Response to Comment on "Reconstructing Past Climate from Noisy Data"

Hans von Storch,[1]* Eduardo Zorita,[1] Julie M. Jones,[1] Fidel Gonzalez-Rouco,[2] Simon F. B. Tett[3]

We implemented a proxy-based method for reconstructing temperatures in the past millennium in simulations with two climate models using the pseudoproxy approach. We show results for detrended and nondetrended calibration using white-noise and red-noise pseudoproxies with realistic noise levels. In all cases, the method underestimates the low-frequency variability of the simulated Northern Hemisphere temperature.

In their comment (*1*), Wahl *et al*. (WRA06) analyze reconstructions of Northern Hemisphere temperature (NHT) based on real-proxy data [MBH98 (*2*)] and our reconstructions in the surrogate climate simulated by the model European Centre Hamburg 4–Hamburg Ocean Primitive Equation–G (ECHO-G) in the past millennium, denoted as pseudoreconstructions based on pseudoproxies [VS04 (*3*)]. They indicate that, contrary to MBH98, VS04 used a detrended calibration and that the validation skill of the reconstruction method should have alerted VS04 about such differences in the calibration process. Finally, WRA06 support the use of nondetrended calibration in the design of statistical climate reconstruction methods.

First, WRA06 are correct in that we implemented the method with detrended rather than nondetrended calibration, and that therefore our original analysis did not test the specific reconstruction method of MBH98. We show here that this difference is, however, not relevant for our main conclusion, although it does affect the magnitude of the reduction in variance. Second, the validation statistics used by MBH98, the reduction of error (RE), is not able to reject the implementation with detrended calibration. Finally, we argue that a nondetrended calibration is only permissible in very special circumstances, which are not fulfilled by many proxy indicators, and that if nondetrended calibration is used, the reconstruction method has to be tested with pseudoproxies containing random long-term trends.

The key issue from WRA06 is whether the differences in the final NHT pseudoreconstruction brought about by the use of detrended or nondetrended calibration are important or not. We answer this question by implementing the MBH98 reconstruction method with detrended and nondetrended calibration and seeing whether the bias of the MBH98 method, reported by VS04, is slightly or considerably reduced. This implementation of the MBH98 method with nondetrended calibration has already been published by two different groups in three papers (*4*–*6*), with coincident results.

For illustration, Fig. 1 shows pseudoreconstructions of the NHT in two climate simulations, with the ECHO-G model (for the last millennium) and the third Hadley Centre coupled model (HadCM3) (for the last 250 years), using the MBH98 method with detrended and nondetrended calibration. We assume a realistic, even optimistic, level of white noise in the pseudoproxies of 75% (*4*) and use a pseudoproxy network colocated with the complete proxy network of MBH98. To illustrate the uncertainty range, the median and the 5% to 95% range of a set of 100 Monte Carlo (MC) realizations are shown.

[1]Institute for Coastal Research, GKSS Research Centre, Geesthacht, Germany. [2]Department of Astrophysics and Atmospheric Physics, Universidad Complutense, Madrid 28040, Spain. [3]UK Meteorological Office, Hadley Centre (Reading Unit), Meteorology Building, University of Reading, Reading, RG6 6BB, UK.

*To whom correspondence should be addressed. E-mail: storch@gkss.de

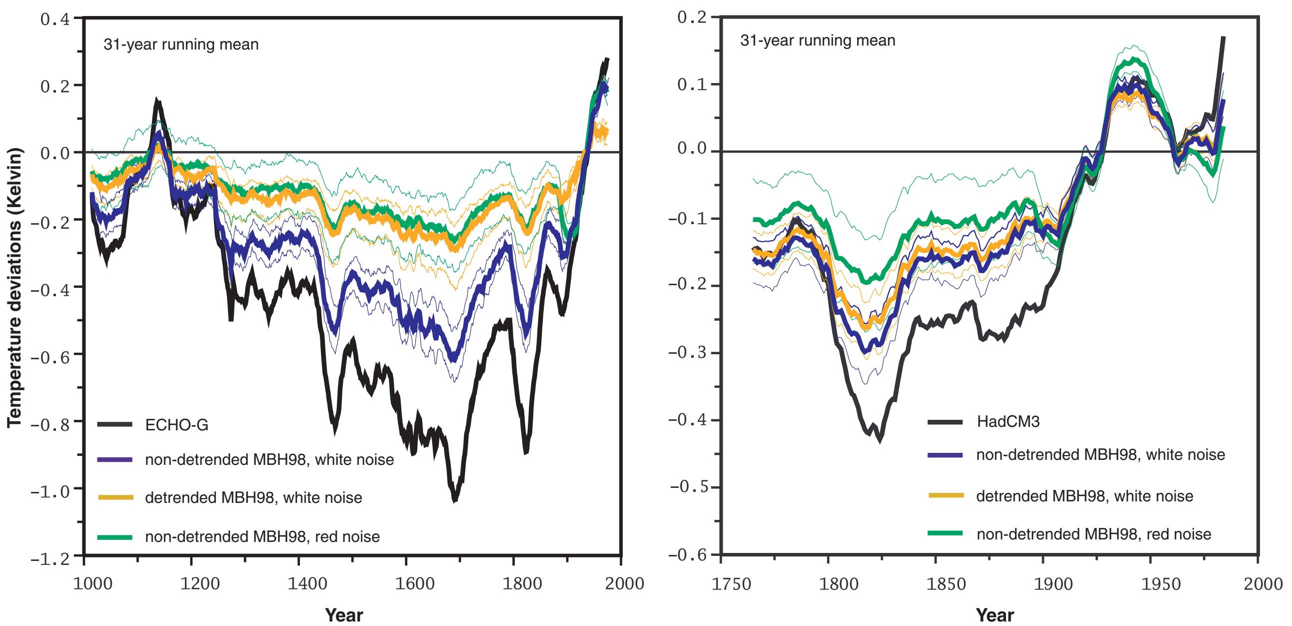

Fig. 1. Northern Hemisphere temperature deviations from the 1900 to 1998 mean, simulated and pseudoreconstructed from a network of pseudoproxies and three implementations of the MBH98 reconstruction method (*2*): with detrended and nondetrended calibration using white-noise pseudoproxies with 75% noise variance; and, additionally, with nondetrended calibration and red-noise pseudoproxies with the same amount of total noise variance, constructed from a AR-1 process with 0.7 1-year autocorrelation. One hundred Monte Carlo realizations of the noise were used to estimate the median and the 5% to 95% range. Two climate models were used, ECHO-G (**left**) and HadCM3 (**right**). Scale on the right is half that on the left.

In the ECHO-G simulation, the bias in the pseudoreconstruction relative to the target NHT is smaller in the case of nondetrended calibration, but still considerable. In this case, the 20th-century trend in the target NHT is replicated. The 5% to 95% range is consistent with the latest estimation of the real uncertainty, for example, around the Late Maunder Minimum (~1700 A.D.) of about ±0.08 K (*7*). It is evident from Fig. 1 that the difference between the target and the pseudoreconstructed NHT is clear and significant. We have also tested the MBH98 method with nondetrended calibration, but allowing for random trends in the pseudoproxies to mimic possible nonclimatic trends. This is achieved by constructing the pseudoproxies with red noise (AR-1 process with 1-year lag autocorrelation of 0.7; the total level of noise remains unchanged). The bias of the pseudoreconstructed NHT increases again relative to the pseudoreconstructions with white noise (Fig. 1).

In the HadCM3 simulation, with a smaller 20th-century trend probably due to the anthropogenic aerosol forcing, the bias is smaller and the differences between detrended and nondetrended calibration are also smaller. However, all MC pseudoreconstructions clearly underestimate the variability of the target NHT. Therefore, the bias resulting from the reconstruction method is still present, and our main original conclusion in VS04 remains valid.

MBH98 used the RE statistics (which at the time they termed "resolved variance") to validate the skill of the reconstruction method with real proxy indicators. They report a RE value for the NHT of 0.7 to 0.8, both in the calibration (1901 to 1980) and in the validation (1856 to 1900) periods, in the case of their complete proxy network. The RE statistic is calculated with annual means of the observed and reconstructed NHT, and therefore it is only partially influenced by a mismatch in the long-term trend. The 2σ range of the validation-period RE statistic in the ECHO-G pseudoreconstructions with white-noise pseudoproxies is 0.65 to 0.81 (nondetrended calibration) and 0.26 to 0.61 (detrended calibration). Both statistics can be considered high, and it would be hard to reject, on these grounds alone, a statistical method that yields such a validation RE. For instance, for the proxy network available back to 1600, MBH98 report a RE value for the NHT of 0.53 in the validation period. Therefore, from these RE values alone both methods should be considered valid. The corresponding ranges in the HadCM3 pseudoreconstructions are 0.52 to 0.74 (nondetrended calibration) and 0.57 to 0.76 (detrended calibration).

It is commonly accepted that proxy indicators may contain nonclimatic trends. This is particularly true with tree-ring data (*8*), which were intensively used in the study by MBH98. The calibration and validation of any statistical method using nondetrended data are dangerous, because the nonclimatic trends are interpreted as a climate signal. Only in the case that the trend in the proxy indicators can be ascertained to be of climate origin is a nondetrended calibration and validation permissible. In realistic circumstances, however, it can lead to an overfitting and lack of skill outside the calibration period. In this respect, the observed and reconstructed NHT shown in figure 1A in (*1*) only agree in the period with a large linear trend (centered in 1930). In the validation period, in contrast, the correlation between the (5-year-smoothed) reconstructed and observed NHT in the validation period 1856 to 1900 is 0.23. This low correlation skill in the validation period has been recently acknowledged (*9*). Furthermore, whenever the observed NHT deviates from the centennial linear trend (e.g., around 1950) the reconstructed NHT does not follow the observed temperature. In our opinion, these are indications of a dangerous nondetrended calibration.

References and Notes

1. E. R. Wahl, D. M. Ritson, C. M. Ammann, *Science* **312**, 529 (2006); www.sciencemag.org/cgi/content/full/312/5773/529b.
2. M. E. Mann, R. S. Bradley, M. K. Hughes, *Nature* **392**, 779 (1998).
3. H. von Storch *et al.*, *Science* **306**, 679 (2004).
4. G. Bürger, U. Cubasch, *Geophys. Res. Lett.* **32**, L23711 (2005).
5. G. Bürger, I. Fast, U. Cubasch, *Tellus* **58A**, 227 (2006).
6. H. von Storch, E. Zorita, *Geophys. Res. Lett.* **32**, L20701 (2005).
7. S. Gerber *et al.*, *Clim. Dyn.* **20**, 281 (2003).
8. K. R. Briffa, T. J. Osborn, *Science* **284**, 926 (1999).
9. E. R. Wahl, C. M. Ammann, *Climatic Change,* in press.
10. This study was accomplished within the European Union project SO&P (EVK2-CT-2002-00160). F.G.-R. was funded by the project Spect (CGL2005-06097/CLI, Spain). S.T. was funded by the Government Met. Research (GMR) contract and SO&P. Computer time for the HadCM3 simulation was funded by the UK Department for Environment, Food, and Rural Affairs under Climate Prediction Program Contract PECD 7/12/37S.

4 November 2005; accepted 28 March 2006
10.1126/science.1121571