Reply to comment by Rutherford et al. on "Erroneous Model Field Representations in Multiple Pseudoproxy Studies: Corrections and Implications"†

Jason E. Smerdon * and Alexey Kaplan

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

Daniel E. Amrhein

Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program in Oceanography, Cambridge, MA

* *Corresponding author address:* Jason E. Smerdon, Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W, P.O. Box 1000, Palisades, NY 10964. E-mail: jsmerdon@ldeo.columbia.edu

†LDEO contribution number XXXX

We are pleased that *Rutherford et al.* (2012, hereinafter R12) confirm the errors that were identified and discussed in *Smerdon et al.* (2010, hereinafter S10). These errors were associated with the processing of the millennium-length NCAR CCSM1.4 (*Ammann et al.* 2007) and the GKSS ECHO-G (*González-Rouco et al.* 2003) simulations by *Mann et al.* (2005) and *Mann et al.* (2007, hereinafter M07). R12 also clarify that related papers published after M07 were not affected by the errors described in S10. We welcome this news, but note that these later publications (even those published after S10) made no attempts to correct the earlier results or to indicate that the new results were free of the previous errors. Below we respond to several additional arguments raised by R12.

R12 emphasize a distinction between the two versions of the regularized expectation maximization (RegEM) method (*Schneider* 2001). They imply that RegEM using truncated total least squares (RegEM-TTLS) is a better climate field reconstruction (CFR) method than RegEM using ridge regression (RegEM-Ridge), the latter of which was used by S10 to illustrate some of the consequences of the model-processing errors. We first note that any CFR method could have been used to demonstrate the errors discovered by S10, making methodological distinctions in this context immaterial. Secondly, it is true that RegEM-TTLS has been shown in pseudoproxy studies to better reconstruct the Northern Hemisphere (NH) mean, but both of the RegEM methods are meant to reconstruct temperature *fields*. Spatial reconstruction skill therefore is a fundamental measure of their methodological performance. To date, the only comprehensive comparisons of the spatial skill of multiple CFR methods did not find RegEM-TTLS to be a clear frontrunner (*Smerdon et al.* 2011; *Li and Smerdon* 2012). To the contrary, RegEM-TTLS performs similarly or worse than other multivariate regression methods in several spatial skill metrics and all of the evaluated methods have

important spatial errors. The advocacy of one multivariate linear CFR method over another is therefore premature.

R12 additionally present the use of the incorrectly oriented CCSM1.4 field as an experiment that serendipitously indicates RegEM-TTLS to be "*rather robust.*" It is true that the reconstructions performed by M05 and M07 can be reinterpreted as experiments with a different sampling scheme (a point made by S10). The claim of robustness, however, requires qualification: the statistics reported in lines three and four of R12's Table 1 are similar only because they are NH averages. The spatial performance of RegEM-TTLS and other CFR methods is nevertheless strongly dependent on the distribution of the pseudoproxy network (*Smerdon et al.* 2011). For instance, regions of large correlation coefficients calculated between known and reconstructed model fields tend to be concentrated in areas of dense pseudoproxy sampling, while most areas without pseudoproxy data perform poorly in this metric. Any perceived robustness of the RegEM-TTLS results therefore only holds for NH-averaged statistics, while regional skill statistics (e.g. for Niño3) would expose important differences between experiments with correct and incorrect sampling.

Regarding the Niño3 assessment statistics specifically, R12 dismiss the significance of these incorrect numbers in M07 by arguing that the paper "*focused on the Northern Hemisphere mean and the overall field reconstruction.*" This assertion ignores the fact that the reconstructed Niño3 index was one of only two diagnostics used by M07 to validate the spatial skill of the RegEM-TTLS method. The method has since been used by *Mann et al.* (2009a) and *Mann et al.* (2009b) to derive real-world CFRs. Reconstructed Niño3 and other regional indices played crucial roles in analyses and conclusions of both studies, and *Mann et al.* (2009a), for instance, cited M07 as the study in which their method "*has been rigorously*

tested." Despite the evident importance of these Niño3 statistics in M07, no subsequent publications have corrected them. Prior to S10, this omission resulted in an unexplained disparity between the Niño3 reconstruction skill in the CCSM1.4 and ECHO-G experiments reported by M07.

Finally, we regrettably point out that R12 continue to perpetuate inconsistencies in the presentation of their pseudoproxy experiments. According to the R12 description, the ECHO-G reconstruction in the top panel of their Figure 1 and the corresponding numbers in Table 1 (second line) should not be different from what is presented by *Rutherford et al.* (2008), but these results indeed differ. There also are inconsistencies in the validation intervals attributed to several experiments summarized in Table 1 of R12. Lines one and three in this table are taken from Table 1 in M07, which reported in its caption a validation interval from 850-1855 CE. The caption in the R12 table, however, attributes the same numbers to validation intervals that extend to the year 1899. These inconsistencies thus continue to undermine the utility of the presented pseudoproxy results and make it difficult to interpret them in future pseudoproxy studies.

Maintaining consistent and correctly documented records of pseudoproxy tests is critical for evaluating CFR methods. The advantage of such tests lies in their ability to serve as common testbeds on which reconstruction methods can be systematically evaluated and compared (*Smerdon* 2012). This advantage is lost if pseudoproxy experiments are inaccurately described or incorrectly executed. Timely corrections to pseudoproxy tests are therefore vital for avoiding the perpetuation of errors and inconsistencies in the published literature, and to improve testing and development of methods for reconstructing climate fields during the Common Era.

Acknowledgments.

This research was supported in part by the National Science Foundation by grant ATM-0902436 and by the National Oceanic and Atmospheric Administration by grant NA07OAR4310060. LDEO contribution number XXXX.

REFERENCES

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas, 2007: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. *Proc. Nat. Acad. Sci. USA*, **104**, 3713-3718, doi:10.1073-pnas.0605064.103.

González-Rouco, F., H. von Storch, and E. Zorita, 2003: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. *Geophys. Res. Lett.*, **30**, 21, 2116, doi:10.1029/2003GL018264.

Li, B., and J.E. Smerdon, 2012: Defining spatial assessment metrics for evaluation of paleoclimatic field reconstructions of the Common Era, *Environmetrics*, in press.

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2005: Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J. Climate*, **18**, 4097-4107.

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of

proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.

Mann, M.E., Z. Zhang, S. Rutherford, R.S. Bradley, M.K. Hughes, D. Shindell, C. Ammann, G. Faluvegi, F. Ni, 2009a: Global Signatures and Dynamical Origins of the Little Ice Age and the Medieval Climate Anomaly. *Science*, 326, 5957, 1256-1260, DOI: 10.1126/science.1177303, 2009.

Mann, M.E., J.D. Woodruff, J.P. Donnelly, and Z. Zhang, 2009b: Atlantic hurricanes and climate over the past 1,500 years. *Nature*, 460, 880-883, doi:10.1038/nature08219.

Rutherford, S., M. E. Mann, E. Wahl, and C. Ammann, 2008: Reply to comment by Jason E. Smerdon et al. on "Robustness of proxy-based climate field reconstruction methods". *J. Geophys. Res.*, **113**, D18107, doi:10.1029/2008JD009964.

Rutherford, S., M. E. Mann, E. Wahl, and C. Ammann, 2012: Comment on "Erroneous Model Field Representations in Multiple Pseudoproxy Studies: Corrections and Implications". *J. Climate*, in review.

Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, **14**, 853-887.

Smerdon, J.E., A. Kaplan, and D.E. Amrhein, 2010: Erroneous Model Field Representations in Multiple Pseudoproxy Studies: Corrections and Implications. *J. Climate*, **23**, 5548-5554, doi:10.1175/2010JCLI3742.1.

Smerdon, J.E., A. Kaplan, E. Zorita, J.F. Gonzlez-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., 2012: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, *WIREs Climate Change*, **3**, 63-77, doi:10.1002/wcc.149.