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

ABSTRACT

Rutherford et al. (2012) confirm the errors that were identified and discussed in *Smerdon et al.* (2010) that either invalidated or required the reinterpretation of quantitative results from pseudoproxy experiments presented in *Mann et al.* (2005), *Mann et al.* (2007) and several subsequent papers. These errors have a strong influence on the spatial skill assessments of climate field reconstructions, despite their small impacts on skill statistics averaged over the Northern Hemisphere. On the basis of spatial performance, RegEM-TTLS (*Mann et al.* 2007) cannot be considered a preferred reconstruction technique (*Smerdon et al.* 2011; *Li and Smerdon* 2012), making methodological distinctions in the current context unnecessary. It is also noted that important skill statistics for the Ninõ3 region presented by *Mann et al.* (2007) have yet to be corrected.

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. This is an important clarification, particularly for papers published after S10, given that none of these later publications explicitly corrected the earlier results or indicated that the new results were free of 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 (see *Smerdon* 2012, for a review), 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 methods for global temperature CFRs 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 also claim that similar results are obtained from pseudoproxy experiments using the correctly and incorrectly oriented CCSM1.4 field. This claim 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; *Werner et al.* 2012; *Annan and Hargreaves* 2012). Any perceived similarity between results presented by M05, M07 and R12 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 as demonstrated in S10.

Regarding the M07 Niño3 assessment statistics specifically, R12 dismiss the significance of these incorrect numbers by arguing that they were "*not relevant to the main conclusions*" of the paper. Yet the reconstructed Niño3 index was one of only two diagnostics used by M07 to validate the spatial skill of RegEM-TTLS, a spatially-explicit field reconstruction method. *Mann et al.* (2009a) and *Mann et al.* (2009b) subsequently used the method to derive real-world CFRs in which reconstructed Niño3 and other regional indices played crucial roles. Despite this, no subsequent publications, including the present R12 comment, have corrected the erroneous Niño3 validation statistics from M07. Prior to S10, this omission caused a confusing disparity between the Niño3 reconstruction skill in the M07 CCSM1.4 and ECHO-G experiments.

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 (see *Smerdon* 2012, for a review). This advantage can only be realized if pseudoproxy experiments are accurately described and correctly executed. Timely corrections to pseudoproxy tests are therefore vital for avoiding the perpetuation of errors and inconsistencies in the published literature.

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.

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

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.

Werner, J. P., J. Luterbacher, and J. E. Smerdon, 2012: A Pseudoproxy Evaluation of Bayesian Hierarchical Modeling and Canonical Correlation Analysis for Climate Field Reconstructions over Europe, *WIREs Climate Change*, **3**, 63-77, doi:10.1002/wcc.149.