The Annals of Applied Statistics
2011, Vol. 5, No. 1, 65–70
DOI: 10.1214/10-AOAS398D
Main article DOI: 10.1214/10-AOAS398

DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE?

BY GAVIN A. SCHMIDT, MICHAEL E. MANN AND SCOTT D. RUTHERFORD

NASA Goddard Institute for Space Studies, Pennsylvania State University and Roger Williams University

McShane and Wyner (2011) (henceforth MW) analyze a dataset of "proxy" climate records previously used by Mann et al. (2008) (henceforth M08) to attempt to assess their utility in reconstructing past temperatures. MW introduce new methods in their analysis, which is welcome. However, the absence of both proper data quality control and appropriate "pseudoproxy" tests to assess the performance of their methods invalidate their main conclusions.

We deal first with the issue of data quality. In the frozen 1000 AD network of 95 proxy records used by MW, 36 tree-ring records were not used by M08 due to their failure to meet objective standards of reliability. These records did not meet the minimal replication requirement of at least eight independent contributing tree cores (as described in the Supplemental Information of M08). That requirement yields a smaller dataset of 59 proxy records back to AD 1000 as clearly indicated in M08. MW's inclusion of the additional poor-quality proxies has a material affect on the reconstructions, inflating the level of peak apparent Medieval warmth, particularly in their featured "OLS PC10" ($K = 10$ PCs of the proxy data used as predictors of instrumental mean NH land temperature) reconstruction. The further elimination of four potentially contaminated "Tiljander" proxies [as tested in M08; M08 also tested the impact of removing tree-ring data, including controversial long "Bristlecone pine" tree-ring records. Recent work [cf. Salzer et al. (2009)], however, demonstrates those data to contain a reliable long-term temperature signal], which yields a set of 55 proxies, further reduces the level of peak Medieval warmth (Figure 1(a); cf. Figure 14 in MW; see also Supplementary Figures S1–S2 [Schmidt, Mann and Rutherford (2011a, 2011b)]).

The MW "OLS PC10" reconstruction has greater peak apparent Medieval warmth in comparison with M08 or any of a dozen similar hemispheric temperature reconstructions [Jansen et al. (2007)]. That additional warmth, as shown above, largely disappears with the use of the more appropriate dataset. Using their reconstruction, MW nonetheless still found recent warmth to be unusual in a long-term context: they estimate an 80% probability that the decade 1997–2006 is

Received September 2010.

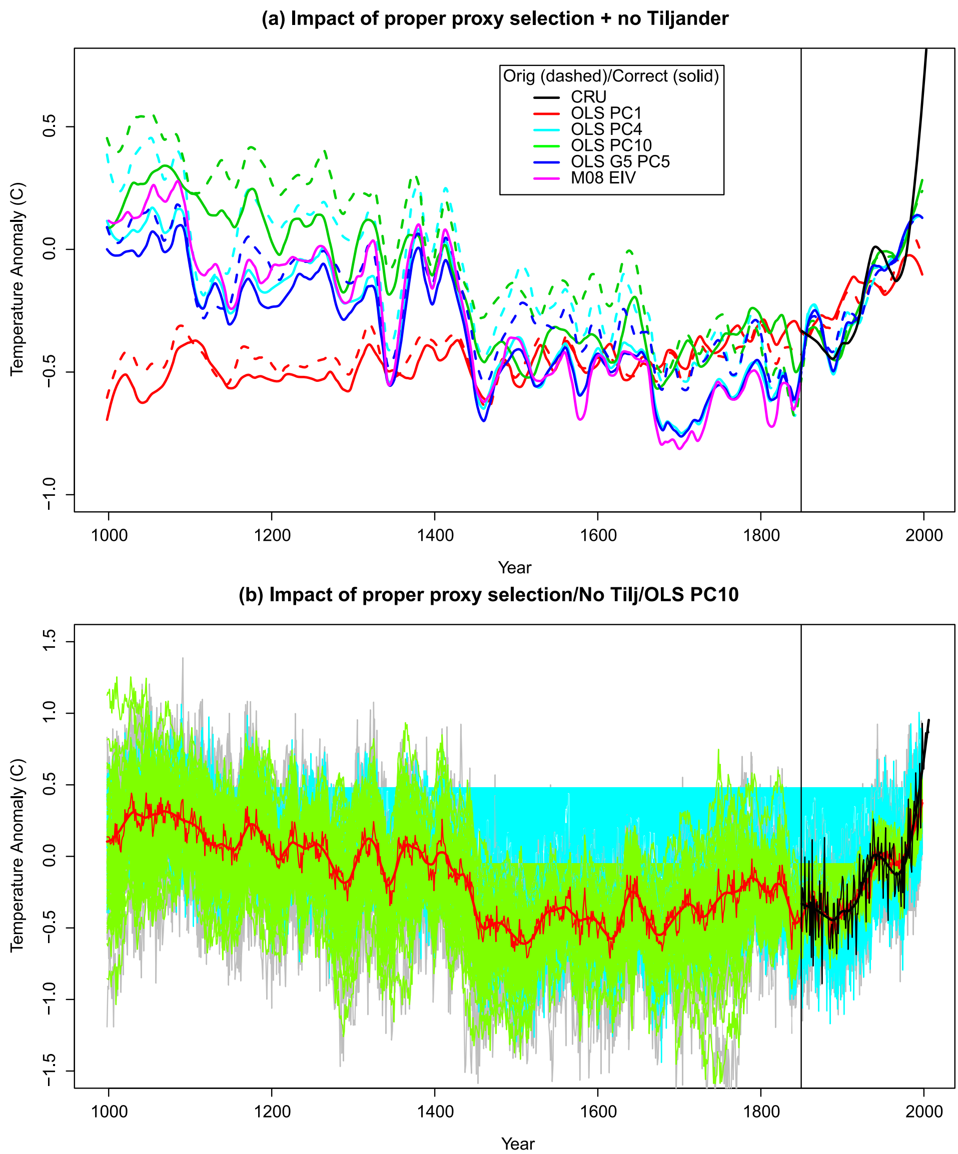

FIG. 1. *Reconstructions of mean Northern Hemisphere land temperatures over the past millennium for various methodological choices* (*cf. MW Figure* 14). (a) *Results using the M*08 *frozen AD* 1000 *network of* 59 *minus* 4 *"Tiljander" proxy records* (*corresponding results based on all* 59 *records are shown in Supplementary Figure S*1). *Shown for comparison are the original MW results and the Mann et al*. (2008) *"EIV" decadal "CRU" NH land temperature reconstruction based on the identical proxy data. The OLS reconstructions have been filtered with a loess smoother* (span = 0.05) *to emphasize low-frequency* (*greater than* 50 *year*) *variations. Associated annual reconstructions are shown in Supplementary Figure S*2. (b) *Comparison of Monte Carlo ensemble* (*and mean*) *reconstructions using "OLS PC*10*" as in MW Figure* 16. *Labeled reconstructions are in color, grey lines are the total set of MW reconstructions after allowing for uncertainties in the coefficients*.

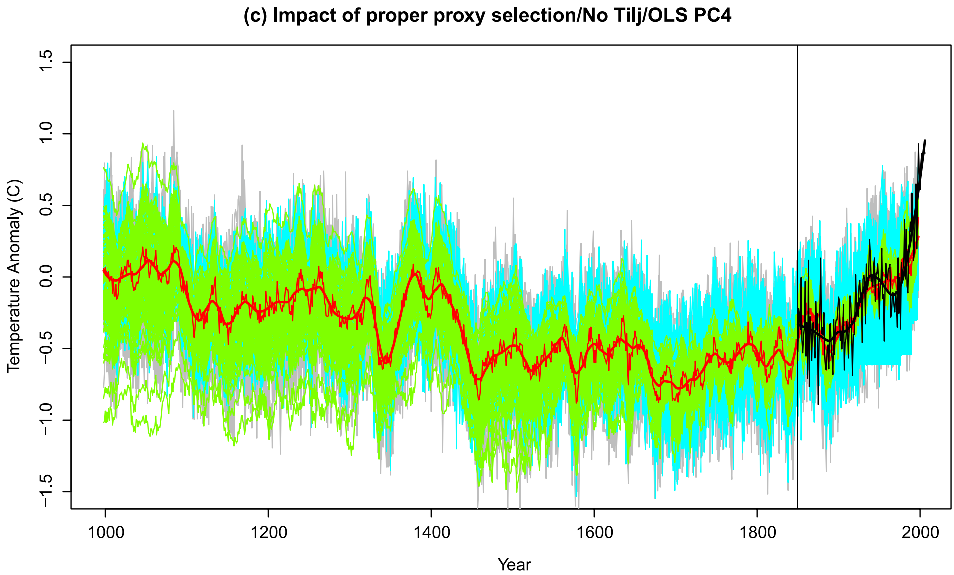

FIG. 1. (c) *As in* (b) *above but instead using "OLS PC*4."

warmer than any other for at least the past 1000 years. Using the more appropriate 55-proxy dataset with the same ($K = 10$) estimation procedure, we calculate a higher probability of 86% that recent decadal warmth is unprecedented for the past millennium [Figure 1(b)].

However $K = 10$ principal components is almost certainly too large, and the resulting reconstruction likely suffers from statistical over-fitting. Objective selection criteria applied to the M08 AD 1000 proxy network (see Supplementary Figure S4), as well as independent "pseudoproxy" analyses discussed below, favor retaining only $K = 4$ ("OLS PC4" in the terminology of MW). Using this reconstruction, we observe a very close match [e.g., Figure 1(a)] with the relevant M08 reconstruction and we calculate considerably higher probabilities up to 99% that recent decadal warmth is unprecedented for at least the past millennium [Figure 1(c)]. These posterior probabilities imply substantially higher confidence than the "likely" assessment by M08 and IPCC (2007) (a 67% level of confidence). Indeed, a probability of 99% not only exceeds the IPCC "very likely" threshold (90%), but reaches the "virtually certain" (99%) threshold. However, since these posterior probabilities do not take into account potential systematic issues in the source data, are sensitive to methodological choices, and vary by a few percent depending on the MCMC realization, we maintain that a "likely" conclusion is most consistent with the balance of evidence [IPCC (2007)].

There are additional methodological weaknesses in the techniques employed by MW that require discussion. MW mix incommensurate (decadal vs. annual resolution) proxy data in their procedure, a problem that is avoided by the "hybrid" frequency band calibration method used by M08. Using a version of the proxy data that was consistently low-pass filtered to retain only decadal features shows even better agreement with the M08 reconstruction (supplementary Figure S3).

Furthermore, methods using simple Ordinary Least Squares (OLS) regressions of principal components of the proxy network and instrumental data suffer from known biases, including the underestimation of variance [see, e.g., Hegerl et al. (2006)]. The spectrally "red" nature of the noise present in proxy records poses a particular challenge [e.g., Jones et al. (2009)]. A standard benchmark in the field is the use of synthetic proxy data known as "pseudoproxies" derived from long-term climate model simulations where the true climate history is known, and the skill of the particular method can be evaluated [see, e.g., Mann et al. (2007); Jones et al. (2009) and numerous references therein]. (We note that the term "pseudoproxy" was misused in MW to instead denote various noise models.) In contrast to the MW claim that their methods perform "fairly similarly," these tests show dramatic differences in model performance (Figure 2). Indeed, the various flavors of OLS and, particularly, the "Lasso" method (used only in the first half of MW), suffer from serious underestimation biases in comparison with, for example, the hybrid RegEM approach of M08 (see also Table S1).

Taken together, these points demonstrate that any conclusions regarding the utility of proxies in reconstructing past climate drawn by MW were, at best, overstated. Assessing the skill of methods that do not work well (such as Lasso) and concluding that no method can therefore work well, is logically flawed. Additional problems exist in their assessment procedure—reducing the size of the hold out periods to 30 years from 46 years in M08, for instance, makes it more difficult to meaningfully diagnose statistical skill.

Problems in climate research, such as statistical climate reconstruction, require sophisticated statistical approaches and a thorough understanding of the data used. Moreover, investigations of the underlying spatial patterns of past climate changes, rather than simply hemispheric mean temperature estimates, are most likely to provide insights into climate dynamics [e.g., Mann et al. (2009), Schmidt (2010)]. Further progress in this area will most likely arise from continuing collaboration between the statistics and climate science communities, such as fostered since 1996 by the joint NSF/NCAR Geophysical Statistics Project.

Acknowledgments. We thank Sonya Miller for substantial technical assistance. The JAGS/rjags code used in the Bayesian modeling was adapted from http://probabilitynotes.wordpress.com/.

Supplementary figures and tables, data used, and scripts for performing all analyses are all available at: http://www.meteo.psu.edu/~mann/supplements/AOAS/

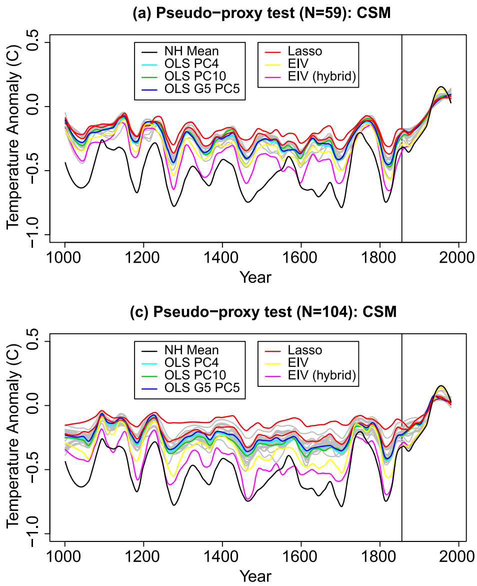

FIG. 2. *Pseudoproxy tests of reconstruction methodologies used by MW and comparison with the hybrid and nonhybrid RegEM EIV methods used by M*08. *The pseudoproxy networks are defined by a randomly selected set of gridboxes using two different coupled ocean-atmosphere general circulation model* (*OAGCM*) *simulations subjected to estimated natural and anthropogenic forcing over the past millennium. Pseudoproxies are constructed assuming "red" proxy noise* [AR(1) *with* $\rho = 0.32$] *yielding mean signal-to-noise amplitude ratio of* SNR $= 0.4$, *characteristics which are consistent with estimates from actual proxy data networks* [*see Mann et al*. (2007)]. *All reconstructions use a calibration interval of* 1856–1980. *Figure shows results for a* 59-*location network including* (a) *NCAR CSM and* (b) *GKSS simulations and a network with* 104 *locations for* (c) *CSM and* (d) *GKSS. Labeled reconstructions are in color, grey lines are the total set of MW reconstruction techniques. Note that uncertainties are reduced for the larger network, where the underestimation bias becomes negligible for the hybrid RegEM EIV method.*

SUPPLEMENTARY MATERIAL

Supplement A: Supplemental figures (DOI: 10.1214/10-AOAS398DSUPPA; .pdf). Additional figures S1–4 and Table S1.

Supplement B: Code and data for producing all figures and results in the paper (DOI: 10.1214/10-AOAS398DSUPPB; .zip).

REFERENCES

HEGERL, G. C., CROWLEY, T. J., HYDE, W. T. and FRAME, D. J. (2006). Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* **440** 1029–1032.

IPCC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, eds.). Cambridge Univ. Press, Cambridge.

JANSEN, E., OVERPECK, J., BRIFFA, K. R., DUPLESSY, J.-C., JOOS, F., MASSON-DELMOTTE, V., OLAGO, D., OTTO-BLIESNER, B., PELTIER, W. R., RAHMSTORF, S., RAMESH, R., RAYNAUD, D., RIND, D., SOLOMINA, O., VILLALBA, R. and ZHANG, D. (2007). Palaeoclimate. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, eds.). Cambridge Univ. Press, Cambridge.

JONES, P. D., BRIFFA, K. R., OSBORN, T. J., LOUGH, J. M., VAN OMMEN, T. D., VINTHER, B. M., LUTERBACHER, J., WAHL, E. R., ZWIERS, F. W., MANN, M. E., SCHMIDT, G. A., AMMANN, C. M., BUCKLEY, B. M., COBB, K. M., ESPER, J., GOOSSE, H., GRAHAM, N., JANSEN, E., KIEFER, T., KULL, C., KÜTTEL, M., MOSLEY-THOMPSON, E., OVERPECK, J. T., RIEDWYL, N., SCHULZ, M., TUDHOPE, A. W., VILLALBA, R., WANNER, H., WOLFF, E. and XOPLAKI, E. (2009). High-resolution paleoclimatology of the last millennium: A review of current status and future prospects. *Holocene* **19** 3–49.

MANN, M. E., ZHANG, Z., HUGHES, M. K., BRADLEY, R. S., MILLER, S. K., RUTHERFORD, S. and NI, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millenia. *Proc. Natl. Acad. Sci. USA* **105** 13252–13257.

MANN, M. E., RUTHERFORD, S., WAHL, E. and AMMANN, C. (2007). Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.* **112** D12109.

MANN, M. E., ZHANG, Z., RUTHERFORD, S., BRADLEY, R. S., HUGHES, M. K., SHINDELL, D., AMMANN, C., FALUGEVI, G. and NI, F. (2009). Global signatures and dynamical origins of the "Little Ice Age" and "Medieval Climate Anomaly." *Science* **326** 1256–1260.

MCSHANE, B. B. and WYNER, A. J. (2011). A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Statist.* **5** 5–44.

SALZER, M. W., HUGHES, M. K., BUNN, A. G. and KIPFMUELLER, K. F. (2009). Recent unprecedented tree-ring growth in bristlecone pine at the highest elevations and possible causes. *Proc. Natl. Acad. Sci. USA* **106** 20348–20353.

SCHMIDT, G. A. (2010). Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change. *J. Quaternary Sci.* **25** 79–87.

SCHMIDT, G. A., MANN, M. E. and RUTHERFORD, S. D. (2011a). Supplement to "Discussion on A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable?" DOI: 10.1214/10-AOAS398DSUPPA.

SCHMIDT, G. A., MANN, M. E. and RUTHERFORD, S. D. (2011b). Supplement to "Discussion on A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable?" DOI: 10.1214/10-AOAS398DSUPPB.

G. A. SCHMIDT
NASA GODDARD INSTITUTE FOR SPACE STUDIES
NEW YORK, NEW YORK
USA
E-MAIL: Gavin.A.Schmidt@nasa.gov

M. E. MANN
DEPARTMENT OF METEOROLOGY AND EARTH AND ENVIRONMENTAL SYSTEMS INSTITUTE
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA
USA
E-MAIL: mann@psu.edu

S. D. RUTHERFORD
DEPARTMENT OF ENVIRONMENTAL SCIENCE
ROGER WILLIAMS UNIVERSITY
BRISTOL, RHODE ISLAND
USA