AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2015GL065265

Key Points:

- Reconstructed spatial patterns of temperature are highly method dependent
- CFRs are more consistent with GCM simulations when an updated network is used
- Results based on a single CFR should be treated with caution

Supporting Information:

- Texts S1–S5, Figures S1–S19, and Tables S1–S3

Correspondence to:
J. Emile-Geay,
julieneg@usc.edu

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

Received 8 JUL 2015
Accepted 14 AUG 2015
Accepted article online 18 AUG 2015
Published online 3 SEP 2015

Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation

Jianghao Wang[1], Julien Emile-Geay[1], Dominique Guillot[2], Nicholas P. McKay[3], and Bala Rajaratnam[4]

[1]Department of Earth Sciences, University of Southern California, Los Angeles, California, USA, [2]Department of Statistics, Stanford University, Stanford, California, USA, [3]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona, USA, [4]Department of Earth System Science and the Woods Institute for the Environment, Stanford University, Stanford, California, USA

Abstract Climate field reconstructions (CFRs) enable spatially resolved estimates of past climates, providing important insights about climate variability over the Common Era. In particular, a reconstructed "La Niña-like" pattern during the transition from the Medieval Climate Anomaly (MCA) to the Little Ice Age has been widely tied to medieval droughts in southwest North America. This pattern is now used as a key benchmark for global climate model simulations of the last millennium, which have yet to reproduce it. Here we test the pattern's robustness by using four different CFR methods and two proxy networks. With the older network, we find the reconstructed patterns to be highly method-dependent, with the La Niña-like pattern not reproduced by two of the CFR methodologies. With the updated proxy network, a globally uniform MCA emerges with all methods, in agreement with simulations from the Paleoclimate Modelling Intercomparison Project Phase 3 ensemble. Our results caution against drawing dynamical interpretations from a single CFR and affirm the importance of developing CFRs through improved statistical methodology and community-driven proxy syntheses.

1. Introduction

Given observations of past climates (proxy records) and a well-calibrated statistical model, climate field reconstructions (CFRs) may provide rich spatial information about a climate variable (e.g., surface temperature), yielding insights into the underlying dynamics [*Jones et al.*, 2009]. Furthermore, CFRs allow detailed evaluations of global climate model (GCM) simulations of the climate of the past millennium [*Braconnot et al.*, 2012; *Masson-Delmotte et al.*, 2013]. In a prominent study, *Mann et al.* [2009] (hereafter M09) found that the differences in sea surface temperature (SST) between the Medieval Climate Anomaly (MCA, 950–1250 A.D.) and the Little Ice Age (LIA, 1400–1700 A.D.) displayed a La Niña-like pattern (cold eastern tropical Pacific, warm western Pacific). Such a pattern has been linked to many hydroclimate perturbations worldwide [*Seager et al.*, 2005], in particular droughts over southwest North America on interannual to multidecadal scales [*Cook et al.*, 2007], including medieval megadroughts [*Seager et al.*, 2007a, 2007b; *Cook et al.*, 2014]. This pattern has therefore been used as a key diagnostic to evaluate GCM simulations of the last millennium [*Graham et al.*, 2011; *Gonzalez-Rouco et al.*, 2011; *Fernández-Donado and et al.*, 2013; *Landrum et al.*, 2013; *Masson-Delmotte et al.*, 2013]. Climate models, however, are not able to reproduce this SST pattern, and the model-data discrepancies have been mostly attributed to model deficiencies [*Diaz et al.*, 2011; *Fernández-Donado and et al.*, 2013; *Bothe et al.*, 2013; *Landrum et al.*, 2013]. The robustness of reconstructed spatial patterns has been extensively studied in pseudoproxy experiments [*Smerdon et al.*, 2008, 2010, 2011; *Li and Smerdon*, 2012; *Annan and Hargreaves*, 2012; *Dannenberg and Wise*, 2013; *Steiger et al.*, 2014; *Wang et al.*, 2014; *Evans et al.*, 2014; *Smerdon et al.*, 2015] but not with real-world proxy networks. Here for the first time, we evaluate the spatial performance of proxy-based CFR methodologies with real proxy records and assess the extent to which they affect model-data comparisons over the Common Era.

2. Experimental Design

To facilitate a direct comparison to M09, we employ the same input data. The calibration target temperature field is the HadCRUT3v surface temperature data set (1850–1995 A.D.; see *Brohan et al.* [2006] and Figure S1 in the supporting information), and the proxy network is identical to that used by M09. The M09 proxy data set (Figure 1a) is a global multiproxy network, including data from tree ring, ice core, coral, speleothem,

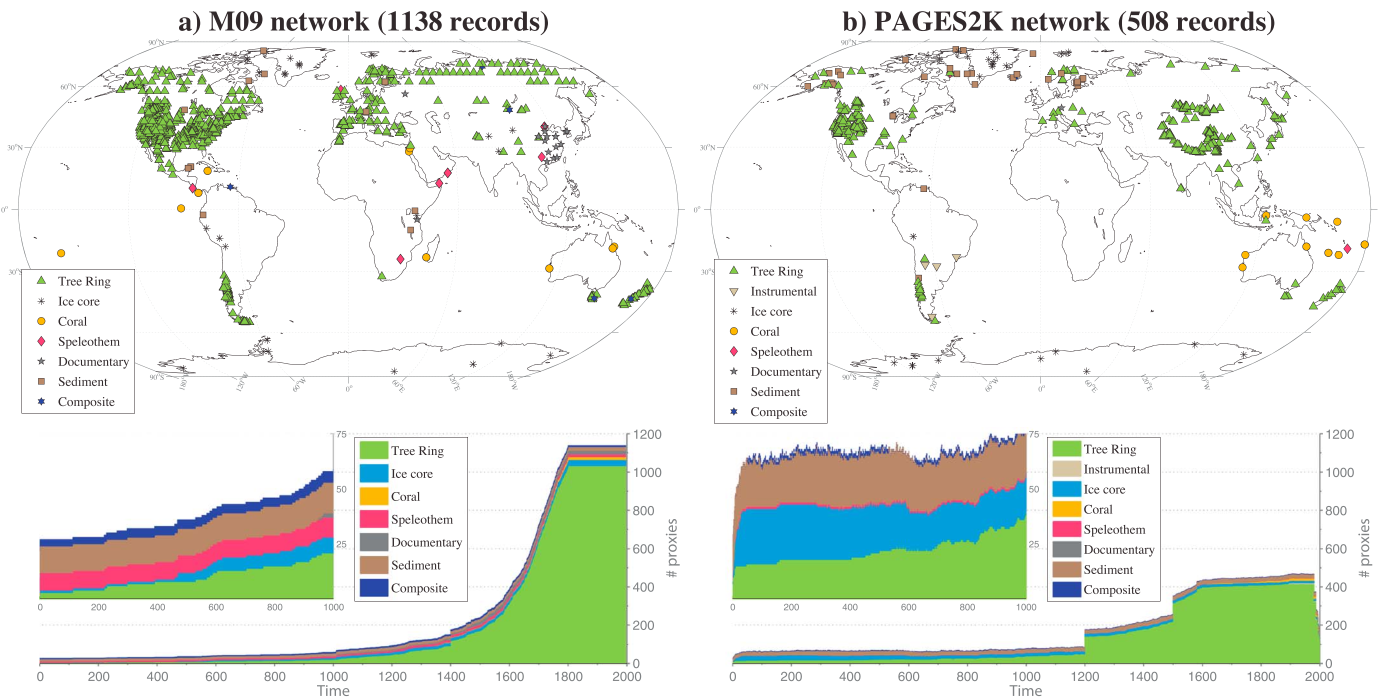

Figure 1. Spatial distribution of two proxy networks used in this study: (a) M09 proxy network and (b) PAGES2k phase 1 network. Top maps show the spatial distribution of proxies in each network, on which color markers separate proxy by type. Bottom color blocks show the temporal availability of each proxy class.

documentary, and marine and lacustrine sedimentary archives [*Wahl et al.*, 2010]. The network contains 1138 annual or lower resolution records, with a strong sampling bias toward the Northern Hemisphere (965 records). Most proxies are terrestrial records, dominated by total ring width measurements from North America and middle- to high-latitude Eurasia (Figure 1a, top). The majority of the records span only the past millennium, with a few low-resolution sediment cores spanning the entire Common Era (Figure 1a, bottom).

To isolate the effect of methodological differences on the reconstructed patterns, we apply a suite of four CFR methods to the same input data, including the regularized EM algorithm [*Schneider*, 2001] (RegEM) with truncated total least squares (TTLS) regression (*RegEM-TTLS*), the *Mann et al.* [2009] implementation of RegEM-TTLS (hereafter *M09-TTLS*), canonical correlation analysis [*Smerdon et al.*, 2010] (CCA), and Gaussian graphical models embedded within the EM algorithm [*Guillot et al.*, 2015] (GraphEM). All methods considered here are variants of multivariate linear regression and employ different strategies to regularize the ill-posed estimation problem of proxy-based reconstructions (Text S2.1). It is important to note that all of these methodologies are based on defensible statistical choices, so there is no a priori reason to favor one method over the other. They have all been evaluated in pseudoproxy contexts [*Mann et al.*, 2008; *Smerdon et al.*, 2010, 2011; *Guillot et al.*, 2015; *Wang et al.*, 2014], and their relative performance is found to vary given different input data. A comprehensive review of the statistical challenges of paleoclimate reconstruction is provided in *Tingley et al.* [2012]; the pros and cons of these four CFR methods, in particular, are detailed in *Wang et al.* [2014].

Our reconstructions were calibrated against the annual HadCRUT3v target over 1850–1995 A.D. This narrows down the selection of input data to proxies available at least beyond 1850 A.D. in order to be temporally calibrated. Finally, due to a paucity of data at the beginning of the Common Era, and our focus on the MCA-LIA differences, we restrict our analysis to the period 850–1995 A.D.

3. Divergence in Reconstructed Spatial Patterns

The reconstructed northern hemispheric (NH) temperature reveals a relatively warm MCA in all CFRs (Figure S5), consistent with previous studies [*Lamb*, 1965; *Hughes and Diaz*, 1994; *Broecker*, 2001; *Jones et al.*, 2009; *Masson-Delmotte et al.*, 2013]. The spatial patterns of these temperature anomalies, however, vary substantially among CFR methods (Figure 2). No compelling common spatial feature can be identified across all

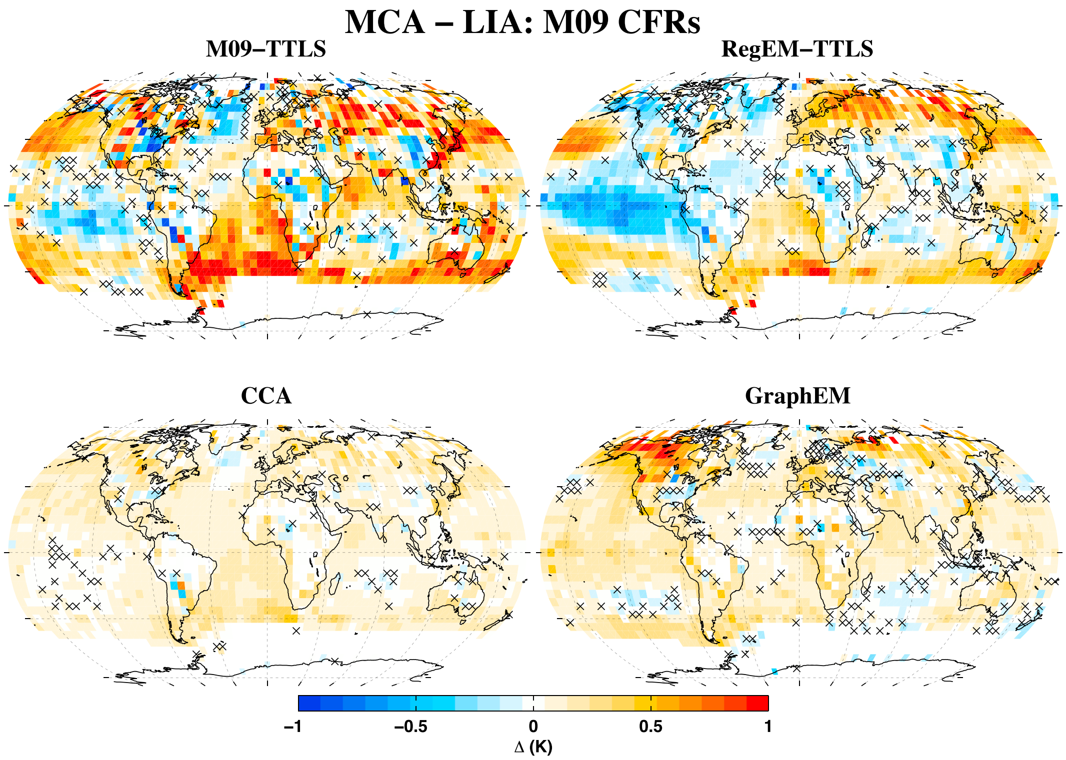

Figure 2. Reconstructed temperature difference between the Medieval Climate Anomaly (MCA 950–1250 A.D.) and the Little Ice Age (LIA 1400–1700 A.D.). Grid cells marked with cross symbols represent insignificant ($p \geq 0.05$) temperature difference between the MCA and LIA.

reconstructions. For the transition from the MCA to the LIA, both TTLS-based reconstructions show a prominent La Niña-like pattern in the tropical Pacific region, as in M09. This pattern is not reproduced in the CCA or the GraphEM-based reconstructions, both of which display an El Nino-neutral pattern, with a relatively uniform warming over the globe. The two TTLS-based reconstructions also display substantial differences due to the details of their implementations. Even in data-rich North America, the RegEM-TTLS and M09-TTLS reconstructions show nearly opposite temperature anomaly patterns.

Which of these spatial patterns, if any, is reliable? Is it possible to identify a "best" method and rely on that CFR? To answer this question, we assessed the statistical skill of the four CFRs, i.e., how well predictions match observations (Text S3.1). Whereas the skill of reconstructing global/hemispheric means is comparable across the different methods, spatial skill varies substantially (Figures S7–S12 and Table S3). Effectively, none of the CFRs is unequivocally more skillful than the others with this proxy network. The vast intermethod differences in spatial patterns and poor reconstruction skill therefore call into question the robustness of the La Niña-like pattern found in M09.

4. Sensitivity to Predictor Selection

Although the fragility of reconstructed spatial patterns has not received much attention in real-world CFRs, the subject has been extensively studied in pseudoproxy experiments (PPEs) [*Smerdon et al.*, 2008, 2010, 2011; *Li and Smerdon*, 2012; *Annan and Hargreaves*, 2012; *Dannenberg and Wise*, 2013; *Steiger et al.*, 2014; *Wang et al.*, 2014; *Evans et al.*, 2014; *Smerdon et al.*, 2015]. In particular, a recent PPE study [*Wang et al.*, 2014] affirms that all four CFR techniques used here can produce skillful NH mean temperature reconstructions with a M09-like pseudoproxy network but that spatial skill varies widely among methods (Text S3.2). In noisy settings, where proxies are poorly correlated with temperature, GraphEM and CCA tend to perform less well than M09-TTLS and RegEM-TTLS (Figure S13). On the other hand, using proxies that correlate well with observed temperatures, CCA and GraphEM outperform M09-TTLS and RegEM-TTLS in most cases (Figure S14).

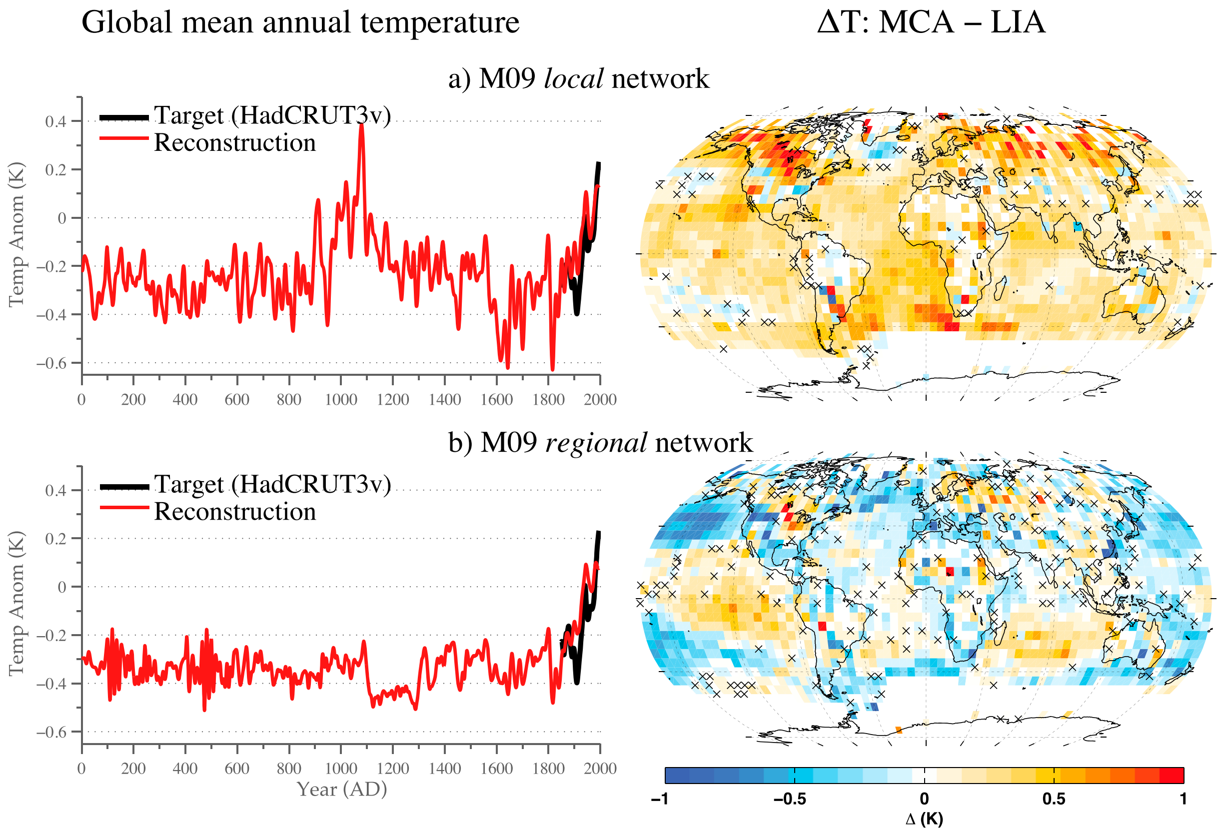

Figure 3. Reconstructed global mean and temperature difference between the Medieval Climate Anomaly (MCA 950–1250 A.D.) and the Little Ice Age (LIA 1400–1700 A.D.), using the M09-TTLS method, applied to two M09 screened networks: (a) local network and (b) regional network. Grid cells marked with cross symbols represent insignificant ($p \geq 0.05$) temperature difference between the MCA and LIA.

This finding leads us to explore the sensitivity of CFR methods to data quality, which we do via screening of the M09 proxies (Text S1.2). Proxies are known to respond to more than one climate variable [e.g., *Cook et al.*, 2004; *Jones et al.*, 2009] and to contain nonclimatic noise. Therefore, we consider two screening criteria: one assumes a purely local temperature-proxy relationship (*local* network, Figure S2, top) and the other recognizes that proxies may reflect regional temperature, via teleconnections (*regional* network, Figure S2, bottom). In both scenarios, proxies are selected based on their Pearson correlations to temperature grid cells in the HadCRUT3v data set. Their statistical significance is established via a nonparametric test taking persistence into account [*Ebisuzaki*, 1997].

Both screened networks were then used to derive CFRs using the M09-TTLS method and were compared to the reconstruction based on the full M09 proxy network (1138 proxies). Again, considering the MCA-LIA temperature difference, the reconstruction using the local network shows a globally warm MCA (Figure 3a). The amplitude of warming during the MCA appears to be even larger than during the late twentieth century. On the other hand, in the reconstruction obtained from the regional network, the MCA appears to be colder than the LIA in most regions, with an El Niño-like pattern (Figure 3b). Neither reconstruction shows the La Niña-like pattern found with the full M09 network (Figure 2, M09-TTLS reconstruction). This further underscores the fragility of reconstructed spatial patterns: even with a single CFR method, minor changes in screening criteria (thus data composition) can substantially influence CFR results.

5. Reconstruction With the PAGES2k Network

Which of the two factors, CFR methodological choice or proxy data composition, has a more fundamental impact on the reconstructions? Until now, we have only based our study on the M09 proxy network, which has an average proxy-temperature correlation lower than 0.3 [*Wang et al.*, 2014, Figures 2 and 3] (see also Figure S2, local network). The value is close to the low end of proxy-temperature correlations usually considered:

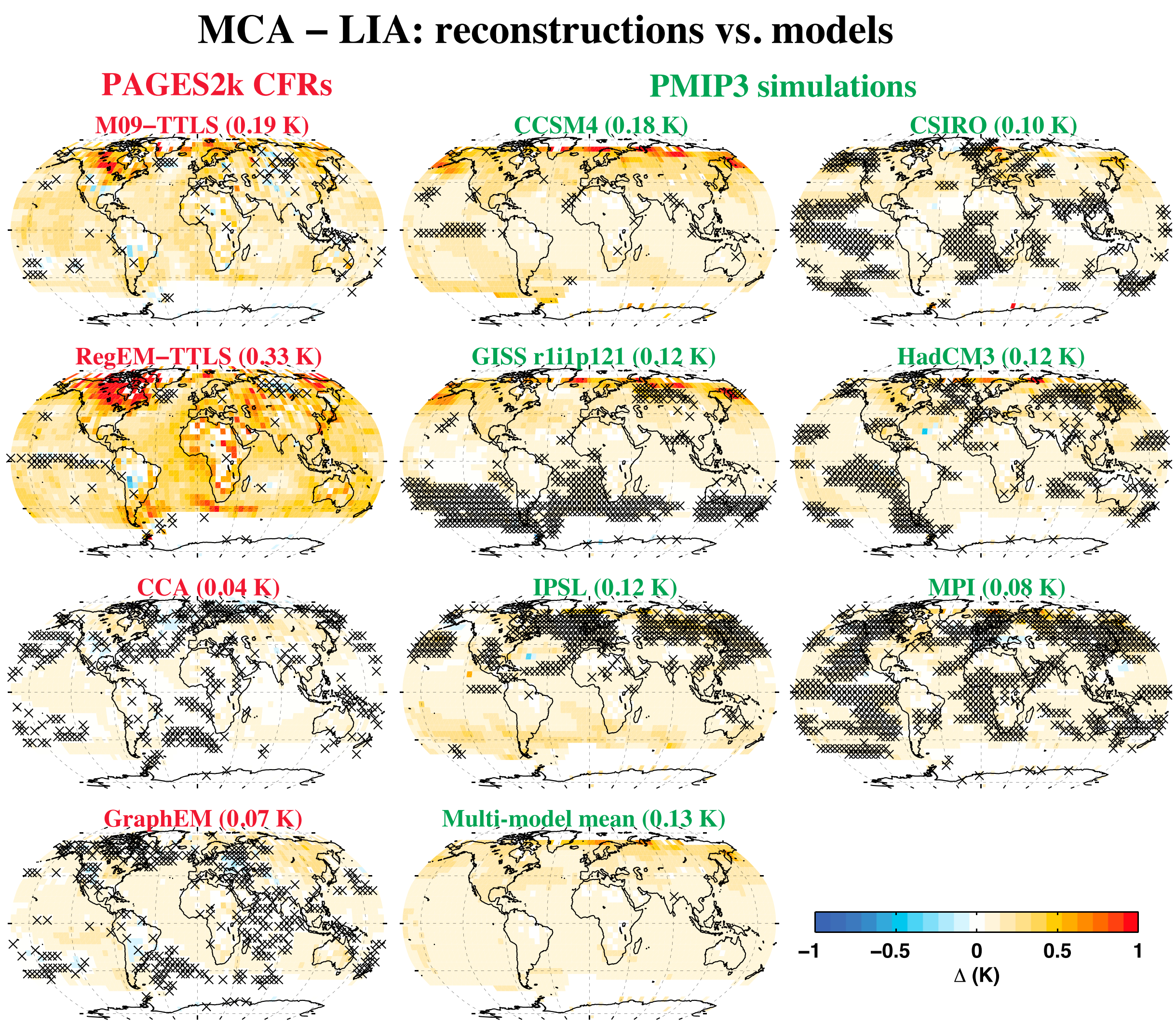

Figure 4. Reconstructed and simulated temperature difference between the Medieval Climate Anomaly (MCA 950–1250 A.D.) and the Little Ice Age (LIA 1400–1700 A.D.). Note that all models are regridded to have 5°× 5° resolution, and grid cells outside the domain of the *Mann et al*. [2009] reconstruction are taken out of the analysis to ensure direct comparability between models and reconstructions. Grid cells marked with cross symbols represent insignificant ($p \geq 0.05$) temperature difference between the MCA and LIA. Numbers above each map denote the (latitudinally) weighted global average values of the MCA-LIA temperature difference.

traditionally, an absolute correlation $|\rho| = 0.45$ or, equivalently, a *signal-to-noise ratio* (SNR) of 0.5 is thought to be reflective of real-world proxies' relationship with temperature [see *Smerdon*, 2012, and references therein] (see also Text S1.2 and equation (S1)). As mentioned above, PPEs have shown that large intermethod discrepancies exist in M09-like noisy settings; on the other hand, common features may emerge from reconstructions based on different CFR techniques when quality-controlled data are used [*Wang et al.*, 2014]. To further probe the role of data quality on CFRs, we now apply the four CFR methods described above to an updated proxy compilation, the 2k Network of the International Geosphere-Biosphere Programme Past Global Changes (PAGES) project (hereinafter *PAGES2k*, www.pages-igbp.org/workinggroups/2k-network) [*PAGES2k Consortium*, 2013]. In doing so, we obtain a suite of new CFRs to compare with the reconstructions based on the M09 proxy network.

The PAGES2k project aims to produce the next-generation database of temperature-sensitive proxy climate records of the last two millennia. The full network is composed of 508 time series of tree rings, pollen, corals, lake and marine sediments, glacier ice, speleothems, and historical documents. The data cover seven continental-scale regions (Figure 1b), including Antarctica, the Arctic, Asia, Australia, Europe, North America, and South America. As in *PAGES2k Consortium* [2013], suitable proxy records from Africa are currently too sparse for a reliable temperature synthesis and are therefore excluded in this study. The PAGES2k network is notably different from the M09 network in two aspects: (1) more than 70 records span the Common Era (compared to 40 in the M09 network) and (2) each proxy was identified by paleoclimate proxy experts from each region to have a demonstrated sensitivity to annual or warm-season temperature [*PAGES2k Consortium*, 2013]. Although the PAGES2k network is composed of significantly fewer records than the M09 network, its overall quality is higher, especially prior to 1750 A.D. (Figures S3 and S4). Notably, 360 of the 508 PAGES2k records were not used in the M09 reconstruction, and 35 are updated versions of the records included in their data set. In this study, a subset of the PAGES2k network is used for reconstruction, where proxies must have constant temporal availability over 1850–1995 A.D. in order to be calibrated against instrumental temperature. Further, we require the instrumental overlap to be longer than 90 years to allow for a meaningful calibration. As in the M09 reconstruction, the HadCRUT3v mean annual temperature (MAT) is the target field. It should be noted that many of the PAGES2k records are predominantly warm-season proxies, which may introduce a bias when targeting the MAT for reconstructions. This study targets the MAT to enable direct comparisons to the M09 reconstruction.

Results based on the PAGES2k network show more intermethod similarities than when the same methods were applied to the M09 network (Figure 2 versus Figure 4, left column). Remarkably, for the reconstructed MCA-LIA temperature difference, no La Niña-like pattern was found in the tropical Pacific with any of the methods. Although the magnitude of temperature difference between the two periods varies among reconstructions, the MCA was found to be globally warmer than the LIA in all four CFRs (Figure 4, left column). Furthermore, a relatively uniform MCA warmth emerges consistently in CCA and GraphEM-based CFRs, whether the M09 or the PAGES2k network is used. In line with pseudoproxy results [*Wang et al.*, 2014], this suggests that CCA and GraphEM are less sensitive to changes in data composition. This could be because they make fewer assumptions about noise structure than TLS-based methods, which assume noise properties to be known in attempts to correct for their biasing effects [*Van Huffel and Vandewalle*, 1991].

6. Comparison With PMIP3 Simulations

Results based on the PAGES2k network offer insights into the evaluations of climate models. To our knowledge, most existing data-model comparison studies have adopted the M09 reconstruction as a benchmark [e.g., *Gonzalez-Rouco et al.*, 2011; *Goosse et al.*, 2012; *Fernández-Donado and et al.*, 2013; *Masson-Delmotte et al.*, 2013; *Bothe et al.*, 2013]. For temporal comparisons, broad agreement was found between GCM simulations and reconstructions, in that simulated NH temperatures mostly lie within the uncertainties of existing reconstructions. However, the range of individual results is very wide: −0.5°C to 1.5°C for reconstructions and −0.3 to 0.3°C for models [*Masson-Delmotte et al.*, 2013]. For spatial comparisons, the MCA-LIA temperature difference is frequently used as a metric to assess model simulations. No model has thus far been able to simulate a La Niña-like pattern. Herein we present an updated comparison by including four CFRs based on the PAGES2k network and six model simulations from the Paleoclimate Modelling Intercomparison Project Phase 3 (PMIP3) [*Braconnot et al.*, 2012]; see Text S4 for details.

Consistent with previous studies [*Fernández-Donado and et al.*, 2013; *Masson-Delmotte et al.*, 2013], we see an overall agreement between reconstructed and simulated NH temperature (Figure S19). The magnitude of temperature variability, however, varies between individual reconstructions and model simulations. On average, reconstructions display much smaller amplitudes of variation (∼0.4°C) than model simulations (0.4°C to 1.2°C, model dependent), but the reconstructed magnitudes of change are more compatible with instrumental data from HadCRUT3v (Figure S19, thick black line). Much of the difference in variability coincides with periods of strong volcanic events [*Mann et al.*, 2005; *Brohan et al.*, 2006; *Anchukaitis et al.*, 2013]. Whereas both simulated and reconstructed temperatures respond markedly to volcanic forcing, PMIP3 models simulate much more cooling (∼0.5°C) than is suggested by the CFRs (∼0.2°C).

For the MCA to LIA temperature difference, the La Niña-like pattern suggested by M09 reconstructions is not reproduced by any of the PMIP3 model simulations or PAGES2k CFRs (Figure 4). In contrast, all simulations and reconstructions display a nearly uniformly warm MCA compared to the LIA. At regional scales, models simulate a weak Arctic amplification, which is partly captured by the M09-TTLS- and RegEM-TTLS-based reconstructions. Overall, the globally warm pattern is similar across model simulations and reconstructions. The magnitude of temperature changes, however, varies among reconstructions and simulations. GCMs simulate a warming of 0.13°C on average, considerably smaller than the warming in the TTLS-based reconstructions (0.2°C–0.3°C). The GraphEM- and CCA-based reconstructions, on the other hand, show smaller temperature differences between the MCA and the LIA (< 0.1°C). Compared with results based on the M09 network, the PAGES2k results suggest that more data-model similarity can be reached when regional expertise is involved in developing a data set. The reduced data-model discrepancies also have important implications for diagnosing climate models: the MCA-LIA La Niña-like pattern may just be an artifact of a particular CFR method applied on a specific set of proxies. Indeed, in pseudoproxy experiments, *Smerdon et al.* [2015] have shown the tendency for some methods and models to enhance the difference between the MCA-LIA, implying that some methods can artificially produce MCA-LIA differences in the tropical Pacific that are on the order of what is reported by M09.

7. Implications

Our results show that the spatial patterns produced in CFRs are greatly method dependent, at least when using the M09 network. This calls into question the robustness of the La Niña-like pattern characteristic of the transition from the MCA to the LIA in M09. Our results suggest that researchers should be cautious when drawing dynamical interpretations from a single reconstruction based on a single network. We also suggest that the MCA-LIA temperature anomaly pattern should not be used as a diagnostic for assessing paleoclimate model simulations until further consensus is reached.

Although a La Niña-like SST pattern persisting for much of the MCA would be consistent with the abundant evidence for medieval megadroughts [*Cook et al.*, 2007], we note that many possible mechanisms have been evoked to cause megadroughts. For instance, droughts in southwestern North America may also be forced from the tropical Atlantic [*Sutton and Hodson*, 2005] or by warming of the Indo-Pacific Warm Pool [*Graham et al.*, 2011]. Furthermore, a recent reconstruction of central equatorial Pacific SST [*Emile-Geay et al.*, 2013b] suggests a warm MCA in the region—as warm as the late twentieth century and warmer than the LIA. To test if the state of the tropical Pacific is indeed related to medieval megadroughts, one should reconstruct the pattern during the MCA, rather than the MCA-LIA difference (Figures S16 and S17). Finally, recent studies have suggested that megadroughts need not be forced but may arise from internal dynamics [e.g., *Coats et al.*, 2013a, 2015; *Stevenson et al.*, 2014]. These elements weaken the case for a La Niña-like MCA and point to alternate scenarios to explain medieval hydroclimate. We also note that most of the existing CFRs, including those presented here, rely heavily on land-based proxies and the stationarity of teleconnections. Climate model studies have shown that ocean-to-land teleconnections have the potential to vary significantly on multidecadal and longer timescales [e.g., *Coats et al.*, 2013b]. If true, the existence of such nonstationary teleconnections should discourage the use of many moisture-sensitive Southwest U.S. tree rings to infer temperature variability in the tropical Pacific. Additionally, reducing uncertainties about medieval patterns of tropical sea surface temperature will require the synthesis of longer and high-resolution marine records, while the interpretation of many terrestrial hydrological proxies may be more rigorously probed via proxy system models [*Evans et al.*, 2013; *Dee et al.*, 2015].

It is clear that the proxy network itself is an important component of intermethod reproducibility. For the first time, the PAGES2k network was used to derive four reconstructions of the global temperature field. Results show much less dependence on specific methods, and no CFR shows a La Niña-like pattern for the MCA-LIA transition. Using this data set, a globally warm MCA was found in all CFRs, consistent with model simulations from the CMIP5/PMIP3 ensemble. Though the first incarnation of this network critically lacks marine records, the convergence is encouraging and in line with recent PPE results showing that CFRs are less sensitive to the choice of statistical methods when the signal-to-noise ratio is reasonably high, i.e., SNR $\geq$ 0.5, equivalent to an absolute proxy-temperature correlation greater than 0.45 [*Wang et al.*, 2014]. The globally homogenous MCA warmth in PAGES2k CFRs and CMIP5/PMIP3 simulations provides new hypotheses to test: was the global warmth a response to radiative forcing or a product of internal variability? Are GCMs and proxy-based CFRs getting convergent results for similar reasons or by coincidence?

Finally, many community-wide efforts are afoot to improve surface temperature reconstructions [*PAGES2k Consortium*, 2014], including building a uniform open-access proxy network (PAGES2k Consortium, A global, quality-controlled, multiproxy database for temperature reconstructions of the Common Era, *Nature Scientific Data*, manuscript in preparation, 2015), combining proxies of different resolutions, and continued improvement of statistical techniques. In agreement with previous studies [*Smerdon et al.*, 2011; *PAGES2k Consortium*, 2014; *Tingley et al.*, 2012], we recommend employing as many methods as possible to ensure robust conclusions and only evaluating GCM simulations against those features that are robust to methodological choices.

Acknowledgments
The authors acknowledge NSF award AGS1003818, PAGES funding for the PAGES2k Climate Reconstruction Methods Workshop, and computational resources from the USC High-Performance Computing Center. Bala Rajaratnam also acknowledges funding from NSF DMSCAREER-1352656, AGS1003823, and DMS1106642. We thank Yuxin Zhou for downloading and regridding the PMIP3-CMIP5 data. We also thank all the scientists who contributed paleoclimate data to the PAGES2k network, including the PAGES2k regional working groups. Code for all reconstruction methods, as well as reconstructed fields, is available at http://climdyn.usc.edu/Publications.html. M09 proxy data are available at http://www.ncdc.noaa.gov/paleo/pubs/mann2009b/mann2009b.html, PAGES2k data are available at https://www.ncdc.noaa.gov/paleo/study/14188, and PMIP3-CMIP5 data are available at http://pcmdi9.llnl.gov/esgf-web-fe/.

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

Anchukaitis, K. J., R. D. D'Arrigo, L. Andreu-Hayles, D. Frank, A. Verstege, A. Curtis, B. M. Buckley, G. C. Jacoby, and E. R. Cook (2013), Tree-ring-reconstructed summer temperatures from northwestern North America during the last nine centuries, *J. Clim.*, *26*(10), 3001–3012, doi:10.1175/JCLI-D-11-00139.1.

Annan, J. D., and J. C. Hargreaves (2012), Identification of climatic state with limited proxy data, *Clim. Past*, *8*(4), 1141–1151, doi:10.5194/cp-8-1141-2012.

Bothe, O., J. Jungclaus, and D. Zanchettin (2013), Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble, *Clim. Past*, *9*(6), 2471–2487, doi:10.5194/cp-9-2471-2013.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao (2012), Evaluation of climate models using palaeoclimatic data, *Nat. Clim. Change*, *2*(6), 417–424, doi:10.1038/nclimate1456.

Broecker, W. S. (2001), Paleoclimate: Was the medieval warm period global?, *Science*, *291*(5508), 1497, doi:10.1126/science.291.5508.1497.

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.

Coats, S., J. E. Smerdon, R. Seager, B. I. Cook, and J. F. González-Rouco (2013a), Megadroughts in southwestern North America in ECHO-G millennial simulations and their comparison to proxy drought reconstructions, *J. Clim.*, *26*(19), 7635–7649, doi:10.1175/JCLI-D-12-00603.1.

Coats, S., J. E. Smerdon, B. I. Cook, and R. Seager (2013b), Stationarity of the tropical pacific teleconnection to North America in CMIP5/PMIP3 model simulations, *Geophys. Res. Lett.*, *40*, 4927–4932, doi:10.1002/grl.50938.

Coats, S., J. E. Smerdon, B. I. Cook, and R. Seager (2015), Are simulated megadroughts in the North American southwest forced?, *J. Clim.*, *28*(1), 124–142, doi:10.1175/JCLI-D-14-00071.1.

Cook, B. I., J. E. Smerdon, R. Seager, and E. R. Cook (2014), Pan-continental droughts in North America over the last millennium, *J. Clim.*, *27*(1), 383–397, doi:10.1175/JCLI-D-13-00100.1.

Cook, E. R., C. A. Woodhouse, C. M. Eakin, D. M. Meko, and D. W. Stahle (2004), Long-term aridity changes in the western United States, *Science*, *306*, 1015–1018, doi:10.1126/science.1102586.

Cook, E. R., R. Seager, M. A. Cane, and D. W. Stahle (2007), North American drought: Reconstructions, causes, and consequences, *Earth Sci. Rev.*, *81*, 93–134, doi:10.1016/j.earscirev.2006.12.002.

Dannenberg, M. P., and E. K. Wise (2013), Performance of climate field reconstruction methods over multiple seasons and climate variables, *J. Geophys. Res. Atmos.*, *118*, 9595–9610, doi:10.1002/jgrd.50765.

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

Diaz, H. F., R. Trigo, M. K. Hughes, M. E. Mann, E. Xoplaki, and D. Barriopedro (2011), Spatial and temporal characteristics of climate in medieval times revisited, *Bull. Am. Meteorol. Soc.*, *92*(11), 1487–1500, doi:10.1175/BAMS-D-10-05003.1.

Ebisuzaki, W. (1997), A method to estimate the statistical significance of a correlation when the data are serially correlated, *J. Clim.*, *10*(9), 2147–2153.

Emile-Geay, J., K. M. Cobb, M. E. Mann, and A. T. Wittenberg (2013b), Estimating central equatorial Pacific SST variability over the past millennium. Part II: Reconstructions and uncertainties, *J. Clim.*, *26*, 2329–2352, doi:10.1175/JCLI-D-11-00511.1.

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

Evans, M. N., S. Tolwinski-Ward, D. 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.

Fernández-Donado, L., J. González-Rouco, C. Raible, C. Ammann, D. Barriopedro, E. García-Bustamante, J. Jungclaus, S. Lorenz, J. Luterbacher, and S. Phipps (2013), Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Clim. Past*, *9*, 393–421, doi:10.5194/cp-9-393-2013.

Gonzalez-Rouco, J., L. Fernández-Donado, C. Raible, D. Barriopedro, J. Lutherbacher, J. Jungclaus, D. Swingedouw, J. Servonnat, E. Zorita, and S. Wagner (2011), Medieval Climate Anomaly to Little Ice Age transition as simulated by current climate models, *PAGES News*, *19*(1), 7–8.

Goosse, H., E. Crespin, S. Dubinkina, M.-F. Loutre, M. E. Mann, H. Renssen, Y. Sallaz-Damaz, and D. Shindell (2012), The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly", *Clim. Dyn.*, *39*(12), 2847–2866, doi:10.1007/s00382-012-1297-0.

Graham, N., C. Ammann, D. Fleitmann, K. Cobb, and J. Luterbacher (2011), Support for global climate reorganization during the "Medieval Climate Anomaly", *Clim. Dyn.*, *37*(5–6), 1217–1245, doi:10.1007/s00382-010-0914-z.

Guillot, D., B. Rajaratnam, and J. Emile-Geay (2015), Statistical paleoclimate reconstructions via Markov random fields, *Ann. Applied. Statist.*, *1*, 324–352, doi:10.1214/14-AOAS794.

Hughes, M. K., and H. F. Diaz (1994), Was there a "Medieval Warm Period", and if so, where and when?, *Clim. Change*, *26*(2), 109–142, doi:10.1007/BF01092410.

Jones, P., et al. (2009), High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects, *The Holocene*, *19*(1), 3–49, doi:10.1177/0959683608098952.

Lamb, H. H. (1965), The early medieval warm epoch and its sequel, *Palaeoecology*, *1*(0), 13–37.

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

Li, B., and J. E. Smerdon (2012), Defining spatial comparison metrics for evaluation of paleoclimatic field reconstructions of the Common Era, *Environmetrics*, *23*(5), 394–406, doi:10.1002/env.2142.

Mann, M. E., M. A. Cane, S. E. Zebiak, and A. Clement (2005), Volcanic and solar forcing of the tropical Pacific over the past 1000 years, *J. Clim.*, *18*(3), 447–456, doi:10.1175/JCLI-3276.1.

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

Mann, M. E., Z. Zhang, S. Rutherford, R. S. Bradley, M. K. Hughes, D. Shindell, C. Ammann, G. Faluvegi, and F. Ni (2009), Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*(5957), 1256–1260, doi:10.1126/science.1177303.

Masson-Delmotte, V., et al. (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 383–464, Information From Paleoclimate Archives, Cambridge Univ. Press, Cambridge, U. K., and New York.

PAGES2k Consortium (2013), Continental-scale temperature variability during the past two millennia, *Nat. Geosci.*, *6*(5), 339–346, doi:10.1038/ngeo1797.

PAGES2k Consortium (2014), A community-driven framework for climate reconstructions, *Eos Trans. AGU*, *95*(40), 361–362, doi:10.1002/2014EO400001.

Schneider, T. (2001), Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *J. Clim.*, *14*, 853–871, doi:10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2.

Seager, R., Y. Kushnir, C. Herweijer, N. Naik, and J. Velez (2005), Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856–2000, *J. Clim.*, *18*(19), 4065–4088, doi:10.1175/JCLI3522.1.

Seager, R., N. Graham, C. Herweijer, A. Gordon, Y. Kushnir, and E. Cook (2007a), Blueprints for medieval hydroclimate, *Quat. Sci. Rev.*, *26*(19–21), 2322–2336, doi:10.1016/j.quascirev.2007.04.020.

Seager, R., et al. (2007b), Model projections of an imminent transition to a more arid climate in southwestern North America, *Science*, *316*(5828), 1181–1184, doi:10.1126/science.1139601.

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

Smerdon, J. E., A. Kaplan, and D. Chang (2008), On the origin of the standardization sensitivity in RegEM climate field reconstructions, *J. Clim.*, *21*(24), 6710–6723, doi:10.1175/2008JCLI2182.1.

Smerdon, J. E., A. Kaplan, D. Chang, and M. N. Evans (2010), A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium, *J. Clim.*, *23*(18), 4856–4880, doi:10.1175/2010JCLI3328.1.

Smerdon, J. E., A. Kaplan, E. Zorita, J. F. González-Rouco, and M. N. Evans (2011), Spatial performance of four climate field reconstruction methods targeting the Common Era, *Geophys. Res. Lett.*, *38*, L11705, doi:10.1029/2011GL047372.

Smerdon, J. E., S. Coats, and T. R. Ault (2015), Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era, *Clim. Dyn.*, 1–22, doi:10.1007/s00382-015-2684-0.

Steiger, N. J., G. J. Hakim, E. J. Steig, D. S. Battisti, and G. H. Roe (2014), Assimilation of time-averaged pseudoproxies for climate reconstruction, *J. Clim.*, *27*(1), 426–441, doi:10.1175/JCLI-D-12-00693.1.

Stevenson, S., A. Timmermann, Y. Chikamoto, S. Langford, and P. DiNezio (2014), Stochastically generated North American megadroughts, *J. Clim.*, *28*(5), 1865–1880, doi:10.1175/JCLI-D-13-00689.1.

Sutton, R. T., and D. L. R. Hodson (2005), Atlantic Ocean forcing of North American and European summer climate, *Science*, *309*(5731), 115–118, doi:10.1126/science.1109496.

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*(0), 1–22, doi:10.1016/j.quascirev.2012.01.012.

Van Huffel, S., and J. Vandewalle (1991), *The Total Least Squares Problem: Computational Aspects and Analysis, Frontiers in Applied Mathematics*, vol. 9, SIAM, Philadelphia, Pa.

Wahl, E. R., D. M. Anderson, B. A. Bauer, R. Buckner, E. P. Gille, W. S. Gross, M. Hartman, and A. Shah (2010), An archive of high-resolution temperature reconstructions over the past 2+ millennia, *Geochem. Geophys. Geosyst.*, *11*, Q01001, doi:10.1029/2009GC002817.

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