



Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments

Jason E. Smerdon*

Millennium-length, forced transient simulations with fully coupled general circulation models have become important new tools for addressing uncertainties in global and hemispheric temperature reconstructions targeting the Common Era (the last two millennia). These model simulations are used as test beds on which to evaluate the performance of paleoclimate reconstruction methods using controlled and systematic investigations known as pseudoproxy experiments (PPEs). Such experiments are motivated by the fact that any given real-world reconstruction is the product of multiple uncontrolled factors, making it difficult to isolate the impact of one factor in reconstruction assessments and comparisons. PPEs have established a common experimental framework that can be systematically altered and evaluated, and thus test reconstruction methods and their dependencies. Although the translation of PPE results into real-world implications must be done cautiously, their experimental design attributes allow researchers to test reconstruction techniques beyond what was previously possible with real-world data alone. This review summarizes the development of PPEs and their findings over the last decade. The state of the science and its implications for global and hemispheric temperature reconstructions is also reviewed, as well as near-term design improvements that will expand the utility of PPEs. © 2011 John Wiley & Sons, Ltd.

How to cite this article:

WIREs Clim Change 2012, 3:63–77. doi: 10.1002/wcc.149

INTRODUCTION

The late 20th and early 21st centuries have emerged as a period of rapid and pervasive environmental changes.^{1,2} Multiple anthropogenic activities have been identified as the leading causes of these changes, principal among them is the burning of fossil fuels and the subsequent release of greenhouse gases into the Earth's system.³ Understanding the impacts of these emissions on the Earth's carbon cycle and climate is of vital importance to human societies and the natural world on which they depend, but a rigorous understanding of these impacts and an ability to anticipate their emergence in the future cannot be achieved by modern observations alone—the potential changes are too vast and complex, the relevant timescales are too varied. A thorough understanding of contemporary climate change is thus dependent on

our ability to comprehensively characterize the Earth's climate system on multiple timescales and over many different states in the past.

Paleoclimatology offers a glimpse into historical climates over a range of timescales and mean states, while allowing us to test and inform our developing hypotheses about the climate system and climatic change.⁴ Among the many time periods of relevance, the Common Era (CE; the last two millennia) is an important target because the abundance of high-resolution paleoclimatic proxies (tree rings, corals, ice cores, etc.) over this time period allow seasonal to annual reconstructions on regional to global spatial scales.⁵ The CE also spans the rise and collapse of many human civilizations, making paleoclimatic information during this time period critical for understanding the complicated relationships between climate and organized societies.^{6–8}

Studies of CE climate have already proved vital to understanding basic properties of the climate

*Correspondence to: jsmerdon@ldeo.columbia.edu

Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA

system and the vulnerabilities of present societies to climate variability that is not fully captured in the modern observational record. For example, megadroughts of extreme duration and intensity throughout much of Mexico and the American Southwest have been tied to small and prolonged decreases in sea surface temperatures in the tropical Pacific (see Ref 9 for a review), while historical drought vulnerabilities also have been characterized in areas of Southeast Asia and Indonesia.^{7,8} The El Niño-Southern Oscillation phenomenon also has been shown to vary on multidecadal timescales¹⁰ and to couple importantly with volcanic activity,^{11,12} while understanding its secular behavior has been identified as important for climate projections of the 21st century.¹³ Despite these and many other successes, however, there remain outstanding questions about our understanding of CE climate and its implications for the future.

Perhaps one of the more widely debated areas of CE paleoclimatology involves the reconstruction of global or hemispheric climate using networks of climate proxies derived from multiple proxy records (see Ref 5 for a review). One of the principal issues of this debate surrounds the amount of reconstructed temperature variability during the CE on decadal and longer time scales, particularly as it relates to the magnitude, phasing, and ubiquity of the Medieval Climatic Anomaly and Little Ice Age.^{14–19} Although much progress has been made to understand how various reconstructions may accurately represent the characteristics of these past epochs, there remain important unanswered questions about reconstruction uncertainties. These questions are tied to understanding the connections between climate and proxy responses across different spectral domains, the response of proxies to multiple environmental variables, the role of teleconnections and noise in the calibration data, and the impact of specific proxy networks and methodological choices on derived reconstructions—questions that are ultimately fundamental to the success of efforts to reconstruct past climatic variability during the CE.^{5,20,21}

One important new tool for addressing questions about the performance of CE reconstruction methods is millennium-length, forced transient simulations with coupled general circulation models (CGCMs). These model simulations are used as test beds on which to evaluate the performance of a given reconstruction method using controlled and systematic experiments—an approach that has come to be known as pseudoproxy experiments (PPEs).²² This approach was originally investigated using an unperturbed control simulation to explore the

selection of optimal proxy locations in global temperature reconstructions.²³ The more complete pseudoproxy approach in which predictor data are perturbed to mimic the characteristics of real proxy data was later explored using instrumental data in multiple settings.^{22,24,25} Since these early publications, an impressive and growing number of PPE studies and subsequent discussions have emerged in the literature,^{26–69} as well as one recent summary of PPE work in the larger context of high-resolution paleoclimatology of the CE.⁵ The following review builds off of these previous efforts by summarizing the last decade or more of PPE studies, surveying their implications for global and hemispheric temperature reconstructions of the CE, and by suggesting near-term design improvements that will increase the utility of PPE applications.

CONSTRUCTION OF PPEs

The motivation for PPEs stems from the fact that real-world reconstructions are derived from many different methods, calibration choices, and proxy networks. Uncertainty in any given real-world reconstruction is therefore a combined result of the employed method, the adopted calibration data and calibration time interval, the spatial and temporal sampling of the proxy network, the actual climate-proxy connection of each proxy record used for the reconstruction, and the presence of stochastic noise in both the climate and proxy data. If the objective is to isolate the impact of one of these factors, it is difficult to do so from comparisons between available real-world reconstructions. PPEs have allowed some of the above challenges to be circumvented by adopting a common framework that can be systematically altered and evaluated, and thus test reconstruction methods and their dependencies. Moreover, PPEs provide a much longer, albeit synthetic, validation period than that which can be achieved with instrumental data, and thus methodological evaluations can extend to lower frequencies and longer time scales. Although one must always be mindful of how PPE results translate into real-world implications, these collective design attributes allow researchers to test reconstruction techniques beyond what was previously possible with real-world data alone.

The primary data of PPEs have been derived from a limited number of forced transient CGCM simulations spanning the last 500 years before present or more^{34,35,70,71}; in some cases model control runs^{23,50,67} or the instrumental record^{22,24,25,60–63}, have been used as alternatives. The approach of PPEs described subsequently is to extract a portion

of a spatiotemporally complete CGCM field in a way that mimics the available proxy and instrumental data used in real-world reconstructions. The principal experimental steps proceed as follows: (1) pseudo-instrumental and pseudoproxy data are subsampled from the complete CGCM field from locations and over temporal periods that approximate their real-world data availability; (2) the time series that represent proxy information are perturbed with noise to simulate the temporal (and in some cases spatial) noise characteristics present in real-world proxy networks; and (3) reconstruction algorithms are applied to the sampled pseudo-instrumental data and pseudoproxy series to produce a reconstruction of the climate simulated by the CGCM. The culminating fourth step is to compare the derived reconstruction

to the known model target outside of the calibration interval as a means of evaluating the skill of the applied method and the uncertainties expected to accompany a real-world reconstruction product.

Pseudoproxy Networks

The adopted pseudoproxy networks in almost all PPEs to date have approximated the Mann et al.⁷² (hereinafter MBH98) multiproxy network.^{28,29,42,45,54–59,64–67} A recent study⁵⁹ has additionally tested a distribution approximating the updated network of Mann et al.⁷³ (see Figure 1 for a comparison of pseudoproxy distributions approximating the MBH98 and Mann et al.⁷³ networks). Exceptions to the above choices include some

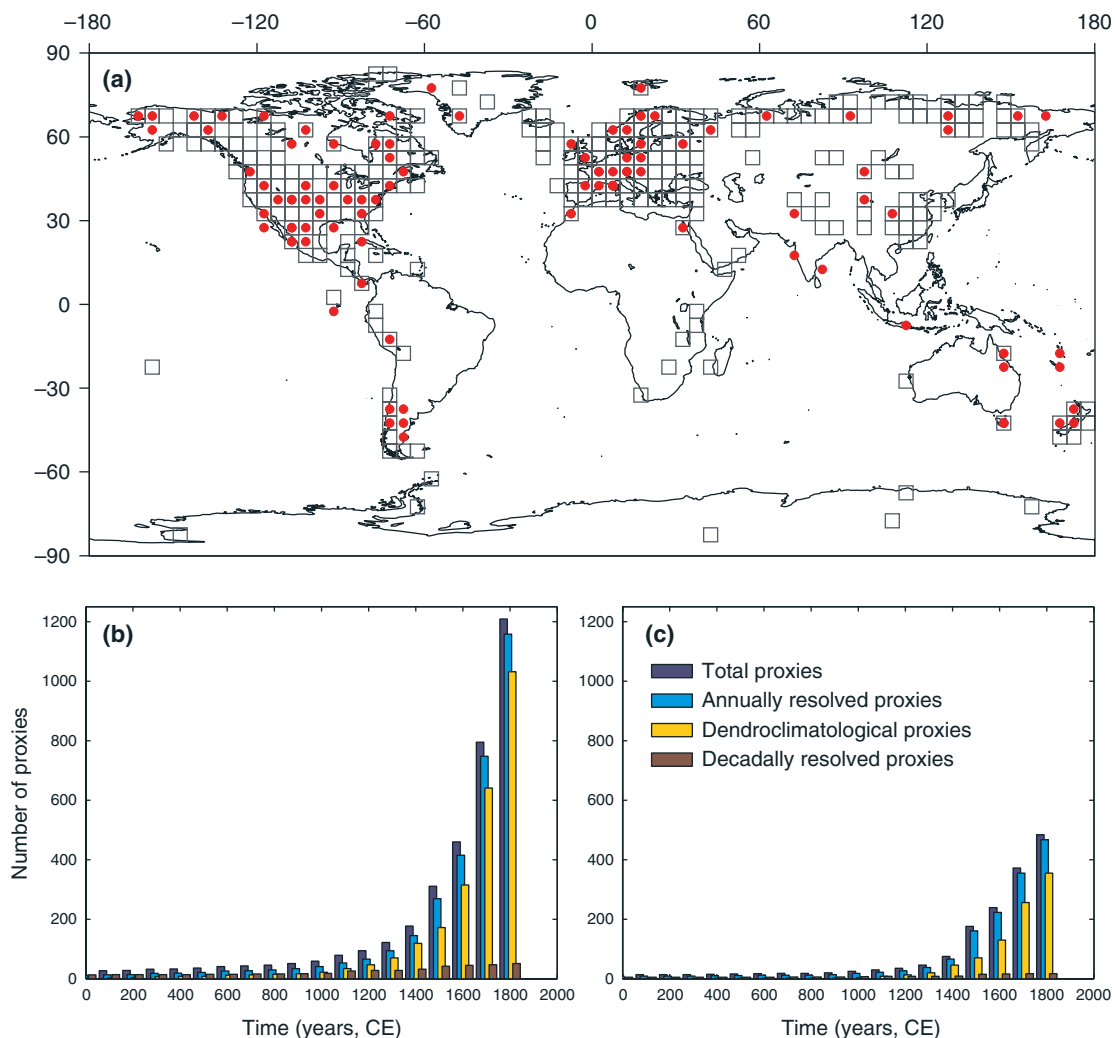


FIGURE 1 | (a) Comparison of two pseudoproxy networks that approximate the most populated nests in the multiproxy networks of Mann et al.⁷² (MBH98, red dots) and Mann et al.⁷³ (gray squares). Bottom plots show the total and categorized proxy abundances in the global multiproxy network of Mann et al.⁷³ for: (b) the full dataset; and (c) the culled dataset as screened and used by Mann et al.⁷³ The plotted abundances are as published in the Supporting Information of Mann et al.⁷³

studies that have tested additional proxy network distributions,^{37,39,41,48,49,61,62} random subsampling of the MBH98 network distribution,⁴⁰ random and semi-random sampling of the target field,^{42,45,50,53,60,64} or noise-only experiments.^{38,46,47,63} In almost all cases, the number of pseudoproxies in the sampled network have been constructed to be temporally invariant. Notable exceptions have tested multiple temporally invariant distributions that approximate densely or sparsely populated nests in specific multiproxy networks.^{29,45} In those PPEs approximating the MBH98 distribution, pseudoproxy availability is typically assumed to be constant during the entire reconstruction period and reflects the availability of the most populated nest in the multiproxy network. While this feature of current PPEs allows a straightforward interpretation of spatial sampling impacts, it is a simplification of real-world networks that rapidly lose records back in time (Figure 1).

Pseudoproxy Noise

Perhaps the most important decision for the design of PPEs is the characteristics of the signal and noise that make up the pseudoproxy network—a choice that typically has the largest impact on the performance of the evaluated methods.^{29,40,45,54,58,59,64–66} Most PPE studies have sampled the modeled temperature field at specific locations to construct a baseline, or no-noise, pseudoproxy network. Various colors of noise with different levels of variance are then added to the sampled temperature time series to establish collections of pseudoproxy networks with multiple signal-to-noise ratios (SNRs). The most widely applied choice of noise has been Gaussian white noise at SNR values of 1.0, 0.5, and 0.25 by standard deviation (see Box 1 for the relationships between SNR and other common noise conventions used in PPEs).^{26,40,42,45,54,56,58,59,64–66} In some cases Gaussian red noise has been used by adding red noise time series with uniform characteristics to the entire pseudoproxy network^{25,31,34,66}; blue noise has also been used in several experiments.^{22,45} An additional study has used residual analyses to approximate the noise character of proxies at specific locations, resulting in mixed pseudoproxies with multiple noise characteristics throughout the employed network.²⁹ To demonstrate the process of pseudoproxy construction, Figure 2 plots a selected temperature time series from the ECHO-G ERIK2 millennium simulation.³⁵ The signal has been normalized and is shown above a white noise time series that has been scaled to have variances of 1, 4, and 16. The combined signal-plus-noise time series are also shown in Figure 2 and represent the resulting pseudoproxy time series used in PPEs.

BOX 1

NOISE CLASSIFICATIONS IN PPEs

Various conventions have been used for classifying the level of noise in pseudoproxy studies. The most common include the signal-to-noise ratio (SNR) by standard deviation, the percent noise by variance (PNV), or the correlation between the signal and the signal-plus-noise time series (r). For white noise, these conventions can be expressed interchangeably. If the level of noise is expressed as the variance of a pure white noise (NVAR) time series relative to a normalized signal time series (variance of one), the expressions relating SNR, PNV, and r are as follows:

$$\text{SNR} = (\text{NVAR})^{-1/2}$$

$$\text{PNV} = (\text{NVAR}) / (1 + \text{NVAR}) = 1 / (\text{SNR}^2 + 1)$$

$$r = (1 + \text{NVAR})^{-1/2} = \text{SNR} / (1 + \text{SNR}^2)^{-1/2} \\ = (1 - \text{PNV})^{1/2}$$

Note that the correlation coefficient represents the mean of the distribution derived from many noise draws for finite time series. Using the common levels of NVAR employed in PPEs (1, 4, and 16) the respective values of SNR, PNV, and r are 1, 0.5, and 0.25; 50, 80, and 94%; and 0.707, 0.447, and 0.243.

While a range of SNRs is typically employed in PPEs, it is generally assumed that SNRs on the order of 0.5–0.25 (by standard deviation) are representative of the actual noise level in real-world proxy records. This estimate is based on local correlations between instrumental grid-point temperatures and proxies during their interval of overlap. One assessment⁴⁵ has evaluated local temperature correlations for the MBH98 network and an additional tree-ring network that has a spatial distribution over extratropical Eurasia and North America. The analysis estimated 0.4 to be a conservative value of the mean SNR in the two datasets, corresponding to a correlation between proxy and local climate of 0.37 (assuming white noise perturbations to the proxy climate signal). A later study²⁹ has performed a similar analysis of three proxy nests in the MBH98 network, although the approach differed from Mann et al.⁴⁵ by determining the highest correlation between the proxies and all locations in the temperature field. This later analysis estimated minimum proxy correlations of 0.3 and a mean near 0.47. This latter estimate appears consistent

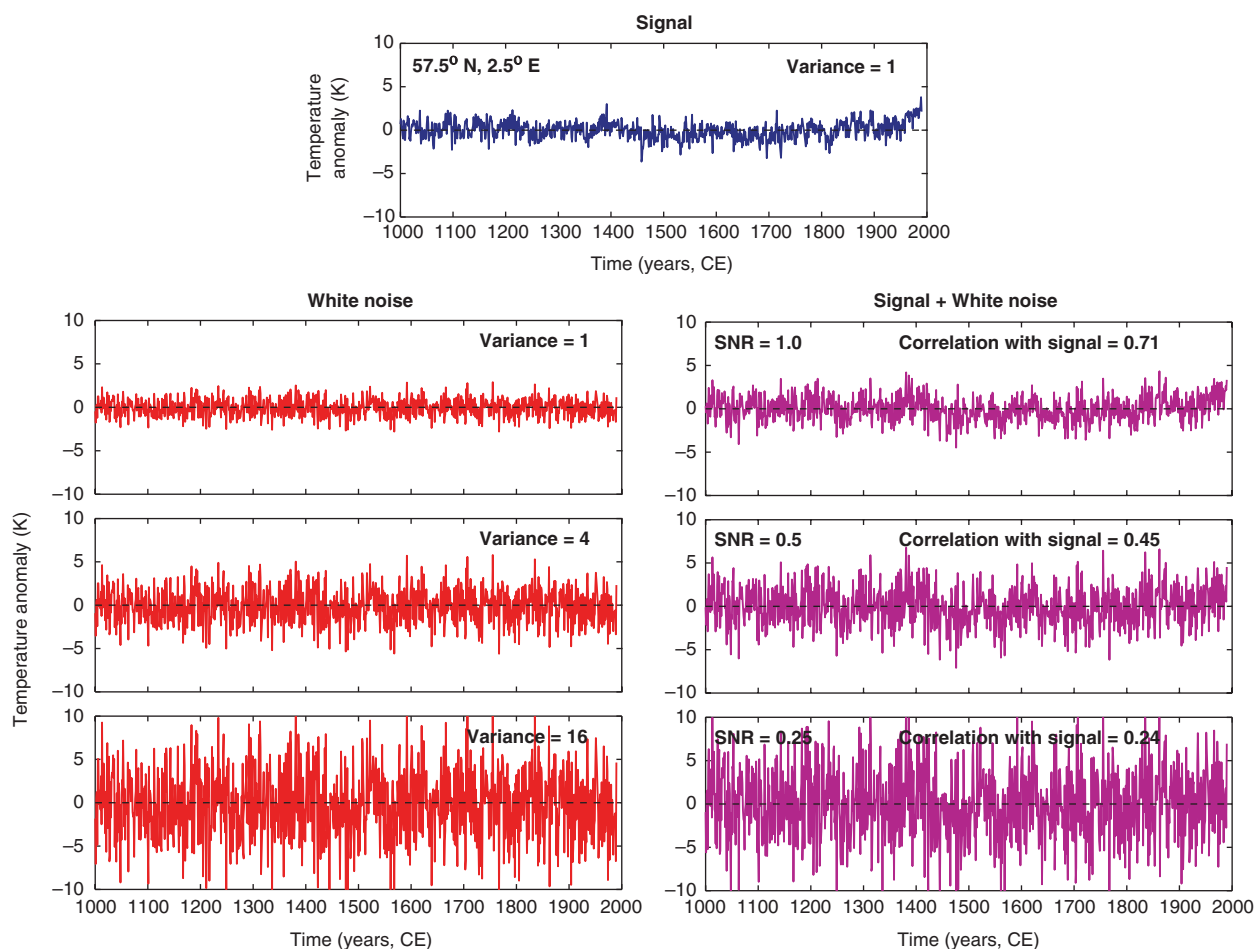


FIGURE 2 | A ‘signal’ time series sampled from the annual surface temperature field of the ECHO-G ERIK2 millennial simulation³⁵ after interpolating to an even 5° latitude–longitude grid (the geographic center of the sampled grid is 57.5°N , 2.5°E). The time series has been normalized to have a standard deviation of one. Shown on the left below the sampled temperature signal is a single white noise time series that has been scaled to have variances of 1, 4, and 16. The sum of the pure-noise and temperature time series is shown in the lower right panels and represents the pseudoproxy time series used in pseudoproxy experiments (PPEs). Signal-to-noise ratios (SNRs) and correlations between the original signal and the signal-plus-noise series are also shown in the lower right-hand plots.

with Mann et al.,⁴⁵ given that it did not apply geographic constraints on the proxy locations and thus estimated maximized proxy-temperature correlations. Mann et al.⁴⁵ also estimated lag-one autocorrelation coefficients for the proxies in the MBH98 network and found $\rho = 0.32$ to be a representative mean value. This estimate compares with the peak distribution of a more recent assessment⁴⁷ of autocorrelations in an updated multiproxy network,⁷³ although the values of the estimated lag-one coefficients in McShane and Wyner⁴⁷ ranged from approximately $\rho = -0.25$ to 1. Regardless of the estimated noise characteristics, almost all studies have used individual noise realizations for each pseudoproxy in the network. Recent studies have noted the importance of employing ensemble experiments that test the dependence of results on both multiple pseudoproxy

noise realizations²⁹ and the spatiotemporal character of the underlying model field employed in the PPE.^{29,59}

TESTING RECONSTRUCTION METHODS WITH PPEs

It is common to separate global or hemispheric temperature reconstruction methods into two categories. The first involves index methods that target indices such as northern hemisphere (NH) or global mean temperatures; the second comprises climate field reconstruction (CFR) methods that target hemispheric or global patterns, i.e., spatial maps of temperature change expressed on a grid. The methodology at the heart of most currently applied index or CFR techniques is based on the application of multivariate

linear regression to calibrate a network of proxies during their period of overlap with the instrumental temperature record. This calibration typically incorporates matrix factorizations, i.e., eigenvalue decompositions, of the target field, proxy matrix, or both. Within this framework, there are four main choices that are typically involved in the application of a given method: (1) whether to cast the problem as a regression of proxies on climate, or vice versa; (2) the degree of rank reduction of the target field after its eigenvalue decomposition (this choice only applies to CFR methods); (3) the degree of rank reduction of the proxy matrix after its eigenvalue decomposition; and (4) the type of regularization used to estimate the regression coefficient matrix—a choice that is often necessary because most CE reconstruction problems are underdetermined (again applicable to CFR methods only). Some additional choices of recent discussion are inclusion of errors in both the predictors and predictand,^{37,45} local proxy calibration,^{32,61} and iterative solutions that cast CFRs as matrix completion problems.^{42,45,74} Most of the discussion in the literature about reconstruction methods centers on these various choices and their importance, which PPEs have been designed, in part, to test.

Testing Reconstructions of NH and Global Means

The reconstruction of NH or global mean temperature indices has been a central focus of large-scale reconstruction studies,^{20,21} and consequently a central focus of PPEs. Reconstruction of these large-scale indices has been done using both index and CFR methods. In cases where CFR methods are used, the computed mean index is derived from a composite average of the reconstructed field. Within index reconstruction methods, two groups of approaches generally exist. Composite-plus scale (CPS) methods average collections of proxy networks into a single index that is then scaled to have the same mean and variance of the target time series over a common interval of overlap. The averaging may be performed with or without weighting based on factors such as the geographic location of the proxies or the correlations between the proxies and local or large-scale temperature indices. The second group of index methods similarly derives a time series estimate from a weighted composite of the proxies, but the weights and scaling are derived from a regression of the proxies on the target temperature index, or vice versa. This regression may be performed using all of the proxies, or from the leading principal components of the proxy network after it is factored.

The first PPEs using a millennial CGCM simulation tested the MBH98 CFR technique for its ability to produce skillful NH mean temperature indices.^{64,67} These studies showed losses of low-frequency variance throughout the reconstruction period, despite subsequent discussions to clarify the methodological approach^{65,75} and to demonstrate that the results were not dependent on the employed millennial simulation.^{44,45,51,55,69,76} Since these first publications and discussions, there have been many PPE studies that have yielded mixed results regarding the skill of NH and global means derived from index and CFR methods. Multiple PPE studies have shown specific methods to successfully reconstruct the NH mean temperature index using representative noise levels and proxy distributions.^{27,32,37,40,45} Most of the regression-based methods tested in these studies used a form of total least squares regression to allow consideration of noise in both proxy and instrumental data, which may play an important role in their NH mean skill. Several of these studies also have applied the regressions in filtered low-frequency spectral domains, which may be another important factor in their skill.^{31,52} A state-space time series model using a Kalman filter algorithm also has been shown to skillfully reconstruct NH means in PPEs, and to compare well with truncated total least squares methods.⁴⁰ Within the same study, a broad range of index methods were also tested, and, with few exceptions, they were shown to skillfully reproduce NH means, particularly at decadal and lower frequencies.⁴⁰ This observation led the authors to conclude that data and sampling differences are the more likely explanation of differences between reconstructions of real-world global and hemispheric temperature means. Despite the above collections of studies reporting demonstrated successes for index reconstructions, multiple PPE studies^{27–29,31,40,48,49,54,56,58,59} also have shown a range of methods to yield reconstructions with significant variance losses and mean biases conditional on different PPE designs, model simulations or ensemble experiments. Current understanding thus reflects multiple methods and PPE designs in which the derived NH mean reconstructions may or may not contain important shortcomings that are relevant to interpretations of CE climate.

Resolving the above differences requires more specific attention to the source of skill in applied methodologies and more direct assessments of how given PPE designs translate into real-world conditions. This point is well illustrated by PPEs that have tested the regularized expectation maximization (RegEM) method⁷⁴ using two different regularization approaches. In one case, RegEM using ridge regression

has been shown in PPEs to produce NH means with significant mean biases and variance losses.^{29,54,56,58,59} In contrast to these results, RegEM employing truncated total least squares has been shown to faithfully reconstruct NH mean temperatures even at high noise levels^{45,59} (see also Figure 4). Despite these different performances in PPEs, both RegEM versions produce real-world NH mean temperature reconstructions that are very similar.^{42,45} Moreover, both of the real-world RegEM reconstructions agree well with the original MBH98 reconstruction,⁴⁵ which was derived using a method that also has been shown to suffer from variance losses and mean biases in PPEs.^{59,65,66} Understanding these apparent contradictions between PPE and real-world reconstruction results is thus important and speaks to the need to better characterize the source of reconstruction skill and how PPE and real-world conditions compare (note that important quantitative corrections regarding some of the above RegEM studies^{42,43,45,54,56} were necessary because they included model data that were incorrectly processed^{51,55,57}).

Testing Reconstructions of NH and Global Fields

The promise of CFR methods is in their ability to estimate spatial patterns of temperature variability and thus provide dynamical insights.^{19,26} Despite this utility and increasing interests in CFR applications, it nevertheless is not widely appreciated that very few large-scale CFRs actually exist (note that there are many regional CFRs for multiple climate variables,^{7,25,77–81} but the discussion herein is specific to hemispheric and global temperature reconstructions). Figure 3 plots the summary of NH surface temperature reconstructions for the CE from Assessment Report Four (AR4) of the Intergovernmental Panel on Climate Change.²¹ As noted in Figure 3, only 2 of the 12 reconstructions are derived from CFRs. Since the publication of the AR4, only one additional large-scale CFR has been published.¹⁹ Research to produce these temperature CFRs is thus still nascent. Derivations of CFRs can be complicated by the fact that they attempt to reconstruct more spatial locations than the number of temporal observations in the calibration interval (the regression problem is underdetermined), and are more likely dependent on climate teleconnections and proxy network distributions than index reconstructions. Much work therefore remains to refine CFR methods and uncertainties, while expanding the proxy networks used to produce them.⁵⁹

For those PPEs that have evaluated CFR methods, very few of them have comprehensively

assessed their spatial skill. Some studies have reported summaries of field statistics or provided spatial plots of limited assessment metrics,^{42,45,50} but the primary evaluations of CFR methods in PPEs to date have focused on their ability to derive skillful NH or global mean indices. Such evaluations are insufficient for assessing the spatial performance of CFRs. Figure 4 plots results from a PPE testing the RegEM method using truncated total least squares exactly as it was applied in its original CFR application.^{45,59} The upper panel in the plot compares the reconstructed and known model time series for the NH mean surface temperature in the NCAR CCSM1.4 millennial integration.⁷⁰ As has been previously reported,⁴⁵ the method reconstructs the NH mean temperature index well, but Figure 4 also shows the grid-point correlation coefficients and mean biases during the reconstruction interval for the derived CFR. These spatial measures of merit indicate that despite the successful reconstruction of the NH mean, correlation coefficients can drop below 0.1 and mean biases can approach more than 1°C within various regions of the reconstructed field. Further evaluating the field skill associated with CFR methods is thus critical for assessing the robustness of their reconstructed spatiotemporal information, which is ultimately the principal motivation for deriving CFR products.

The few PPEs that have directly assessed spatial skill of CFR methods have reported significant variations in regional performance. A PPE study restricted to Europe and the North Atlantic yielded regions of bias in reconstructions from two CFR methods of over 1°C, particularly at SNRs approaching real-world levels.⁴⁹ These biases were present even though reduction of error statistics stayed positive throughout the field in all experiments performed (reduction of error values below zero are considered failed validations²⁰). Similarly, a pseudoproxy analysis of reconstructed postvolcanic temperature anomalies showed one CFR method to reproduce well the pattern of European temperature responses after 15 tropical volcanic eruptions.⁵ The amplitude of the reconstructed temperature response, however, was significantly reduced relative to the model target. In another regional PPE study over North America using instrumental data, regionally variable field skill was again demonstrated in experiments approaching realistic noise levels.^{61,62} In what the authors termed medium (SNR = 0.5) and hard (SNR = 0.33) experiments, derived reconstructions from two different CFR methods yielded coefficients of determination approaching 0 and coefficients of efficiency approaching –1 in multiple regions of the reconstructed field (positive values are again the threshold for successful validation in the latter

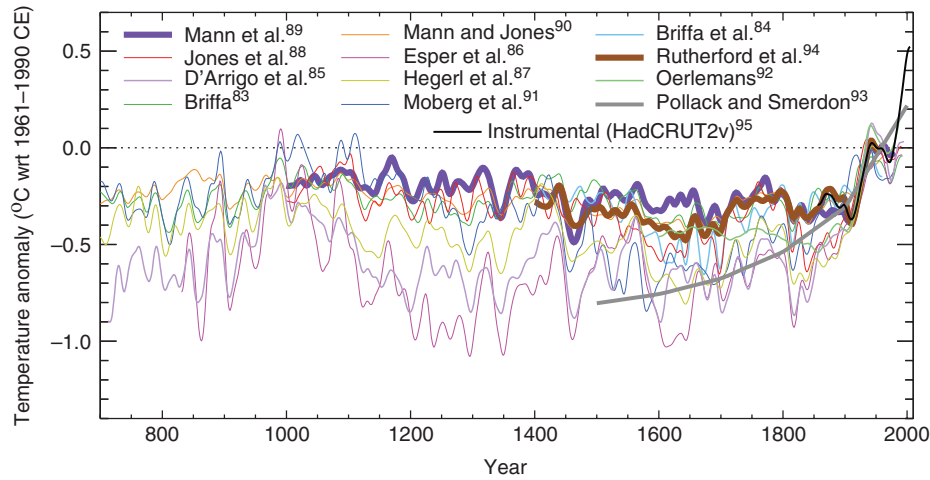


FIGURE 3 | Index and climate field reconstruction (CFR; **bolded**) reconstructions of NH temperature variations during the Common Era (CE) using multiple climate-proxy networks and methods; the HadCRUTv2 instrumental temperature record⁸² is shown in black. All series have been smoothed using a Gaussian-weighted filter to remove fluctuations on time scales less than 30 years. All temperatures represent anomalies from their 1961 to 1990 mean^{83–95}. Reconstructions derived from CFRs have been bolded by the author. (Reprinted with permission from Figure 6.10(b) in Ref 21. Copyright 2007 Cambridge University Press).

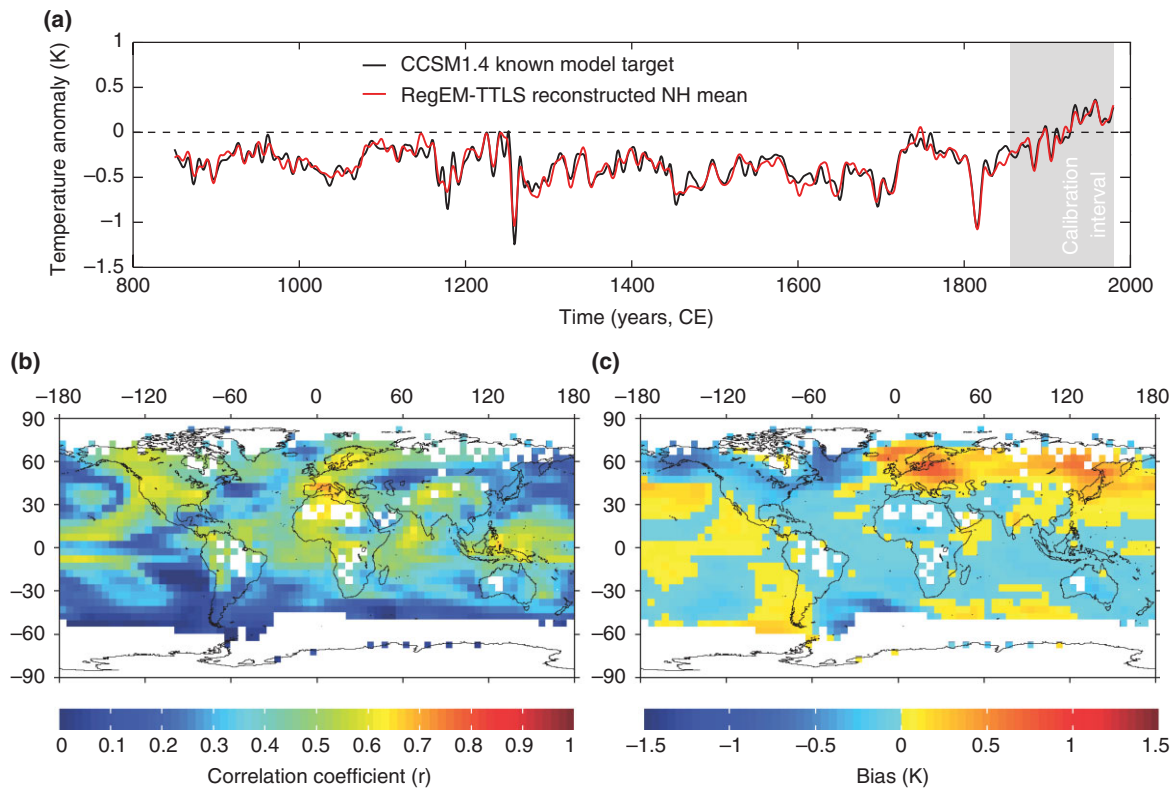


FIGURE 4 | Pseudoproxy results derived from an experiment using the NCAR CCSM1.4 millennial simulation, a pseudoproxy sampling scheme approximating the most populated nest in the multiproxy network of Mann et al.⁷² (MBH98), a signal-to-noise ratio (SNR) of 0.5, and a representative instrumental temperature mask.⁵⁹ Both pseudoproxies and the instrumental temperatures were sampled from the annual surface temperature field of the CCSM1.4 simulation, after properly interpolating to a 5° latitude–longitude grid.⁵⁷ Results are derived from the hybrid RegEM truncated total least squares method using a calibration interval from 1856 to 1980 Common Era (CE).⁵⁹ Panel (a) compares the low-pass filtered NH mean temperature anomalies (relative to the calibration interval) derived from the applied climate field reconstruction (CFR) method and the known model mean. Panel (b) plots the local correlation coefficients computed between the reconstruction and known model field during the reconstruction interval (850–1855 CE). Panel (c) is the same as in (b), but for the difference between the reconstructed and known model means, i.e., the reconstruction bias.

metric²⁰). These regional results are consistent with hemispheric and global PPE assessments that have also shown spatially variable skill statistics for multiple commonly applied CFR methods.^{56,58,59} In the most comprehensive of these assessments,⁵⁹ four CFR methods were tested using two different millennial simulations and two multiproxy distributions as the basis of global PPEs. At realistic noise levels, these experiments again yielded spatially variable verification statistics: minimum grid-point correlation coefficients ranged between 0 and 0.1, some regional biases exceeded 1°C and ratios between reconstructed and known model standard deviations fell below 0.3 in multiple regions. While some performance characteristics differed, all four methods and associated variants performed quite similarly—consistent with the fact that all employed methods were based on the same underlying multivariate linear model. It additionally was noted that field skill tended to concentrate in densely sampled pseudoproxy regions, an observation that argues for the need to expand the spatial sampling of current multiproxy networks. Finally, methodological performance was shown to be partially dependent on the underlying model field, suggesting that PPE assessments of field skill should be mindful of model dependencies. Collectively, current assessments of CFR spatial performance indicate the need to more fully vet the field skill of contemporary methods using multiple models and PPE designs, while more directly connecting PPE results to the specific characteristics of real proxies and climate fields.

FUTURE DESIGN IMPROVEMENTS IN PPEs

Current PPEs have implemented designs that largely represent best-case scenarios for methodological assessments. These are useful to the degree that even these scenarios suggest certain challenges for the current generation of reconstruction methods that ultimately must be addressed and evaluated. Nevertheless, real-world proxy records have signal and noise characteristics that are unique to the physical, chemical or biological system from which they are measured. As a result, specific proxy series likely contain nonlinear, multivariate, and nonstationary signal and noise characteristics²⁰ that can be seasonally dependent.⁸¹ None of these characteristics are fully captured in currently employed PPEs, making their representations of signal and noise perhaps the largest idealization. Improvements to these current pseudoproxy designs nevertheless can be partially addressed with available tools. Li et al.⁴¹ have employed idealized models of tree-ring, pollen, and borehole temperature proxies to

test their relative contributions to NH mean temperature reconstructions. Similarly, multiple studies have specifically modeled geothermal climate data using millennial CGCM simulations and tested the performance of contemporary borehole temperature inversion techniques.^{36,96} Ongoing work to model the connection between climate and proxy systems,^{36,97–100} or proxy relevant variables,¹⁰¹ therefore introduces the capability to more realistically model pseudoproxy records from the output of millennial CGCM simulations. This already has been accomplished for some tree-ring networks, the results of which are available in the public domain for testing and evaluation.¹⁰² Continuing efforts therefore will advance PPEs beyond the annual temperature-plus-noise models that currently are used to construct pseudoproxy time series, and thus emulate the multivariate and potentially nonlinear connections between proxies and climate expected to exist in real-world proxies. Moreover, a multivariate focus on pseudoproxy construction will motivate new designs for PPEs that allow testing of other reconstruction targets such as precipitation⁸⁰ or drought metrics.^{7,9,77}

Another immediately possible PPE improvement is connected to the fact that most studies have adopted temporally invariant pseudoproxy networks that reflect the spatial distribution of the most populated nests in available multiproxy networks. Such a convention overestimates the level of multiproxy sampling during the earlier period of the CE (Figure 2), except in those studies that explicitly test smaller networks with more uniform temporal sampling.^{37,48} PPE designs that test either the declining number of predictors in real-world multiproxy networks or use multiple temporally invariant networks that represent high and low-population nests are therefore warranted (see Refs 29 and 45 for examples of pseudoproxy studies that explicitly tested nests corresponding to temporal subsets of multiproxy networks). Contributions from specific proxy types and their spatiotemporal distributions also should be tested by constructing subsets of pseudoproxy networks representative of specific kinds of proxies (following, for example, the work of Li et al.⁴¹). Such proxy-specific experiments would more realistically capture real-world conditions than the temporally invariant pseudoproxy networks used in most current PPEs, as well as quantify contributions to reconstruction skill from specific kinds of proxy archives and their locations.

Several studies have also shown the importance of evaluating PPEs in the context of ensembles based on both multiple noise realizations in pseudoproxy networks and on the underlying spatiotemporal characteristics of the target field.^{29,59} For instance,

differences between methods tested on individual noise realizations may not be statistically significant when uncertainties due to random errors are incorporated. This has been shown for NH mean temperature estimates,²⁹ but such ensemble work has not been done in the context of spatial CFR performance. Future work to evaluate field skill in ensembles of CFRs is therefore important. This is further supported by the recent demonstration of model dependencies in PPE results.⁵⁹ The direct addition of noise to the target data or techniques such as phase randomizing²⁹ of the target field are means of achieving ensembles from single models, which can and should be incorporated into future PPEs. The upcoming public release of multiple millennium-length, forced transient simulations from coupled CGCMs as part of the Coupled Model Intercomparison Project Phase 5¹⁰³ provides an additional and timely opportunity to greatly enhance the number of model simulations used in PPEs. Incorporating these simulations into ensemble PPEs will test reconstruction skill over a range of underlying covariance structures in the model fields, which will help characterize common methodological performance while separating model-dependent behavior. Studies that connect these ensemble results to PPEs that use the observational temperature record instead of simulated model fields will further link PPE results to real-world conditions by identifying those model-based PPEs that best characterize the spatial performance of real-world CFRs. This insight will provide guidance on how to interpret model-based PPEs as analogs for real-world reconstructions and therefore allow them to be more directly interpreted for things such as optimal proxy-site selections, representations of regional high and low reconstruction field skill, and as measures of low-frequency fidelity.

CONCLUSIONS

PPEs always will be only approximations of reality. While they provide flexible and controlled experimental designs, care is required when interpreting their results for implications regarding actual reconstructions of CE climate. In this regard, it is helpful to draw an analogy to the use of animal models in medical research.¹⁰⁴ Such lines of inquiry allow controlled and systematic experiments that never would

be possible with human subjects and thus greatly expand medical research progress and capabilities. Nevertheless, research using animal models is complicated by the fact it ultimately must be interpreted for its applicability to humans. It therefore is necessary to establish an experimental design that carries relevance for human application, while steps subsequent to animal-model experiments are often vital for determining how to translate their results for applicability to human patients. PPEs can be viewed in a similar light. There are no true control experiments for reconstructing past climate, and real-world data availability limits the extent of methodological validation. Both of these realities restrict the range of reconstruction evaluations that can be accomplished with real data alone, thus giving rise to the utility of PPEs. It nevertheless is important to design PPE frameworks to more accurately represent the real-world challenges inherent in CE reconstruction problems. Direct and rigorous connections between PPE results and the spatiotemporal characteristics of the actual climate also need to be further explored. Attention to both of these issues will strengthen PPE relevance and ultimately advance climate reconstructions of the CE.

Despite the potential for further improvements, PPEs have already yielded important insights into the performance of reconstruction methods and the possible uncertainties in real-world reconstructions. They have revealed potentially important deficiencies in contemporary reconstruction methods and the likely sources of skill in others, while allowing direct and systematic comparisons between a suite of applied reconstruction techniques. This latter point will become ever more important as emerging methods are applied to the climate reconstruction problem.^{40,41,61,62,105} The growing number of studies^{38,45,47,57–63,102} making reconstruction codes and PPE data publicly available will further drive success by compounding the degree to which methods are tested and increasing the creativity with which PPEs are designed. The collective result will be rigorously tested methods and an enhanced understanding of the uncertainties in real-world reconstruction products, all of which will inform the manner in which reconstructions are used to answer important questions about our past climate and its lessons for the future.

ACKNOWLEDGMENTS

I am grateful to three anonymous reviewers, all of whom strengthened this contribution with their comments and suggestions, and to the editors for their invitation to compose this review. This work was supported in

part by NSF grant ATM0902436 and by NOAA grants NA07OAR4310060 and NA10OAR4320137. LDEO contribution 7511.

REFERENCES

1. Zalasiewicz J, Williams M, Smith AG, Barry TL, Coe AL, Bown PR, Brenchley P, Cantrill D, Gale A, Gibbard P, et al. Are we now living in the Anthropocene? *GSA Today* 2008, 18:4–8. doi:10.1130/GSAT01802A.1.
2. Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Klein Tank A, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, et al. Observations: surface and atmospheric climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2007.
3. Hegerl GC, Zwiers FW, Braconnot P, Gillett NP, Luo Y, Marengo Orsini JA, Nicholls N, Penner JE, Stott PA. Understanding and attributing climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2007.
4. Schmidt GA. Enhancing the relevance of paleoclimatic model/data comparisons for assessments of future climate change. *J Quaternary Sci* 2010, 25:79–87. doi:10.1002/jqs.1314.
5. Jones PD, Briffa KR, Osborn TJ, Lough JM, van Ommen TD, Vinther BM, Luterbacher J, Wahl ER, Zwiers FW, Mann ME, et al. High resolution paleoclimatology of the last millennium: a review of current status and future prospects. *Holocene* 2009, 19: 3–49.
6. Cullen HM, deMenocal PB, Hemming S, Hemming G, Brown FH, Guilderson T, Sirocko F. Climate change and the collapse of the Akkadian empire: evidence from the deep-sea. *Geology* 2000, 28:379–382.
7. Cook ER, Anchukaitis KJ, Buckley BM, D'Arrigo RD, Jacoby GC, Wright WE. Asian monsoon failure and megadrought during the last millennium. *Science* 2010, 328:486–489.
8. Buckley BM, Anchukaitis KJ, Penny D, Fletcher R, Cook ER, Sano M, Le Canh Nam, Wichienkeo A, Ton That Minh, Truong Mai Hong. Climate as a contributing factor in the demise of Angkor, Cambodia. *Proc Nat Acad Sci USA* 2010, 107:6748–6752.
9. Cook ER, Seager R, Cane MA, Stahle DW. North American drought: reconstructions, causes, and consequences. *Earth-Sci Rev* 2007, 81:93–134.
10. Li J, Xie S-P, Cook ER, Huang GD, Arrigo R, Liu F, Ma J, Zheng X-T. Interdecadal modulation of El Niño amplitude during the past millennium. *Nature Clim Change* 2011, 1:114–118.
11. Mann ME, Cane MA, Zebiak SE, Clement A. Volcanic and solar forcing of the tropical Pacific over the past 1000 years. *J Climate* 2005, 18:447–456.
12. Emile-Geay J, Seager R, Cane MA, Cook ER, Haug GH. Volcanoes and ENSO over the past millennium. *J Climate* 2008, 21:3134–3148. doi:10.1175/2007JCLI1884.1.
13. Vecchi GA, Wittenberg AT. El Niño and our future climate: where do we stand? *Wiley Interdiscipl Rev Climate Change* 2010, 1:260–270.
14. Hughes MK, Diaz HF. Was there a medieval warm period, and if so, where and when? *Climatic Change* 1994, 26:109–142.
15. Broecker WS. Paleoclimate: was the medieval warm period global? *Science* 2001, 291:1497–1499.
16. Mann ME. The value of multiple proxies. *Science* 2002, 297:1481–1482.
17. Bradley RS, Hughes MK, Diaz HF. Climate in medieval time. *Science* 2003, 302:404–405.
18. Mann ME, Ammann CM, Bradley RS, Briffa KR, Crowley TJ, Hughes MK, Jones PD, Oppenheimer M, Osborn TJ, Overpeck JT, et al. On past temperatures and anomalous late 20th-century warmth. *Eos Trans Amer Geophys* 2003, 84:256–258.
19. Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F. Global signatures and dynamical origins of the little ice age and the medieval climate anomaly. *Science* 2009, 326:1256–1260. doi:10.1126/science.1177303.
20. North GR, Biondi F, Bloomfield P, Christy JR, Cuffey KM, Dickinson RE, Druffel ERM, Nychka D, Otto-Bliesner B, Roberts N, et al. *Surface Temperature Reconstructions for the Last 2,000 Years*. Washington DC: The National Academies Press; 2006, 196.
21. Jansen E, Overpeck J, Briffa KR, Duplessy J-C, Joos R, Masson-Delmotte V, Olago D, Otto-Bliesner B, Peltier WR, Rahmstorf S, et al. Palaeoclimate. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2007.

22. Mann ME, Rutherford S. Climate reconstruction using pseudoproxies. *Geophys Res Lett* 2002, 29: 1501. doi:10.1029/2001GL014554.
23. Bradley RS. In: Jones PD, Bradley RS, Jouzel J, eds. *Are There Optimum Sites for Global Paleo-Temperature Reconstructions? Climatic Variations and Forcing Mechanisms of the Last 2000 Years*, NATO ASI Series, vol I 41. Berlin, Heidelberg: Springer-Verlag; 1996, 603–624.
24. Evans MN, Kaplan A, Cane MA, Villalba R. Globality and optimality in climate field reconstructions from proxy data. In: Markgraf V, ed. *Interhemispheric Climate Linkages*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2001, 53–72.
25. Evans MN, Kaplan A, Cane MA. Pacific sea surface temperature field reconstruction from coral $\delta^{18}\text{O}$ data using reduced space objective analysis. *Paleoceanography* 2002, 17:1007. doi:10.1029/2000PA000590.
26. Ammann C, Wahl E. The Importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climate Change* 2007, 85:71–88. doi:10.1007/s10584-007-9276-x.
27. Ammann CM, Genton MG, Li B. Technical note: correcting for signal attenuation from noisy proxy data in climate reconstructions. *Climate Past* 2010, 6:273–279. doi:10.5194/cp-6-273-2010.
28. Bürger G, Fast I, Cubasch U. Climate reconstruction by regression—32 variations on a theme. *Tellus* 2006, 58A:227–235.
29. Christiansen B, Schmith T, Thejll P. A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. *J Climate* 2009, 22:951–976. doi:10.1175/2008JCLI2301.1.
30. Christiansen B, Schmith Thejll P. A surrogate ensemble study of sea level reconstructions. *J Climate* 2010, 23:4306–4326.
31. Christiansen B, Schmith T, Thejll P. Reply. *J Climate* 2010, 23:2839–2844.
32. Christiansen B. Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided? *J Climate* 2011, 24:674–692.
33. Dmitriev EV, Chavro AI. Possible causes of the underestimation of paleoclimate low-frequency variability by statistical methods. *Izvestiya, Atmos Ocean Phys* 2006, 42:586–597. doi:10.1134/S0001433806050057.
34. González-Rouco F, von Storch H, Zorita E. Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. *Geophys Res Lett* 2003, 30:2116. doi:10.1029/2003GL018264.
35. González-Rouco JF, Beltrami H, Zorita E, von Storch H. Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. *Geophys Res Lett* 2006, 33:L01703. doi:10.1029/2005GL024,693.
36. González-Rouco JF, Beltrami H, Zorita E, Stevens MB. Borehole climatology: a discussion based on contributions from climate modeling. *Climate Past* 2009, 5:97–127. doi:10.5194/cp-5-97-2009.
37. Hegerl GC, Crowley T, Allen M, Hyde WT, Pollack H, Smerdon J, Zorita E. Detection of human influence on a new 1500 yr climate reconstruction. *J Climate* 2007, 20:650–666. doi:10.1175/JCLI4011.1.
38. Kaplan A. Discussion of: a statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann App Stat* 2011, 5:47–51.
39. Küttel M, Luterbacher J, Zorita E, Xoplaki E, Riedwyl N, Wanner H. Testing a European winter surface temperature reconstruction in a surrogate climate. *Geophys Res Lett* 2007, 34:L07710. doi:10.1029/2006GL027907.
40. Lee TCK, Zwiers FW, Tsao M. Evaluation of proxy-based millennial reconstruction methods. *Climate Dynam* 2008, 31:263–281. doi:10.1007/s00382-007-0351-9.
41. Li B, Nychka DW, Ammann CM. The value of multiproxy reconstruction of past climate. *J Am Stat Assoc* 2010, 105:883–895. doi:10.1198/jasa.2010.ap09379.
42. Mann ME, Rutherford S, Wahl E, Ammann C. Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J Climate* 2005, 18:4097–4107.
43. Mann ME, Rutherford S, Wahl E, Ammann C. Reply to comments on “testing the fidelity of methods used on proxy-based reconstructions of past climate” by Smerdon and Kaplan. *J Climate* 2007, 20:5671–5674.
44. Mann ME, Rutherford S, Wahl E, Ammann C. Reply to Comments on “testing the fidelity of methods used on proxy-based reconstructions of past climate” by Zorita et al. *J Climate* 2007, 20:3699–3703.
45. Mann ME, Rutherford S, Wahl E, Ammann C. Robustness of proxy-based climate field reconstruction methods. *J Geophys Res* 2007, 112:D12109. doi:10.1029/2006JD008272.
46. McIntyre S, McKittrick R. Hockey sticks, principal components, and spurious significance. *Geophys Res Lett* 2005, 32:L03710. doi:10.1029/2004GL021750.
47. McShane BB, Wyner AJ. A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann Appl Stat* 2011, 5:5–44.
48. Moberg A, Mohammad R, Mauritzen T. Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction. *Climate Dynam* 2008, 31:957–971. doi: 10.1007/s00382-008-0392-8.
49. Riedwyl N, Küttel M, Luterbacher J, Wanner H. Comparison of climate field reconstruction techniques:

- application to Europe. *Climate Dynam* 2009, 32: 381–395.
50. Rutherford S, Mann ME, Delworth TL, Stouffer RJ. Climate field reconstruction under stationary and non-stationary forcing. *J Climate* 2003, 16:462–479.
 51. Rutherford S, Mann ME, Wahl E, Ammann C. Reply to comment by Jason E. Smerdon et al. on “robustness of proxy-based climate field reconstruction methods.” *J Geophys Res* 2008, 113:D18107. doi:10.1029/2008JD009964.
 52. Rutherford SD, Mann ME, Ammann CM, Wahl ER. Comments on: “a surrogate ensemble study of climate reconstruction methods: stochasticity and robustness.” *J Climate* 2010, 23:2832–2838.
 53. Schmidt GA, Mann ME, Rutherford SD. Discussion of: a statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann App Stat* 2011, 5:67–70.
 54. Smerdon JE, Kaplan A. Comments on “testing the fidelity of methods used in proxy-based reconstructions of past climate”: the role of the standardization interval. *J Climate* 2007, 20:5666–5670.
 55. Smerdon JE, González-Rouco JF, Zorita E. Comment on “robustness of proxy-based climate field reconstruction methods” by Michael E. Mann et al. *J Geophys Res* 2008, 113:D18106. doi:10.1029/2007JD009542.
 56. Smerdon JE, Kaplan A, Chang D. On the standardization sensitivity of RegEM climate field reconstructions. *J Climate* 2008, 21:6710–6723.
 57. Smerdon JE, Kaplan A, Amrhein DE. Erroneous model field representations in multiple pseudoproxy studies: corrections and implications. *J Climate* 2010, 23:5548–5554.
 58. Smerdon JE, Kaplan A, Chang D, Evans MN. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J Climate* 2011, 24:1284–1309. doi:10.1175/2010JCLI4110.1.
 59. Smerdon JE, Kaplan A, Zorita E, González-Rouco JF, Evans MN. Spatial performance of four climate field reconstruction methods targeting the common era. *Geophys Res Lett* 2011, 38:L11705. doi:10.1029/2011GL047372.
 60. Smerdon JE. Discussion of a statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable?. *Ann App Stat* 2011, 5:76–79.
 61. Tingley MP, Huybers P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: development and applications to paleoclimate reconstruction problems. *J Climate* 2010, 23:2759–2781.
 62. Tingley MP, Huybers P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: comparison with the regularized expectation maximization algorithm. *J Climate* 2010, 23:2782–2800.
 63. Tingley MP. Spurious predictions with random time series: the Lasso in the context of paleoclimatic reconstructions. Discussion of: a statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann App Stat* 2011, 5:83–87.
 64. von Storch H, Zorita E, Jones JM, Dimitriev Y, González-Rouco F, Tett SFB. Reconstructing past climate from noisy data. *Science* 2004, 306:679–682.
 65. von Storch H, Zorita E, Jones JM, Dimitriev Y, González-Rouco F, Tett SFB. Response to comment on “reconstructing past climate from noisy data.” *Science* 2006, 312:529c.
 66. von Storch H, Zorita E, González-Rouco F. Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation. *Int J Earth Sci* 2009, 98:67–82.
 67. Zorita E, Gonzalez-Rouco F, Legutke S. Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model. *J Climate* 2003, 20:1378–1390.
 68. Zorita E, von Storch H. Methodical aspects of reconstructing non-local historical temperatures. *Memor Soc Astronom Ital* 2005, 76:794–801.
 69. Zorita E, González-Rouco JF, von Storch H. Comment on “testing the fidelity of methods used in proxy-based reconstructions of past climate” by Mann et al. *J Climate* 2007, 20:2693–3698.
 70. Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA. Solar influence on climate during the past millennium: results from transient simulations with the NCAR climate system model. *Proc Nat Acad Sci USA* 2007, 104:3713–3718. doi:10.1073/pnas.0605064.103.
 71. Stendel M, Mogensen IA, Christensen JH. Influence of various forcings on global climate in historical times using a coupled atmosphere-ocean general circulation model. *Climate Dynam* 2006, 26:1–15.
 72. Mann ME, Bradley RS, Hughes MK. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* 1998, 392:779–787.
 73. Mann ME, Zhang Z, Hughes MK, Bradley RS, Miller SK, Rutherford S, Ni F. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc Nat Acad Sci USA* 2008, 105:13252–13257.
 74. Schneider T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J Climate* 2001, 14:853–887.

75. Wahl ER, Ritson DM, Ammann CM. Comment on “reconstructing past climate from noisy data.” *Science* 2006, 312:529b.
76. Rahmstorf S. Testing climate reconstructions. *Science* 2006, 312:1872.
77. Cook E, Krusic P. *The North American Drought Atlas*. Boulder, CO: NOAA Paleoclimatology; 2004.
78. Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H. European seasonal and annual temperature variability, trends and extremes since 1550. *Science* 2004, 303:1499–1503.
79. Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H. Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. *Climate Dynam* 2002, 18:545–561.
80. Neukom R, Luterbacher J, Villalba R, Küttel M, Frank D, Jones PD, Gosjean M, Esper J, Lopez L, Wanner H. Multi-centennial summer and winter precipitation variability in southern South America. *Geophys Res Lett* 2010, 37:L14708. doi:10.1029/2010GL043680.
81. Pauling A, Luterbacher J, Wanner H. Evaluation of proxies for European and North Atlantic temperature field reconstructions. *Geophys Res Lett* 2003, 30:1787. doi:10.1029/2003GL017589.
82. Jones PD, Osborn TJ, Briffa KR, Folland CK, Horton B, Alexander LV, Parker DE, Rayner NA. Adjusting for sampling density in grid-box land and ocean surface temperature time series. *J Geophys Res* 2001, 106:3371–3380.
83. Briffa KR. Annual climate variability in the Holocene: interpreting the message of ancient trees. *Quat Sci Rev* 2000, 19:87–105.
84. Briffa KR, Osborn TJ, Schweingruber FH, Harris IC, Jones PD, Shiyatov SG, Vaganov EA. Low-frequency temperature variations from a northern tree ring density network. *J Geophys Res* 2001, 106:2929–2941.
85. D’Arrigo R, Wilson R, Jacoby G. On the long-term context for late twentieth century warming. *J Geophys Res* 2006, 111:D03103. doi:10.1029/2005JD006352.
86. Esper J, Cook ER, Schweingruber FH. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* 2002, 295:2250–2253.
87. Hegerl GC, Crowley TJ, Hyde WT, Frame DJ. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* 2006, 440:1029–1032.
88. Jones PD, Briffa KR, Barnett TP, Tett SFB. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control-run temperatures. *The Holocene* 1998, 8:455–471.
89. Mann ME, Bradley RS, Hughes MK. Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. *Geophys Res Lett* 1999, 26:759–762.
90. Mann ME, Jones PD. Global surface temperatures over the past two millennia. *Geophys Res Lett* 2003, 30:1820. doi:10.1029/2003GL017814.
91. Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlen W. Highly variable Northern Hemisphere temperatures reconstructed from low and high-resolution proxy data. *Nature* 2005, 433:613–617.
92. Oerlemans J. Extracting a climate signal from 169 glacier records. *Science* 2005, 308:675–677.
93. Pollack HN, Smerdon JE. Borehole climate reconstructions: Spatial structure and hemispheric averages. *J Geophys Res* 2004, 109:D11106. doi:10.1029/2003JD004163.
94. Rutherford S, Mann ME, Osborn TJ, Bradley RS, Briffa KR, Hughes MK, Jones PD. Proxy-based Northern Hemisphere surface temperature reconstructions: sensitivity to method, predictor network, target season, and target domain. *J Climate* 2005, 18:2308–2329.
95. Zhang Z, Mann ME, Cook ER. Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. *Holocene* 2004, 14:502–516.
96. Stevens MB, González-Rouco JF, Beltrami H. North American climate of the last millennium: underground temperatures and model comparison. *J Geophys Res* 2008, 113:F01008. doi:10.1029/2006JF000705.
97. Evans MN, Reichert BK, Kaplan A, Anchukaitis KJ, Vaganov EA, Hughes MK, Cane MA. A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J Geophys Res* 2006, 111:G03008. doi:10.1029/2006JG000166.
98. Anchukaitis KJ, Evans MN, Kaplan A, Vaganov EA, Hughes MK, Grissino-Mayer HD, Cane MA. Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophys Res Lett* 2006, 33:L04705. doi:10.1029/2005GL025050.
99. Tolwinski-Ward SE, Evans MN, Hughes MK, Anchukaitis KJ. An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dynam* 2010, 36:2419–2439. doi:10.1007/s00382-010-0945-5.
100. Thompson DM, Ault TR, Evans MN, Cole JE, Emile-Geay J. Comparison of observed and simulated tropical trends using a forward model of coral $\delta^{18}O$. *Geophys Res Lett* 2011, 38:L14706. doi:10.1029/2011GL048224.
101. Schmidt GA, LeGrande AN, Hoffmann G. Water isotope expressions of intrinsic and forced variability in

- a coupled ocean-atmosphere model. *J Geophys Res* 2007, 112:D10103. doi:10.1029/2006JD007781.
102. NOAA Paleoclimatology, Paleoclimate Reconstruction Challenge. Available at: <http://www.ncdc.noaa.gov/paleo/pubs/pr-challenge/pr-challenge.html>. (Accessed November 9, 2011).
 103. World Climate Research Programme, Coupled Model Intercomparison Project. Available at: <http://cmip-pcmdi.llnl.gov/cmip5/>. (Accessed November 9, 2011).
 104. Bateson P, Biggs P, Cuthbert A, Cuthill I, Testing M, Keverne EB, King S, Page C, Petersen O, Rothwell N, et al. *The use of non-human animals in medical research: a guide for scientists*. London, UK: Royal Society Animals in Research Committee, The Royal Society; 2004, 28.
 105. Widmann M, Goosse H, van der Schrier G, Schnur R, Barkmeijer J. Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate Past* 2010, 6:627–644.