Supporting Information

Evans, Smerdon, Kaplan, Tolwinski-Ward, and González-Rouco, *Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors*, submitted to Geophys. Res. Lett.

1 On the construction of VS-Lite pseudoproxies

VS-Lite is a simple but realistic model for simulating tree-ring width variations as a thresholded minimization function of monthly temperature and soil moisture, scaled by insolation and integrated over a prescribed pre-season and growing season to produce an annual (growing season) ring-width estimate (*Tolwinski-Ward et al.*, 2011). Soil moisture, in turn, is estimated within VS-Lite via the Climate Prediction Center (CPC) Leaky Bucket model (*Huang et al.*, 1996) and input monthly temperature and precipitation. The strengths and weaknesses of VS-Lite as a descriptor of actual tree-ring width data have been established (*Tolwinski-Ward et al.*, 2011), and an algorithm for tuning the model for specific applications has been published (*Tolwinski-Ward et al.*, 2013). Code for both the model and parameter estimation is available for download from the NOAA/National Climatic Data Center (ftp://ftp.ncdc.noaa.gov/pub/data/paleo/softlib/vs-lite/). We used VS-Lite version 2.5 to produce results for this manuscript. The parameter choices used to develop the VS-Lite pseudoproxies are described below.

1.1 VS-Lite parameter choices

VS-Lite has 4 environmental parameters; 8 additional parameters must be specified to run the embedded CPC/leaky bucket moisture model (*Huang et al.*, 1996; *Tolwinski-Ward et al.*, 2011). An important part of this study was therefore to determine whether the results are sensitive to parameter estimation. Previous work had suggested that results are not sensitive to specification of the CPC/leaky bucket moisture model parameters (*Tolwinski-Ward et al.*, 2011), so we performed three classes of environmental parameter estimation experiments: (1) environmental parameters set to ± 2 standard deviations above and below the monthly mean of gridpoint temperature and precipitation in the calibration period; (2) default parameters held constant at all locations (*Tolwinski-Ward et al.*, 2011); and (3) environmental parameters tuned using the method of *Tolwinski-Ward et al.* (2013) to local warm season (NH: May–Oct; SH: Nov-Apr) monthly temperature and precipitation over the CFR calibration period (1856–1990). For all three tuning experiments, we examined basic statistics (probability density function, mean, median, mode, skewness, number of missing (non-growth) values) of the resulting VSL simulations before adding random noise at SNR=0.5 to form the VSL pseudoproxies. We also compared the parameterized VS-Lite simulations to the original input T and P series at each site and by comparing their first EOF and principal component to that of input T and P series.

The median 1000–1855 C.E. correlation across all sites of the three parameter specification experiments was highly significant (median p<0.001), but with shared variance between climatological and default simulations, at $\approx 10\%$, smaller than shared variance between default and tuned simulations ($\approx 59\%$) or between tuned and climatological simulations
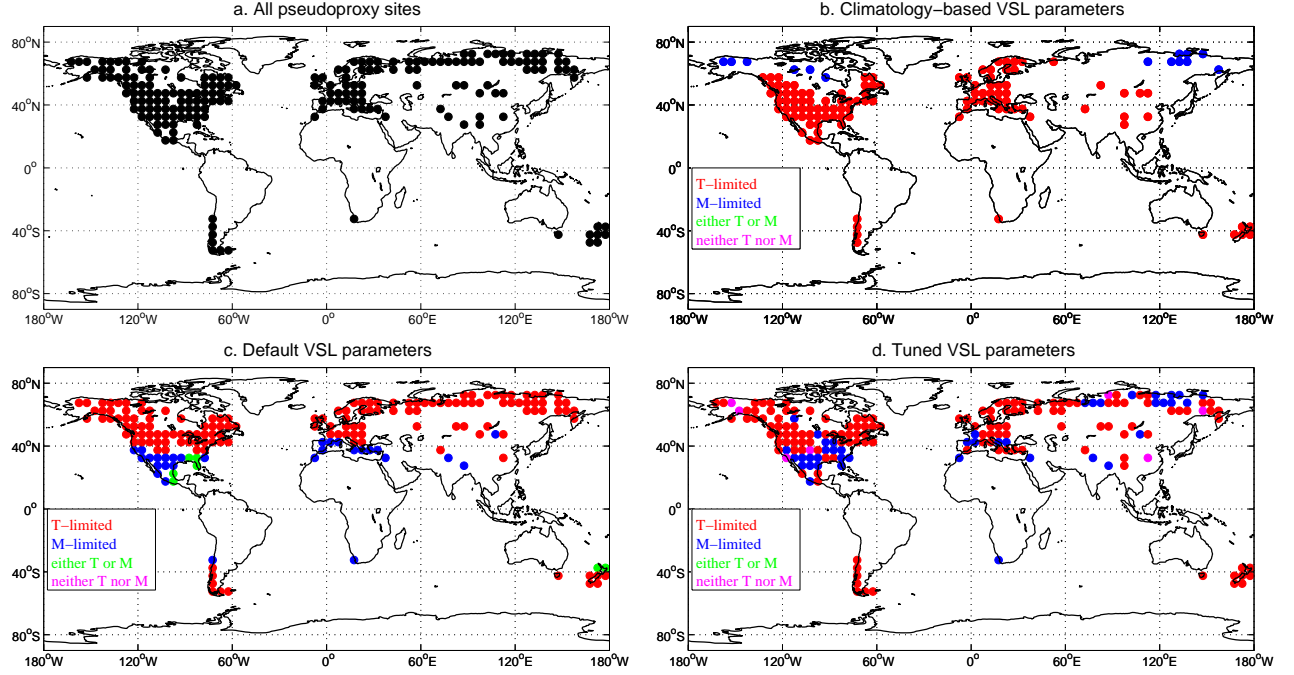
Figure 1: Simulated growth limitation of VS-Lite pseudoproxies by location. (a): All sites in the pseudoproxy network. (b): With parameters tuned to the $\pm 2\sigma$ values of T and calculated M within the NH (SH) Apr-Sep (Oct-Mar) growing season, moisture-limited (blue), temperature-limited (red), either moisture or temperature (green), and neither moisture nor temperature limited (magenta) sites are evaluated as such if $>95\%$ of simulated years at the site are limited by the respective growth function. Note that by this definition not all sites will fall into one of these categories because of the $>95\%$ probability threshold. (c): as in (b), except for default VS-Lite parameters (*Tolwinski-Ward et al.*, 2011). (d): as in (b), except for tuned VS-Lite parameters *Tolwinski-Ward et al.* (2013).

(≈39%). In all experiment classes, simulated ring-width series produced some simulations with highly skewed and/or apparently threshholded responses that are inconsistent with actual observational ring width statistics (*Mann et al.*, 2008). Somewhat surprisingly, probability density functions for simulations using default parameters had smaller skewness and more Gaussian structure than tuned or climatological simulations. In the default and tuned parameter experiments, we were able to obtain spatially realistic patterns of sensitivity to moisture or temperature (Fig 1c,d), with primarily temperature dependent simulations in temperate and subpolar regions, and primarily moisture dependent simulations for semiarid subtropical regions. However, in the climatologically-tuned simulations, we were unable to obtain spatially realistic patterns of sensitivity to moisture or temperature (Fig. 1b). We also found that the climatological specification of parameters produced about 53% T_1 estimates below freezing and about 21% M_2 estimates above a volume/volume soil moisture of 0.6, both of which are implausible. These results suggest that appropriate parameterization is a key component of the PPE design. Based on these results, we decided that the tuned or default parameters produced the most realistic and therefore most appropriate simulations for the PPEs presented in this study. Median default and tuned parameters are given in Table 1. Once white noise was added to the simulations to produce SNR=0.5 pseudoproxies, we found that the CFRs produced from default and tuned parameters produced statistics very similar to one another (Fig. 2).

Parameter	Default	Tuned 25th percentile	Tuned 50th percentile	Tuned 75th percentile
$T_1(^oC)$	4.2	2.7	4.4	6.2
$T_2(^oC)$	10.6	11.4	12.9	16.5
$M_1(v/v)$	0.028	0.031	0.069	0.089
$M_2(v/v)$	0.34	0.32	0.43	0.47

Table 1: Comparison of default(*Tolwinski-Ward et al.*, 2011) and tuned parameter sets. For N=191 sets of 4 tuned parameters, the 25th, 50th and 75th percentile values across all N parameter estimates is shown.

1.2 Uncertainty specification

In addition, specification of uncertainty in the VS-Lite inputs is less straightforward than for direct climatic measures such as temperature or precipitation. While the latter may be used to create pseudoproxies by adding random variance to extracted climatic timeseries, the former realistically requires uncertainty in both the climatic timeseries input as well as in the proxy system model (VS-Lite) itself. Since we wished to create pseudoproxies differing primarily in their mapping from surface climate conditions, we first created noise-free VS-Lite simulations, analogous to extracting noise-free temperature timeseries from model simulations in prior PPE work (*Smerdon*, 2012). Let us denote by X these noise-free timeseries (T, P, T+P, VSL), extracted or calculated from the ECHO-G output, which are standardized to zero mean and unit variance. X is the "signal". Let $\sigma_\epsilon^2 < 1$ be the variance of independent and normally-distributed random timeseries ϵ ("noise"). Then the pseudoproxy D with zero mean, unit variance, and SNR by standard deviation $\sqrt{1 - \sigma_\epsilon^2}/\sigma_\epsilon$
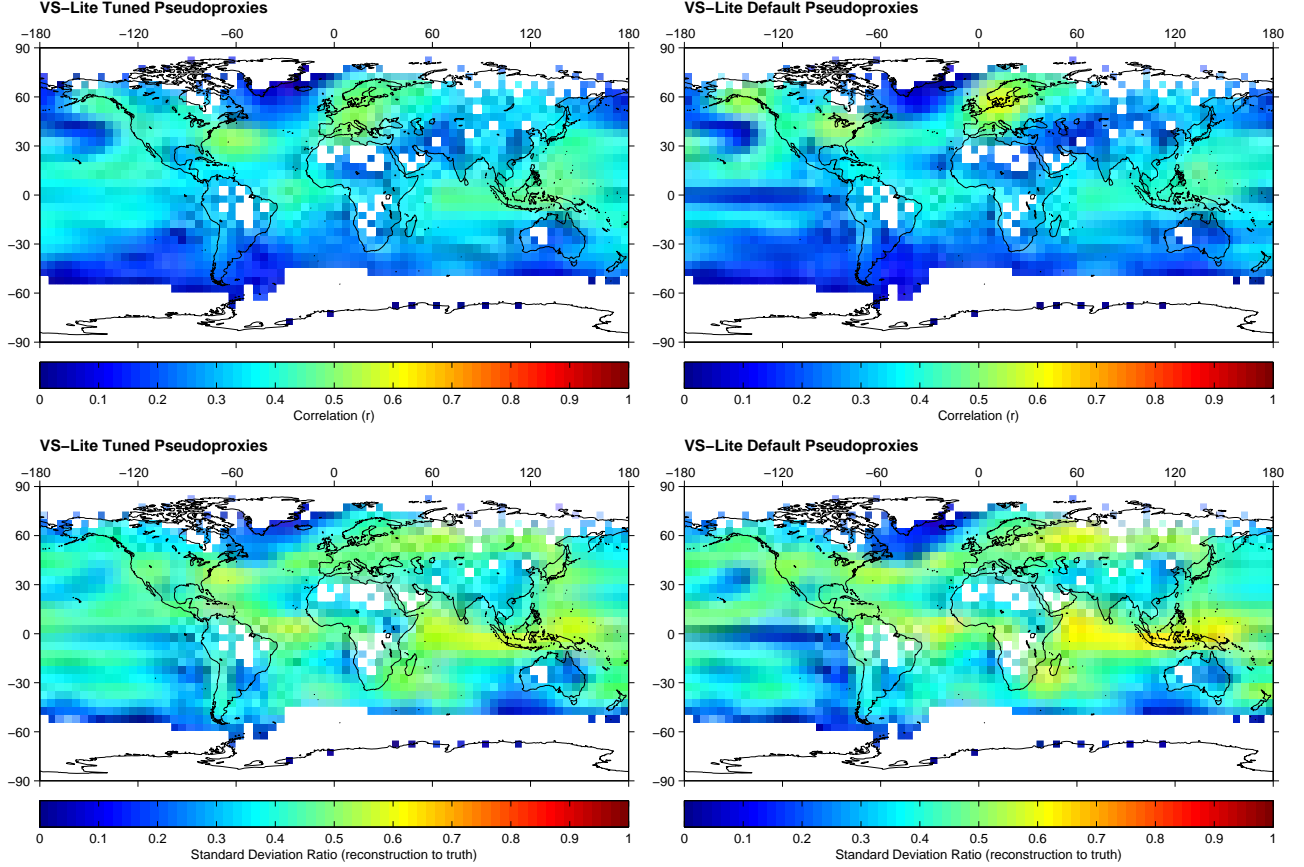
Figure 2: Sensitivity of results to plausible VS-Lite parameter sets. Tuned (left panels) vs. untuned (right panels) VSL-pseudoproxy derived CFR skill (upper row) and standard deviation ratio maps (lower row).

73 is (*Tolwinski-Ward et al.*, 2013):

$$D = \sqrt{(1 - \sigma_\epsilon^2)}X + \epsilon$$

74 For $\sigma_\epsilon = \sqrt{4/5}$, SNR by standard deviation is $\sqrt{1 - \sigma_\epsilon^2}/\sigma_\epsilon = \sqrt{1 - 4/5}/\sqrt{4/5} = 0.5$.

1.3 Observing network

76 A third source of uncertainty is in the observational network. Dendrochronological indicators
77 (tree ring width or maximum latewood density) are found in only 191 of the 283 pseudoproxy
78 gridpoints mapped from (*Mann et al.*, 2008) at 5° x 5° resolution (Fig. 1). We therefore had
79 to make a choice: develop climate field reconstructions based on a realistic dendrochronolog-
80 ical network, or preserve a direct network comparison to the results of *Smerdon et al.* (2011)
81 and others. We made the former choice, recognizing that results presented in this work are
82 less skillful than and can only be qualitatively compared to prior work on the *Mann et al.*
83 (2008) network.

4

2 Choice of last millennium simulation

We have used the ECHO-G "ERIK2" last millennium simulation (*González-Rouco et al.*, 2006, 2009) as the basis of our PPE design. There are now multiple last-millennium simulations available from fully coupled GCMs (*Taylor et al.*, 2012), all of which can serve as the basis of a given PPE framework. The ECHO-G ERIK2 simulation is chosen here in part for continuity, given that *Smerdon et al.* (2011) used the ECHO-G and CCSM 1.4 last-millennium simulations to perform similar PPEs. Although the earlier results are not directly comparable because we have emulated a subset of the *Mann et al.* (2008) proxy network that reflects the availability of only dendroclimatic records, the qualitatively similar PPE designs allow meaningful comparisons between the current results using the ECHO-G simulation and those from *Smerdon et al.* (2011). The CCSM1.4 last-millennium simulation was not used herein because the simulated precipitation fields, which are necessary for the current PPE construction, were not available publicly. Moreover, PPEs using the ECHO-G last-millennium simulation in *Smerdon et al.* (2011) yielded reconstructions that generally validated better than those constructed from the CCSM1.4 simulation, including performance associated with the Northern Hemisphere mean. This improved performance baseline allows for greater separation between reconstruction results based on the more diverse pseudoproxy networks that were constructed for this study.

3 Choice of CFR algorithm

The CCA method was used exclusively for the PPEs performed herein. Multiple CFR methods have now been tested and compared in pseudoproxy contexts (*Smerdon*, 2012; *Smerdon et al.*, 2011; *Wang et al.*, 2014; *Tingley et al.*, 2012). Although small differences exist among methods, the current collection of PPEs reported in the literature indicate that the field performance of available CFR methods, including CCA, is similar. We therefore have employed a single CFR method for our purposes, because our focus is on the effects of differences in the pseudoproxy network design, not on methodological differences across multiple CFR techniques. All of the pseudoproxy constructions used in this study will be made available for further testing across different methodological applications, at http://one.geol.umd.edu/www/data/.

4 Random and systematic error in reconstructed fields
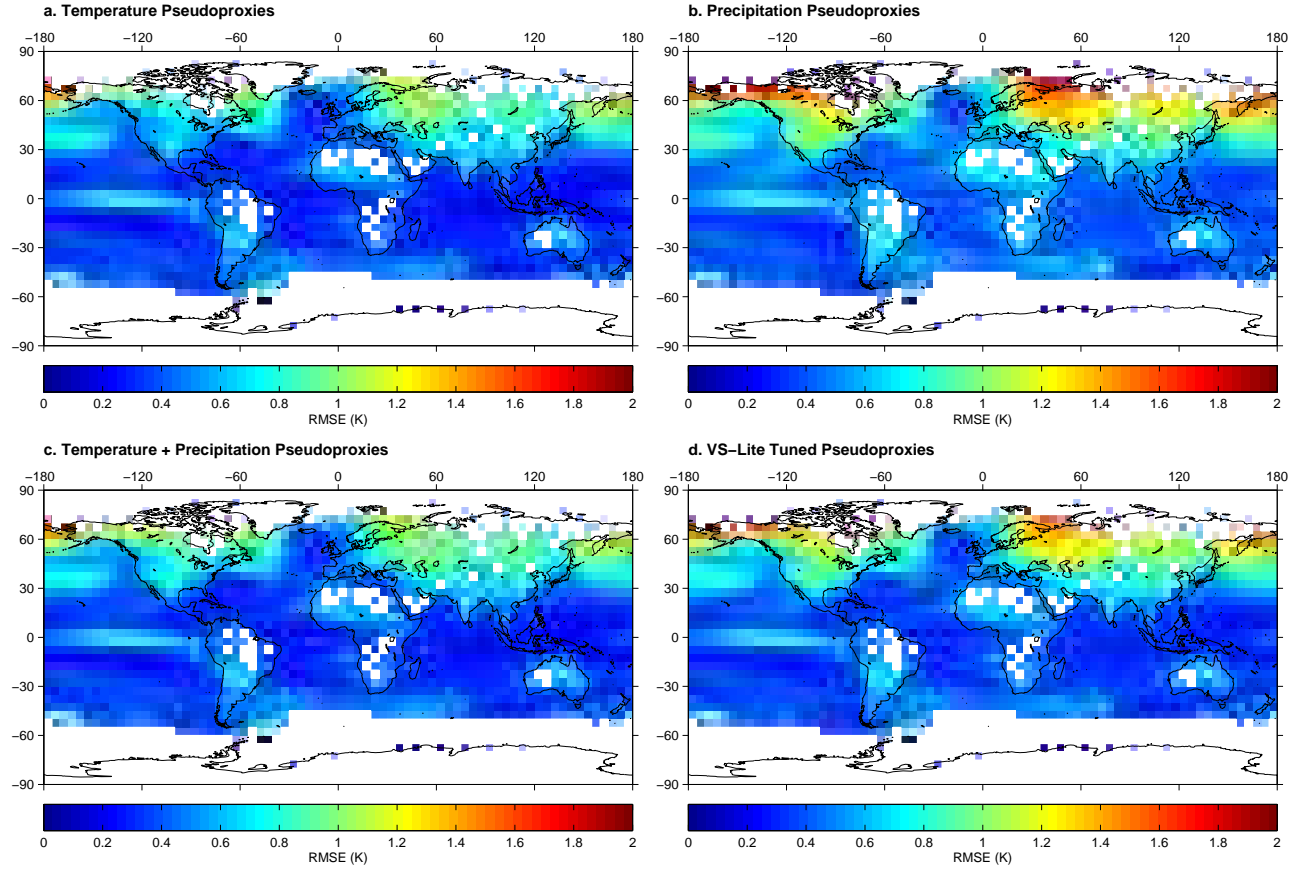
4.1 Root-mean-square error (RMSE)

Figure 3: (Supplemental Materials) Root-mean-square error (RMSE) fields for: (a) Temperature-only, (b) precipitation-only, (c) precipitation/temperature, and (d) VS-Lite pseudoproxy-based reconstruction of mean annual temperature.
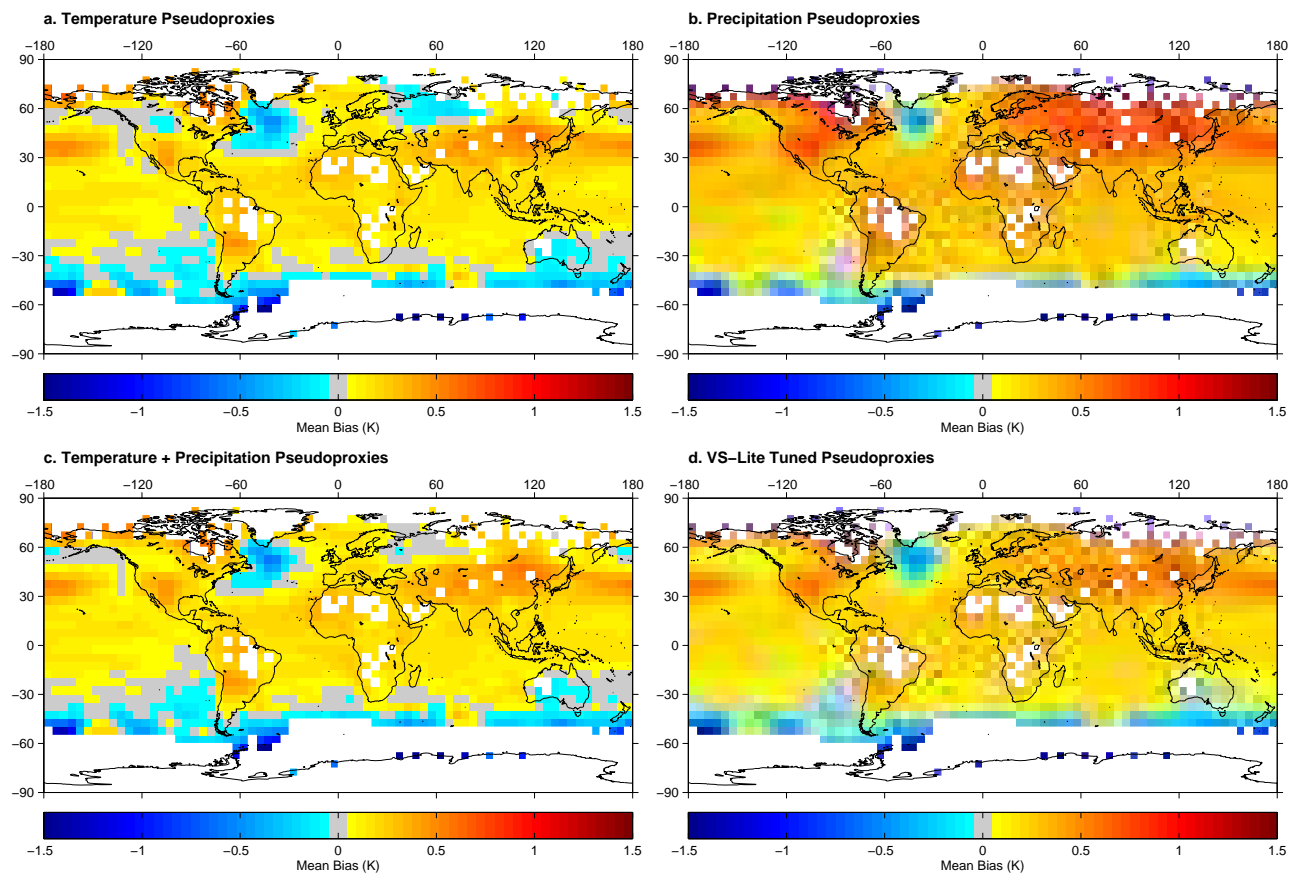
4.2   Mean bias

Figure 4: (Supplemental Materials) Mean bias fields for: (a) Temperature-only, (b) precipitation-only, (c) precipitation/temperature, and (d) VS-Lite pseudoproxy-based reconstruction of mean annual temperature.

4.3   Resolved fraction of high-frequency amplitude variance

a. Temperature Pseudoproxies

b. Precipitation Pseudoproxies

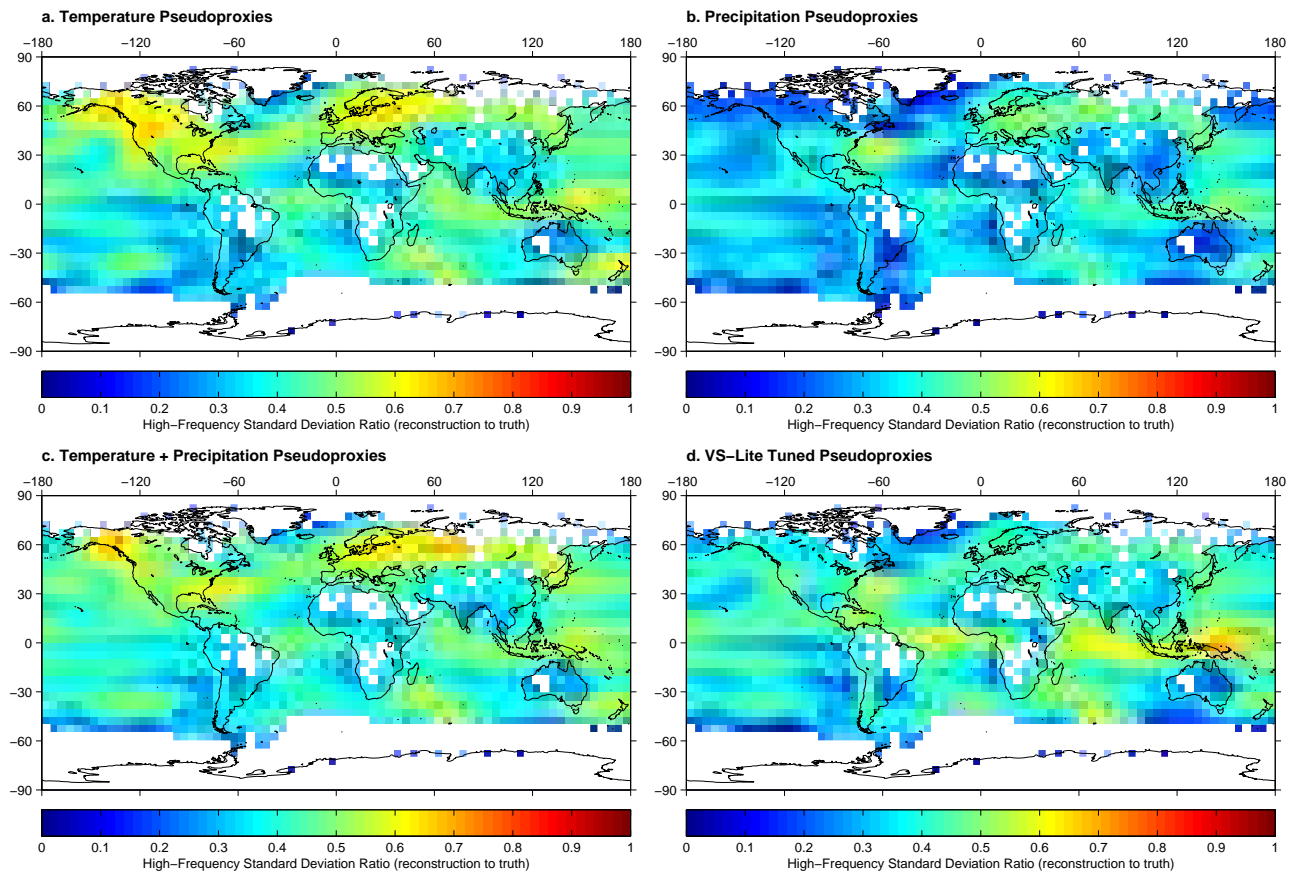**c. Temperature + Precipitation Pseudoproxies**

d. VS–Lite Tuned Pseudoproxies

Figure 5: (Supplemental Materials) Highpass amplitude ratio ($\sigma_{CFR}/\sigma_{truth}$, f>1/20y) fields for: (a) Temperature-only, (b) precipitation-only, (c) precipitation/temperature, and (d) VS-Lite pseudoproxy-based reconstruction of mean annual temperature.

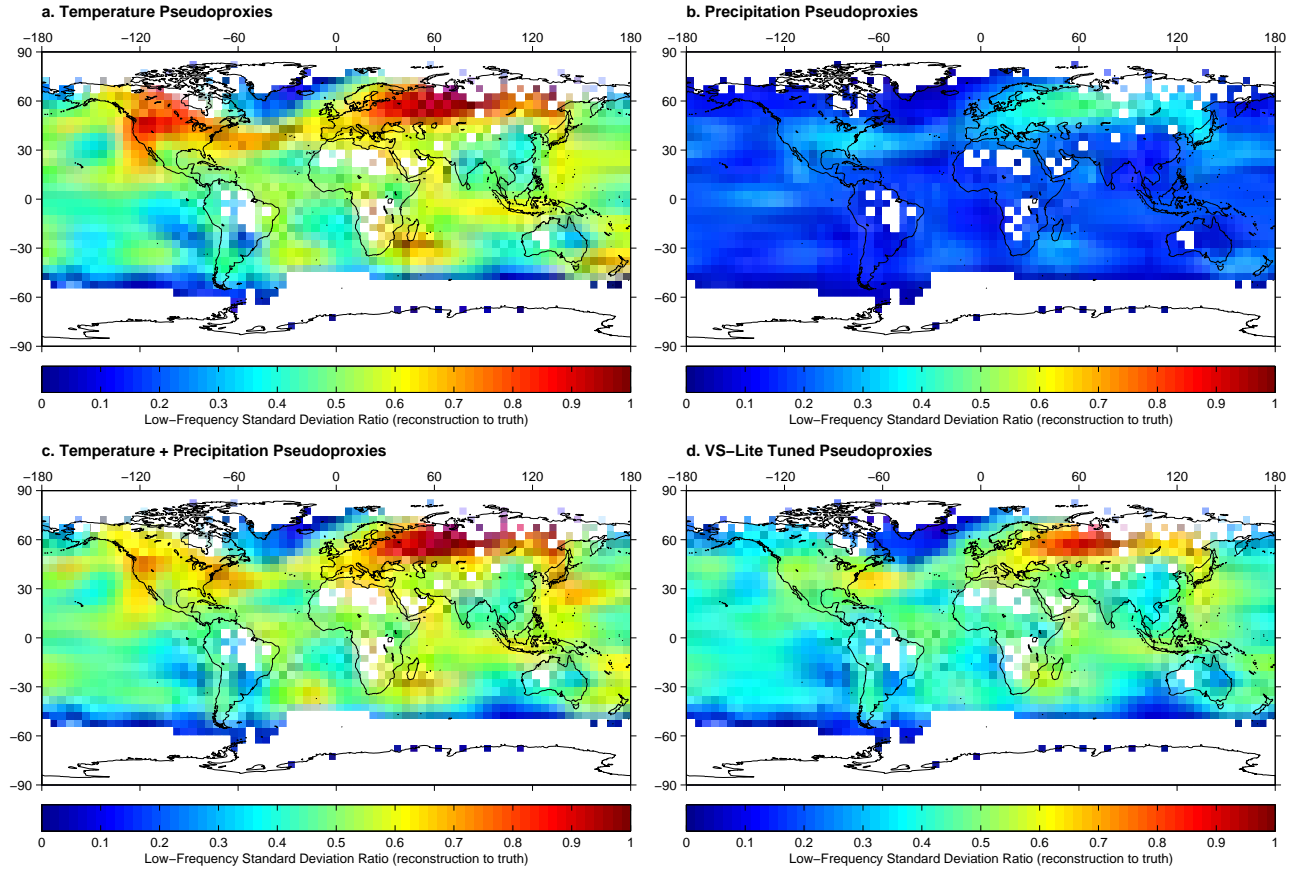# 4.4 Resolved fraction of low-frequency amplitude variance

Figure 6: Lowpass amplitude ratio ($\sigma_{CFR}/\sigma_{truth}$, f<1/20y) fields for: (a) Temperature-only, (b) precipitation-only, (c) precipitation/temperature, and (d) VS-Lite pseudoproxy-based reconstruction of mean annual temperature.
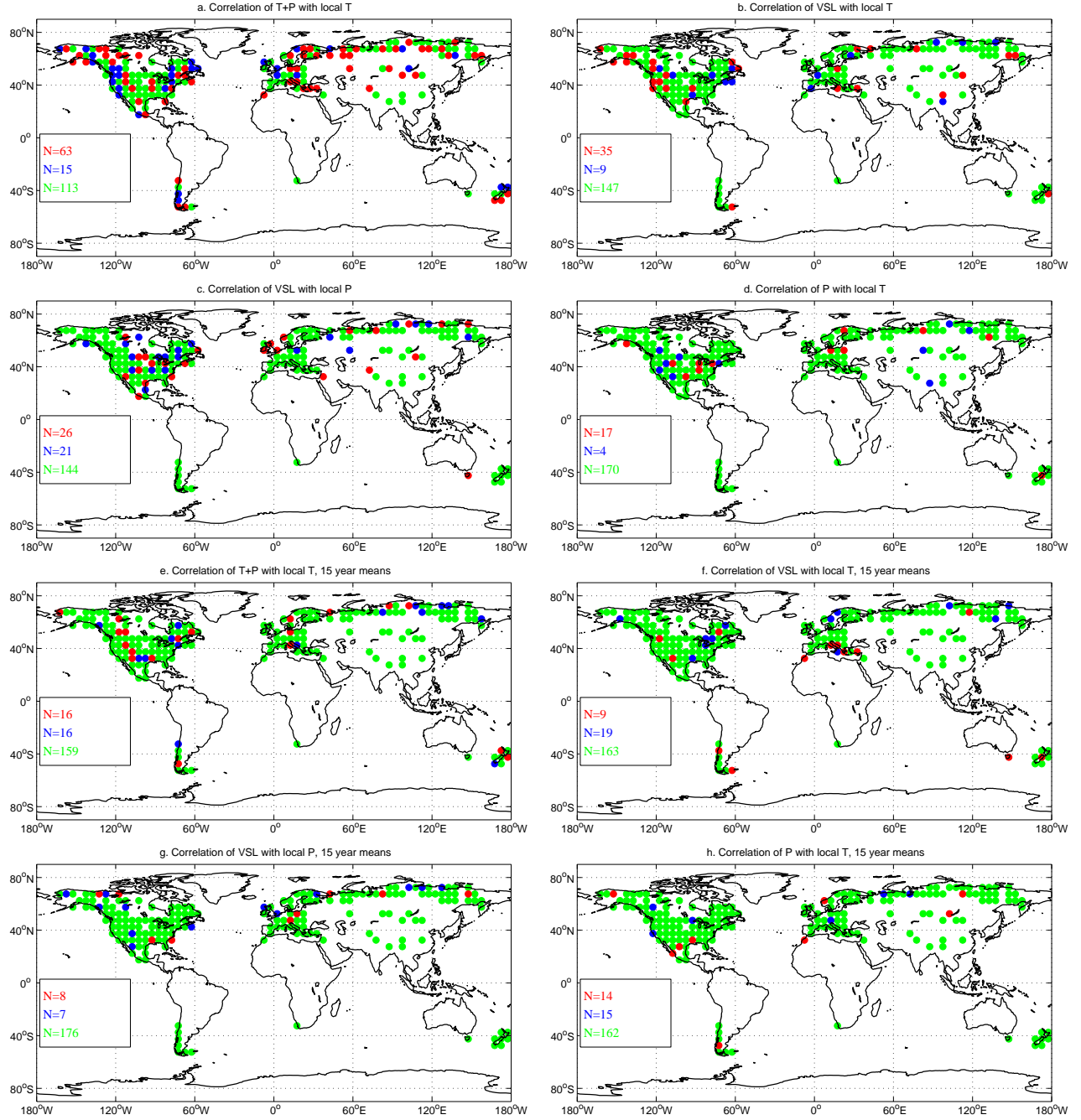
9

5 Local correlations between pseudoproxies

Figure 7: a: Correlation of T and T+P pseudoproxies (red: p≤0.05, blue: 0.05<p≤0.10, green: p>0.10) for SNR=0.5 case, 1721-1855 validation interval, N=135. b: As in (a), except for correlation of T and VSL pseudoproxies. c: As in (a), except for correlation of VSL and P pseudoproxies. d: As in (a), except for correlation of T and P pseudoproxies. Panels e,f,g,h : as in panels a,b,c,d, except for correlations of 15-year averages for the same period (N=9).

6 Truncation parameters for CCA-based reconstructions

PPE	d_{cca}	d_p	d_T
T	8	21	26
P	4	26	6
T+P	12	30	26
VSL	7	25	10

Table 2: Truncation parameters for CCA-based reconstructions. d_{cca} = number of EOF patterns of calibrated pseudoproxy-target correlation. d_p = number of EOF patterns resolved in the pseudoproxy data. d_T = number of EOF patterns resolved in the calibration target temperature field, following *Smerdon et al.* (2011).

References

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.

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.

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.

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, 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. Clim.*, *24*, 1284–1309, doi: 10.1175/2010JCL14110.1.

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.

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

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

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.

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.