

Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors

M. N. Evans,^{1,2} J. E. Smerdon,³ A. Kaplan³, S. E. Tolwinski-Ward⁴, J. F.

González-Rouco⁵

¹Department of Geology, University of Maryland, College Park, MD USA

²Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD USA

³Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY USA

⁴AIR Worldwide Corporation, Boston, MA USA

⁵Facultad de Ciencias Físicas, Instituto de Geociencias (UCM-CSIC), Universidad Complutense, Madrid, Spain

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2014GL062063

Climate field reconstructions (CFRs) of the global annual surface air temperature (SAT) field and associated global area-weighted mean annual temperature (GMAT) are derived in a collection of pseudoproxy experiments for the past millennium. Pseudoproxies are modeled from temperature (T), precipitation (P), T+P, and VS-Lite (VSL), a nonlinear and multivariate proxy system model for tree-ring widths. Spatial patterns of reconstruction skill and spectral bias for the T+P and VSL-derived CFRs are similar to those previously shown using temperature-only pseudoproxies, but demonstrate overall degraded skill and spectral bias for SAT reconstruction. Analysis of GMAT spectra nevertheless suggests that the true GMAT frequency spectrum is resolved by those pseudoproxies (T, T+P, and VSL) that contain some temperature information. The results suggest that mixed temperature and moisture-responding paleoclimate data may produce actual GMAT reconstructions with skill, error and spectral characteristics like those expected from univariate and linear temperature responders, but spatially-resolved CFR results should be analyzed cautiously.

1. Introduction

Pseudoproxy experiments (PPEs; see *Smerdon* [2012] for a comprehensive review) are synthetic evaluations of the uncertainties associated with the reconstruction of actual climate indices and climate fields. Because the PPE construct is systematic, albeit with realistically-chosen observing sites, observational uncertainty, and reconstruction methodology, PPE results can be used to assess the likely impact of climate field reconstruction (CFR) uncertainties via controlled experimentation. PPE-based research has shown that observing network [*Evans et al.*, 1998, 2001, 2002], space [*Smerdon et al.*, 2011a], time and frequency [*von Storch et al.*, 2004; *Mann et al.*, 2005; *von Storch et al.*, 2009; *Wang et al.*, 2014], methodology [*von Storch et al.*, 2009; *Smerdon et al.*, 2011b], paleodata type [*Evans et al.*, 2002], environmental response [*Franke et al.*, 2013], error model [*von Storch et al.*, 2009], reconstruction target [*Smerdon et al.*, 2011a; *Christiansen*, 2011; *Emile-Geay et al.*, 2013], and possibly forcing mechanism [*Wang et al.*, 2014] can all substantially influence the skill and uncertainty characteristics of CFRs and derived indices.

One of the most important sets of annual-resolution paleodata are tree-ring width chronologies (e.g. *Evans et al.* [2013], *Wang et al.* [2014]). Although selected for use in CFRs to reflect a univariate and linear proxy system response to temperature [*Mann et al.*, 1999], trees generally represent a nonlinear and multivariate sensor of environmental variations [*Fritts*, 1976; *Cook and Kairiukstis*, 1990; *Vaganov et al.*, 2006, 2011; *Franke et al.*, 2013]. Studies have examined the sensitivity of actual CFRs to inclusion of tree-ring observations [*Mann et al.*, 1998; *Evans et al.*, 2002; *Mann et al.*, 2008], but CFR skill and uncertainty arising from use of tree-ring width chronologies (or, for that mat-

ter, any nonlinear or non-univariate proxy system) have not been systematically isolated and explored in the PPE context. Here we assess CFR uncertainty in a PPE in which pseudoproxies are modeled using information other than surface temperature, including a proxy system model for tree-ring width variations [Evans *et al.*, 2006a; Graham and Wahl, 2011]. Although a vast simplification of tree physiology, this model, VS-Lite, is realistically multivariate and nonlinear, and has been validated against actual tree-ring width chronology data [Vaganov *et al.*, 1999; Evans *et al.*, 2006b; Anchukaitis *et al.*, 2006; Tolwinski-Ward *et al.*, 2011a, b; Breitenmoser *et al.*, 2014]. We use the PPE construct to infer how the skill and uncertainty attributes of real-world temperature CFRs may depend on the linearity and dimensionality of the proxy system models on which they are based.

2. Experimental Design

As our reconstruction target we used annually-averaged surface temperature fields from the last millennium “ERIK-2” run (1000–1990 C.E.; González-Rouco *et al.* [2006, 2009]) of the ECHO-G AOGCM [Legutke and Voss, 1999; Zorita *et al.*, 2005], interpolated to 5 x 5 degree spatial resolution [Smerdon *et al.*, 2011b], with an observational mask that approximates data availability from the gridded temperature product of Brohan *et al.* [2006] (Fig. 1; for further details, see Mann *et al.* [2008] and Smerdon *et al.* [2011b]). Although CMIP5/PMIP3 simulations are now available [Taylor *et al.*, 2012], use of the “ERIK-2” simulation in these experiments allows direct comparison with other PPEs in the literature that use only mean annual surface air temperature to construct the pseudoproxies; see Supporting Information (hereinafter SI) for more information. ECHO-G

“ERIK-2” simulated precipitation and temperature fields differ systematically from climatological observations in mean, annual cycle and variance at the gridpoint scale over some regions in which we simulate tree-ring widths, but they are also broadly realistic at the large scale [Min *et al.*, 2005a, b], as expected from GCM output [Zorita *et al.*, 1995; von Storch, 2010]. Therefore, we did not correct ECHO-G output for bias relative to observed climatology prior to construction of the pseudoproxies. We used the model output to create pseudoproxies at 191 grid cells in which dendrochronological data (tree-ring width or maximum latewood density) were available in the Mann *et al.* [2008] multiproxy network. Four sets of pseudoproxies were derived from ECHO-G output: (1) annually averaged temperature (T); (2) annually-averaged precipitation (P); (3) equally weighted sum of annual temperature and annual precipitation (T+P); and (4) tree-ring width chronologies simulated by VS-Lite (VSL), for which inputs were site latitude and monthly temperature and precipitation (SI Section 1.2).

VS-Lite is a simple but realistic model for simulating tree-ring width variations as a thresholded minimization function of temperature and soil moisture, scaled by insolation and integrated over a prescribed pre-season and growing season [Tolwinski-Ward *et al.*, 2011a]. Soil moisture, in turn, is estimated within VS-Lite via the Climate Prediction Center (CPC) Leaky Bucket model [Huang *et al.*, 1996] and input temperature and precipitation. We implemented VS-Lite using both tuned (SI; Tolwinski-Ward *et al.* [2013]) and default [Tolwinski-Ward *et al.*, 2011a] parameters. Tolwinski-Ward *et al.* [2013] found that the use of biased inputs may also bias estimation of the VS-Lite parameters; the authors also found that the skill of VS-Lite simulations may be sensitive to both moisture

and temperature parameters depending on climatology of the location. However, results of VS-Lite simulations with tuned and default parameters are quite similar (1000–1855 C.E. median correlation over 191 raw simulations = 0.77; 80% of correlations within the interval 0.39-0.97). When evaluated by location for moisture and temperature limitations, both tuned and default simulations produce results that are realistic (SI Fig. 1). As in the real world, most simulated high latitude sites are temperature-limited; moisture-limited sites are found in semi-arid regions, and there are few sites that are either moisture or temperature-limited [Babst et al., 2013; St. George and Ault, 2014], consistent with prior expectations based on principles of dendrochronology and observations [Fritts, 1976].

To produce the CFRs, we added Gaussian uncorrelated noise with mean zero and variance 4/5 to all centered, noise-free candidate series (T, P, T+P, VSL) scaled to a variance of 1/5. This construct (see SI) results in pseudoproxies with zero mean, unit variance, and signal-to-noise ratio (SNR) by standard deviation [Smerdon, 2012] of 0.5, with the added noise term representing the local observational error with respect to surface air temperature. For T-based pseudoproxies, this may be interpreted simply as measurement uncertainty, but for T+P, P, and VSL, the noise term also represents the observational uncertainty arising from the pseudoproxy formulation itself. It may not, however, represent sources of variation not directly related to climate [von Storch et al., 2009]. We used canonical correlation analysis (CCA; Smerdon et al. [2011b]; SI) to produce each CFR, with pseudoproxy calibration over the period 1856–1990 C.E. We report validation statistics for the full reconstruction period 1000–1855 C.E. The skill of reconstructed temperature fields is evaluated using correlation, root-mean-squared error (RMSE), bias, and

amplitude ratio (reconstructed amplitude/true amplitude, as measured by standard deviation of the fields), all of which are measured relative to the “true” fields known from the complete and original ECHO-G surface temperatures. For each PPE, we also constructed area-weighted global mean annual temperature (GMAT) timeseries and their spectra for comparison with the true GMAT timeseries and spectrum.

3. CFR skill and uncertainty

Correlation fields for the four PPEs are shown in Figure 2; Table 1 gives pointwise-mean and area-weighted global mean summary statistics. As long as temperature information is part of the pseudoproxy construction, validation-period skill, bias, and resolved amplitude patterns for gridded fields (Figs. 2,3; SI Figs. 3,4) and GMAT timeseries (Table 1) are similar, albeit weaker, than those reported by *Smerdon et al.* [2011a] for a larger multiproxy PPE network (283 locations vs. the subset of 191 sites studied here, at which tree ring-width or maximum latewood density are found; Fig. 1), for which pseudoproxies were based solely on temperature (*Smerdon et al.* [2011a], SI Fig. S4, Table S4). The CFR derived solely from precipitation (Fig. 2b; Table 1) has relatively poor skill and uncertainty characteristics. These results partly arise from the number of coupled CCA patterns resolved for each PPE (SI Table 2; *Smerdon et al.* [2011b]): 8 (12,7) spatial patterns are calibrated in the T (T+P, VSL) PPEs, but only 4 patterns are calibrated in the P-based PPE.

For all experiments, and consistent with the results of *Smerdon et al.* [2011a], skill and amplitude recovery is generally highest in regions having dense observational coverage (compare Fig. 1 to Figs. 2 and 3). This is a result that has previously been shown across

reconstruction methodologies [Smerdon *et al.*, 2011a], but likely deteriorates as observing network density decreases over time [Wang *et al.*, 2014]. Not surprisingly, given the PPE target of surface temperature, the highest levels of regional skill are found in densely-observed parts of North America and Europe (Fig. 1) for the temperature-based PPE (Fig. 2). The T+P PPE produces almost as much skill in these regions (Fig. 2a,c). The VSL-based PPE (Fig. 2d) produces a similar but overall weaker skill pattern, but not as weak as that resulting from use of precipitation-based pseudoproxies (Fig. 2a). These relative performance characteristics are also found in the resolved amplitude (Fig. 3), RMSE (SI Fig. 3), mean bias (SI Fig. 4), and frequency-band-resolved variance (SI Figs. 5, 6) fields. Locally (SI Fig. 7), these results are consistent with limited correlation between gridpoint temperature and precipitation timeseries, the increased complexity of the pseudoproxy formulation in VSL, and the “lossy” nature of information via the transformation of T and P into simulated ring widths [Tolwinski-Ward *et al.*, 2014]. Although the T+P candidate linearly retains approximately 50% of the variance in the T candidate, and P contains very little information correlated with T, VS-Lite’s trivariate growth function is dependent on nonlinear and bivariate transformation of T and P into soil moisture, nonlinear scaling by insolation, and subannually-resolved, thresholded dependence on temperature and soil moisture [Tolwinski-Ward *et al.*, 2011a]. The net effect of these features appears to remove VSL further from a linear correlation with the target field at the pseudoproxy site than T+P, but not as far as P (SI Fig. 7).

We would expect that the common spatial variations in skill, error and bias, which are observed across all PPEs, to be in part generated from teleconnections between densely

and sparsely-observed regions, and therefore somewhat dependent on the climate model and simulation on which the PPE is based [Smerdon *et al.*, 2011a; Wang *et al.*, 2014]. With the climate model basis and the observing network as constants in this set of experiments, T, T+P and VSL-based PPEs indicate some skill in regions without dense observations – for instance, the Indo-Pacific region and the tropics more generally – while the P-based PPE doesn't show skill in these same regions. This suggests that remote CFR skill derives from recovery of large-scale patterns by the reconstruction methodology, and from resolved large-scale patterns in SAT in particular. Given that there are large-scale patterns in CFR skill, bias, RMSE and amplitude recovery for all PPEs (Figs. 2,3; SI Figs. 3–6), however, the results suggest that even precipitation or moisture-dominated paleo-observations may produce validated non-local CFR skill, perhaps because of the covariance of temperature and precipitation patterns in the climate system, here as represented in a climate simulation. Examination of the amplitude patterns over lowpass and highpass frequency bands (SI Figs. 5,6) suggests that much of the similarity arises from resolution of low-frequency ($<1/20$ cyc/yr) signals, with only limited skill arising from the correlation of T+P, P, and VSL pseudoproxies with local T (SI Fig. 7).

Spectral analysis of detrended reconstructed GMAT for the validation interval 1000–1855 C.E., and comparison to ECHO-G 'truth' (black line) suggests that PPE skill in the spectral domain is a function of frequency. All PPEs underestimate the low-frequency variance and overestimate variance at frequencies higher than about 0.4 cyc/yr (Fig. 4), thereby underestimating the true 'redness' of the ECHO-G power spectrum [von Storch *et al.*, 2004, 2009]. However, for frequencies less than about $1/5$ cyc/yr, spectra for T,

T+P and VSL-based CFRs are indistinguishable within uncertainty (Fig. 4), which suggests that partial temperature responders, even if realistically nonlinear and multivariate, may nevertheless produce accurate spectra of reconstructed GMAT [Franke *et al.*, 2013]. Examination of the resolved amplitude patterns for lowpass and highpass frequency bands (SI Figs. 5,6) suggests that much of the skill in the T, T+P and VSL-based PPEs arises from resolution of low-frequency ($<1/20$ cyc/yr) signals. In contrast, but consistent with the results of Franke *et al.* [2013] for precipitation proxies, the resolved variance in the GMAT reconstruction for the P-based CFR is significantly smaller than truth for this frequency interval, and fails to resolve low ($<1/5$ cyc/yr) frequencies (Fig 4; SI Figs. 5,6). A similar result is observed for the spectrum of GMAT reconstructed from pseudoproxies constructed using the soil moisture model [Huang *et al.*, 1996] embedded within VS-Lite, but with even less total resolved variance than observed for the P-based CFR (results not shown). Together these results suggest that skill from mixed temperature and moisture responders for reconstruction of the SAT field may be biased in amplitude as a function of both frequency and observing network. Regional-level analysis of actual CFR results, especially outside of densely observed regions in space and/or time (e.g. Mann *et al.* [2009]) should therefore proceed with caution and additional validation prior to interpretation [von Storch *et al.*, 2009; Smerdon *et al.*, 2011a; Franke *et al.*, 2013; Wang *et al.*, 2014].

4. Conclusion

PPE results using pseudoproxies that realistically (if crudely) mimic the multivariate and nonlinear response of tree-ring width variations to environmental forcing suggest that CFRs based largely on tree-ring widths should have spatial skill and error similar in pat-

tern but substantially poorer in quality than CFRs produced from idealized univariate linear temperature responders. However, as long as the responder contains some temperature information, GMAT reconstructions derived from such indicators should accurately resolve the low frequency spectral characteristics of the true climate. More generally, the results presented here may also have implications for the expected skill and uncertainty for surface temperature CFRs based on other proxy systems having mixed moisture and temperature responses, such as the stable isotopic composition of various marine and terrestrial archives, or for multiproxy-based GMAT reconstructions based on collections of paleodata with combined multivariate responses (e.g. *Mann et al.* [1998]; *Cook et al.* [1999]; *Evans et al.* [2002]; *Mann* [2002]; *Cook et al.* [2004]; *Mann et al.* [2008]). These results may not take into account many systematic observational uncertainties in actual paleodata [*Evans et al.*, 2013] and CFR construction [*Emile-Geay et al.*, 2013], such as chronological uncertainty (e.g. *Anchukaitis and Tierney* [2012]; *Comboul et al.* [2014]), response seasonality (e.g. *St. George and Ault* [2014]), or standardization or signal filtering choices (e.g. *Melvin and Briffa* [2008]). We also caution that these results may be sensitive to the choice of realistic proxy system model parameters (e.g. *Anchukaitis et al.* [2012]). In the real world, the low-frequency signals in the paleo-observations themselves are difficult to validate against direct observations; doing so for spatially-resolved CFRs, given the results presented here, is only more so. Nevertheless, a strategy that includes further improvements in the observing network extent, reconstruction methods, observational SNR properties, and interpretive modeling of tree-ring width and other pa-

leodata should continue to improve the quality and utility of real-world CFRs for analysis of regional-to-global scale surface temperature variations.

Acknowledgments. We are grateful to two anonymous reviewers whose comments helped improve this paper. Work was supported by grants NSF/ATM0902715 to MNE and NSF/ATM0902436 to JES and AK; MNE and JES also acknowledge support from NOAA grant NA10OAR431037. The pseudoproxies used in this study will be made available for further testing across different methodological applications, at <http://one.geol.umd.edu/www/data/> and the NOAA/National Climatic Data Center (<http://www.ncdc.noaa.gov>). Code for both VS-Lite and its environmental parameter estimation is available for download from the NOAA/National Climatic Data Center (<ftp://ftp.ncdc.noaa.gov/pub/data/paleo/softlib/vs-lite/>)

References

- Anchukaitis, K. J., and J. E. Tierney (2012), Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, *Clim. Dyn.*, pp. 1423–1441, doi: 10.1007/s00382-012-1483-0.
- Anchukaitis, K. J., M. N. Evans, A. Kaplan, E. A. Vaganov, H. D. Grissino-Mayer, M. K. Hughes, 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, L04,705, doi: 10.1029/2005GL025050.
- Anchukaitis, K. J., P. Breitenmoser, K. R. Briffa, A. Buchwal, U. Buntgen, E. Cook, R. D'Arrigo, J. Esper, M. Evans, D. Frank, B. Gunnarson, M. Hughes, A. Kirilyanov,

C. Korner, P. Krusic, B. Luckman, T. M. Melvin, M. Salzer, A. Shashkin, C. Timmreck, E. Vaganov, and R. Wilson (2012), Tree rings and volcanic cooling, *Nature Geoscience*, *5*, 836–837, doi:10.1038/ngeo1645.

Babst, F., B. Poulter, V. Trouet, K. Tan, B. Neuwirth, R. Wilson, M. Carrer, M. Grabner, W. Tegel, T. Levanic, M. Panayotov, C. Urbinati, O. Bouriaud, P. Ciais, and D. Frank (2013), Site- and species-specific responses of forest growth to climate across the European continent, *Glob. Ecol. Biogeogr.*, *22*, 706–717, doi: 10.1111/geb.12023.

Breitenmoser, P., S. Brönnimann, and S. Frank (2014), Forward modeling of tree-ring width and comparison with a global network of tree-ring chronologies, *Clim. Past.*, *10*, 437–449, doi:10.5194/cpd-10-437-2014.

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*, D12,106, doi: 10.1029/2005JD006548.

Christiansen, B. (2011), Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?, *J. Clim.*, *24*, 674–692.

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

Cook, E. R., and L. Kairiukstis (Eds.) (1990), *Methods of Dendrochronology: Applications in the environmental sciences*, Kluwer, Boston, 394pp.

Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland (1999), Drought reconstructions for the continental United States, *J. Clim.*, *12*, 1145–1162.

Cook, E. R., C. A. Woodhouse, C. M. Eakin, D. M. Meko, and D. Stahle (2004), Long-term aridity changes in the western United States, *Science*, *306*(5698), 1015–1018.

Emile-Geay, J., K. M. Cobb, M. E. Mann, and A. T. Wittenberg (2013), Estimating central equatorial Pacific SST variability over the past millennium. Part I: Methodology and validation, *J. Clim.*, *26*, 2302–2328, doi: <http://dx.doi.org/10.1175/JCLI-D-11-00510.1>.

Evans, M. N., A. Kaplan, and M. A. Cane (1998), Optimal sites for coral-based reconstruction of sea surface temperature, *Paleoceanography*, *13*, 502–516.

Evans, M. N., A. Kaplan, M. A. Cane, and R. Villalba (2001), Globality and optimality in climate field reconstructions from proxy data, in *Interhemispheric Climate Linkages*, edited by V. Markgraf, pp. 53–72, Cambridge University Press, Cambridge, UK.

Evans, M. N., A. Kaplan, and M. A. Cane (2002), Pacific sea surface temperature field reconstruction from coral $\delta^{18}\text{O}$ data using reduced space objective analysis, *Paleoceanography*, *17*, 10.1029/2000PA000,590.

Evans, M. N., A. Kaplan, J. E. Smerdon, and M. K. Hughes (2006a), The spectral, multivariate and non-stationary nature of the paleoproxy-climate connection: Implications for paleoclimate reconstructions, in *Transactions, EOS*, vol. 87(52) Fall Meet. Suppl., pp. Abstract U11B–02.

Evans, M. N., B. K. Reichert, A. Kaplan, K. J. Anchukaitis, E. A. Vaganov, M. K. Hughes, and M. A. Cane (2006b), A forward modeling approach to paleoclimatic interpretation

of tree-ring data, *J. Geophys. Res.*, *111*, G03,008, doi:10.1029/2006JG000166.

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

Franke, J., D. C. Frank, C. C. Raible, J. Esper, and S. Brönniman (2013), Spectral biases in tree-ring climate proxies, *Nature Climate Change*, *3*, 360–364, doi:10.1038/nclimate1816.

Fritts, H. C. (1976), *Tree Rings and Climate*, Academic Press, New York.

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*, L01,703, doi:10.1029/2005GL024693.

González-Rouco, J. F., H. Beltrami, E. Zorita, and M. B. Stevens (2009), Borehole climatology: a discussion based on contributions from climate modeling, *Clim. Past*, *5*, 97–127, doi:10.5194/cp-5-97-2009.

Graham, N. E., and E. R. Wahl (2011), Paleoclimate Reconstruction Challenge: Available for participation, *PAGES/CLIVAR Newsl.*, *19(2)*, 71–72.

Huang, J., H. M. van den Dool, and K. P. Georgakakos (1996), Analysis of model-calculated soil moisture over the United States (1931-1993) and applications to long-range temperature forecasts, *J. Clim.*, *9*, 1350–1362.

Legutke, S., and R. Voss (1999), The Hamburg atmosphere-ocean coupled circulation model ECHO-G, DKRZ, *Tech. rep.*, 18, Dtsch. Klimarechenzentrum.

Mann, M. E. (2002), The value of multiple proxies, *Science*, *297*, 1481–1482.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1998), Global temperature patterns over the past five centuries: Implications for anthropogenic and natural forcing of climate, *Nature*, *392*, 779–787.

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

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. Clim.*, *18*(20), 4097–4107.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, and S. Rutherford (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proc. Nat. Acad. Sci.*, *105*(36), 13,252–13,257, doi:10.1073/pnas.0805721105.

Mann, Z., M. E. Zhang, S. Rutherford, M. K. Bradley, R. S. Hughes, D. Shindell, C. Ammann, G. Faluvegi, and F. Ni (2009), Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*, 1256–1260, doi:10.1126/science.1177303.

Melvin, T. M., and K. R. Briffa (2008), A “signal-free” approach to dendroclimatic standardisation, *Dendrochronologia*, *26*, 71–86, doi: 10.1016/j.dendro.2007.12.001.

Min, S., S. Legutke, A. Hense, and W. Kwon (2005a), Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G - I. Near surface temperature, precipitation and sea level pressure, *Tellus*, *57A*, 605–621.

Min, S., S. Legutke, A. Hense, and W. Kwon (2005b), Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G - II. El Niño Southern Oscillation and North Atlantic Oscillation, *Tellus*, 57A, 622–640.

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

Smerdon, J. E., A. Kaplan, E. Zorita, J. F. González-Rouco, and M. N. Evans (2011a), Spatial performance of four climate field reconstruction methods targeting the Common Era, *Geophys. Res. Lett.*, 38, L11,705, doi: 10.1029/2011GL047372.

Smerdon, J. E., A. Kaplan, D. Chang, and M. N. Evans (2011b), A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium, *J. Clim.*, 24, 1284–1309, doi: 10.1175/2010JCL14110.1.

St. George, S., and T. R. Ault (2014), The imprint of climate within Northern Hemisphere trees, *Quat. Sci. Rev.*, 89, 1–4, doi: 10.1016/j.quascirev.2014.007.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the Experiment Design, *Bull. Amer. Meteor. Soc.*, 93, 485–498, doi: 10.1016/BAMS-D-11-00094.1.

Tolwinski-Ward, S. E., M. N. Evans, M. K. Hughes, and K. J. Anchukaitis (2011a), An efficient forward model of the climatic controls on interannual variation in tree-ring width, *Clim. Dyn.*, 36, 2419–2439, doi: 10.1007/s00382-010-0945-5.

Tolwinski-Ward, S. E., M. N. Evans, M. K. Hughes, and K. J. Anchukaitis (2011b), Erratum to: An efficient forward model of the climatic controls on interannual variation in tree-ring width, *Clim. Dyn.*, doi: 10.1007/s00382-011-1062-9.

Tolwinski-Ward, S. E., K. J. Anchukaitis, and M. N. Evans (2013), Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width, *Clim. Past.*, *9*, 1481–1493, doi: 10.5194/cp-9-1481-2013.

Tolwinski-Ward, S. E., M. P. Tingley, M. N. Evans, D. W. Nychka, and M. K. Hughes (2014), Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response, *Clim. Dyn.*, doi: 10.1007/s00382-014-2139.

Vaganov, E. A., M. K. Hughes, A. V. Kirilyanov, F. H. Schweingruber, and P. P. Silkin (1999), Influence of snowfall and melt timing on tree growth in subarctic Eurasia, *Nature*, *400*, 149–151.

Vaganov, E. A., M. K. Hughes, and A. V. Shashkin (2006), *Growth dynamics of conifer tree rings: Images of past and future environments*, Ecol. Studies 183, Springer-Verlag, Berlin/Heidelberg/New York.

Vaganov, E. A., K. J. Anchukaitis, and M. N. Evans (2011), How well understood are the processes that create dendroclimatic records? A mechanistic model of the climatic control on conifer tree-ring growth dynamics, in *Tree Rings and Climate: Sharpening the Focus, Developments in Environmental Research*, vol. 11, edited by M. K. Hughes, T. W. Swetnam, and H. F. Diaz, chap. 3, pp. 37–75, Springer, New York.

von Storch, H. (2010), Climate models and modeling: an editorial essay, *WIREs Clim. Change*, *1*, 305–310, doi: 10.1002/wcc.12.

von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, J. F. González-Rouco, and S. F. B. Tett (2004), Reconstructing past climate from noisy data, *Science*, *306*, 679–682.

von Storch, H., E. Zorita, and J. 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, doi: 10.1007/s00531-008-0349-5.

Wang, J., J. Emile-Geay, D. Guillot, J. E. Smerdon, and B. Rajaratnam (2014), Evaluating climate field reconstruction techniques using improved emulation of real-world conditions, *Clim. Past.*, *10*, 1–19, doi:10.5194/cp-10-1-2014.

Zorita, E., J. F. González-Rouco, and H. von Storch (1995), Inconsistencies at the interface of climate impact studies and global climate research, *Meteorol. Zeitschrift*, *4*, 71–80.

Zorita, E., J. F. González-Rouco, H. von Storch, J. P. Montávez, and F. Valero (2005), Natural and anthropogenic modes of surface temperature variations in the last thousand years, *Geophys. Res. Lett.*, *32*, L01,703, doi:10.1029/2004GL021563.

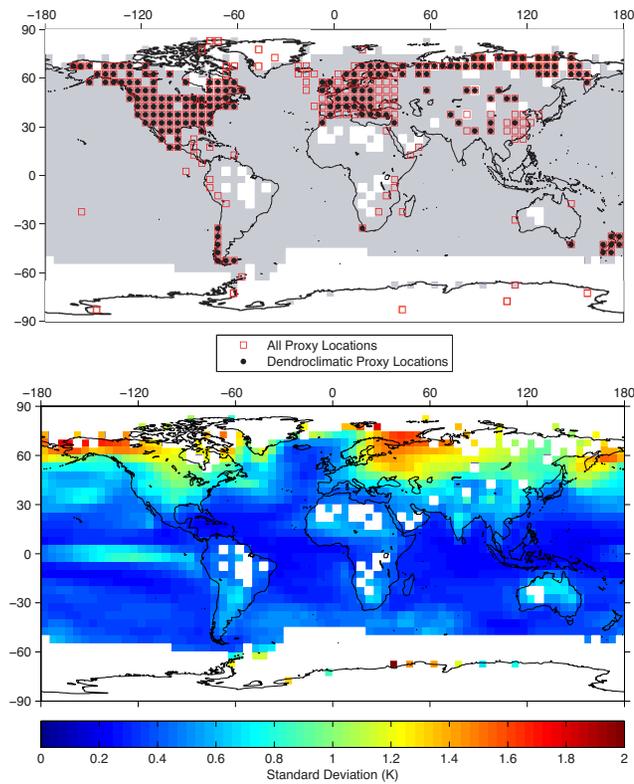


Figure 1. Top: Map of all grid cells containing climate proxy data in the *Mann et al.* [2008] network on the 5x5 degree grid (open red squares) and those that just contain dendroclimatic data (closed black circles). Grey shading indicates reconstruction target field masked by actual surface air temperature availability [*Brohan et al.*, 2006]. Bottom: Map of the standard deviation of mean annual surface temperature from the “ERIK-2” ECHO-G simulation [*González-Rouco et al.*, 2006, 2009], based on the reconstruction validation interval 1000–1855 C.E.

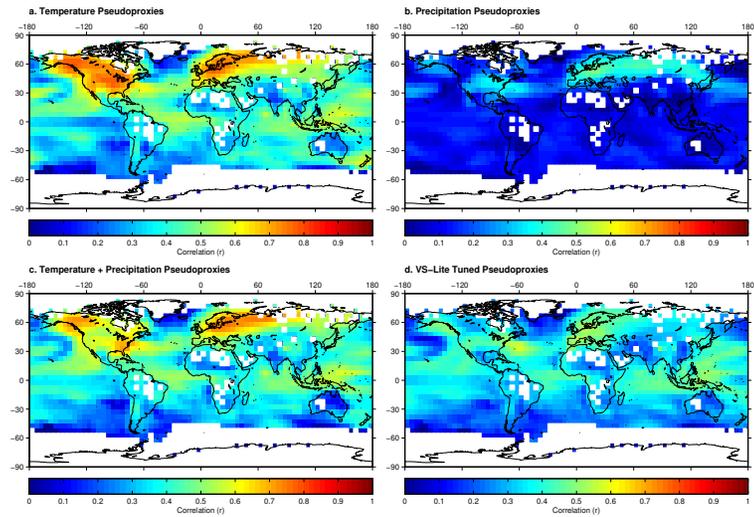


Figure 2. CFR correlation with true field for: (a) Temperature-only (T), (b) precipitation-only (P), (c) temperature + precipitation (T+P), and (d) VS-Lite pseudoproxy-based reconstruction of mean annual temperature (VSL).

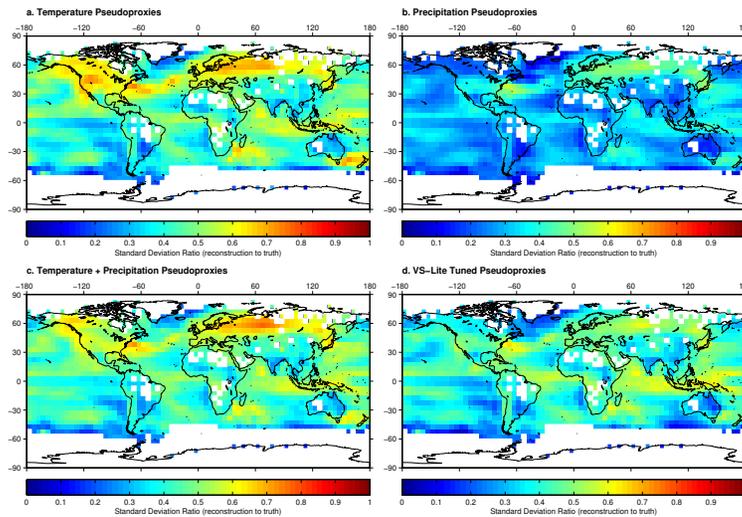


Figure 3. CFR standard deviation ratios ($\sigma_{CFR}/\sigma_{truth}$) for: (a) Temperature-only (T), (b) precipitation-only (P), (c) temperature + precipitation (T+P), and (d) VS-Lite (VSL) pseudoproxy-based reconstruction of mean annual temperature.

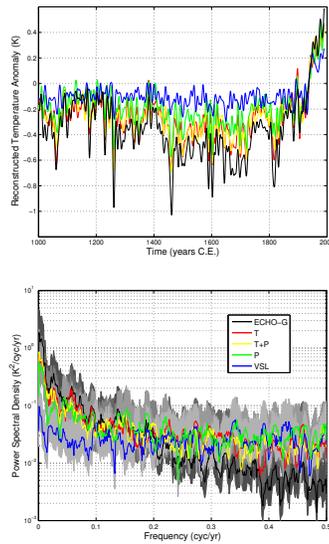


Figure 4. Top: Timeseries of ECHO-G global mean air temperature (GMAT) and reconstructed GMAT derived from temperature (T), temperature+precipitation (T+P), precipitation (P), and VS-Lite (VSL) pseudoproxies. Series have been smoothed with a decadal-timescale filter for clarity. Bottom: Multitaper power spectra for: T, P+T, P, and VSL GMAT for detrended timeseries for the validation interval 1000–1855 C.E. Dark grey shading are 95% amplitude confidence intervals for the ECHO-G spectrum, and light grey shading are 95% confidence intervals for the other spectral estimates.

Table 1. Summary statistics for PPE experiments

PPE ^a	r(GMAT) ^b	r(Field) ^c	$\sigma_{CFR}/\sigma_{true}$ ^d	Bias ^e	RMSE ^f
T	0.78(<0.0001)	0.41	0.45	0.11	0.53
P	0.26(0.13)	0.14	0.30	0.29	0.67
T+P	0.77(<0.0001)	0.39	0.45	0.12	0.54
VSL	0.65(0.0003)	0.33	0.41	0.19	0.60

^a Pseudoproxy experiments, as described in Section 2

^b Correlation between reconstructed and true area-weighted GMAT; p-values for GMAT correlations in parentheses given estimated effective degrees of freedom = 21, 21, 21 and 24 for T, P, T+P and VSL-based area-weighted GMAT from the respective PPEs.

^c Point-wise correlation between reconstructed and true surface temperature fields

^d mean gridpoint amplitude ratio, reconstructed/true

^e mean gridpoint bias, reconstructed-true

^f mean gridpoint root-mean-square error, reconstructed-true