

Auxiliary Material

Spatial Performance of Four Climate Field Reconstruction Methods Targeting the Common Era
J. E. Smerdon, A. Kaplan, E. Zorita, J. F. González-Rouco, and M. N. Evans

S1. Experimental Design

The pseudoproxy experiments (PPEs) performed in this study use two millennial-length Atmosphere-Ocean General Circulation Model (AOGCM) integrations: the NCAR CCSM 1.4 paleo run [Ammann *et al.*, 2007; hereinafter CCSM] and the GKSS ECHO-g ERIK2 run [González-Rouco *et al.*, 2006; hereinafter ECHO-g]. The annual means of the modeled temperature fields are interpolated to 5° latitude-longitude grids and comprise the grid from which all samplings are performed [Smerdon *et al.*, 2008]. These interpolated model fields do not suffer from processing problems previously described by Smerdon *et al.* [2010b]. Pseudoproxies are sampled from the 104 temperature grid points that approximate the actual proxy locations in the most populated nest of the Mann *et al.* [1998; hereinafter MBH98] proxy network (Fig. S1); all pseudoproxies are taken as available for the entire reconstruction interval. This framework is consistent with all previous PPEs that have tested the performance of the four methods considered in this manuscript [e.g. von Storch *et al.*, 2004, 2006; Mann *et al.*, 2005, 2007; Smerdon and Kaplan, 2007; Smerdon *et al.*, 2008, 2010a; Christiansen *et al.*, 2009]. Nevertheless, recent index and field reconstructions have employed an updated multiproxy network containing a total of 1,209 proxies [Mann *et al.*, 2008, 2009]. This network was screened to achieve ‘optimal’ reconstructions, yielding a total of 484 proxies in the most populated proxy nest (1800-55 C.E.). By 1500 C.E. only 176 screened proxies were used and by 1400 C.E. only 75 (Fig. S2). The MBH98 distribution is therefore representative of sampling approximating the updated optimal network by 1400 or 1500 C.E., thus underestimating spatial sampling after this time period and overestimating sampling prior to the period. The MBH98 sampling network is thus a reasonable intermediate approximation of the spatial sampling of the Mann *et al.* [2008] network in PPEs that choose to impose constant sampling density with time, as we do here and as all previous studies have done. Furthermore, sampling densities in the Mann *et al.* [2008] multiproxy network have largely increased in previously dense areas of the MBH98 network, thus not increasing the spatial extent of the proxy distribution significantly (Fig. S1). We nevertheless perform additional PPEs later in this Auxiliary Material that test CFR performance with a pseudoproxy network approximating the most populated nest in the Mann *et al.* [2008] multiproxy network. This approximation was achieved by sampling all grid cells that contained at least one record in the unscreened Mann *et al.* [2008] global network, yielding a total of 283 sampled grid cells (Fig. S1).

All methods use the same network of white-noise pseudoproxies. Each of the sampled pseudoproxy series is perturbed at four white-noise levels: signal-to-noise ratios (SNRs) of infinity (no noise), 1.0, 0.5 and 0.25, by standard deviation. The AOGCM field is also subsampled to approximate available instrumental temperature data in the Brohan *et al.* [2006] instrumental field resulting in a total of 1,732 grid cells in the global field [Mann *et al.*, 2008]. All tested methods are calibrated from 1856-1980 C.E.; all validation statistics are calculated during the reconstruction intervals from 850-1855 C.E. and 1000-1855 C.E. for the CCSM and

ECHO-g simulations, respectively. The above conventions are simplifications of real-world conditions. The noise in real proxies is typically multivariate (i.e. sensitive to climate variables in addition to temperature), non-stationary, and autocorrelated [e.g., *Jacoby and D'Arrigo* 1995; *Briffa et al.*, 1998; *Esper et al.*, 2005; *Evans et al.*, 2002; *Anchukaitis et al.*, 2006], while proxy sensitivity is typically seasonally dependent [e.g. *Pauling et al.*, 2003]. The modeled climates are considered to reasonably mimic real-world field statistics, but important features such as the strength and character of teleconnections vary across simulations and can be different from observations. Despite these assumptions and simplifications, the adopted experimental setup can be considered a best-case scenario for real-world conditions, whereas additional modifications to the PPE framework to more fully mimic real-world proxies will only degrade the results [e.g. *von Storch et al.*, 2004, 2006; *Mann et al.*, 2007].

S2. Methods

S2.1. *Mann et al.* [1998]

The MBH98 method was applied as emulated by *von Storch et al.* [2006]. The implementation of the method has been previously verified with the original proxy and instrumental temperature data used by MBH98 [*von Storch et al.*, 2009]. No detrending was performed during the calibration interval [*von Storch et al.*, 2006]. The number of EOF-PC pairs retained from the pseudo-instrumental temperature field was estimated by the N-rule [*Overland and Preisendorfer*, 1982]. Because the number of pseudoproxies in the network is kept constant for the whole reconstruction period, the number of retained EOFs is also constant within each experiment, and is 8 and 9 for the CCSM and ECHO-g experiments, respectively. These numbers are also independent of the noise level in the pseudoproxies.

S2.2. RegEM-TTLS

RegEM reconstructions were performed using truncated total least squares (RegEM-TTLS) for regularization [*Mann et al.*, 2007]. Our implementation of the RegEM-TTLS algorithm was originally validated using the codes and data made publicly available by *Mann et al.* [2007]. For these experiments (not shown), objective selection of the truncation parameter was performed in the high-frequency domain using a linear fit to the log-eigenvalue spectrum. The leading eigenvalues that were found to lie above the estimated linear function were retained. Within the low-frequency domain, the number of leading eigenvalues was set so that 50% of the target field variability was maintained. Each of the above selection criteria is equivalent to those applied by *Mann et al.* [2007]. Reconstructed fields are not provided at the *Mann et al.* [2007] public website. We therefore did not verify our implementation based on derived fields, but instead used the publicly provided decadal filtered NHM time series for verification. Derived reconstructions produced filtered NHM time series that matched the archived time series within a correlation coefficient of 0.99 or better.

RegEM CFR methods have typically employed hybrid-spectral approaches in which reconstructions are independently performed in high and low-frequency domains (usually split at the 20-year period) and then combined to form a full CFR [*Rutherford et al.*, 2005; *Mann et al.*, 2005, 2007]. Differences between hybrid and non-hybrid reconstructions have been reported to be minimal [*Rutherford et al.*, 2005; *Mann et al.*, 2005, 2007], although recent discussions have debated the importance of hybrid calibrations on the skill of derived RegEM-TTLS reconstructions [*Rutherford et al.*, 2010; *Christiansen et al.*, 2010]. Unless otherwise noted, all

CFRs performed in our study are non-hybrid reconstructions to provide consistent comparisons with the other non-hybrid methods that we have tested. In section S3 we explicitly compare the performance of the hybrid and non-hybrid RegEM-TTLS methods and show that most of the gross spatial features of the CFRs are similar.

Non-hybrid RegEM-TTLS CFRs were calculated using a linear fit to the log-eigenvalue spectrum to determine the truncation parameter in the same manner that was advocated by *Mann et al.* [2007] for the high-frequency component of their hybrid reconstructions. For the hybrid reconstructions, a linear fit to the log-eigenvalue spectrum was again used to determine the truncation parameter for the high-frequency reconstructions, while the low-frequency truncation was determined by selecting the eigenvalue rank yielding 33% of the cumulative variance in the low-frequency field. This reduction in the retained cumulative variance from 50%, as originally adopted by *Mann et al.* [2007], has been advocated in later publications by *Rutherford et al.* [2010] and *Mann et al.* [2009]. The selected truncation parameters for all of the RegEM-TTLS CFRs are given in Table S1, as well as the rank reduction of the target field. In all hybrid and non-hybrid reconstructions, a value of 10^{-4} was used for the stagnation tolerance and the inflation parameter was set to 1.

S2.3. Ridge Regression

We apply standard ridge regressions [*Hoerl and Kennard*, 1970] for the ridge regression CFRs in this study. The application of a single ridge regression is a break from earlier studies that have used ridge regression as the form of regularization in the iterative RegEM algorithm (RegEM-Ridge). The application of RegEM-Ridge for the purpose of Common Era CFRs has been discussed in detail in various publications [*Schneider* 2001; *Mann et al.*, 2005; *Smerdon and Kaplan* 2007; *Lee et al.*, 2007; *Smerdon et al.*, 2008, 2010a; *Christiansen et al.*, 2009]. Our motivation for applying a standard ridge regression in the present context is twofold: 1) a single ridge regression is simpler and more computationally efficient than the RegEM-Ridge algorithm; and 2) for the current problem, in which a uniform block comprises the missing values in the data matrix, the RegEM-Ridge solution converges to the single ridge regression result. We have confirmed this latter finding using PPEs (results not shown). The value of the ridge parameter for the single ridge regressions was determined in the same manner applied by *Schneider* [2001] to select the ridge parameter for RegEM-Ridge, namely by minimization of the generalized cross validation (GCV) function [*Golub et al.*, 1979]. GCV selections of the ridge parameter in the ridge regression CFRs at SNRs of infinity, 1.0, 0.5 and 0.25 were 0.63, 1.24, 1.44, 2.18 and 0.63, 1.19, 1.37, 1.47 for the CCSM and ECHO-g PPEs, respectively. Given the uniform block matrix of missing values in the PPE context, each ridge regression used the singularly selected ridge parameter for all missing values.

S2.4. Canonical Correlation Analysis

The canonical correlation analysis (CCA) method was applied as described in *Smerdon et al.* [2010a]. Dimensions of the proxy and instrumental fields were both reduced by eigenvalue truncation, as were the number of retained canonical coefficients. These dimensional reductions were selected based on ‘leave-half-out’ cross-validation statistics, as described by *Smerdon et al.* [2010a]. The CCA dimensional selections are given in Table S2 for the minimum cross-validation root mean square error (RMSE), as well as those achieved for preferred dimensions taken as the first local minimum of the RMSE space to guard against artificial skill [*Smerdon et*

al., 2010a]. These preferred dimensions are used in all of the CCA CFRs in this study. In those cases where the preferred dimensional selections were different from those of the absolute minimum RMSE, the mean RMSE is increased by only several thousandths of a Kelvin degree.

S2.5. Assessment Metrics

We plot or summarize four different statistics to assess reconstruction skill: 1) the Pearson's correlation coefficient; 2) the difference between the reconstruction and target means, i.e. the bias; 3) the ratio between the reconstruction and target standard deviations; and 4) the RMSE. We use RSME as a summary statistic that combines both the resolved variance and the bias into a single assessment. The coefficient of efficiency (CE) and reduction of error (RE) statistics are often used in the paleoclimate literature as additional statistical validation measures. Advocates of these statistics point out that RE and CE measure both the resolved variance and reconstructed mean in derived reconstructions (e.g., *Wahl and Ammann* [2007]), a characteristic that is also shared by the RMSE statistic adopted in this study. The chosen assessment metrics collectively characterize the variance and the fidelity of local means in the derived reconstructions and transparently decompose the uncertainty of the CFRs into individual error categories. While not all of the methods include straightforward derivations of error estimates, these out-of-sample-assessment metrics collectively define the residuals that are often used to derive empirically based uncertainty estimates for paleoclimatic reconstructions.

S3. Comparisons of Hybrid and Non-Hybrid RegEM-TTLS CFRs

A precursory evaluation of our emulation of the RegEM-TTLS method is possible based on skill results published by *Mann et al.* [2007], in which skill statistics were reported for their versions of the hybrid RegEM-TTLS method. These results were affected by processing errors [*Smerdon et al.*, 2010b], resulting in PPEs that use pseudoproxy and target distributions unrepresentative of the realistic distributions used in the present study. Nevertheless, the relative behavior of our RegEM-TTLS CFRs is consistent with the mean field r^2 (called multivariate r^2 by *Mann et al.* [2007]) and the r^2 for the NHMs, as they were reported by *Mann et al.* [2007] (Table S3). The overall effect of the different sampling schemes appears to be slightly reduced skill in the CFRs that we have performed based on realistic sampling schemes.

Comparisons between the hybrid and non-hybrid RegEM-TTLS reconstructions yield mixed results that do not necessarily advocate for one methodological version over the other. The correlation coefficients and standard deviation ratio patterns are broadly similar for both versions, while the mean biases are significantly improved for the hybrid version (Fig. S3 and Table S4). Non-hybrid RegEM-TTLS CFRs outperform the hybrid counterparts (or are minimally worse) in mean field correlations, standard deviation ratios and RMSE for all noise levels (Table S4), except for the highest noise case in which the hybrid version is modestly more skillful. Note that mean standard deviation ratios in both sets of CFRs indicate preservation of variance in regions of small correlation coefficients, thus larger mean standard deviation ratios are not necessarily a good measure of CFR success. Except for an SNR of 1.0, the hybrid CFRs outperform the non-hybrid CFRs in Northern Hemisphere mean correlations (NHMCs). The reduced mean biases noted in the overall field, are also more evident in the hybrid version of the NHM time series (Fig. S7). This difference may or may not be a consistent difference between the two methodological options and should be tested with more realizations of pseudoproxy noise and model simulations. Nevertheless, it should be underscored that even in the case of the

hybrid RegEM-TTLS CFR, in which skillful NHM reconstructions are evident, the field statistics presented in Fig. S3 and Table S4 indicate that evaluation of NHM performance is insufficient for a full assessment of CFR skill.

With the exception of the differences in the mean biases in the hybrid and non-hybrid CFRs, there appears to be only modest differences between the two methodological choices. Comparisons between either of the RegEM-TTLS versions are therefore appropriate in the context of our study, and our choice to use non-hybrid versions of the RegEM-TTLS method is warranted. It is also worth noting that with the exception of the mean biases, in which either the hybrid or non-hybrid RegEM-TTLS CFRs outperform the other methods at almost all noise levels, the hybrid RegEM-TTLS CFRs are less skillful than all other methods in both the mean field correlations and the overall error in the field for all noise levels except at an SNR of 0.25. This again underscores the limited value of using NHMs as indicators of field skill, while suggesting that none of the evaluated methods can be necessarily considered to have characteristics that are significant improvements over another. This discussion also points to the need to further investigate how best to compare the field skill of CFRs, as well as determine robust spatial skill metrics for CFR assessments.

S4. MBH98 vs. *Mann et al.* [2008] Pseudoproxy Networks

We tested the impact of an updated pseudoproxy network based on the distribution of the *Mann et al.* [2008] multiproxy network using CCA CFRs. The selected dimensions for the CCA reconstructions using either the MBH98 or *Mann et al.* [2008] pseudoproxy networks are given in Table S2. Field statistic comparisons and NHM time series are given for the two CCA PPEs in Figs. S4 and S8 and in Table S4.

In all cases, CCA CFRs that used pseudoproxies representing the most populated nest in the *Mann et al.* [2008] network yielded improved reconstruction skill relative to the equivalent MBH98 pseudoproxy CFRs. Correlation coefficients over the more densely sampled areas of North America and Eurasia showed significant improvements and further reflect the observation that the largest correlation coefficients concentrate in highly sampled regions. NHMCs also showed marked increases, although the mean biases and variance losses evident in the CCSM NHMs derived from the CCA CFRs were not wholly mitigated by the updated network (Fig. S8). Despite improvements, however, the updated network did not eliminate all of the spatial errors observed in CFRs that used the MBH98 network (Fig. S4 and Table S4). Large areas of the Southern Hemisphere, for instance, still appear poorly resolved in the CFRs using the updated network (Fig. S4). Heterogeneous variance losses and mean biases are also still evident (Fig. S4). While the improvements in the CFRs that use the updated network are therefore promising, particularly in the NH, the richer *Mann et al.* [2008] network appears unlikely to mitigate all of the spatial errors in the CFRs as we have characterized them. This finding likely stems from the fact that increased proxy densities in the new network have occurred primarily in previously densely sampled areas. Expanding multiproxy networks to undersampled regions such as the ocean basins and Southern Hemisphere therefore should be an important goal of the paleoclimate community. It should also be noted that PPEs are useful for the general conclusion that spatial skill can concentrate in densely sampled areas, but they cannot be used for optimal site selections because the spatial statistics in model fields (e.g. teleconnections) are not necessarily representative of the real-world climate. Optimal site selection studies are nevertheless possible

with proxy data [e.g. Evans et al. 1998] or with instrumental data alone, and represent a useful evaluation of how to improve CFRs through increased spatial sampling.

Auxiliary Material References

- Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, *Proc. Nat. Acad. Sci. USA*, *104*, 3713-3718, doi:10.1073-pnas.0605064.103.
- Anchukaitis, K. J., M. N. Evans, A. Kaplan, E. A. Vaganov, M. K. Hughes, H. D. Grissino-Mayer, and M. A. Cane (2006), Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought, *Geophys. Res. Lett.*, *33*, L04705, doi:10.1029/2005GL025050
- Briffa, K., F. Schweingruber, P. D. Jones, and T. Osborn (1998), Reduced sensitivity of recent tree growth to temperature at high northern latitudes, *Nature*, *391*, 678–682.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Christiansen, B. T. Schmith, and P. Thejll (2009), A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness, *J. Climate*, *22*(4), 951-976, doi:10.1175/2008JCLI2301.1.
- Christiansen, B., T. Schmith, and P. Thejll (2010), Reply, *J. Climate*, *23*(10), 2839-2844, doi:10.1175/2010JCLI3281.1.
- Esper, J., R. J. S. Wilson, D. C. Frank, A. Moberg, H. Wanner, and J. Luterbacher (2005), Climate: Past ranges and future changes. *Quat. Sci. Rev.*, *24*, 2164-2166.
- Evans, M., A. Kaplan, and M. Cane (1998), Optimal sites for coral-based reconstruction of sea surface temperature, *Paleoceanography*, *13*, 502-516.
- Evans, M. N., A. Kaplan, and M. A. Cane (2002), Pacific sea surface temperature field reconstruction from coral $d^{18}O$ data using reduced space objective analysis. *Paleoceanography*, *17*(1), doi:10.1029/2000PA000590.
- Golub, G. H., M. Heath, G. Wahba (1979), Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, *21*(2), 215-223.
- González-Rouco, J. F., H. Beltrami, E. Zorita, and H. von Storch (2006), Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, *Geophys. Res. Lett.*, *33*, L01703, doi:10.1029/2005GL024693.
- Hoerl, A. E., and R. W. Kennard (1970), Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, *12*, 55-67.
- Jacoby, G. C., and R. D'Arrigo (1995), Tree-ring width and density evidence of climatic and potential forest change in Alaska, *Global Biogeochem. Cycles*, *9*, 227–234
- Lee, T. C. K., F. W. Zwiers, and M. Tsao (2007), Evaluation of proxy-based millennial reconstruction methods. *Climate Dyn.*, *31*, 263-281.
- 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., S. Rutherford, E. Wahl, and C. Ammann (2005), Testing the fidelity of methods used in proxy-based reconstructions of past climate, *J. Climate*, *18*, 4097-4107.
- Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *J. Geophys. Res.*, *112*, D12109,

- doi:10.1029/2006JD008272.
- 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. Nat. Acad. Sci. USA*, *105*, 36, 13252-13257.
- Mann, M.E., Z. Zhang, S. Rutherford, R. S. Bradley, M. K. Hughes, D. Shindell, C. Ammann, G. Faluvegi, F. Ni, (2009), Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*(5957), 1256-1260.
- Overland, J. E., and R. W. Preisendorfer (1982), A significance test for principal components applied to a cyclone climatology, *J. Climate*, *110*, 1-4.
- Pauling, A., J. Luterbacher, and H. Wanner (2003), Evaluation of proxies for European and North Atlantic temperature field reconstructions, *Geophys. Res. Lett.*, *30*(15), 1787, doi:10.1029/2003GL017589.
- Rutherford, S., M. E. Mann, T. J. Osborn, R. S. Bradley, K. R. Briffa, M. K. Hughes, and P. D. Jones (2005), Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to methodology, predictor network, target season, and target domain, *J. Climate*, *18*, 2308-2329.
- Rutherford, S.D., M.E. Mann, C.M. Ammann, and E.R. Wahl (2010), Comments on: “A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness” by Christiansen, Schmith and Thejll, *J. Climate*, *23*(10), 2832-2838, doi:10.1175/2010JCLI3146.1.
- Schneider, T. (2001), Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *J. Climate*, *14*, 853-887.
- Smerdon, J.E., and A. Kaplan (2007), Comments on “Testing the fidelity of methods used in proxy-based reconstructions of past climate”: The role of the standardization interval, *J. Climate*, *20*, 22, 5666-5670.
- Smerdon, J.E., A. Kaplan, and D. Chang (2008), On the standardization sensitivity of RegEM climate field reconstructions, *J. Climate*, *21*(24), 6710-6723.
- Smerdon, J.E., A. Kaplan, and D.E. Amrhein (2010b), Erroneous Model Field Representations in Multiple Pseudoproxy Studies: Corrections and Implications, *J. Climate*, *23*, 5548-5554, doi: 10.1175/2010JCLI3742.1.
- Smerdon, J.E., A. Kaplan, D. Chang, and M.N. Evans (2011), A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium, *J. Climate*, *24*, 1284-1309.
- Wahl, E. R., and C. M. Ammann (2007), Robustness of the Mann, Bradley, Hughes reconstruction of surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence, *Clim. Change*, *85*, 33-69.
- von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett (2004), Reconstructing past climate from noisy data, *Science*, *306*, 679-682.
- von Storch, H., E. Zorita, J. M. Jones, F. González-Rouco, and S. F. B. Tett (2006), Response to comment on “Reconstructing past climate from noisy data,” *Science*, *312*, 529c.
- von Storch, H., E. Zorita, and F. González-Rouco (2009), Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, *Int. J. Earth Sci.*, *98*, 67-82.

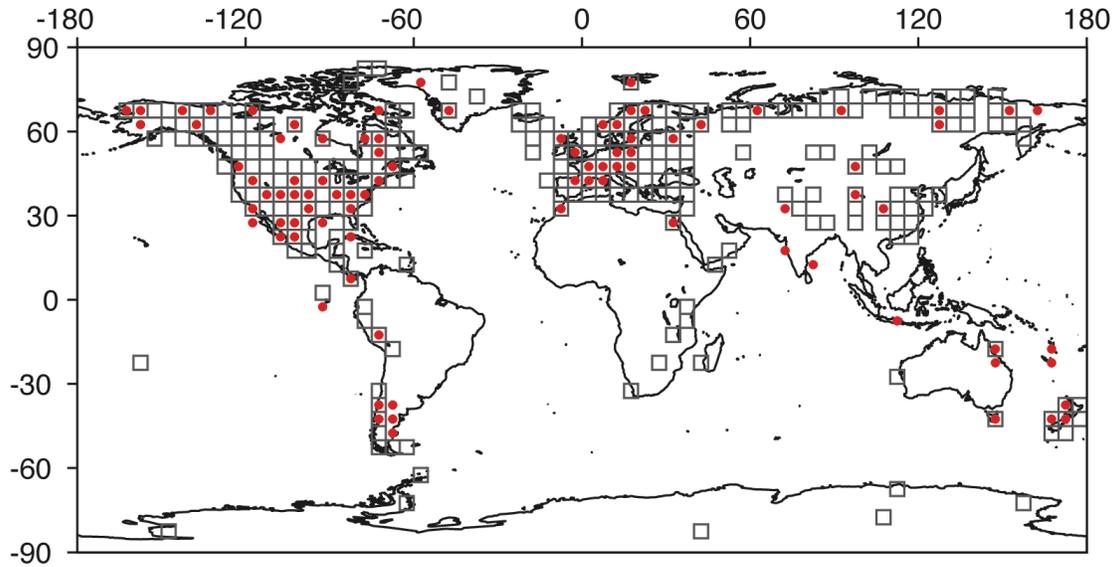


Figure S1. Comparison of the pseudoproxy networks used in this study to approximate the most populated nests in the MBH98 (red dots) and *Mann et al.* [2008] (grey squares) multiproxy networks.

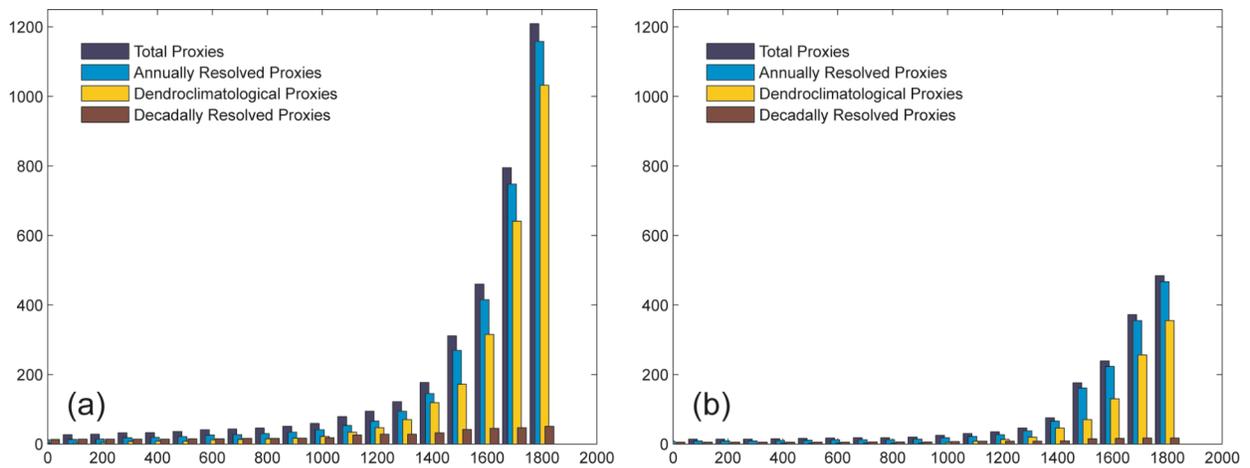


Figure S2. Total and categorized proxy abundances in the global *Mann et al.* [2008] proxy network for: (a) the full dataset; and (b) the culled dataset as screened and used by *Mann et al.* [2008]. The plotted abundances are as published in the Supporting Information by *Mann et al.* [2008].

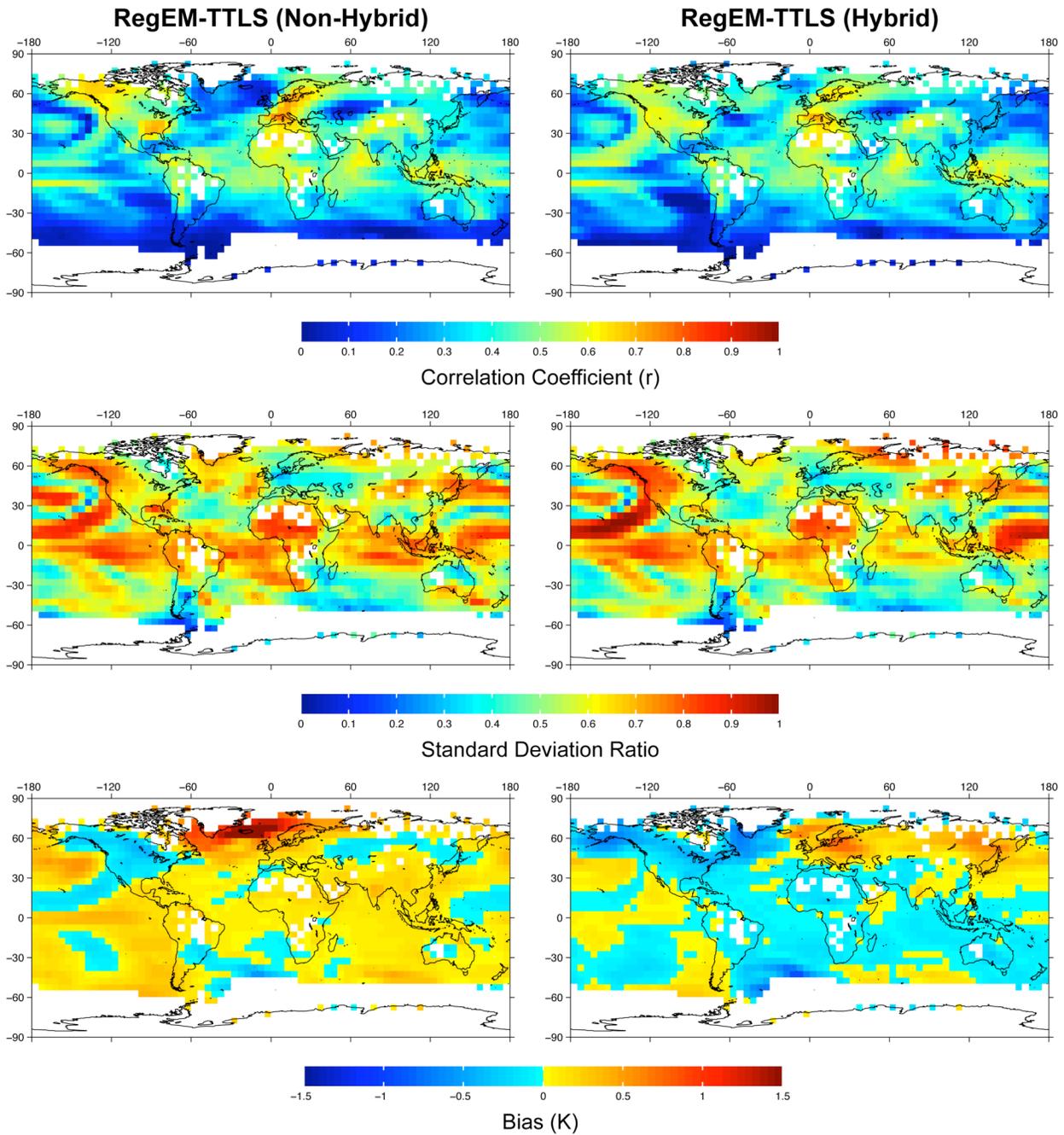


Figure S3. Field statistic comparisons for CFRs derived using the non-hybrid (left column) and hybrid (right column) versions of the RegEM-TTLS method and the CCSM PPE framework. Field statistics are computed as shown in Figs. 1-3 of the main article and correspond to correlation coefficients (top row), standard deviation ratios (middle row) and mean biases (bottom row).

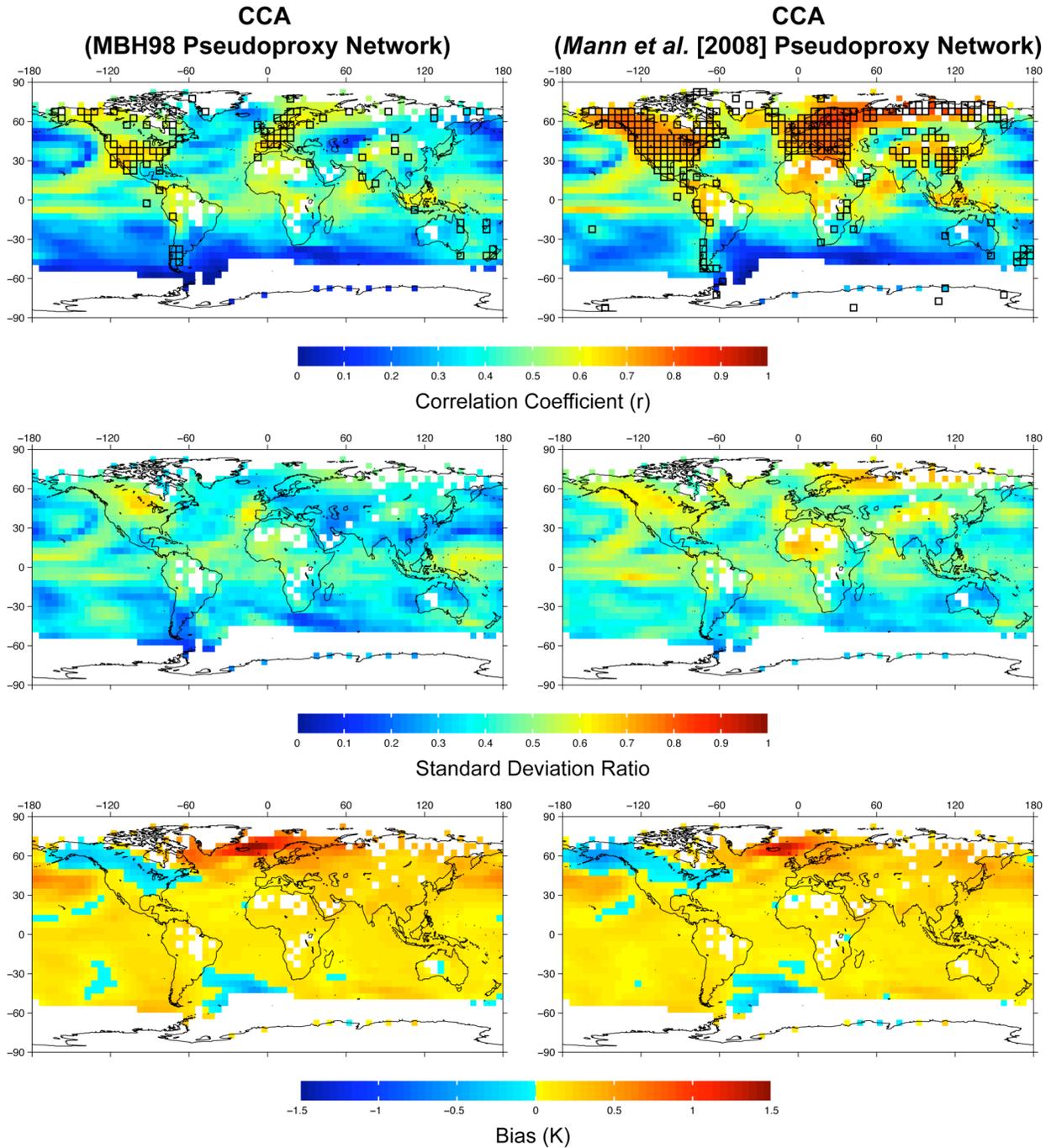


Figure S4. Field statistic comparisons for CCA CFRs and pseudoproxy networks approximating the MBH98 (left column) and *Mann et al.* [2008] (right column) multiproxy networks in the CCSM PPE framework. Field statistics are computed as shown in Figs. 1-3 of the main article and correspond to correlation coefficients (top row), standard deviation ratios (middle row) and mean biases (bottom row).

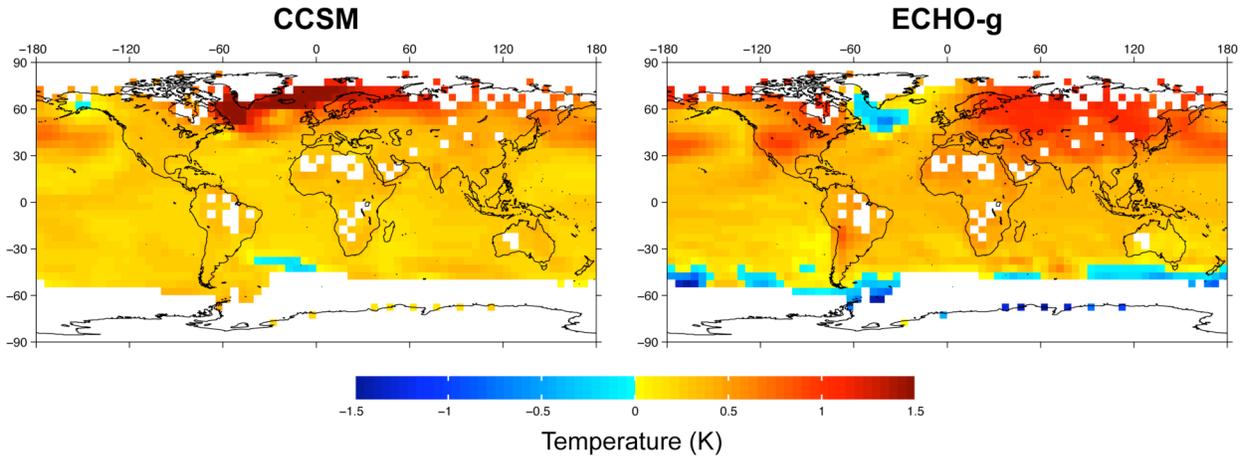


Figure S5. Mean temperature differences between the calibration and reconstruction interval means in the CCSM (left panel) and ECHO-g (right panel) simulations.

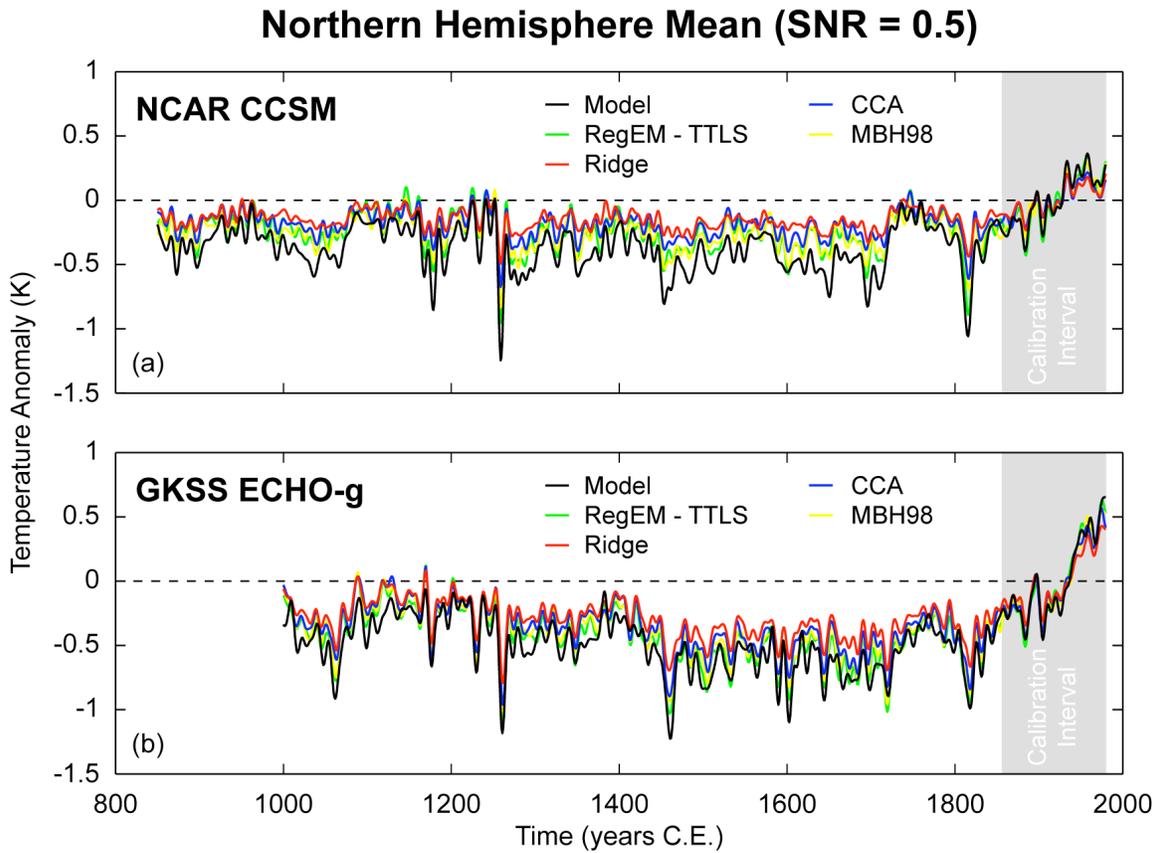


Figure S6. Area-weighted NHM time series computed from CFRs derived from the four non-hybrid methods, the two model experiments, and pseudoproxies with an SNR of 0.5.

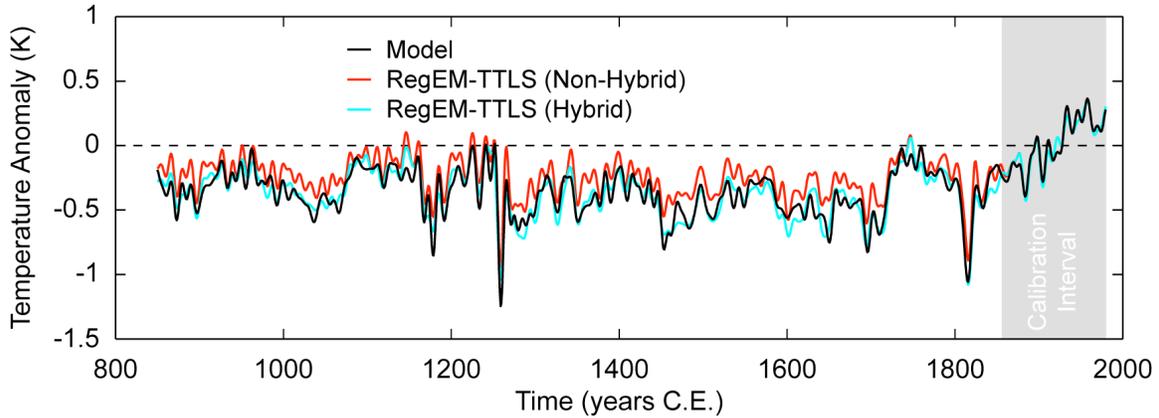


Figure S7. Area-weighted NHM time series computed from hybrid and non-hybrid versions of the RegEM-TTLS method and the CCSM PPE framework.

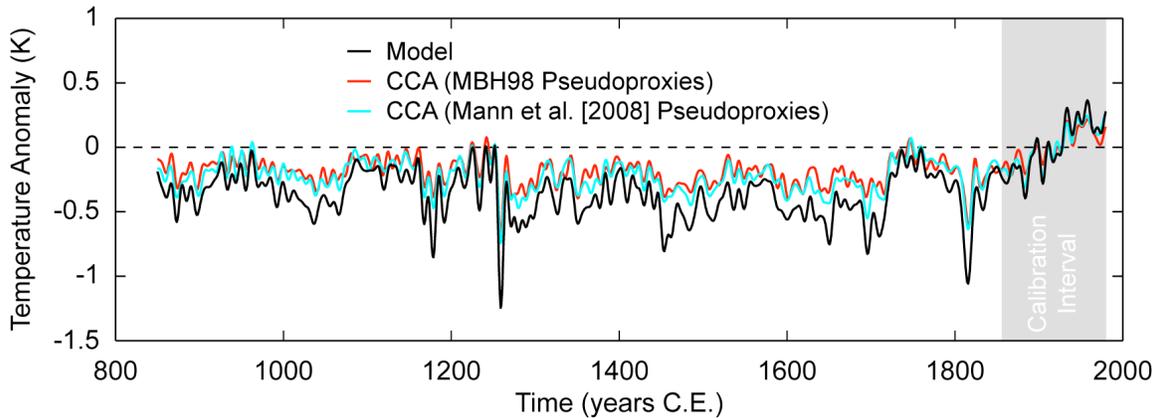


Figure S8. Area-weighted NHM time series computed from CCA CFRs and pseudoproxy networks approximating the MBH98 and *Mann et al.* [2008] multiproxy networks in the CCSM PPE framework.

Table S1. Truncation parameters used in the RegEM-TTLS CFRs performed in this study^a.

SNR	Non-Hybrid		Hybrid	
	Full Spectrum	Instrumental PCs Retained	High-Frequency	Low-Frequency
Infinity	6 (4)	6	6	1
1.0	6 (4)	6	5	1
0.5	4 (2)	4	4	2
0.25	4 (2)	4	4	3

^aAll numbers are for CCSM reconstructions, except those shown in parentheses, which report the selections for the ECHO-g reconstructions. The total instrumental PCs retained are shown for the hybrid reconstructions, while the total number of retained PCs for the non-hybrid reconstructions are equal to the given values of the truncation parameters in all cases.

Table S2. Dimensional selections and associated calibration period RMSE values for all CCA CFRs performed in this study^a.

SNR	Absolute Minimum		Preferred Dimensions	
	d_{cca}, d_p, d_t	RMSE	d_{cca}, d_p, d_t	RMSE
<i>CCSM (MBH98 Pseudoproxy Network)</i>				
Inf	20,23,49	0.53	15,15,49	0.53
1.0	18,20,50	0.58	13,23,20	0.58
0.5	10,19,10	0.64	7,18,10	0.64
0.25	5,29,5	0.68	2,28,3	0.68
<i>CCSM (Mann et al. [2008] Pseudoproxy Network)</i>				
Inf	37,37,49	0.47	33,33,48	0.47
1.0	36,36,50	0.52	27,28,49	0.52
0.5	32,34,49	0.59	19,35,30	0.59
0.25	6,44,6	0.66	6,44,6	0.66
<i>ECHO-g (MBH98 Pseudoproxy Network)</i>				
Inf	18,21,45	0.47	18,21,45	0.47
1.0	23,25,32	0.52	4,32,4	0.54
0.5	16,24,19	0.59	3,41,3	0.59
0.25	4,38,4	0.67	4,38,4	0.67

^aThe preferred dimensions were those used for all of the analyzed CCA CFRs herein and were selected as the first local minimum in the RMSE space [Smerdon et al., 2010a].

Table S3. CCSM verification statistics (850-1855 C.E.) computed for RegEM-TTLS CFRs in Mann et al. [2007] and in this study^a.

SNR	Mann et al. [2007]		This Study (Hybrid)		This Study (Non-Hybrid)	
	Mean-Field r^2	r^2 for NH Means	Mean-Field r^2	r^2 for NH Means	Mean-Field r^2	r^2 for NH Means
Infinity	0.30	0.87	0.21	0.76	0.24	0.76
1.0	0.23	0.86	0.19	0.69	0.22	0.71
0.5	0.19	0.83	0.15	0.65	0.15	0.58
0.25	0.06	0.34	0.07	0.40	0.07	0.27

^aNumbers are not directly comparable because of problems in the Mann et al. [2007] pseudoproxy network and target field as reported by Smerdon et al. [2010b]. The relative behavior in these statistics is nevertheless consistent, while suggesting that the updated PPE framework in this study that reflects real-world sampling distributions slightly reduces the skill in the hybrid RegEM-TTLS method tested by Mann et al. [2007].

Table S4. Verification statistics computed for all reconstructions performed in this study^a.

SNR	NHMC	Mean Field Correlation	Mean STD Ratio	Mean Bias (K)	Mean RMSE (K)
<i>MBH98</i>					
Inf	0.874 (0.932)	0.504 (0.553)	0.586 (0.604)	0.004 (-0.005)	0.530 (0.494)
1.0	0.847 (0.899)	0.474 (0.517)	0.532 (0.578)	0.056 (0.018)	0.543 (0.507)
0.5	0.739 (0.813)	0.391 (0.444)	0.503 (0.575)	0.099 (0.050)	0.588 (0.538)
0.25	0.437 (0.562)	0.226 (0.280)	0.451 (0.573)	0.199 (0.146)	0.674 (0.620)
<i>RegEM-TTLS (Non-Hybrid)</i>					
Inf.	0.873 (0.926)	0.460 (0.506)	0.614 (0.584)	0.018 (-0.024)	0.561 (0.514)
1.0	0.842 (0.900)	0.433 (0.481)	0.631 (0.620)	0.027 (-0.017)	0.571 (0.524)
0.5	0.759 (0.818)	0.350 (0.408)	0.616 (0.572)	0.108 (-0.001)	0.632 (0.546)
0.25	0.517 (0.640)	0.229 (0.307)	0.622 (0.618)	0.172 (0.040)	0.707 (0.585)
<i>RegEM-TTLS (Hybrid)</i>					
Inf.	0.874	0.429	0.578	-0.019	0.575
1.0	0.833	0.402	0.570	-0.017	0.579
0.5	0.808	0.359	0.583	-0.026	0.607
0.25	0.632	0.244	0.680	-0.017	0.666
<i>Ridge Regression</i>					
Inf.	0.947 (0.960)	0.649 (0.667)	0.699 (0.695)	0.017 (-0.011)	0.417 (0.413)
1.0	0.874 (0.912)	0.524 (0.554)	0.477 (0.537)	0.103 (0.041)	0.518 (0.483)
0.5	0.757 (0.804)	0.385 (0.433)	0.344 (0.431)	0.179 (0.126)	0.603 (0.546)
0.25	0.425 (0.597)	0.178 (0.258)	0.157 (0.326)	0.279 (0.253)	0.691 (0.639)
<i>CCA</i>					
Inf.	0.922 (0.953)	0.581 (0.628)	0.676 (0.713)	0.007 (-0.005)	0.470 (0.442)
1.0	0.870 (0.895)	0.507 (0.518)	0.549 (0.561)	0.060 (0.009)	0.521 (0.513)
0.5	0.772 (0.783)	0.390 (0.425)	0.392 (0.504)	0.142 (0.065)	0.597 (0.547)
0.25	0.391 (0.567)	0.177 (0.280)	0.214 (0.447)	0.254 (0.162)	0.681 (0.602)
<i>CCA (Mann et al. [2008] Pseudoproxies)</i>					
Inf.	0.972	0.642	0.781	-0.002	0.394
1.0	0.940	0.566	0.597	0.035	0.450
0.5	0.867	0.463	0.451	0.116	0.532
0.25	0.638	0.281	0.311	0.202	0.639

^aUnless otherwise noted in the subheadings, all reconstructions are non-hybrid and performed using the MBH98 pseudoproxy network. All numbers given outside of parentheses are for CCSM experiments (850-1855 C.E. validation period); numbers in parentheses are for the ECHO-g experiments (1000-1855 C.E. validation period). For all PPEs that used the MBH98 pseudoproxy network, the numbers given in bold black (grey) correspond to the best (second best) performance in each metric category for each noise level. In cases of ties, numbers from both methods have been bolded.