstructed from low and high-resolution proxy data, *Nature*, *433*, 613–617.

Nash, J., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.

PAGES2K Consortium (2013), Continental-scale temperature variability during the past two millennia, *Nat. Geosci.*, *6*(5), 339–346.

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

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.

Tardif, R., G. J. Hakim, and C. Snyder (2014), Coupled atmosphere–ocean data assimilation experiments with a low-order model and CMIP5 model data, *Clim. Dyn.*, *45*, 1415–1427.

Thompson, D. M., T. R. Ault, M. N. Evans, J. E. Cole, and J. Emile-Geay (2011), Comparison of observed and simulated tropical climate trends using a forward model of coral $\delta^{18}O$, *Geophys. Res. Lett.*, *38*, L14706, doi:10.1029/2011GL048224.

Tierney, J., N. Abram, K. Anchukaitis, M. Evans, C. Giry, K. Kilbourne, C. Saenger, H. Wu, and J. Zinke (2015), Tropical sea-surface temperatures for the past four centuries reconstructed from coral archives, *Paleoceanography*, *30*, 226–252. doi:10.1002/2014PA002717.

Tingley, M. P., and P. Huybers (2013), Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, *Nature*, *496*(7444), 201–205, doi:10.1038/nature11969.

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam (2012), Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quat. Sci. Rev.*, *35*, 1–22, doi:10.1016/j.quascirev.2012.01.012.

Tolwinski-Ward, S., M. Tingley, M. Evans, M. Hughes, and D. Nychka (2014), Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response, *Clim. Dyn.*, *44*, 791–806.

Tolwinski-Ward, S. E., M. N. Evans, M. K. Hughes, and K. J. Anchukaitis (2010), An efficient forward model of the climate controls on interannual variation in tree-ring width, *Clim. Dyn.*, *36*(11–12), 2419–2439, doi:10.1007/s00382-010-0945-5.

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.

Wang, J., J. Emile-Geay, D. Guillot, N. P. McKay, and B. Rajaratnam (2015), Fragility of reconstructed temperature patterns over the common era: Implications for model evaluation., *Geophys. Res. Lett.*, *42*, 7162–7170, doi:10.1002/2015GL065265.

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.

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