Spatial performance of four climate field reconstruction methods targeting the Common Era

J. E. Smerdon,[1] A. Kaplan,[1] E. Zorita,[2] J. F. González-Rouco,[3] and M. N. Evans[4]

Received 5 March 2011; revised 26 April 2011; accepted 29 April 2011; published 15 June 2011.

[1] The spatial skill of four climate field reconstruction (CFR) methods is investigated using pseudoproxy experiments (PPEs) based on two millennial-length general circulation model simulations. Results indicate that presently available global and hemispheric CFRs for the Common Era likely suffer from spatial uncertainties not previously characterized. No individual method produced CFRs with universally superior spatial error statistics, making it difficult to advocate for one method over another. Northern Hemisphere means are shown to be insufficient for evaluating spatial skill, indicating that the spatial performance of future CFRs should be rigorously tested for dependence on proxy type and location, target data and employed methodologies. Observed model-dependent methodological performance also indicates that CFR methods must be tested across multiple models and conclusions from PPEs should be carefully evaluated against the spatial statistics of real-world climatic fields.
Citation: 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.

1. Introduction

[2] Hemispheric and global reconstructions of temperature indices or fields provide estimates of climate variability prior to widespread availability of instrumental records, validation fields for general circulation models (GCMs), and estimates of climate sensitivity that help constrain climate projections for the 21st century [e.g., *Jansen et al.*, 2007]. Regional subsets from climate field reconstructions (CFRs) also have been used to characterize climate system dynamics [e.g., *Mann et al.*, 2009]. Despite the promise of these endeavors, methodological studies have demonstrated the potential for reconstructions to underestimate past temperature variability [e.g., *von Storch et al.*, 2004, 2006; *Lee et al.*, 2008; *Smerdon and Kaplan*, 2007; *Smerdon et al.*, 2011; *Christiansen et al.*, 2009], although several methods have been shown to successfully reconstruct Northern Hemisphere mean (NHM) temperatures in synthetic experiments [*Hegerl et al.*, 2007; *Mann et al.*, 2007]. All of these studies, however, have focused almost exclusively on characterizations of NHM reconstructions. Here we specifically compare the spatial characteristics of derived temperature CFRs using four methods in identical synthetic experiments and show that all methods yield CFRs with important spatial errors.

[3] Pseudoproxy experiments (PPEs), in which synthetic tests are constructed from subsamples of spatiotemporally-complete climate model output, have allowed controlled evaluations of CFR methods. *Von Storch et al.* [2004, 2006] used PPEs to characterize warm biases and variance losses in NHM temperatures estimated from *Mann et al.*'s [1998] CFR method (hereinafter MBH98). Canonical correlation analysis (hereinafter CCA) [see *Smerdon et al.*, 2011, and references therein] and the regularized expectation maximization (RegEM) method using ridge regression [*Schneider*, 2001] have also been shown to have similar shortcomings as CFR methods in PPEs [*Smerdon and Kaplan*, 2007; *Smerdon et al.*, 2011; *Christiansen et al.*, 2009]. In contrast, a RegEM application using truncated total least squares (hereinafter RegEM-TTLS) was used to successfully reconstruct NHM temperatures in a PPE context [*Mann et al.*, 2007].

[4] Here we explicitly examine the spatial performance of four CFR methods: MBH98, RegEM-TTLS, ridge regression and CCA. These methods include the full suite of multivariate linear methods most commonly used for global and hemispheric temperature CFRs, and our experiments are the first comparison of the spatial skill of these methods in identical PPEs.

2. Methodology

[5] PPEs are constructed from the surface air temperature fields of the National Center for Atmospheric Research Community Climate System Model 1.4 (hereinafter CCSM) [*Ammann et al.*, 2007] and the Hamburg Atmosphere-Ocean Coupled Circulation Model ERIK2 (hereinafter ECHO-g) [*González-Rouco et al.*, 2006] millennial simulations, according to *Smerdon et al.*'s [2010] corrections. Pseudoproxy distributions approximate the multi-proxy locations of the most populated nest in the MBH98 network (Figure 1 and Figure S1 of the auxiliary material).[1] A pseudoproxy network approximating the updated *Mann et al.*'s [2008] network is also tested and presented in the auxiliary material (Figures S1 and S2). Noise perturbations added to the pseudoproxies are drawn from Gaussian white-noise distributions and described in the auxiliary material. All results shown herein represent experiments using pseudoproxy signal-to-noise ratios (SNRs) of 0.5 by standard deviation, while the auxiliary material provides a summary of experiments using four different SNRs. Typical proxy records are estimated

[1] Lamont-Doherty Earth Observatory, Palisades, New York, USA.
[2] Institute for Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany.
[3] Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Madrid, Spain.
[4] Department of Geology, University of Maryland, College Park, Maryland, USA.

Copyright 2011 by the American Geophysical Union.
0094-8276/11/2011GL047372

[1] Auxiliary materials are available in the HTML. doi:10.1029/2011GL047372.

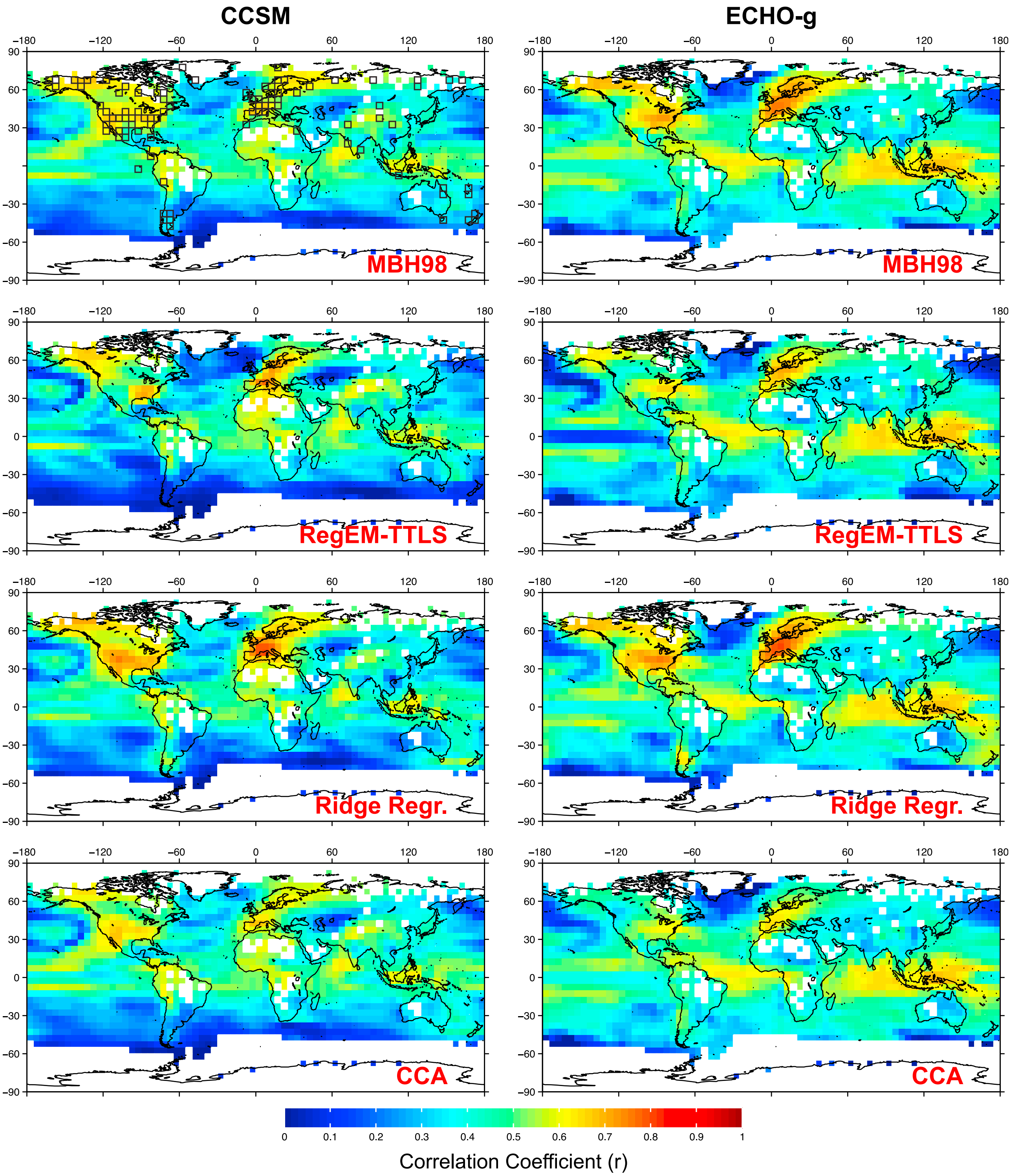

Figure 1. Local correlation coefficients for the four non-hybrid CFR methods using pseudoproxies with SNRs of 0.5 and the (left) CCSM and (right) ECHO-g model fields. All methods use the same pseudoproxies, target field, and intervals for calibration (1856–1980 CE). (top left) Grid-point locations of the pseudoproxies used in all the PPEs, which approximate the distribution of the most populated nest in the MBH98 network.

to have SNRs in the range of 0.5–0.25 [e.g., *Mann et al.*, 2007]. The instrumental field has been masked to mimic the availability of global temperature data as determined by *Mann et al.* [2008]. This collective PPE design is a simplification of real-world conditions, and represents a best-case scenario (see further discussion in the auxiliary material).

[6] The MBH98 method was applied as emulated by *von Storch et al.* [2006]. RegEM-TTLS was applied as by *Mann et al.* [2007]; we test the hybrid and non-hybrid versions of

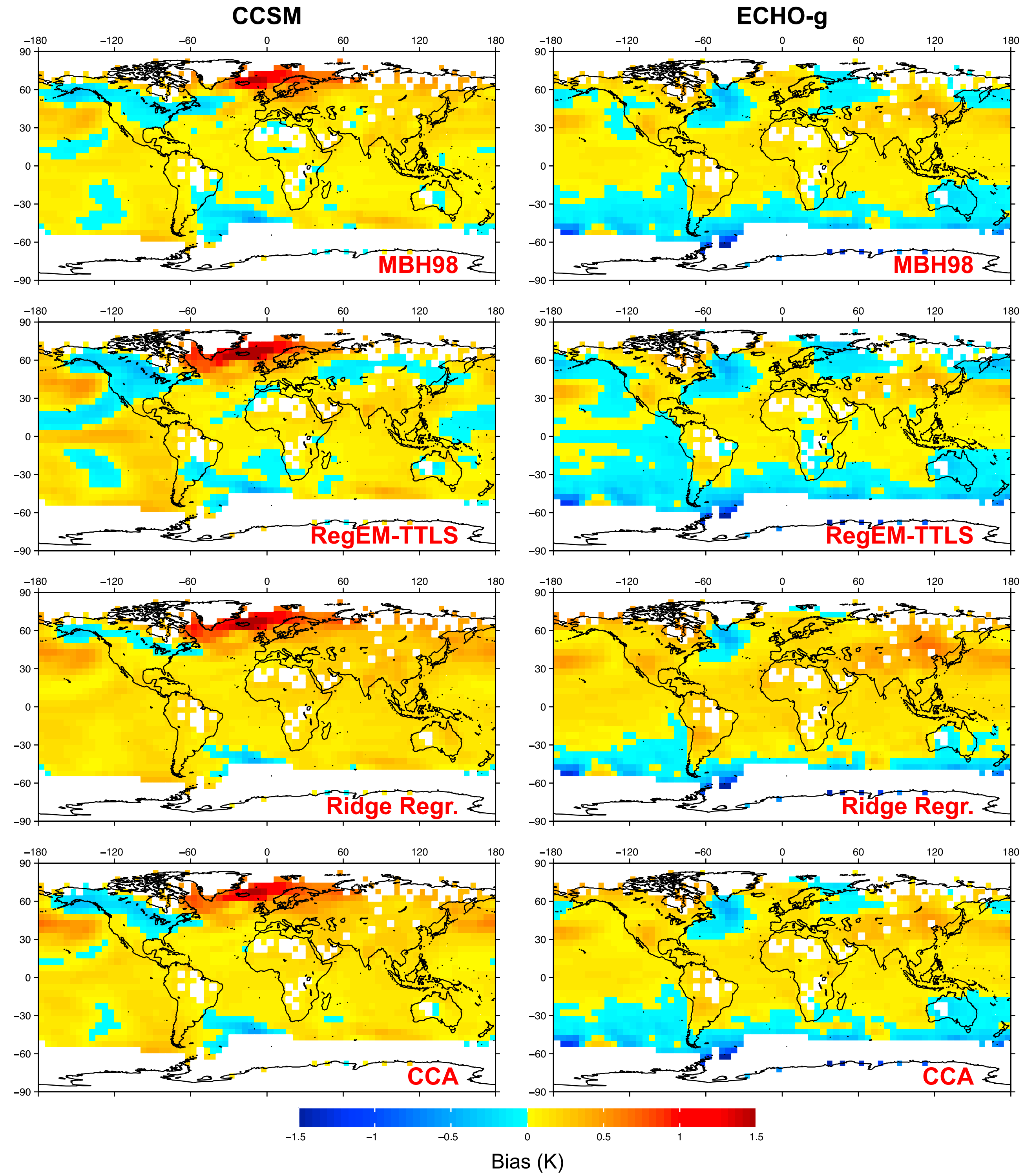

Figure 2. Same as in Figure 1, but for biases, that is, mean differences between the four non-hybrid CFRs and the true model field during the reconstruction interval.

RegEM-TTLS and find only small differences between the two choices (see results presented in the auxiliary material). Ridge regressions [*Hoerl and Kennard*, 1970] were done according to the standard formulation using singular value decomposition [*Hansen*, 1997]; minimization of the generalized cross validation function was used to select the ridge parameter [*Golub et al.*, 1979]. CCA was applied as by *Smerdon et al.* [2011]. See the auxiliary material for a detailed description of the applied methods.

3. Analysis and Results

[7] We evaluate field skill using spatially resolved correlation coefficients, biases, variance ratios and root mean

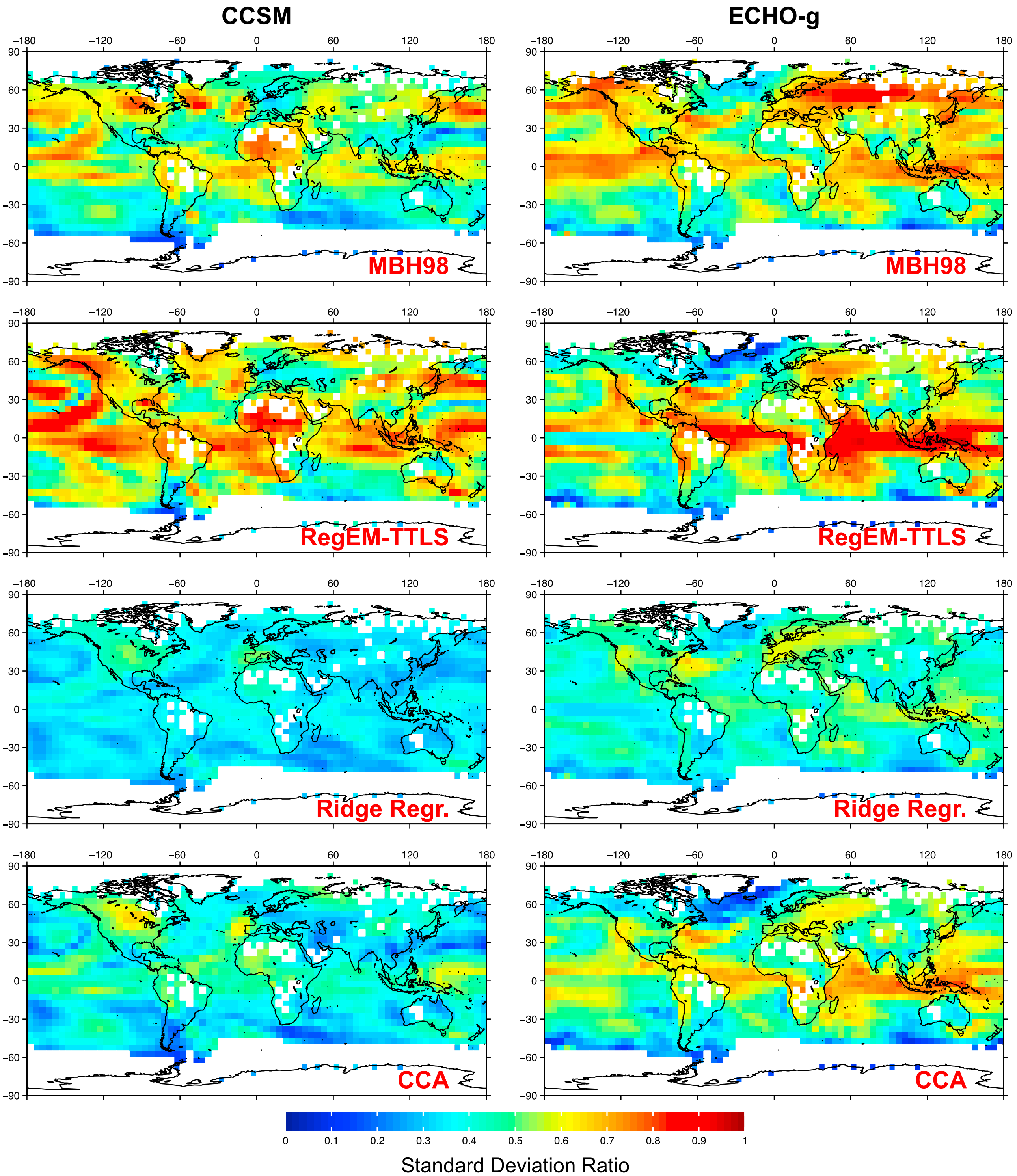

Figure 3. Same as in Figure 1, but for the ratio between the sample standard deviations estimated in the reconstruction interval of the four non-hybrid CFRs and the true model fields.

squared errors (see auxiliary material for further discussion). Correlation coefficients for all CFRs yield consistent spatial patterns in each of the CCSM and ECHO-g PPEs (Figure 1), although the patterns differ across the two models. The hybrid RegEM-TTLS CFR yields a pattern similar to the non-hybrid version (Figure S3), except in some tropical areas. In all CFRs, large correlation coefficients generally coincide with high-density pseudoproxy sampling, while low values occur over most extratropical oceans and the sparsely sampled Southern Hemisphere. This tendency also occurs in CFRs using the richer multiproxy distribution of *Mann et al.* [2008] (Figure S4): correlation coefficients increase in densely sampled areas, but are low outside of them. In some cases, relatively large correlation coefficients occur in tropical areas that are not adjacent to any pseudoproxy locations, reflecting sampled teleconnections between the tropics and midlatitudes.

[8] Spatially variable mean biases are present in all CFRs (Figures 2, S3, and S4) and largely compare to the spatial patterns in the differences between the calibration and verification interval means (Figure S5). The overall mean biases for the globe (Table S4) indicate that the MBH98 and RegEM-TTLS methods are least biased, but all methods yield CFRs with regional means or NGMs different from the target. Notably, these biases are one statistic of the hybrid RegEM-TTLS CFRs that show a marked improvement over the non-hybrid methods tested herein, even though the mean RMSE in the hybrid RegEM-TTLS CFRs is not necessarily reduced relative to other methods (Table S4). The general problem with large biases is particularly obvious in the CCSM PPEs over the North Atlantic, where regional biases greater than 1°C are observed for all methods (Figure 2). While these biases cannot be assumed to occur in real-world CFRs (nor are they nearly as large in the ECHO-g PPEs), the possibility of such regional biases requires further caution when interpreting relative warm and cold regions in real-world CFRs.

[9] Variance losses are expected for regression-based CFR methods that blend signal and error variances as a characteristic of formulation [e.g., *von Storch et al.*, 2004]. All derived CFRs suffer variance losses (Table S4), the patterns of which vary appreciably between methods and climate models, and are spatially heterogeneous (Figures 3, S3, and S4). Ridge regression and CCA display similar patterns (Figure 3), although variance losses are larger for the former method. These two methods also exhibit the well-behaved characteristic of preserving more variance in regions where correlation coefficients are largest. Such behavior is less prevalent in the MBH98 and RegEM-TTLS (hybrid and non-hybrid) CFRs, which enhance variance in small correlation coefficient areas, that is, preserved variance patterns do not match well the correlation coefficient patterns. The hybrid method does not yield a systematic improvement in these preserved variance patterns (Figure S3 and Table S4), while improved spatial sampling does (Figure S4 and Table S4).

[10] CFR skill is generally consistent across the two model PPEs, but some important differences exist. Skill is universally higher for all methods in every reported assessment metric for the ECHO-g PPEs, relative to the CCSM PPEs (Table S4). In particular, while all of the CFRs have some skill in the tropics, the skill is much higher for the ECHO-g CFRs, indicating enhanced sampling of tropical teleconnections in the ECHO-g vs. the CCSM PPEs. This difference appears to have significant impact on the NHM estimates, which are reconstructed more skillfully by all methods in the ECHO-g PPEs (Figure S6 and Table S4). Note that theses differences in the reconstructed NHMs appear more significant across the two model simulations than for those observed between the hybrid and non-hybrid RegEM-TTLS methods (Figure S7) or the two different pseudoproxy sampling schemes (Figure S8).

4. Conclusions

[11] The spatial performance of four CFR methods identifies some limits on the ability of currently employed multivariate linear CFR methods to extract information from sparse and noisy observations. No single method produced CFRs with universally advantageous characteristics, making it difficult to advocate for one method over another. NHMs were insufficient for characterizing spatial uncertainties in CFRs, indicating that spatially-resolved error metrics are necessary for evaluating CFR field skill.

[12] Our results further suggest that CFR skill will improve with new proxy sampling in currently undersampled regions and with denser replication elsewhere. Model dependencies also indicate the importance of evaluating CFR methods with multiple model-based PPEs, while rigorous comparisons between employed model fields and observed climate fields appear essential for determining the applicability of PPEs to assessments of real-world CFRs. Collectively, our findings should guide future efforts to improve large-scale CFRs through the application of new methodologies, expanded proxy networks and robust quantification of uncertainties.

[13] **Acknowledgments.** Supported in part by NSF grants ATM0902436 and ATM0902715, by NASA grant NNX09AF44G, by NOAA grants NA07OAR4310060 and NA10OAR4320137, and by the European Project Millennium. Supplementary materials can be accessed at http://www.ldeo.columbia.edu/~jsmerdon/2011_grl_supplement.html. LDEO contribution 7471.

[14] The Editor thanks the anonymous reviewer for their assistance in evaluating this paper.

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. Natl. Acad. Sci. U. S. A.*, *104*, 3713–3718, doi:10.1073/pnas.0605064103.

Christiansen, B., T. Schmith, and P. Thejll (2009), A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness, *J. Clim.*, *22*, 951–976, doi:10.1175/2008JCLI2301.1.

Golub, G. H., M. Heath, and G. Wahba (1979), Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, *21*, 215–223, doi:10.2307/1268518.

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*, L01703, doi:10.1029/2005GL024693.

Hansen, P. C. (1997), *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, *SIAM Monogr. Math. Model. Comput.*, 247 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Hegerl, G. C., et al. (2007), Detection of human influence on a new, validated 1500-year climate reconstruction, *J. Clim.*, *20*, 650–666, doi:10.1175/JCLI4011.1.

Hoerl, A. E., and R. W. Kennard (1970), Ridge regression: Biased estimation for non-orthogonal problems, *Technometrics*, *12*, 55–67, doi:10.2307/1267351.

Jansen, E., et al. (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*, edited by S. Solomon et al., pp. 433–497, Cambridge Univ. Press, Cambridge, U. K.

Lee, T. C. K., F. W. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Clim. Dyn.*, *31*, 263–281, doi:10.1007/s00382-007-0351-9.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779–787, doi:10.1038/33859.

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., et al. (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proc. Natl. Acad. Sci. U. S. A.*, *105*, 13,252–13,257, doi:10.1073/pnas.0805721105.

Mann, M. E., et al. (2009), Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*, 1256–1260, doi:10.1126/science.1177303.
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.
Smerdon, J. E., and A. Kaplan (2007), Comments on "Testing the fidelity of methods used in proxy-based reconstructions of past climate": The role of the standardization interval, *J. Clim.*, *20*, 5666–5670, doi:10.1175/2007JCLI1794.1.
Smerdon, J. E., A. Kaplan, and D. E. Amrhein (2010), Erroneous model field representations in multiple pseudoproxy studies: Corrections and implications, *J. Clim.*, *23*, 5548–5554, doi:10.1175/2010JCLI3742.1.
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:10.1175/2010JCLI4110.1.
von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett (2004), Reconstructing past climate from noisy data, *Science*, *306*, 679–682, doi:10.1126/science.1096109.
von Storch, H., E. Zorita, J. M. Jones, F. González-Rouco, and S. F. B. Tett (2006), Response to comment on "Reconstructing past climate from noisy data," *Science*, *312*, 529c, doi:10.1126/science.1121571.

M. N. Evans, Department of Geology, University of Maryland, College Park, MD 20742, USA.

J. F. González-Rouco, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Ciudad Universitaria, E-28040 Madrid, Spain.

A. Kaplan and J. E. Smerdon, Lamont-Doherty Earth Observatory, PO Box 100, 61 Rte. 9W, Palisades, NY 10964, USA. (jsmerdon@ldeo.columbia.edu)

E. Zorita, Institute for Coastal Research, Helmholtz-Zentrum Geesthacht, Max-Planck-Strasse 1, D-21502 Geesthacht, Germany.