Mem. S.A.It. Vol. 76, 794

Memorie della

Methodical aspects of reconstructing non-local historical temperatures

E. Zorita and H. von Storch

Institute for Coastal Research, GKSS Research Center, Geesthacht, Germany

Abstract. The performance of two methods to reconstruct Northern Hemisphere temperature histories of the past millennium is analyzed by applying them to the known development of an extended climate model simulation. The MBH-method underestimates low-frequency variability significantly, whereas Moberg's method operates satisfactorily. Some caveats apply.

1. Introduction: Reconstruction approaches

When reconstructing historical temperatures from proxies, in principle three different approaches are possible:

1. Physical inversion. i.e., inverting the equations which describe the process of the proxy formation. The most prominent example of this approach is related to the inversion of borehole temperatures (Majorowicz and Skinner 2001)
2. Inflation of time series thought to represent a magnitude proportional to the temperature history, so that their variance/trend during times with instrumental data coincides with the variance/trend of the instrumental data (Moberg et al. 2005).
3. Regression, i.e, an empirically determined linear (or possibly even non-linear) relationship which maps proxy data on temperature (Briffa et al. 2004)

Of these three, the first is conceptually the most promising and academically satisfying, as it operates with first principles and is not loaded with fitting needs and pitfalls. However, in many cases such inverted "forward" models can not be constructed, simply because adequate "forward" models do not yet exist.

Therefore, the other two methods, making use of empirical knowledge, have to be used in many cases. The former, favored by Moberg et al. (2005) assumes that the variability in the proxy data is entirely related to the variability of the relevant physical quantity (in Moberg's case to temperature). In that case, low-frequency variability is considered separately.[1] During the instrumental period the low-frequency variance is more strongly related to the trend. Statistical support to the assumption that the trends in the proxy data and in the instrumental data match can not be given, as there is only one realization of the paired proxy/instrumental trend. Instead physical arguments need to be invoked. The third method acknowledges that the proxy data do

Send offprint requests to: H. von Storch

[1] Moberg et al. (2005) have constructed two independent temperature histories, one for short term variations and one for low-frequency developments. We are concerned only with the latter.

not only include effects of the varying physical environment but possibly other non-negligible factors. Only part of the proxy-variance is related to non-local climate variability; the contamination of the climate signal in the proxy data in general depends on the time scale (Briffa and Osborn 1999). Also, only part of the non-local climate variability is archived in the proxy data; this part does not need to be stationary in time (Schmutz et al. 2000).

Regression assumes the validity of a linear equation[2]

$$y_t = \alpha \cdot x_t + \eta_t. \tag{1}$$

Here, y_t is the it predictand, i.e., the quantity which needs to be reconstructed (e.g., northern hemisphere temperature), x_t the *predictor*, i.e., the quantity used to derive the temperature, i.e, proxy data. α is a coefficient which is to be determined, and η_t is the part of the predictand, which can not be described by the predictor – it is usually considered noise. The link (1) is considered the correct link between predictand and predictor.

The coefficient α is obtained by fitting pairs of samples of y_t and x_t so that $\text{Var}(\hat{y}_t - y_t) = \text{Var}(\eta_t)$ is minimized (least square fit, see von Storch and Zwiers (2002)), but also other criteria can be used (e.g., the OLS algorithm, Thejll and Schmith (2005)). This fit operates – as always in statistics – by analyzing time series, which contain many independent "events"; by comparing the synchronous developments in both y_t and x_t the existence of a linear link as in (1) is concluded. To do so in a reliable way, the presence of many such independent events is required. Therefore, low-frequency variations and trends are usually subtracted before deriving fitting equation (1) to the data.

Then, this α is used to derive an estimate

$$\hat{y}_t = \alpha \cdot x_t \tag{2}$$

of y_t at times when y_t is unknown. The estimator $\hat{y}_t$ is favorable in terms of the error in predicting the predictand (von Storch 1999), but the variance of the estimator $\hat{y}_t$ is smaller than the variable y_t to be determined. Thus, in general, there is a loss of variance. This is meaningful, as it reflects the insight that only part of the variability of the predictand is related to the variability of the predictor, so that part of the variability of the predictand is described by the *noise* unrelated to the predictor. For the purpose of reconstructing past developments, this loss of variance is an unfortunate but unavoidable by-product of this set-up. Of course, the question is how large this loss of variability is (Wahl and Ammann 2005).

In case of **inflation**, α is determined so that $\text{Var}(\hat{y}_t) = \text{Var}(y_t)$. In this case, equation (1) is *no longer valid*. The inflation method enforces that the predicted series has the same variance as the original one, i.e., error term in (1) is assumed to be zero, or insignificant. This is a significant assumption, which will be valid only under special conditions. If the error term is not negligible, the inflation method is associated with larger errors than regression if properly fitted (von Storch 1999).

The **Mann, Bradley and Hughes reconstruction** (Mann et al. 1998a,b) of millennial temperature variations is the best known among a series of different reconstructions, after it had been acknowledged by the Third assessment Report of the Intergovernmental Panel on Climate Change (IPCC; (Houghton et al. 2001)) as a reasonable and useful analysis. The reconstruction depicts the history of the Northern Hemisphere temperature, or global temperature development as a "hockey stick", with a long, weakly downward bending "shaft" until about 1850-1900, and a steeply upward pointing "blade" in industrial times. Using the names of the authors, the method is often referred to as MBH.

The MBH method is a hybrid method, bringing together elements of multiple regression and inflation. The regression is fitted to proxy and instrumental data during the industrial times – the fit is done by inversion of a regression from characteristic temperature patterns on proxies, and is then inverted. This procedure improves the robustness of the estimation. The method contains also a *scaling*, which means that the resulting variance is

[2] For the sake of formal simplicity, we argue here only with one predictor and not with multiple predictors

inflated as to fit the variance during industrial times. A-priori it is not clear, how the errors of the method on different time scales may look like, as analytical insights are not easily at hand.

The same is true for the inflation method (Moberg et al. 2005). Therefore, the performance of the methods needs to be addressed in the virtual laboratory of a quasi-realistic climate model. In the two following sections, we will deal with these assessments.

Figure 1 displays the estimated temperature evolution suggested by MBH and by Moberg and colleagues for the past 1000 years. MBH reconstructs variations with a year-to-year resolution, while Moberg aims only at variations on time scales of 80 years and longer. Obviously, two different stories are told about the long-term temperature history of the Northern Hemisphere.

2. Climate model simulations - a laboratory for testing the performance of methods

The success of reconstruction methods depends on many aspects; the quality and informational contents of the proxy data, the reconstruction method, the validity of assumptions about stationarity and representativity. Simulations with quasi-realistic climate models, i.e., models with maximum complexity, which do not equal reality in complexity but at least approximate it to some extent, may serve as a test-bed for assessing the performance of such methods

The usage of such models for testing complex reconstructions methods have been suggested by Mann and Rutherford (2002); Zorita et al. (2003); von Storch et al. (2004). The general idea is to construct **pseudo-proxies** $S(M)$ from the simulation output M. Then, the reconstructions method R is applied to the pseudo proxies. Let us assume that the reconstruction method supposedly generates certain statistics $f(M)$ of M, say the global mean or the amplitudes of certain EOFs coefficients. The method is considered admissible if $R[S(M)] = f(M)$ for all M, or in short, if

$$RS = f. \tag{3}$$

We test with the model output – the real variables M and the pseudo-proxies $S(M)$ – if equation (3) is fulfilled with f = northern hemisphere mean.

The question is, of course, how to construct pseudo proxies. What represents a contamination of the proxies, which we need to simulate? Contaminations may be local weather phenomena not related to the near-global or continental scale parameter of interest, i.e., local short term or regional term weather events, which are not related to the large scale feature of interest (for instance, in case of water accumulation in Greenland ice cores, (Crüger et al. 2004) or in coral compositions, (Crüger et al. 2005). In case of tree rings, it can be the competition of the different species for light and nutrients; stresses related to fire, insects and - in modern times- insecticides ((Briffa 1995; Osborn and Briffa 2000). On longer scales, the process of filtering out the effects of age-dependent growth and the fertilization effect of changing land-use and enhanced atmospheric carbon level are getting significant (Briffa and Osborn 1999; Briffa et al. 1998).

Thus, a variety of different processes affect the archiving of climate variables in proxy data, with different outcomes on different time scales. Modeling the errors associated with proxy data can not cover all aspects; only effects of first order can be take in into account – at least to some rudimentary extent. If the regression method exploits year-to-year variability, the addition of white noise – as in Mann and Rutherford (2002) or von Storch et al. (2004) – is adequate. If, however, low frequency/trend co-variations in proxy-data and instrumental data are exploited, then a more complex error model is needed, which deal with errors on decadal and centennial time scales as well.

In the following we use the outcome of a 1000-year simulation with the climate model ECHO-G as our laboratory to test the reconstruction methods. The model has been sub-

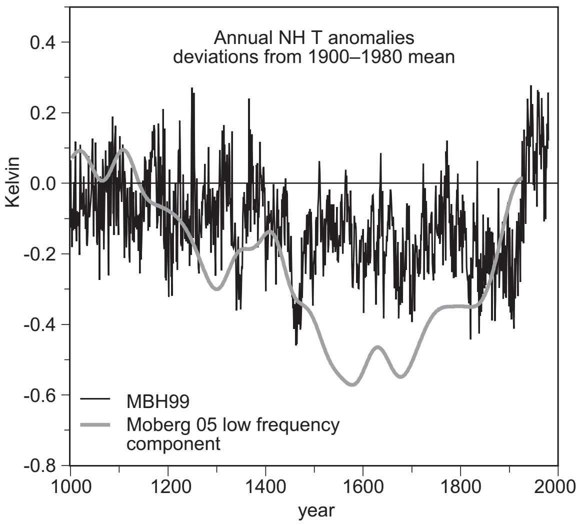

Fig. 1. Reconstruction of Northern Hemisphere temperature from MBH (solid) and Moberg et al. (dashed; only low-frequency part). Note that MBH is estimating the temperature history with annual resolution, whereas Moberg et al. low-frequency reconstructions deals only with time scales of 80 years and longer.

ject to variable solar and volcanic forcing as well as greenhouse gas forcing, which have been estimated from depositions in ice cores and sun-spots observations. The temperature variability of this simulation is relatively large, with a downward development from a warm period around 1100 to an absolute minimum during the Late Maunder Minimum at about 1700, and a recovery until modern times, with accelerating warming in the last decades. The Northern Hemisphere temperature difference between the Late Maunder Minimum and the 1990s is about 1.3K. the model sensitivity (equilibrium global temperature change under CO_2 doubling) is about 2.5K (Figure 2). For further details refer to Zorita et al. (2005). Also a control run without any external forcing is available ((Zorita et al. 2003).

3. Results: The MBH method

The MBH method has been tested in two versions, one with using detrended data during the calibration period, and another one which makes use not only of the year-to-year variability but also of the trends in instrumental and proxy data during the calibration period.

The regression/inflation model of MBH is trained with proxy and instrumental data from 1900-1980. Thus, all data are considered as deviations from the 1900-1980 mean value, even though this period is not representative for the temperature history of the past millennium. This is, however, part of the problem of deriving a transfer function from the limited evidence available in the 20th century.

We have simulated MBH's approach by first selecting a set of grid point temperature time series. These are randomized (i.e., a noise component is added), and then the MBH method is trained with both the local randomized temperature (the pseudo-proxies)and accurate temperature development in the 20th century (von Storch et al. 2004).

Different levels of noise are administered. For comparison, also the case without adding noise is considered. Then Gaussian white and red noise[3] are added. The white noise gives rise to a signal-to-noise ratio $sn = \sqrt{1/3}$ for year-

[3] "Red noise" n_t is generated by an auto-regressive process of first order, i.e, $n_{t+1} = \lambda \cdot n_t + \epsilon_t$; the de-correlation time is $\frac{1+\lambda}{1-\lambda}$. for further details, refer to von Storch and Zwiers (2002).

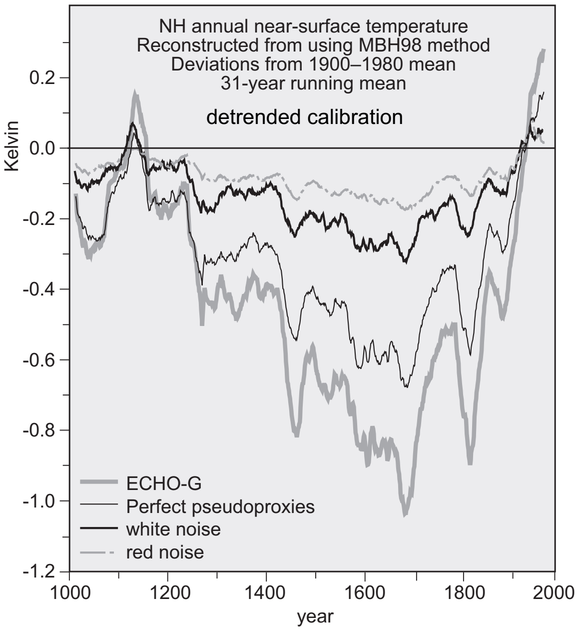

Fig. 2. Testing the MBH method, **without** exploiting the trends during the industrial period, in the laboratory of ECHO-G.Shown are different Northern Hemisphere temperature histories - as simulated in ECHO-G, as reconstructed with non-randomized local temperatures, with white noise pseudo-proxies ($sn = \sqrt{1/3}$ for short term variations, $sn = 2$ for long-term variations) and red noise (de-correlation time of 4 years; $sn = \sqrt{1/3}$ for short term variations, $sn = 1$ for long-term variations)

to-year variations and to $sn = 2$ for centennial scales; the red noise, with a de-correlation time of 4 years, is associated with $sn = \sqrt{1/3}$ and 1.

In the first case (Figure 2), we train the model only the with year-to-year variability, i.e., the trend is subtracted prior to the fit of the MBH regression/inflation model (von Storch et al. 2004). We think this is statistically prudent: the inclusion of trends means to rely on information with one degree-of-freedom, with no option to determine the uncertainty of the link between the trends. It seems, however, that MBH have exploited the trends; therefore we deal with the un-detrended case as well (Figure 3).

In all cases, whether the calibration was done after de-trending the data or not, the MBH-method leads to a marked underestimation of the real low-frequency variability (Figures 2, 3). When the data are not-detrended, the fit during the calibration period exhibits a too weak upward trend. With no randomization, the temperature difference between Late Maunder Minimum and modern times is reduced 0.9K, with white noise to 0.4K and with red noise to about 0.2K (Figure 2). When the fit exploits the link between proxy and instrumental observation trends in the 20th century, the underestimation is a bit weaker, namely 1.1K, 0.8K and 0.6K. (Figure 3). The figures demonstrate that the underestimation is uniform and physically significant.

Likely, the underestimation is related to the fact that all data are centered relative to the 1900-1980 mean, so that the variations in the training period are relatively small compared to the numbers prior to the instrumental period. the effects gets considerably smaller, if the training period considers also data from, for

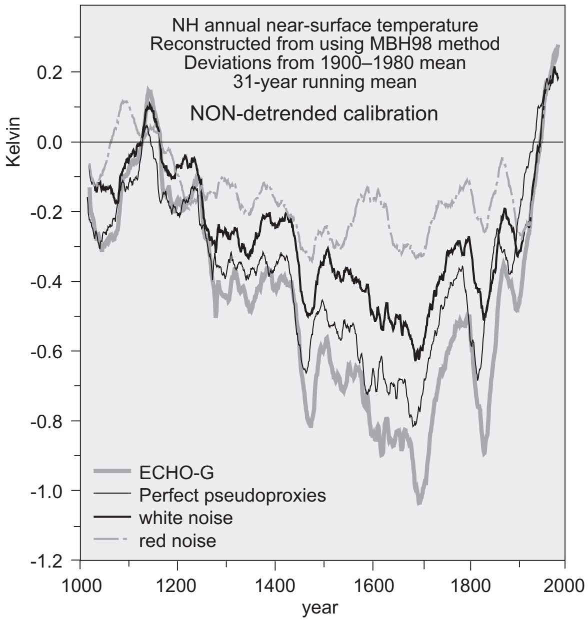

Fig. 3. Testing the MBH method, **with** exploiting the trends during the industrial period, in the laboratory of ECHO-G.Curves as in Figure 2

instance, the Little Ice Age (von Storch et al. 2004).

Obviously, one should not generalize the results here; there are a number of free parameters to be chosen, in particular the level of the noise and the redness of the noise. Also, it is likely that in the real world the noise will not behave nicely Gaussian and stationary.

4. Results: Moberg et al.'s approach

We have simulated Moberg's low-frequency approach by first selecting a series of grid point temperature time series co-located with the low-frequency proxies used by (Moberg et al. 2005). These time series are time-filtered so that only variability longer than 80 years is retained. Then, centennial noise, with a signal-to-noise ratio of 1 is added; this is done in the Fourier-domain. All series are standardized and averaged. Finally the variance is rescaled so that the variance in the period 1856-1990 is set to be equal to the model North Hemisphere temperature variance in the same period. The latter is Fourier-filtered so that only scales longer than 80 years are retained.[4]

The result of this exercise is shown in Figure 4 – several random cases as light lines and an average across all random cases as grey line. The reconstructions do not systematically over- or underestimate low-frequency variability; instead the method operates without obvious biases and reproduces the low-frequency variability faithfully. However, in the periods with lowest temperatures (Late Maunder Minimum, the method shows a slight overestimation of the temperature anomalies.

We have simulated Moberg's low-frequency approach by first selecting a series of grid point temperature time series. These time series are time-filtered so that only variability longer than 80 years is retained. Then,

[4] In Moberg et al. (2005) the instrumental data only variability shorter than 4 years is suppressed. This means that the 80-year and longer variance is compared to variance on a broader scale band – a simulations shows that this leads to an overestimation.

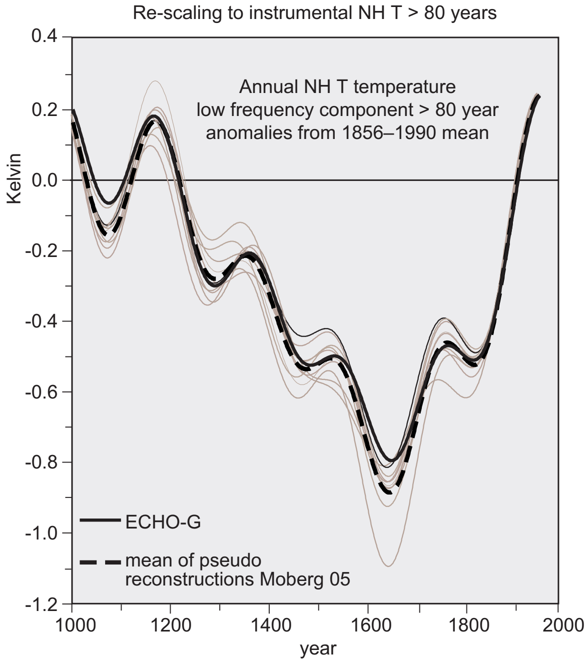

Fig. 4. Testing the Moberg method in the laboratory of ECHO-G. The inflation is done using 80-year filtered data. Solid: Northern Hemisphere temperature simulated in ECHO-G Light: Several reconstructions obtained when feeding Moberg's method with local temperature records, which have been contaminated with centennial noise (signal-to-noise ratio: 1)Dashed: Ensemble average of light lines.

centennial noise, with a signal-to-noise ratio of 1 is added; this is done in the Fourier-domain. All series are standardized and averaged. Finally the low-frequency variance during the 20th century is set to be equal to the model global mean temperature of the last century. The latter is Fourier-filtered so that only scales longer than 80 years are retained.[5]

The result of this exercise is shown in Figure 4 – several random cases as light lines and an average across all random cases as grey line. The reconstructions do not systematically over- or underestimate low-frequency variability; instead the method operates without obvious biases and reproduces the low-frequency variability faithfully.

[5] In Moberg et al. (2005) the instrumental data only variability shorter than 4 years is suppressed. This means that the 80-year and longer variance is compared to variance on a broader scale band – a simulations shows that this leads to an overestimation.

5. Concluding Remarks and Caveats

Because of lack of space, we have dealt only with the best guess (conditional expectations) of the reconstructions, not with the expected error.

The suggested approach deals only with the methodological issues of deriving low-frequency non-local climate variability from a set of proxies, which are *stationarily related to their climatic environment*. In case of the MBH reconstructions other problems with the quality and stationarity of the proxy data may, or may not, prevail (McIntyre and Mckitrick 2005).

In our analysis we have tried to simulate the real situation; this goal has certainly be achieved only to a limited extent. We had to assume that both the errors in the proxy data and the contaminations of the link with instrumental data take the form of stationary Gaussian random variables. Likely, this assumption is not well fulfilled. It is plausible that the process of archiving of climate data in proxy data was disturbed in an irregular, non-random manner by various factors; also the climatic information archived in the proxies will quite possibly have undergone significant variations. This means, that the skill of the reconstructions methods is likely overestimated.

Our reconstructions of the past, based on proxy data, are likely less realistic than we usually are willing to admit.

The best way to – at least: partially – overcome this problem is to replace the statistical methods by inverted process-based models as in case of borehole temperatures, i.e., the first option listed at the begin of this article (Weber and von Storch 1999).

Acknowledgements. We are grateful for useful discussions with Gerd Bürger and Irina Fast, and with Dave Ritson. We enjoyed comments by Michael Mann and Steve McIntyre.

References

Briffa, K.R. 1995, In: H. von Storch and A. Navarra (eds) *Analysis of Climate Variability: Applications of Statistical Techniques*, Springer Verlag, 77

Briffa, K.R., & Osborn, T.J. 1999, Science, 284, 926

Briffa, K.R., Schweingruber, F.H., Jones, P.D., Osborn, T.J., Shiyatov, S.G., & Vaganov, E.A. 1998, Nature, 391, 678

Briffa, K.R., Osborn, T. J., & Schweingruber, F. H. 2004, Global and Planetary Change, 40, 11

Crüger, T., Fischer, H., & von Storch, H. 2004, J. Geophys. Res. 109, D21110, doi:10.1029/2004JD005014

Crüger, T., Kuhnert, H., Pätzold, J., & Zorita, E. 2005, in preparation

Houghton, J.T., Ding, Y., Griggs, D.J., et al. 2001, Climate Change 2001: The Scientific Basis., Cambridge University Press, 881

Majorowicz J.A.,& Skinner, W.R. 2001, Climate Research, 16, 157

Mann, M.E., Bradley, R.S., & Hughes, M.K. 1998, Nature, 392, 779

Mann, M.E., Bradley, R.S., & Hughes, M.K. 1998, Nature, 430, 105

Mann, M.E., & Rutherford, S. 2002, Geophys. Res. Lett., 29, 139, 1

McIntyre, S., & Mckitrick, R. 2005, Energy and Environment, 16, 69

Moberg, A., Sonechkin, D.M., Holmgren, K., et al. 2005, Nature, 433, 613

Osborn, T.J., & Briffa, K.R. 2000, Dendrochronologia, 18, 9

Schmutz, C. J., Luterbacher, J. , Gyalistras, D., et al. 2000, Geophys. Res. Lett., 27, 1135

Thejll, P., & Schmith, T. 2005, J. Geophys. Res., in press

von Storch, H., & Zwiers, F.W. 2002, Statistical Analysis in Climate Research, Cambridge University Press, ISBN 0 521 45071 3, 494

von Storch, H. 1999, J. Climate 12, 3505

von Storch, H., Zorita, E., Jones, J., et al. 2004, Science, 306, 679

Wahl, E.R., & Ammann, C.M. 2005, submitted

Weber, S.L., & von Storch, H. 1999, EOS 80, 380

Zorita, E., González-Rouco J. F., & Legutke, S. 2003, J. Climate 16, 1378

Zorita, E., González-Rouco, F., von Storch, H., Montavez, J.P., & Valero, F. 2005, Geophys. Res. Letters, 32, L08707