NOTES AND CORRESPONDENCE

Comments on "Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate"

EDUARDO ZORITA

GKSS Research Center, Geesthacht, Germany

FIDEL GONZALEZ-ROUCO

Universidad Complutense, Madrid, Spain

HANS VON STORCH

GKSS Research Center, Geesthacht, Germany

(Manuscript received 25 January 2006, in final form 8 September 2006)

ABSTRACT

Mann et el. found that a version of the Regularized Expectation Maximization (RegEM) method to reconstruct the temperatures of the last millennium showed similar results to previous reconstructions in one of their earlier papers. They also tested the RegEM method in the surrogate climate of a simulation with the Climate System Model (CSM) and found no attenuation of the pseudoreconstructed centennial variability of the Northern Hemisphere mean temperature compared to the one simulated by the model. This is in contrast with the results by von Storch et al., who found, in a simulation with ECHO-G model, that the earlier Mann et al. method underestimates the centennial temperature variability of the Northern Hemisphere temperature. The newer paper by Mann et al. explains that this difference is in part due to the unrealistic character of the ECHO-G simulation. However, it is shown here that similar results to those of von Storch et al. are also found in an ECHO-G simulation that closely resembles the CSM simulation used by Mann et al. Therefore, it is argued here that this discrepancy could be related to other factors, probably to the use of a longer calibration period and to the difference between RegEM and the original method by Mann et al.

In a recent paper, Mann et al. (2005, hereafter M05) tested two different statistical methods for the reconstruction of past temperatures in the last millennium. They used the so-called pseudoproxy approach (M05 and references therein), in which the statistical method is implemented in a long climate simulations with state-of-the-art climate models. In this approach pseudoproxies, which play the role of the real proxies, are constructed by contaminating gridpoint simulated temperatures with statistical noise and applying the statistical method to estimate the target temperature. The estimated temperature can be then compared to the simulated temperature, thereby yielding an evaluation of the performance of the reconstruction method.

For a meaningful test of the reconstruction method, the constructed pseudoproxies should represent, to some level of realism, the statistical characteristics of the real proxies, such as their link to the local climate variables they intend to represent. The pseudoproxy network should also mimic the network of real proxy indicators. Finally, the climate simulation should represent a plausible realization of the past climate, considering our present uncertainties in past external forcing and in climate sensitivity. In a previous work, von Storch et al. (2004, hereafter VS04) tested the climate reconstruction method of Mann et al. (1998, hereafter MBH98) in the context of a climate simulation with the ECHAM and the global Hamburg Ocean Primitive Equation (ECHO-G) model and found that this

Corresponding author address: Eduardo Zorita, GKSS Research Center, Max-Planck-Str. 1, 21502 Geesthacht, Germany.
E-mail: eduardo.zorita@gkss.de

DOI: 10.1175/JCLI4171.1

method can considerably underestimate the low-frequency (decadal to centennial) variability of the past Northern Hemisphere temperature (NHT).

M05 used another reconstruction method, a variant of the Regularized Expectation Maximization (RegEM) method, and the same proxy network as in MBH98 and found that the reconstructed NHT is very similar to one obtained by MBH98, thus, by implication, supporting the performance of the MBH98 method. In a second step, M05 evaluate the RegEM method within a simulation of the past millennium with the model Climate System Model (CSM). In their evaluation they find no underestimation of the low-frequency variability of NHT. Therefore, there seems to be a discrepancy between the purported good performance of the MBH98 method, indirectly supported by the RegEM method, and the worse performance of the MBH98 method reported by VS04. M05 rationalize this discrepancy by arguing that the simulation with the model ECHO-G was not suitable for this test because it was performed with an unrealistic estimation of past external forcing, and because this simulation suffered from an initial disequilibrium with the imposed external forcing (Osborn et al. 2006).

We do not agree with this explanation. In the following we will try to clarify the forcing estimates of millennial external forcing used in the ECHO-G simulation and show that the issues raised about the ECHO-G simulation cannot explain the discrepant interpretation of the performance of the MBH98 method.

One assumption in the pseudoproxy approach is that the climate simulation used in the analysis has to represent a plausible realization of the past climate. One problem here is that the past climate is not precisely known. The uncertainties in the simulation of the past climate are related to the uncertainties of the past external forcing, in particular those of the solar irradiance and the volcanic forcing, and to our limited knowledge of climate sensitivity. The different reconstructions of Total Solar Irradiance (TSI) are usually labeled by their implied changes between the Late Maunder Minimum (LMM; 1680–1710) and today. According to the Third Assessment Report of the International Panel on Climate Change (IPCC; Houghton et al. 2001) most reconstructions lie within the range 0.2% and 0.4%. A still larger range (0.15%–0.65%) was also reported shortly before starting this simulation (Bard et al. 2000). The IPCC report states that "the level of understanding is very low" because of "large uncertainty" (Houghton et al. 2001). The simulation with the model ECHO-G in VS04 was performed using the reconstruction of past changes of net solar forcing provided by Crowley (2000), rescaled to transform these variations in forcing to variations of the solar constant used to drive the model. The chosen rescaling implied a change of the solar constant of 0.3% between the recent period (1960–90, in the ECHO-G simulation with mean of 1367.25 W m^{-2}) and the LMM (1680–1710, mean of 1363.23 W m^{-2}). This change is well within the uncertainty range of the IPCC. Other recent climate simulations of past centuries have used a TSI with similar variations: in the simulation with the Third Hadley Centre Coupled Ocean–Atmosphere GCM (HadCM3; Tett et al. 2007) the changes of the TSI are 0.25%, and Bauer et al. (2003) present two simulations with the Climate-Biosphere model CLIMBER with 0.24% and 0.32% changes in the TSI. However, the values used in climate simulations clearly do not cover the range of uncertainty described in the IPCC report. The amplitude of past TSI is still a much-contested topic. Lean et al. (2002), based on a solar model, argue that past TSI changes may have been much lower than assumed so far, and they proposed a value for the LMM TSI changes as low as 0.05%. However, the physical assumptions underlying these results have been contested by Solanki and Krivova (2004). The debate on this issue is certainly not closed (Scafetta and West 2005a; Lean 2006; Scafetta and West 2005b)

Concerning the volcanic forcing, the ECHO-G simulation used the estimation of volcanic radiative forcing provided by Crowley (2000). This is exactly the same as the volcanic forcing used by Bauer et al. (2003). The Crowley estimations have been later revised (Hegerl et al. 2003), but these authors acknowledge that the remaining uncertainties may be as large as 50%.

The ECHO-G simulation also neglected other external forcings, such as anthropogenic tropospheric aerosols and land used changes, that, when present, very likely contribute to cool the global average temperature. Therefore, the recent temperature trends simulated by ECHO-G in this simulations may be too large. However, we point out that the linear trend of the simulated global average near-surface temperature in the period of 1900–90 is 0.62 K $century^{-1}$, compared to a linear trend of 0.57 K $century^{-1}$ derived from the observational dataset provided by the Hadley Centre Climatic Research Unit (HadCRUT2v).

A reconstruction method should be able to perform well not only in one particular realization of the climate of the past millennium, but in a whole range of plausible situations compatible with our present uncertainty of past external forcings. The solar and volcanic forcing used in VS04 is, as explained above, well within the ranges of our present uncertainty. Also, it would seem inconsistent to make the performance of a particular reconstruction method dependent upon the existence

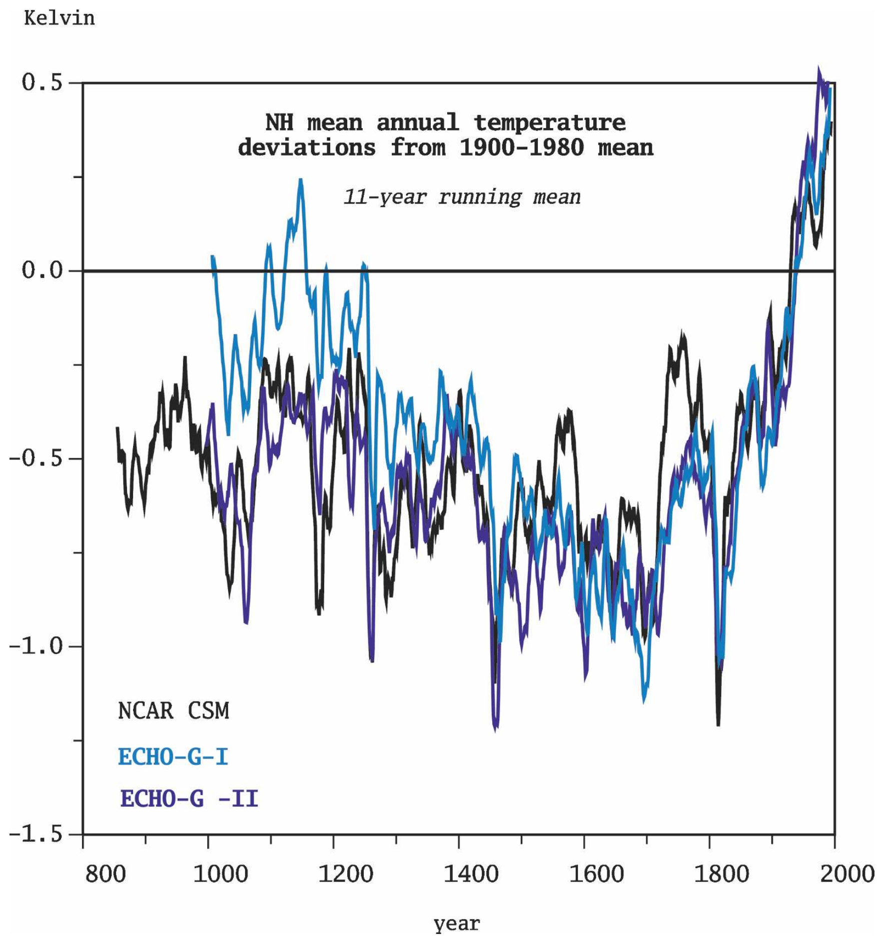

FIG. 1. Northern Hemisphere annual mean temperature in two simulations of the last millennium with the model ECHO-G started from different initial conditions and in one simulation with the CSM (Mann et al. 2005).

or absence of tropospheric aerosol forcing, particularly regarding the large uncertainties in this forcing.

Unfortunately, M05 do not specify which amplitude of solar and volcanic forcing has been used in their simulation. M05 suggest that the simulations with the models ECHO-G and CSM show a clearly different behavior, an assertion that requires further explanation. Although the simulations with the models ECHO-G and CSM have been carried out completely independently from each other, the simulated Northern Hemisphere temperature is not, in the context of testing reconstruction methods, as different as the reading of M05 may suggest (Fig. 1, black and light blue lines, respectively). The ECHO-G simulation covers the period 1000–1990. In the initial four centuries of the millennium, the simulation with ECHO-G shows higher temperatures than the CSM simulation. M05 are probably correct regarding the initial model disequilibrium in the ECHO-G simulation used in VS04, as it has been discussed by Goosse et al. (2005) and Osborn et al. (2006). To equilibrate a coupled ocean–atmosphere model to new external forcing requires probably a simulation spindown time of several centuries, or longer, to drive the various components of the climate system into a quasi-equilibrium state with the new external forcing conditions (assuming that the climate in year A.D. 1000 was indeed in equilibrium with the external forcing at that time). The spindown of the simulation with ECHO-G was started from initial conditions taken from a present-day control simulation and lasted 100 model years. After those 100 yr the external forcing attained the historical forcing at year A.D. 1000, and from that point onward the forcing was changed according to the estimates detailed above (see Fig. 1 in Osborn et al. 2006). Thus, the assertion by M05 that "the simulation was initialized from a very warm twentieth-century state at A.D. 1000" is not correct. However, the initial state before those 100 yr of spindown

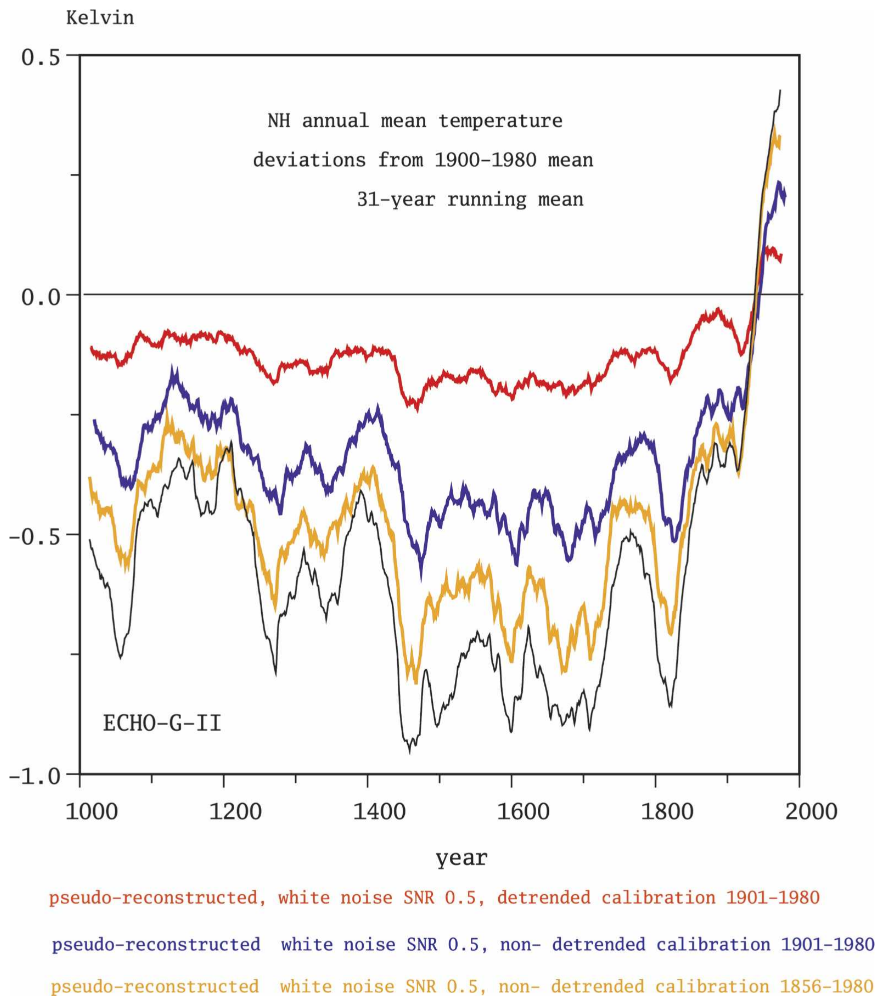

FIG. 2. Pseudoreconstructions of the annual mean NHT in the ECHO-G-II simulation, using the Mann et al. (1998) algorithm and two calibration periods, and using detrended calibration with the Mann et al. (1998) algorithm with one calibration period. The pseudoproxies are created by contamination of grid-cell temperature with white noise. In each pseudoreconstructions all pseudoproxies contain the same amount of white noise, with a local SNR of 0.5 (corresponding to a local correlation between pseudoproxy and temperature of 0.44). The pseudoproxy network is kept constant throughout the simulation, collocated to the full proxy network of Mann et al. (1998).

was likely too warm and this could have influenced the first few centuries of the simulation.

In the period of about 1400–1990, the magnitude of temperature anomalies is similar to that simulated by the CSM model (Fig. 1). It is in this period that the pseudoreconstructions of NHT shown in VS04 deviate more strongly from the target NHT, whereas in the first four centuries the pseudoreconstruction performs much better. The reason for this is probably that the warmer temperatures in the ECHO-G simulation in the first four centuries are closer to the temperatures of the calibration period (1900–80), and any statistical method should perform better when the estimations lie within the range of the calibration. Therefore, this initial disequilibrium in the ECHO-G simulation is probably not related to the reported bias of the MBH98 method reported by VS04, since in this initial period the MBH98 method performs well.

A twin simulation also of the period 1000–1990 (denoted here as ECHO-G-II), with the same external forcing, spun down from colder initial conditions, has now been completed (González-Rouco et al. 2006; Fig. 1, dark blue line). In this new simulation, the NHT around A.D. 1100 is colder than in the original ECHO-G simulation, consistent with the analysis of the original ECHO-G simulation by Osborn et al. (2006). Remarkably, this second simulation with the model ECHO-G is similar to the CSM simulation along the whole millennium (Fig. 1). Interestingly, the warming trends in the last two centuries are of comparable magnitude in spite of ECHO-G not making use of anthropogenic sulfate aerosols and CSM having lower climate sensitivity (M05, p. 4098). Without information about the forcing used by M05, this point cannot be clarified.

We think that differences in the reconstruction bias reported by VS04 and M05 could be explained by factors different from those proposed by M05. First, the reconstruction method is different (MBH98 versus RegEM); second, M05 originally performed a nondetrended calibration of their statistical model for the estimation of the regression parameters, whereas VS04, in contrast, detrended the predictors and predictands in the calibration period (Wahl et al. 2006; von Storch et al. 2006). This step was introduced by VS04 to avoid, in the real world, the regression parameters being influenced by a common, but perhaps not physically related, trend. In the reconstruction step, the original (nondetrended) proxy indicators were used. It has been shown previously that the detrended calibration reduces the low-frequency variance of the pseudoreconstruction (von Storch et al. 2006; Zorita and von Storch 2005). Third, in the most realistic case of pseudoproxies constructed with a signal-to-noise ratio (SNR) of 0.5, M05 used a longer calibration period (1856–1980) than in VS04, who used the period 1900–80 (as originally used in MBH98).

In the following, we describe the influence of the last two factors on the final pseudoreconstruction of NHT by implementing the MBH98 method in the ECHO-G-II simulation, the one more similar to the CSM simulation. Among the cases shown by M05 we consider here the case with an SNR of 0.5, since it corresponds to a realistic correlation of 0.44 between pseudoproxy and local temperature. The results are depicted in Fig. 2. It can be seen that the method, both with detrended and nondetrended calibration in 1900–80, shows a considerable bias, albeit in the nondetrended calibration case this bias is smaller. These conclusions are equivalent to those obtained with the first ECHO-G-I simulation (von Storch et al. 2006). Therefore, the use of the ECHO-G-I simulations cannot explain the discrepancy between M05 and VS04.

The influence of the longer calibration period is also apparent. With nondetrended calibration in the period 1856–1980, the performance of the method clearly improves. We have not implemented the RegEM method, and therefore we cannot ascertain whether the RegEM method also shows this dependency on the length of calibration period with noisier pseudoproxies, but it would be useful to test whether or not this method shows a similar dependency when pseudoproxies with SNR of 0.5 are considered. This could corroborate or eliminate this reason as a source of the discrepancy between M05 and VS04.

Finally, we would like to indicate that, in our opinion, the issue of a detrended or nondetrended calibration of the regression method and the issue of a white or red noise model (or even more complicated noise models) for the pseudoproxies are interrelated. The real proxies could possibly exhibit trends that are not necessarily related to temperature trends. These proxy trends may be caused by other environmental reasons (e.g., nutrients) or other climate forcings (e.g., precipitation). A nondetrended calibration should, therefore, include this possibility in the pseudoproxies. One way to achieve this, when nondetrended calibration is used, is to construct the pseudoproxies with red noise with a certain level of lagged autocorrelation. We will not pursue this aspect here, as it is beyond the frame of this comment (Zorita and von Storch 2005).

Acknowledgments. This work was partially funded by the European Union through the project SO&P (EVK2-CT-2002-00160) and by the Spanish Ministerio de Educación y Ciencia through Grant REN-2002-04584-CLI.

REFERENCES

Bard, E., G. Raisbeck, F. Yiou, and J. Jouzel, 2000: Solar irradiance during the last 1200 years based on cosmogenic nucleids. *Tellus,* **52B,** 985–992.

Bauer, E., M. Claussen, V. Brovkin, and A. Huenerbein, 2003: Assessing climate forcings of the Earth system for the past millennium. *Geophys. Res. Lett.,* **30,** 1276, doi:10.1029/2002GL016639.

Crowley, T. J., 2000: Causes of climate change over the past 1000 years. *Science,* **289,** 270–277.

González-Rouco, 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.

Goosse, H., T. J. Crowley, E. Zorita, C. M. Ammann, H. Renssen, and E. Driesschaert, 2005: Modeling the climate of the last millennium: What causes the differences between simulations? *Geophys. Res. Lett.,* **32,** L06710, doi:10.1029/2005GL022368.

Hegerl, C., T. J. Crowley, S. K. Baum, K.-Y. Kim, and W. Hyde, 2003: Detection of volcanic, solar and greenhouse gas signals in paleo-reconstructions of the Northern Hemispheric temperature. *Geophys. Res. Lett.,* **30,** 1242, doi:10.1029/2002GL016635.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis.* Cambridge University Press, 881 pp.

Lean, J. L., 2006: Comment on “Estimated solar contribution to the global surface warming using the ACRIM TSI satellite composite” by Scafetta and B. J. West. *Geophys. Res. Lett.,* **33,** L15701, doi:10.1029/2005GL025342.

——, Y.-M. Wang, and N. R. Sheeley, 2002: The effect of increasing solar activity on the sun’s total and open magnetic flux during multiple cycles: Implications for solar forcing of climate. *Geophys. Res. Lett.,* **29,** 2224, doi:10.1029/2002GL015880.

Mann, M. E., R. S. Bradley, and M. K. Hugues, 1998: Global scale temperature patterns and climate forcing over the past six centuries. *Nature,* **392,** 779–787.

——, 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.

Osborn, T. J., S. C. B. Raper, and K. R. Briffa, 2006: Simulated climate change during the last 1000 years: Comparing the ECHO-G general circulation model with the MAGGIC simple climate model. *Climate Dyn.,* **27,** 185–197.

Scafetta, N., and B. J. West, 2005a: Estimated solar contribution to the global surface warming using the ACRIM TSI satellite composite. *Geophys. Res. Lett.,* **32,** L18173, doi:10.1029/2005GL023849.

——, and ——, 2005b: Reply to comment by J. L. Lean on “Estimated solar contribution to the global surface warming using the ACRIM TSI satellite composite.” *Geophys. Res. Lett.,* **33,** L15702, doi:10.1029/2006GL025668.

Solanki, S. K., and N. A. Krivova, 2004: Solar irradiance variations: From current measurements to long-term estimates. *Solar Phys.,* **224,** 197–208.

Tett, S. F. B., and Coauthors, 2007: The impact of natural and anthropogenic forcings on climate and hydrology since 1550. *Climate Dyn.,* **28,** 3–34.

von Storch, H., E. Zorita, J. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett, 2004: Reconstructing past climate from noisy data. *Science,* **306,** 679–682.

——, ——, ——, F. González-Rouco, and S. F. B. Tett, 2006: Response to the comment by Wahl, Ritson and Amman “Reconstruction of century scale temperature variations.” *Science,* **312,** 529.

Wahl, E. R., D. M. Ritson, and C. M. Ammann, 2006: Comment on “Reconstructing past climate from noisy data.” *Science,* **312,** 529b.

Zorita, E., and H. von Storch, 2005: Methodological aspects of reconstruction non-local historical temperatures. *Mem. Soc. Astron. Italiana,* **76,** 794–801.