

Comment on “Reconstructing Past Climate from Noisy Data”

Eugene R. Wahl,^{1*} David M. Ritson,² Caspar M. Ammann³

von Storch *et al.* (Reports, 22 October 2004, p. 679) criticized the ability of the “hockey stick” climate field reconstruction method to yield realistic estimates of past variation in Northern Hemisphere temperature. However, their conclusion was based on incorrect implementation of the reconstruction procedure. Calibration was performed using detrended data, thus artificially removing a large fraction of the physical response to radiative forcing.

Retention of century-scale temperature variations in proxy-based climate reconstructions is important for understanding real-world natural climate variability and to estimate climate sensitivity. Both are fundamental benchmarks for climate model simulations used to examine human-induced climate change. A recent study by von Storch *et al.* (VS04) (1) purported to apply “as realistically as possible” the methodology of Mann *et al.* (MBH) (2, 3) to reconstruct Northern Hemisphere surface temperatures from climate model output. Comparing these emulated reconstructions [based on pseudoproxy data constructed by adding white noise to European Centre Hamburg 4–Hamburg Ocean Primitive Equation–G (ECHO-G) surface temperatures at MBH proxy sites] to the actual model temperatures, VS04 found that the MBH-style reconstructions underestimated the amplitude of true simulated northern hemisphere temperatures by a factor of up to three or more [figure 2A in (1); the exact factor depends on the amount of noise included in the pseudoproxies.]. VS04 thus reasoned that MBH could have systematically underestimated past temperature excursions by similar factors. This critique has assumed political importance, being cited in a congressional inquiry concerning the MBH reconstruction (4). It has gone unnoted that the VS04 analysis differed critically from the procedures used by MBH, which bears directly on the validity of the VS04 critique.

MBH (see Fig. 1A) calibrated proxies against time series of dominant instrumental temperatures patterns over 1902 to 1980 in a procedure guaranteeing (by construction) retention of sample mean and variance, and thus the calibration period trend (2, 3). MBH additionally validated the reconstructions over an independent time span, 1854 to 1901 (called the “verification”

period) (2, 3), during which at least mean (low-frequency) tracking of instrumental temperatures must also be demonstrated. Figure 1B shows the corresponding VS04 results, with two pseudoproxy-based estimates of the true model temperatures. The “75% noise” curve is the case from VS04 [figure 2A in (1)] that shows proxy-based reconstructions underestimating the amplitude of true ECHO-G temperatures by more than a factor of three. Although there is strong agreement in MBH between observed and reconstructed temperatures in the 1902 to 1980 calibration period, and good performance in capturing mean temperature during the verification period (Fig. 1A), the results in VS04 are very different (Fig. 1B). Large, systematic amplitude losses appear between the reconstructed and true (simulated) temperatures over both the calibration and verification periods, even though their temporal structures remain similar. In fact, the VS04 results could be closely mimicked by applying scaling factors to the ECHO-G output that reflect the amounts of noise added to construct the pseudoproxies—factors the MBH method would necessarily assimilate in calibration. The systematic amplitude losses in calibration and verification in VS04 indicate highly unsuccessful validation, which would have led to dismissal of the reconstruction results in a real-world paleoclimate analysis and clearly demonstrate a fundamental discrepancy from the MBH algorithm. Therefore, the VS04 results (1) cannot speak to the question of whether (and if so, why) the MBH procedure causes large losses of low frequency variability in climate reconstruction.

A later 2005 conference report by Zorita and von Storch (ZVS05) (5) acknowledged that VS04 had altered the MBH procedure to base their reconstructions on detrended data, training the model on year-to-year variability. ZVS05 showed results for the same analysis using non-detrended data, which calibrate and verify far more realistically [figure 3 in (5)]. These results indicate still some, but much smaller, amplitude loss in the MBH method, at most $\sim 0.2^\circ$ for the perfect pseudoproxy case (which VS04 suggest shows loss of low frequency variance “induced

by the method alone”), in relation to a total excursion of $\sim 1.3^\circ$ over the 1000-year simulation.

What causes the difference in the VS04/ZVS05 results, and is it indeed “statistically prudent” (ZVS05) to use detrended data for calibration [see also various experiments in (6)]? Calibration with detrended data artificially dampens low-frequency climate variations and largely removes effects from the most fundamental physical processes responsible for climatic changes. The MBH reconstruction recombines spatial modes of temperature variability, called “empirical orthogonal functions” or EOFs, which (more or less, given orthogonality) represent physical processes. Some modes can directly influence global/hemispheric mean temperature, e.g., the phase of El Niño–Southern Oscillation (mostly contained in EOF2 in MBH), whereas others are of more regional importance. But over past centuries and the millennium, and particularly over the 20th century, global and hemispheric temperature changes are not simply due to a recombination of internal modes of variability but largely result from externally imposed perturbations to the planet’s energy balance (7). The 20th century warming “trend,” at its core, contains necessary information for the reconstruction algorithm to identify the climate system’s primary response to large-scale radiative forcing. Removing this physical process (contained in MBH EOF1) effectively dismisses a large portion of the central physical mechanism necessary to represent climate in both pre-industrial and recent times.

Statistically, the MBH procedure allows a century-scale trend (such as the radiatively induced warming trend, or a possible linear component in the trend contributed by any other physical mode of variability) to be mathematically separated from other climatic variations. The proxy series will still calibrate against, and add weight to, all of the EOFs retained in the reconstruction with which they have a relationship. Detrending is therefore not statistically required, and in fact, will artificially dampen low-frequency signals associated with any mode of variability that contributes to EOF1 in MBH.

The VS04 results have been interpreted to cast serious doubt on the MBH reconstruction. [Note that a newer method has since been presented and evaluated (8, 9).] However, these results are in large part dependent on a detrending step not used by MBH, which is physically inappropriate and statistically not required. The take-away message for the climate community should be strong encouragement for more vigorous cross-comparisons of the various reconstruction implementations, based on real-world proxy series, model emulations, and simulated modifications to real-world data. Such a step would help eliminate unnecessary confusion that can distract from the crucial contributions of climate change research to important scientific and policy questions.

¹Environmental Studies and Geology Division, Science Center, Alfred University, Alfred, NY 14802, USA. ²Physics Department, Stanford University, Stanford, CA 94305, USA. ³Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80307, USA.

*To whom correspondence should be addressed. E-mail: wahl@alfred.edu

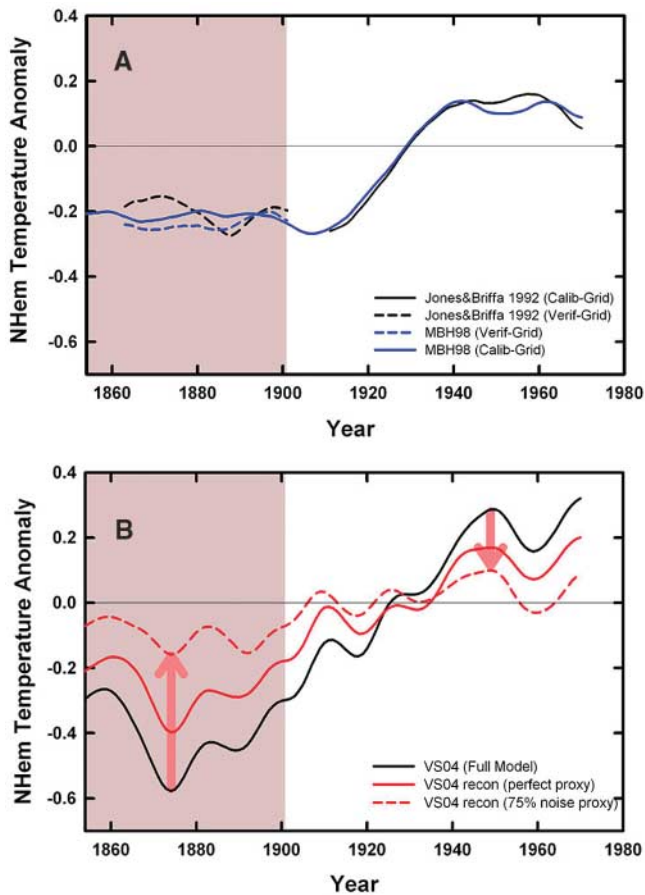


Fig. 1. MBH and VSO4 climate reconstructions. **(A)** Original MBH (2) Northern Hemisphere surface temperature instrumental data and reconstruction over calibration (1902 to 1980) and verification (1854 to 1901, shaded) periods. Curves show moving averages, smoothed with a 21-year Gaussian filter (without end extrapolations). No offset exists between the instrumental and reconstructed series during the calibration period, and the reconstructed mean in the verification period (using the spatially more restricted verification grid) misses the instrumental target by only 0.037° (annual, unsmoothed data) (10). **(B)** Like (A), but for von Storch *et al.* (1) simulated MBH-style reconstructions (using pseudoproxies from ECHO-G model output with no noise—i.e., perfect—and 75% white noise added) compared with actual ECHO-G Northern Hemisphere surface temperatures over 1854 to 1980. The large offsets in the calibration and verification (shaded) periods are highlighted by arrows.

References and Notes

1. H. von Storch *et al.*, *Science* **306**, 679 (2004); published online 30 September 2004 (10.1126/science.1096109).
2. M. E. Mann, R. S. Bradley, M. K. Hughes, *Nature* **392**, 779 (1998).
3. M. E. Mann, R. S. Bradley, M. K. Hughes, *Geophys. Res. Lett.* **26**, 759 (1999).
4. Consult www.house.gov/science/hot/climate%20dispute/index.htm.
5. E. Zorita, H. von Storch, *Mem. S.A. It.* **76**, 794 (2005). <http://sait.oat.ts.astro.it/MSAIT760405/PDF/2005MmSAI..76..794Z.pdf>.
6. G. Bürger, U. Cubasch, *Geophys. Res. Lett.* **32**, L23711 (2005).
7. T. J. Crowley, *Science* **289**, 270 (2000). 10.1126/science.289.5477.270 (2000).
8. S. Rutherford *et al.*, *J. Clim.* **18**, 2308 (2005).
9. M. Mann, S. Rutherford, E. Wahl, C. Ammann, *J. Clim.* **18**, 4097 (2005).
10. The original data are available at <http://hlocene.meteo.psu.edu/shared/research/MANNETAL98> and also as supporting material through the *Nature* online supplementary information for (2), as updated in 2004. Only the original instrumental data should be used for determining the skill of the original MBH reconstructions, because it was against these values, with their restricted but constant grid, that the statistical model of MBH was calibrated. Using a full Northern Hemisphere average based on all available information at each time step for skill analysis would introduce spatial-coverage inconsistencies, and would require recalibration.
11. We thank von Storch *et al.* (1) for providing output data of their simulations.

3 October 2005; accepted 27 February 2006
 10.1126/science.1120866