Final Draft
of the original manuscript:

Storch, H.v.; Zorita, E.; Gonzalez-Rouco, F.:
Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation
In: International Journal of Earth Sciences (2008) Springer

DOI: 10.1007/s00531-008-0349-5

Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation

Hans von Storch and Eduardo Zorita [1]

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

and

Fidel González-Rouco

Universidad Complutense, Madrid, Spain

November 15, 2007

[1]Corresponding author: eduardo.zorita@gkss.de, GKSS Forschungszentrum, Max-Planck-Str.1, 21502 Geesthacht, Germany

Abstract

It is difficult to assess the performance of climate reconstruction methods in pre-instrumental periods. Here, the virtual reality created by a climate simulation of the past millennium with the model ECHO-G is used as a test bed of three methods to reconstruct the annual Northern Hemisphere temperature. The methods are Composite plus Scaling, the inverse regression method of Mann et al. (1998) and a direct principal components regression method. The testing methodology is based on the construction of pseudo-proxies derived from the climate model output, the application of each of these methods to pseudo-proxy timeseries, and the comparison of their result with the simulated mean temperature. Different structures of the noise have been used to construct pseudo-proxies, ranging from to the the simulated grid-point precipitation. Also, one sparse and one denser pseudo-proxy network, co-located with two real networks, have been considered. All three methods underestimate the simulated variations of the Northern Hemisphere temperature, but the Composite plus Scaling method clearly displays a better performance and is robust against the different noise models and network size. The most relevant factor determining the skill of the reconstruction appears to be the network size, whereas the different noise models tend to yield similar results.

1 Introduction

Climate reconstructions of the past few millennia may provide useful information of the amplitude of climate variations at centennial timescales, help to put the recent 20th century global warming into the perspective of natural climate variations (Briffa and Osborn, 2002; Jones and Mann, 2004; National Research Council, 2006), and help to identify with a larger degree of certainty the influence on global climate of anthropogenic emissions of greenhouse gases. Climate models of varying complexities, from energy balance models (EBM) to three dimensional ocean-atmosphere general circulation models (AOGCMs) offer a powerful laboratory in which to test the response of atmospheric thermodynamics and dynamics to available estimates of past external forcing levels (Jones and Mann, 2004). Integrated assessment of climate reconstructions of past climate and forcing variations in combination with model simulations of the past climate evolution provides a basis for a variety of scientific aims like detection and attribution studies of human influences on climate (e.g. Crowley 2000), narrowing estimations of climate sensitivity (e. g. Hegerl et al. 2007), validation of climate reconstruction strategies (e. g. von Storch et al. 2004).

The amplitude of variations of the past Northern Hemisphere temperature is being intensely debated. On one hand, some proxy-based reconstructions (Mann et al., 1999, Jones et al., 1998) show relatively small amplitude of variations, of the order of 0.2 K. On the other hand, more recent reconstructions indicate that the amplitude of variations may have have been larger, lying in the range 0.6-0.8 K (Esper et al., 2002,Moberg et al. 2005, Hegerl et al. 2007). Temperature reconstruction based on borehole temperature profiles seem to support a large temperature change

between present and about 500 years ago. Other evidence against small temperature variations in the past millennium is provided by the reconstructed concentrations of atmospheric CO_2 concentrations (Scheffer et al 2006). These authors argue that small temperature variations would imply a too large sensitivity of CO_2 to temperature changes (of the order of 42 ppm/K) whereas evidence from the Last Glacial Maximum and from carbon-cycle models converge to values closer to 10 ppm/K (Friedlingstein et al. 2006)

The methods applied in the proxy-based reconstructions are also different. Some of them make use of regression-based methodology and others use different variants of a simple proxy-averaging (Jones and Mann 2004; National Research Council, 2006). After calibration with data from the instrumental period, these methods are then applied to estimate past temperature variations from the longer proxy time series. The design of the statistical model is usually subject to some means of validation based on the use of independent data. This step is intended to support the application of the reconstruction model to a time interval different from that used for calibration. For this purpose, a portion of the available instrumental data is used for calibration and a portion is set aside for validation with independent instrumental data. Exceptions for this approach are the reconstructions methods by Hegerl et al. (2007) and Moberg et el. (2005) who involve among others decadally resolved proxy records in their analysis. The number of degrees of freedom (effective sample size) is considerably reduced in such situations of lower temporal resolution. Thus isolation of a short independent period becomes sometimes difficult, and other means of validation have to be undertaken. This problem is not only typical of situations making use of low resolution proxies but also of situations in which even if higher resolution (i. e. yearly) data are available, the methodologies place their emphasis on the low frequency

(decadal, multi-decadal and centennial) time scales present in the data. The idealized limit of such a situation would be the case of time series with a very large linear trend (relative to the interannual variability) in which the strong autocorrelation would virtually reduce the effective sample size to one. The nature of some of the proxies used for climate reconstructions may distort the retrieve the centennial climate variability. Tree-ring width and perhaps also tree ring density depend also on tree age. This effect has to be filtered out to retrieve the climatic signal, but methods for doing may introduce statistical uncertainties at timescales of the typical tree lifespan. Documentary information, when not anchored on the reporting of physical phenomena such as onset of lake freezing, are subject to the subjective interpretation of the chronist about the likelihood of a certain observations in a time framework longer than a typical human lifespan. Therefore, there is a need of testing statistical methods in timescales much longer than the calibration period.

The availability of model paleo-simulations allows for the application of the so called pseudo-proxy approach to by-pass the limitations mentioned above. In this type of methodological validation, the statistical reconstruction method is applied in the virtual reality of simulations with coupled climate models (Mann and Rutherford 2002, Mann et al. 2005, von Storch et al. 2004; Hegerl et al. 2007). In this set-up all variables are perfectly known and the output of the statistical method can be directly compared to the *true (simulated)* target temperature (e.g. the Northern Hemisphere temperature) and the bias and associated uncertainties of the method can be evaluated. In the case of multi-century long simulations, this technique offers the possibility of scanning the performance of the model at lower frequencies (multidecadal and centennial) than can could be possible by the use of roughly a century of instrumental data.

The main problem in this approach is, however, the need for timeseries that play the role of the real proxies. So far, the pseudo-proxies have been produced in this virtual reality simply as a combination of grid-point simulated temperature with the addition of statistical noise with some previously defined properties. The studies mentioned above have used quite simple models of the noise contained in the pseudo-proxies, namely just white noise or red noise generated by an autoregressive process. This is certainly a first step in this line of research and more efforts are needed to quantify and characterize the structure of the noise present in the different types of proxy records. This is, however, not an easy task, since it requires a deep knowledge of the behaviour of the proxy indicators at low-frequencies and their link to the local temperatures.

The pseudo-proxy approach provides a test ground for evaluating the validity of a variety of problems and hypothesis related to the methodology, the statistical properties of proxies or their distribution. An example of this is the influence of a shrinking coverage of the proxy network backward in time (Küttel et al. 2006). Normally, not all proxy indicators are available for the whole period of interest. For instance, in the study by Mann et al. (1999), the network of proxy indicators comprises 112 time series in the most recent period (1820-1980), whereas the first four hundred years of the past millennium are covered by only 12 indicators. Some authors prefer to use a proxy network of constant size, thereby considering only indicators that posses a complete, or almost complete, record in the period of interests (Crowley and Lowery 2000). However, it should be kept in mind that the quality of a single record may diminish backwards in time, if , for instance, the number of individual trees comprised in a chronology also diminishes, or, if the time resolution of the proxy becomes coarser.

In the present study, we intend to address some methodological aspects of the climate re-

construction in the past millennium using the pseudo-proxy approach. The focus lies in three different methods: the inverse-regression method proposed by Mann et al. (1998); the direct principal-components regression method, exemplified in Luterbacher et al. (2004) in their reconstructions of past European temperatures and more frequently used in a number of other studies focused on regional temperature reconstructions; and the so called Composite Plus Scaling Method (CPS, sometimes denoted also as Composite Matching Variance CMV) used by Crowley and Lowery (2000). Note that Moberg et al. (2005) applied a somewhat more sophisticated variant of the CPS method combined with a wavelet analysis.

The present analysis is therefore not exhaustive. Altogether, these three methods have been extensively used in the paleoclimate literature, but other more recent methods have been recently introduced. This is the case of Regularized Expectation Maximization (RegEM), which originally was aimed at filling missing values in a set of observations (Schneider, 2001). Regularized Expectation Maximization is a relatively more complex iterative imputation method with a linear regression in its core. It has recently been also applied for proxy-based climate reconstructions with different variants for the regression core, either ridge regression (Mann et al., 2005; Rutherford et al., 2005) or truncated total least-squares (Mann et al.,2007). The variants used for the core regression seem to have a strong influence on the final results. The code for the ridge-regression variant has been recently revised to avoid a too strong sensitivity to the length of the calibration period (Mann et al, 2005, supplementary information). However, the latest variants of RegEM seem to yield better results than previous methods (Mann et al, 2007; Riedwyl, personal communication, manuscript submitted to Climate Dynamics).

These three methods are tested here with two different pseudo-proxy networks. To keep the

results comparable to one another, we have not used the locations of the real proxy networks used by the different authors in the respective studies with real proxies. Here, only the locations of the complete network used in Mann et al. (1999) and the locations of the smallest network used by Mann et al. (1999) have been considered to define the location of a common pseudo-proxy network for all methods.

The noise models to construct the pseudo-proxies are in a first step a simple white-noise and red-noise model with spatially constant autocorrelation. Additionally, somewhat more sophisticated noise models have been implemented. The pseudo-proxies have been constructed by contamination with grid-point precipitation simulated by the climate model itself. The rationale for this noise model is that some proxy indicators, tree-rings, speleothems, ice-cores, may be representing not only temperature but also precipitation variability as well. This noise model has been included in our analysis not as a outright realistic test of the statistical methods, but just as an scenario to find out what could be the potential influence on the final reconstructions of a certain contamination by a precipitation signal in an attempt to reconstruct temperature. Finally, the noise component of the pseudo-proxies has been modeled by spatially varying autoregressive process and by spatially varying *long-range time-autocorrelation* process. For these type of processes the time-autocorrelation does not decay exponentially with time-lag, as in autoregressive processes, but decays more slowly following a power-law. Processes of this type have been identified in real records (Koutsoyiannis, 2003; Bunde et al. 2005), in particular in hydrological timeseries, although the question whether they are ubiquitous in nature is still debated. Again, this noise model is included here not as a claim that it may represent noise in real proxy indicators, for which there would be no empirical justification, but just to test its consequences for

empirical reconstruction methods. Note that the fact that such processes may have been identified in climatic records does not imply that they may also be able to represent non-climatic noise in proxies.

All these pseudo-reconstructions exercises have been performed in a climate simulation of the past millennium with the model ECHO-G. This is not the same simulation used by von Storch et al. (2004) in their pseudo-proxy analysis of the Mann et al. (1998) reconstruction method, but a simulation with the same model using different initial conditions. It has been alleged that the von Storch et al. (2004) simulation is affected in the initial centuries of the millennium by an artificial drift due to a lack of equilibrium of the global temperature with the external forcing at the start of the simulation in year 1000 A.D. (Osborn et al. 2006). Although the effect of this drift is essentially observed in roughly the first three hundred simulation years, the relevance of this drift for the purpose of testing reconstruction methods is debated. We use here a second simulation with the same model started with initial conditions taken from a time-step of the first ECHO-G simulation, after the model has had five centuries to accommodate to the varying external forcing.

A broad description of the model and the simulation used herein is presented in next section. Section 3 addresses the details of the three evaluated reconstruction methods as well as other methodological aspects like the design of the pseudo-proxies and uncertainty treatment. Finally results are presented in section 4 and discussed along with the conclusions in section 5.

2 Description of model and simulation

The global coupled climate model ECHO-G consists of the spectral atmospheric model ECHAM4 (Roeckner et al. 1999) and the ocean thermodynamic/dynamic sea ice model HOPE-G (Wolf et al. 1997). Both sub-models were developed at the Max-Planck-Institute of Meteorology in Hamburg. In the present setup, the atmospheric model has a horizontal resolution of T30 (approx. 3.75°x 3.75°) and 19 vertical levels. The ocean model HOPE-G has an effective horizontal resolution of approximately 2.8° x 2.8° with 20 vertical levels. In the tropical regions a grid refinement is employed with decreasing meridional grid-point separation, reaching a value of 0.5° at the Equator. This increased resolution allows for a more realistic representation of ENSO events (Min et al. 2005). The climate model is flux-adjusted to avoid climate drift. The flux adjustment is held constant in time and its global average is set to zero.

The model was driven by the following external forcings: total solar irradiance, volcanic forcing, and well-mixed greenhouse gases. The solar forcing was derived from the data provided by Crowley (2000) through transforming effective solar forcing to Total Solar Irradiance (TSI) units to drive the model. There still exists a large uncertainty in the amplitude of past TSI at centennial timescales (IPCC 2001). In this simulation the Crowley data were re-scaled so that the differences between the Late Maunder Minimum (1680-1710) and present (1960-1990) are 0.3 % of the TSI. The volcanic net radiative forcing was translated to changes in an affective solar constant by multiplying with the factor 4 / (1- α), where α is the planetary albedo. The volcanic forcing is thus implemented in this simulation as a global annual reduction in the solar constant. The values provided by Crowley already take into account an e-folding time

in the years following a volcanic eruption. The concentrations of atmospheric carbon dioxide and methane, as provided by Etheridge et al. (1996) and Etheridge et al. (1998) respectively, were derived from ice-core measurements. Concentrations of N2O were used as in previous scenario experiments with this model (Roeckner et al., 1999): fixed 276.7 ppb before 1860 and the historical evolution from 1860 to 1990 AD adjusted from Battle et al. (1996).

As stated in the introduction, the initial conditions for these simulations were taken from a previous simulation of the same model spanning the past millennium (von Storch et al., 2004). The atmospheric and oceanic conditions at the beginning of year 1700 AD of this previous simulations were used to start an spin-down period of 100 model years in which the external forcing values were slowly driven to those estimated for 1000 AD: fifty years of gradual transition from the 1700 AD to the 1000 AD forcing levels and fifty more years of constant conditions fixed at the 1000 AD levels. Since the difference between both forcing states is relatively small, changes in this adaptation period are minor in comparison with those in the simulation used by von Storch et al. (2004) which transited from current industrial values to 1000 AD forcing conditions in 100 model years. After this initial phase the simulation then proceeded with the historical estimations of the external forcing until year 1990. A comparison of the evolution of both millennial simulations is reported in González-Rouco et al. (2006).

The simulated annual mean Northern Hemisphere temperature is here compared with a simulation of the past 1200 years performed with the Climate System Model (CSM) of the National Center for Atmospheric research (Mann et al. 2005) in Figure 1. The NCAR coupled model is not flux-adjusted but the small climate drift in the simulation is subsequently corrected by subtracting a long-term trend estimated from a control simulation (Mann et al. 2005). The North-

ern Hemisphere temperatures evolve in a quite similar fashion along the whole millennium, although the external forcings are presumably different (no information about the CSM simulations was available to us), and the CSM simulations has additionally included the anthropogenic tropospheric aerosol forcing, which is lacking in the ECHO-G simulations. The agreement between both simulation suggests that the flux correction used in the ECHO-G model has not a relevant influence on the simulation of the temperature variations in the past millennium. Differences among a number of millennium simulations can be accounted for by the differences in the prescribed external forcing (Goosse et al., 2005). Figure 1 also evidences that the differences in the initial state only affect the ECHO-G simulations along the first two to three centuries. After 1300 AD differences between both simulations fall within the internal variability experienced in any two simulations with different initial conditions. The results within this paper will focus on this ECHO-G simulation. However, they are also supported by the simulation used by von Storch et al. (2004) in spite of its deviations in the first centuries of the millennium.

3 Statistical reconstruction methods

3.1 MBH method

The MBH method is based on an inverse regression equation between the proxy indicators and the leading Principal Components (PC) of the Northern Hemisphere temperature field (Mann et al. 1998. For the sake of completeness a brief description follows here, the interested reader can consult the original publication (Mann et al. 1998) and also Bürger et al. (2005) for details.

The statistical method used by MBH98 to calibrate the proxy indicators is quite elaborated,

but the core of this method can be condensed as follows. A linear relationship between annual proxy indicators and annual mean instrumental records is assumed:

$$P_i(t) = \sum_j a_{ij} pc_j(t) + \epsilon_i(t) \tag{1}$$

where $P_i(t)$ is the the value of the proxy indicator i at year t, $pc_j(t)$ is the value of the temperature principal component j at year t, a_{ij} are regression coefficients to be estimated, $\epsilon_i(t)$ is the part of the proxy indicator that cannot be linearly explained by the regression equation. The regression coefficients a_{ij} are estimated by minimizing the variance of ϵ_j by ordinary least-square-error minimization.

Once the regression coefficients a_{ij} have been estimated, the MBH98 method estimates the value of the temperature principal components $\hat{pc}_j(t')$ in periods outside the calibration period by, first, finding a value of $\tilde{pc}_j$ that minimises the variance of the residuals ϵ in the following equation (Mann et al 1998).

$$P_i(t') = \sum_j a_{ij} \tilde{pc}_j(t') + \epsilon_i(t') \tag{2}$$

This step can be also re-formulated in terms of the pseudo-inverse matrix containing the regression coefficients (Bürger and Cubasch 2005).

These initial estimations of the principal components are then renormalized so that they conserve the variance of the instrumental principal components in the calibration period.

$$\hat{pc}_j(t) = nc\, \tilde{pc}_j(t) \tag{3}$$

where nc is the renormalization constant. Note that, because of this renormalisation step,

$\hat{pc}_j$ is no longer the optimal solution in the sense of mean-square-error minimisation.

The estimated PCs $\hat{pc}_j$ in the whole period, together with the spatial eigenvectors, can be then back-transformed to reconstruct the temperature field and, by spatial averaging, the Northern Hemisphere mean.

It has been tested whether our implementation of the Mann et al. (1998) algorithm can reasonably replicate the original result. For illustration, our emulation is compared in Figure 2 with the original reconstructed PCs in the period 1820-1980, with the same set of real proxy indicators used by Mann et al. (1999) and the same instrumental temperature Pcs. Although not completely identical to the original result, the replication of the algorithm can be considered as sufficient, with correlations between original and replicated PCs over 0.9 for all four PCs. Other authors that have emulated the MBH98 algorithm also achieve a largely similar results but not a complete agreement (Wahl and Ammann 2007).

3.2 Direct principal-components regression

This approach is based on a more practical view of the regression equations. The goal is to reconstruct the temperature and this is the variable considered as predictand, also in the form of its principal components. The predictors are in this case the proxy indicators:

$$pc_i(t) = \sum_j a_{ij} P_j(t) + \epsilon_i(t) \tag{4}$$

. where the notation is the same as in equation 1. The regression parameters are estimated by least-mean-square error minimization in the calibration period. The estimation of the temperature PCs in the whole millennium is then accomplished by inserting the longer proxy records

into equation 4.

3.3 Composite plus Scaling

This is is a much simpler method than both previous methods. The standardized proxy indicators (mean zero and unit standard deviation unity in the calibration period) are simply averaged. The resulting dimensionless time-series is re-scaled to match the variance of the observed Northern Hemisphere temperature in the calibration period. This variance-matching step can be performed in different frequency bands. For instance, Crowley and Lowery (2000) rescale to match the observed variance at interannual timescales, whereas Moberg et al. (2005) match the variance only at timescales between 80 and 4 years. In the present study, this variance-matching is accomplished with interannual data for the results to be comparable to those obtained with the other two methods, as these methods (direct and inverse regression) also are calibrated with unfiltered (interannual) data. Within this method the final reconstruction is just proportional to the mean of the standardized proxy records, and the only adjustable parameter depends on the variance of target record. The estimation of this variance may be also subject to uncertainties. It has been previous found (Esper et al., 2005) that, in the observational record, the choice of calibration period to estimate the observed variance of the target variable can result in differences in the amplitude of the final reconstruction of up to 50% of the total amplitude.

It has been argued that a simple averaging method as CPS would be more robust than more sophisticated regression methods, as it is free of many of the technical subtleties that may burdened the latter methods (Huybers, 2005). It has been claimed by Mann et al (2005), however, that CPS method as implemented by Moberg et al. (2005) may be prone to overestimating the

low-frequency variance of the Northern Hemisphere temperature.

3.4 Calibration variants

Instrumental and proxy time-series usually exhibit strong trends in the 20th century. These trends may be deterministic, i.e. caused by an external factor that may or may be not common to temperature and proxies, or they may represent extended periods of sustained anomalies that appear in non-stationary stochastic processes or long-memory processes. In both cases, the estimation of regression parameters with trend time-series may be problematic and may give rise to the so called ”spurious regression” (Yule 1926, Granger and Newbold 1974). In these cases more sophisticated statistical models may be needed. In climatology a simple, albeit not perfect, solution to this problem has been sometimes implemented, namely the estimation of regression parameters or the estimation of correlations after linearly detrending the data. Here, the two variants of the calibration step, with detrended and non-detrended data, are shown. The detrending step is, however, debated (Ammann and Wahl, 2007). It has been argued that, although the issue of the loss of degrees of freedom when regression trendy timeseries has to be considered, the detrending step in methods that make use of a Principal Components prefiltering of the target field hinder the representation of the global energy imbalance of the 20th century. However, this argument only only be valid if the calculation of the temperature principal components had been conducted with detrended grid temperatures, which is not the case either in Storch et al. (2004) or in the present study. Here and in Storch et al. (2004) the principal components prefiltering was applied to the original grid-point temperatures and the leading temperature principal component shows a strong trend corresponding to the 20th century warm-

ing. Only in the subsequent proxy calibration is the detrending step introduced. This is also valid here for the direct principal regression method. Once the regression parameters have been estimated, the reconstruction is performed with non-detrended proxy data. For the case of the CP method the calibration variants are performed by accomplishing the variance-matching step with the linearly detrended, and alternatively not detrended, Northern Hemisphere temperature and dimensionless proxy-average series.

3.5 Pseudo-proxy network

Two pseudo-proxy networks, one large and one small, were used to test the performance of the reconstruction methods. The largest network is co-located to the proxy network used in Mann et al. (1998). This pseudo-proxy network is the same as in the study by von Storch et al. (2004) using ECHO-G as laboratory. With the resolution of this model, the pseudo-proxy network comprises 105 grid-points (Figure 3). The second pseudo-proxy network is co-located with the smallest proxy network used in the study by Mann et al. (1999). This network comprises the proxy indicators available for the period 1000-1400 A.D. and consists of 12 proxy indicators. The location of the pseudo-proxies in the model ECHO-G is shown in Figure 3 and comprises also 12 model grid-points.

3.6 Noise models

The pseudo-proxies were generated by taking the annual mean temperature of the corresponding model grid-points and adding statistical noise. The level of non-climatic noise in the real proxies is not known, and therefore a subjective guess has to be made here. For high (interannual)

frequencies the amplitude of non-climatic noise can be estimated from the correlation between the proxy indicator and local instrumental temperature. Usually, these correlations are of the order of 0.5 or lower (Jones and Mann 2004). Higher correlations are sometimes reported, but they are rare. A correlation of 0.5 corresponds to a signal-to-noise ratio of 0.33 , the signal being here defined as the climate signal in the proxy record, and the noise the non-climatic part of the proxy variability. The amplitude of noise can be also labelled as the amount of noise variance as a percentage of the total proxy variance. For a correlation of 0.5, the amount of noise variance is 75%. For white-noise pseudo-proxies a reasonable amount of noise would be the one that matches a local correlation of 0.5 between the pseudo-proxy an the grid-point simulated temperature.

For red-noise pseudo-proxies, this question is not so straightforward. The spectral characteristics of non-climatic noise in real proxies is not known and only educated, somewhat arbitrary, guesses can be made. This guess can also be strongly dependent on the nature of the proxy indicator, and it may well be the case that the spectral characteristics of the noise are different from those of the signal. In this case, the amount of noise present in the proxy indicators will be frequency-dependent. In this study, we construct red-noise pseudo-proxies based on two different assumptions. Within the first, the relative amount of noise is constant for all pseudo-proxies. They are constructed in such a way that that contain approximately 75% percent of the noise variance across the whole frequency range that can be resolved with the length of the climate simulation. The proxies are constructed by fitting the parameters of an autoregressive process to the simulated grid-point temperatures. The order of the autoregressive-process is low (order 2). Random noise using these autoregressive parameters is then generated and the variance of

the noise is adjusted to achieve the desired level of noise variance in the pseudo-proxy. As the autocorrelation of the grid-point temperatures do not exactly match the autocorrelation function of an autoregressive process for all lags, and the autocorrelation function of the grid-point temperature itself varies spatially, the amount of noise variance does not exactly amount to 75% throughout the whole spectral range.

In the second implementation of red-noise pseudo proxies, the spectral characteristics are assumed to vary spatially. In this case, the noise is assumed to be a AR-1 process, but the value of the lag-1 autocorrelation is drawn at random from a beta distribution (parameters 7 and 3), in the interval (0,1). The beta distribution offers some advantages for this purpose. The beta distribution takes values in a bounded interval (whereas for instance, gaussian deviates would have to be truncated to the interval between 0 and unity). Also, the beta distribution is flexible enough to achieve a reasonable approximation to the desired form of the distribution with just two free parameters. The choice of these parameters of the beta distribution results in a mean value of 0.7 and a standard deviation of 0.43 for the lag-1 autocorrelation coefficient. An example of the sample autocorrelation for 100 noise timeseries generated in this way is shown in Figure 4. In this case, the amplitude of noise included in the pseudo-proxies is 75% of the total pseudo-proxy variance at high (interannual) frequencies, but as the spectral characteristics of the noise are drawn at random it cannot be guaranteed that this level of noise is maintained for lower frequencies. In general, for pseudo-proxies were the added noise has a higher autocorrelation than that of the associated grid-point temperature, the relative level of noise will increase with decreasing frequency, and vice versa.

A slightly more sophisticated model for the non-climatic noise, making use of long-range

time autocorrelation, has been also implemented. This model is based on *long-term persistence* process (Hoskings, 1984; Koutsoyiannis, 2003; Cohn and Lins, 2005). This type of processes display an autocorrelation function as a function of lag that decays with a power-law, instead of the more rapid decay characteristic of autoregressive processes, which is exponential in the case of AR-1 processes or, more generally, bounded by an exponential decay for AR-n processes. A long-range autocorrelation behavior has been identified in a number of observed hydrological and climatic records (Bunde et al. 2005).

In such processes the autocorrelation function C follows:

$$C(l) \sim l^{-\gamma} \tag{5}$$

where l is the time-lag. The value of γ is related to the Hurst exponent H and to the fractional-difference parameter d(see e.g. Cohn and Lins, 2005) by

$$\gamma = 2(1-H) \tag{6}$$

$$\gamma = 1-2d \tag{7}$$

.

Long-range-autocorrelation is characterized by values of d between 0 and 0.5, which corresponds to values of γ between zero and unity. Typical reported values for hydrological time series are $d \sim 0.3$, corresponding to $\gamma \sim 0.5$ (Cohn and Lins 2005). Values of d higher than .5 (values of γ lower than 0) give rise to non-stationary processes and values of d lower than 0.5 (values of γ higher than 1) do not produce long-range correlation.

In the present study long-term-persistence noise has been generated by the method proposed by Makse et al. (1995), based on the transformation of the Fourier coefficients of realizations

of white noise. The methods makes use of the relationship -through Fourier transformation- between the spectral density of a timeseries and its autocorrelation function, and of the fact that the Fourier transform of a power-law-decaying autocorrelation function can be calculated analytically. In a first step, a realisation of a gaussian white-noise is generated by standard algorithms. The resulting Fourier coefficients are then re-scaled to the target Fourier transform, and finally the modified Fourier coefficients are back-transformed to the time domain. The resulting timeseries displays the desired form of the autocorrelation function. A long-range-correlation process is described by a single parameter, say γ, and the lag-1 autocorrelation cannot be prescribed independently of γ. However, for comparison purposes, in this application one would like to align the lag-1 autocorrelation of the long-range-autocorrelated noise with that of the AR-1 noise. This is achieved here by mixing the long-range noise with white noise to achieve the desired lag-1 autocorrelation. The form of the autocorrelation function is thereby not changed , and only the general level of the autocorrelations can be adjusted up or down.

The noise in the pseudo-proxies is constructed to have a population-lag-1 autocorrelation of 0.7 and the value of γ is however allowed to vary across the pseudo-proxies. This value is drawn from a beta distribution with parameters (3,3) in the interval (0,1), which yield a mean value of 0.5 and a standard deviation of 0.18 for the γ parameter (corresponding to 0.75 and 0.36 for the Hurst coefficients, respectively). The rationale here is to open the possibility for some of the pseudo-proxies to have long-memory noise. Figure 4 also shows the sample autocorrelation function estimated from 100 long-term persistence timeseries generated with this methodology. As in the previous case, the relative level of noise cannot be maintained constant throughout the whole frequency range, as the spectrum of the noise is defined by the random value of γ.

The amplitude of the noise is such that, as in the previous cases, it accounts for 75% of the pseudo-proxy variance at interannual frequency.

3.7 Estimation of the reconstruction uncertainties

In general, the uncertainties in the reconstruction of the Northern Hemisphere temperature for each particular method are composed of two terms (National Research Council, 2006): uncertainties originating in the estimation of the free parameters inherent to each model and to inadequacies in the model structure, and uncertainties originated in the high-frequency random part of the temperature variability. When considering long timescales the random part is considerably reduced by low-pass filtering, so that at those timescales the uncertainties in the model parameters tend to be the larger contributor.

In this study, 31-year running mean values of the target Northern Hemisphere temperature and of its reconstruction will be considered. Part of the uncertainty of the first kind will be estimated by generating 100 realizations of the pseudo-proxy noise and using these realizations as input for the reconstruction method. The resulting spread in the reconstruction illustrates the spread in the best estimate that could have been obtained with different samples and check whether this range of variation includes the target temperature, i.e. whether or not the statistical method provides biased estimations. In the following figures, the 5%-95% spread is shown together with the median. The median and the spread are calculated for each 31-year moving window, so that the curves shown in the curves do not represent individual reconstruction obtained from a particular noise realization, but an envelope encompassing 90% of the pseudo-reconstructions in the corresponding moving window.

4 Results

4.1 Pseudo-proxies from simple noise models

The results of the three reconstructions methods of the Northern Hemisphere temperature in the ECHO-G simulation for both pseudo-proxy networks from simple white-noise and red-noise models (spatially constant spectral properties) and for both variants of the calibration (detrended or not detrended) are shown in Figure 5, together with the target model-simulated North Hemisphere annual mean temperature. The correlation between pseudo-proxies and local temperature is 0.5. The general characteristics of common to all pseudo-reconstructions is that they all underestimate the past variations of the mean temperature, and the estimated temperature in the centuries previous to the instrumental period is too warm. The bias, however, depends on the noise model, on the calibration variant and on the size of the pseudo-proxy network.

For the complete proxy network, the Monte Carlo uncertainties are considerably smaller than for the smaller pseudo-proxy network. It is with the complete network where the differences among the performance of the methods are, therefore, more clear. In this case, the detrended-calibration variant performs worse than the non–detrended calibration. Also, the assumption of red-noise in the pseudo-proxies deteriorates the performance of all three methods relative to the white-noise model. The best-performing method among these three methods tested is Composite plus Scaling, which with non-detrended calibration and white noise pseudo-proxies is able to yield temperatures very close to the true simulated temperature. For this method, red-noise pseudo-proxies also cause a very minor, almost imperceptible worsen-

ing of the skill, although red-noise pseudo-proxies widen the uncertainty range. In this sense, therefore, CPS seems to be the most robust method among the three tested in this study.

When the pseudo-proxy network is considerably decimated to just 12 proxy indicators, the skill of all three methods deteriorates again, and the Monte Carlo uncertainties become considerably larger compared to the results obtained with the complete network. However, the reduction in skill does not seem to be not very dramatic considering that the size of the network has been reduced by a factor of about ten. It seems, therefore, that at multidecadal timescales a limited number of proxies is able to capture a large part of the signal of the Northern Hemisphere mean temperature.

4.2 Perfect pseudo-proxies

In an ideal set-up, a large high-resolution proxy network should perform best will be always preferred. However, sometimes a choice has to be made between a small network of high-quality proxies (defined here as proxy indicators with a very high correlation with local temperature) and a larger network poorer local correlations. This trade-off between the quality and quantity of the proxies can be explored also in the virtual reality of reconstructions with pseudo-proxies. To explore the skill of a small high-quality network, pseudo-reconstructions with the 12–pseudo-proxy network with no noise (i.e. perfect pseudo-proxies involving directly the model-simulated grid-point temperatures with no noise contamination), has been also investigated for all three reconstructions methods (Figure 5), this time only with the non-detrended calibration variant. It turns out, as expected, that with the small proxy-network the pseudo-reconstructed temperature is closer to the target temperature than with noise-contaminated proxies, either white-noise or

red-noise. The improvement is particularly visible for the two regression-based methods. For the CPS method, which previously displayed the best performance, the improvement is not so dramatic. It seems that this method is also more robust against the presence or of absence of noise in the pseudo-proxies. Actually, the results obtained with CPS are also the most insensitive to the amplitude of the noise (not shown).

Comparing the performance of the three methods in the cases of the large and noisy pseudo-proxy network (Fig 5) with the result obtained with the small network of perfect pseudo-proxies (Fig 6), it turns out that the differences in the pseudo-reconstructions are small, albeit the small network with perfect pseudo-proxies provides very slightly better results. They could be very well inside the uncertainty bounds.

Perfect pseudo-proxies are, however, a limiting unrealistic case. It should be noted that the results with the small network of perfect pseudo-proxy set could be model dependent and also present spatial variability associated to the actual geographical distribution of a few number of proxies, i. e. alternative distributions of 12 proxies could perhaps capture better the large-scale modes of temperature variation. These two aspects have not been constrained in this approach, but the small differences between Figures 4 and 5 suggests that a large network of imperfect pseudo-proxies can be a practical strategy in the real world, provided that the amplitude of the noise remains limited to a local correlation between proxies and local temperature of about 0.5 and does not considerably drop below this value.

4.3 Pseudo-proxies with mixed temperature-precipitation response

A problem that may arise in the use of proxy records to reconstruct past climates is related to the climate variable that the record is meant to represent. For some records this question cannot be answered unambiguously (Jones et al. 1998; National Research Council, 2006). Even if the proxy record is correlated to local temperature in the calibration period, this may not hold for longer timescales. The correlation of the proxy to temperature could stem from an indirect correlation of temperature to another underlying driving variable, for instance, precipitation. In this section we investigate the effect on temperature reconstructions of pseudo-proxies that are constructed as a mixture of local temperature and precipitation, but which are used only to reconstruct temperature. In this sense, the pseudo-proxies are constructed as in the previous sections but replacing the statistical noise by the grid-cell model-simulated precipitation.

Several other set-ups to implement this idea could have been can be considered, but here only a simple zero-order scheme will be tested, in which the grid-point pseudo-proxies are constructed, for each grid-point, by adding temperature and precipitation in such a way that the amount of interannual variance for the whole simulation period is equally shared by both variables. In this simulation this ratio leads to an average absolute value of correlation between local temperature and the pseudo-proxy of 0.67 for the large network of 105 pseudo-proxies and a range of 0.5 and 0.86 for the small network of 12 pseudo-proxies. The amount of non-temperature noise is, therefore, smaller than in the previous cases constructed with stochastic noise. These relatively high correlation values are mainly caused by the tendency of annual precipitation to follow the local temperature at decadal and longer timescales. Note, however,

that these values are not the correlation between local temperature and precipitation, since half of the variability of the pseudo-proxies already represents local temperature. The calibration variant used in this case is also the non-detrended calibration.

The results obtained with the large and small pseudo-proxy network are shown in Figure 7. Qualitatively, there are no significant differences from the results obtained in previous sections. All three methods tend to underestimate past variations of the Northern Hemisphere temperature. Again, the best performing method is CPS. The regression-based methods display clear biases, although smaller than for the previous cases with pseudo-proxies which contain a larger amount of non-temperature noise.

4.4 Pseudo-proxies with spatially varying stochastic noise

Figure 8 shows the pseudo-reconstructions of the Northern Hemisphere temperature using pseudo-proxies containing stochastic noise with spatially varying spectral characteristics, for the large and small networks. In both cases the CPS method performs clearly better than the regression-based methods. Again, the use of very few proxy indicators degrades the performance of all three methods, in particular for direct principal-components regression. As expected the uncertainty ranges are larger for the small network for the three methods. When comparing the results obtained in this subsection with those obtained with noise with spatially constant autocorrelation (Fig. 5), some changes can be observed. The performance of the three methods is slightly poorer when the noise is spatially heterogeneous, and the uncertainty bounds have, unsurprisingly, increased. However, the qualitative picture remains the same as with spatially homogeneous noise (Fig. 5). This qualitative picture remains also valid when long-term-persistence

pseudo-proxies are used. In this case, the smaller pseudo-proxy network also leads to a larger warm bias and larger uncertainty ranges than with the larger network. A direct comparison with the case with spatially heterogeneous autoregressive noise is hampered by the fact that the it is difficult to obtain a similar spread in the spectral characteristics, so that the apparently larger uncertainty range in the case of the small long-memory proxy network than with autoregressive noise should not be pursued too far.

4.5 Influence of proxy seasonality and proxy location

Real proxies usually do not directly record annual mean temperature but are more sensitive to climate conditions in one season than in the rest of the year. For instance, trees and other biological systems may respond to growing season temperature or available moisture and ice-core records may reflect more strongly environmental conditions in the season with strongest precipitation rate. Nevertheless they are used to reconstruct annual mean temperatures assuming that seasonal and annual temperatures are strongly correlated. Again, this assumption can be checked in the instrumental period but uncertainty about this connection at centennial timescales is much larger (Jones et al., 2003). The influence on reconstructions of annual temperature using of seasonal proxies can be explored with pseudoreconstructions in the climate simulations. In the climate simulations annual and summer temperatures are in general strongly correlated at multidecadal timescales and here, for brevity, only for the cases of white-noise pseudoproxies and non-detrended calibration. The conclusions can be extended to all other cases -red noise pseudoproxies and detrended calibration. The pseudo-proxies are then constructed by taking the June-August mean grid-pint temperature and adding white noise to reach a correlation to

the June-August grid-point temperature of 0.5. The target temperature is still the annual mean Northern Hemisphere temperature. Therefore, for both regression methods the spatial EOF patterns were calculated from the annual mean temperature field. Figure 10 displays the pseudoreconstructions for the large and small pseudoproxy networks. The differences to the results obtained with annual-temperature pseudoproxies are vary small. In this case, again, all three methods underestimate the low-frequency variations of the target temperature and the use of the small pseudoproxy network yields worse results than the full network.

Another aspect that may influence the performance of the reconstructions is the fact that in the full proxy network the North American continent is overrepresented, especially compared to the Eurasian land mass. To some extent this over-representation is ameliorated in the MBH98 method by a previous principal components analysis of some of the proxy subnetworks in areas with a dense coverage of real proxies, such as in North America (Mann et al. 1998). Nevertheless, it could be interesting to investigate whether a dense Asian pseudo-proxy network would yield results much different than those obtained with the real proxy locations. This has been also investigated in this section by interchanging a subset of the North American and Asian proxy locations in the full proxy network: the pseudo-proxies located in the geographical box (30N-55N;120W-60W) are transferred to (30N-55N;120E-60E) by transforming west longitudes to east longitudes. The opposite transformation was applied to the locations in (30N-55N;120E-60E) (Figure 3). The influence of the pseudo reconstructions obtained with the flipped proxy network have been investigated for the cases with white-noise and red-noise pseudoproxies. The results for the white-noise pseudoproxies are very similar as with the original proxy location. For red-noise pseudoproxies, the pseudo-reconstructions are slightly better

(Figure 11, although qualitatively similar. The differences in this latter case are probably not large enough to claim that Asian proxies could be fundamentally more useful to reconstruct the Northern Hemisphere temperature in the real world. This aspect may be well dependent on the particular climate model. This example, however, illustrates that it should be indeed desirable to achieve a regular proxy coverage over the globe.

5 Discussion and conclusion

Three statistical methods to reconstruct the Northern Hemisphere temperature in the past millennium have been tested in the virtual reality of a simulation with the climate model ECHO-G, under different conditions related to the size of the proxy network and pseudo-proxy noise model. Two of the methods are regression-based and the third method is a simple all-proxy-averaging with re-scaling. The general conclusion extracted from this study is that all methods underestimate the low-frequency variability of the Northern Hemisphere temperature. However, the simple method Composite plus Scaling provides, in the conditions tested in this study, better results and is more robust against changes in the proxy network and noise characteristics. This conclusion is consistent with the results found by Juckes et al. (2007) in an analysis of real reconstructions of the past Northern Hemisphere temperature from real proxy records.

The factors that determine more strongly the skill of the pseudo-reconstruction are, within those tested in this study, the size of the proxy network (with better reconstructions when the network is larger) and the previous detrending of the calibration data (which degrades the quality of the pseudo-reconstruction). In this study, the role of the amplitude of the noise has not been

tested, as it became clear in previous analysis that this factor quite strongly influences the skill of the pseudo-reconstruction (von Storch et al. 2004).

The superiority of the CPS method found here may not have general validity and in other circumstances regression-based methods may display a better performance. For instance, this study indicates that regression-based methods are more sensitive to the size of the proxy-network (compare upper and lower panels in Figure 5). If the focus of the reconstruction method is regional rather than hemispheric and the density of proxy indicators is higher, their performance may surpass the CPS method. When applying the pseudo-proxy method to test the reconstructions of past European temperatures of Luterbacher et al. (2004), who used a direct principal-components regression method, Küttel et al. (2007) found that the full pseudo-proxy network representing a proxy network of 166 indicators in Europe yields a satisfactory pseudo-reconstruction of mean European temperatures. This high number of proxy indicators covers quite effectively the whole target region. Decimation of the pseudo-proxy network to emulate the diminishing number of proxies available backwards in time induces again a deterioration of the pseudo-reconstructions, leading to an underestimation of the low-frequency variability in the model ECHO-G (Küttel et al., 2007).

One of the main uncertainties in pseudo-proxy studies is a valid model for the non-climate noise, in particular at longer timescales. The models used in the present study (white noise, red noise, precipitation noise, and long-term persistence) are admittedly simple, but more complex noise models require a more detailed analysis of the response of the different proxies, particularly of the biological proxy indicators as tree-rings and corals, to modulations of non-climatic environmental and ecological factors at long timescales. This knowledge is so far not generally

available. Within the noise models tested here, this study indicates that the nature of the noise is not the source of large differences in the performance of the methods, but it has to be noted that the choice of the noise-model parameters are just educated guesses, especially for the parameters that drive the low-frequency variability of the proxy indicator, so that further analysis may prompt a revision of this conclusion.

In this study, only the output of one climate simulation for the whole past millennium has been used and the performance of the reconstructions methods could be different when using another climate model. However, the objective of a pseudo-proxy analysis of reconstruction methods should not be limited to identify simulations where the performance of some of the methods is acceptable, but to test the robustness of the reconstruction methods under different, but plausible, realizations of past climates. This line of research is still limited by the number of simulations covering the whole past millennium with state-of- the-art General Circulation Models. To our knowledge only two models have been used for this purpose (shown in Fig 1), so that the assessment of the robustness of reconstructions methods is still limited in this regard.

Acknowledgments

We thank A. Bunde for his help in generating synthetic timeseries with long-range-correlation. The work of H.v.S. and E.Z. is part of the European project Millennium European Climate (contract number 017008, GOCE). F. G-R was partially funded by the Spanish Ministerio de Educación y Ciencia through grant REN-2002-04584-CLI.

References

Ammann CM, and Wahl ER (2007) The importance of geophysical context in statistical evaluations of climate reconstructions procedures. Climate Change 85:71-88

Battle M, Bender M, Sowers T, Tans PP, Butler JH, Elkins JW, Ellis JT, Conway T, Zhang N, Lang P, and Clarke AD, 1996: Atmospheric gas concentrations over the past century measured in air from firn and the South Pole. Nature 383: 231-235

Bunde A, Eichner JF, Kantelhardt JW, and Havlin S (2005) Long-term memory: a natural mechanism for the clustering of extreme events and anomalous residual times in climate records. Phys. Rev. Lett. 94: 048701-1–048701-4.

Briffa KR, and Osborn TJ (2002) Blowing hot and cold, Science 292: 662-663

Briffa KR, Schweingruber FH, Jones PD, Osborn TJ, Shiyatov SG and Vaganov EA (1998) Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. Nature 391: 678-682

Bürger G, Fast I , Cubasch U (2006) Climate reconstruction by regression-32 variations on a theme. Tellus 58A: 227-235

Cohn TA and Lins HF(2005) Nature's style: naturally trendy. Geophys. Res. Lett. 32: L23402 doi:10.1029/2005GL024476

Crowley TJ (2000) Causes of climate change over the past 1000 years. Science 289: 270-277

Crowley TJ, and Lowery TS (2000) How warm was the medieval warm period?. Ambio 29: 51-54

Esper J, Cook ER, and Schweingruber FH (2002) Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. Science 295: 2250-2253

Esper J, Frank D, RJS Wilson, KR Briffa (2005) Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophys. Res. Lett. 32: L07711, doi:10.1029/2004GL021236

Etheridge DM, Steele LP, Langenfelds RL, Francey RJ, Barnola J-M and Morgan VI (1996) Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. J. Geophys. Res. 101 D2: 4115-4128

Etheridge DM, Steele LP, Francey RJ, and Langenfelds RL (1998) Atmospheric methane between 1000 A.D. and present: Evidence of anthropogenic emissions and climatic variability. J. Geophys. Res. 103 D13: 15,979-15,994

Friedlingstein, P., et al. (2006) Climate-carbon cycle feedback analysis, results from the C4MIP model intercomparison, J. Clim 19: 3337-3353

González-Rouco JF, Beltrami H, Zorita E, von Storch H. (2006) Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling. Geophys. Res. Letters 33: L01703, doi:10.1029/2005GL024693

Goosse H, Crowley TJ, Zorita E, Ammann CM, Renssen H, and Driesschaert E (2005) Geophys. Res. Lett. 32: L06710, doi:10.1029/2005GL022368

Granger CWJ, and Newbold P (1974) Spurious regression in econometrics. Journal of Econometrics 74:111-120

Hegerl GC, Crowley TJC, Allen M, Hyde WT, Pollack HN, Smeardon J, and Zorita E (2007) Detection of human influence in a new validated 1500 year temperature reconstruction. J. Climate 20: 650-666

Hoskings J (1984) Modelling persistence in hydrological timeseries using fractional differencing. Water Res. Res. 20: 1898-1908

Huybers P (2005) Comment on ”Hockey sticks, principal components and spurious significance” by S. McIntyre and R. McKitrick. Geophys. Res. Lett. 32: L20705 doi:101029/2005GL023395

Intergovernmental Panel on Climate Change Third Assessment Report, Cambridge University Press, Cambridge, 2001

Jones, PD, Briffa KR, Barnett TP, and Tett SFB (1998) High-resolution paleoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. The Holocene: 8, 455:471

Jones, PD, Briffa KR, Osborn TJ (2003) Changes in Northern Hemisphere annual cycle: implications for paleoclimatology. ys. Res. 108(D18): 4588, doi:10.1029/2003JD003695

Jones PD, and Mann ME (2004) Climate over the past millennia. Rev. Geophys. 42: RG2002, doi:10.1029/2003RG000143

Juckes MN, Allen MR ,Briffa KR, Esper J, Hegerl GC, Moberg A, Osborn TJ, and Weber SL (2007) Clim. Past 3: 591-609

Koutsoyiannis D (2003) Climate change, the Hurst phenomenon and hydrological statistics. Hydrol. Sci. 48: 3-24

Küttel M, Luterbacher J, Zorita E, Xoplaki E, Riedwyl N, and Wanner H (2007) Testing a European winter surface temperature reconstruction in a surrogate climate. Geophys. Res. Lett. 34, L07710, doi:10.1029/2006GL027907

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M,and Wanner H (2004) European seasonal and annual temperature variability, trends and extremes since 1500 A.D. Science 303:1499-1503

Mann ME, Bradley RS, and Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392: 779-787

Mann ME, Bradley RS, and Hughes MK (1999) Northern temperature over the past millennium:inferences, uncertainties and limitations. Geophys. Res. Lett. 26:75-762

Mann ME, Bradley RS, and Hughes MK (2004) Corrigendum: Global-scale temperature patterns and climate forcing over the past six centuries, Nature 430: 105-105

Mann ME and S. Rutherford (2002) Climate reconstructions using pseudoproxies. Geophys. Res. Lett. 29: 139-1-139-4 doi:10.1029/2001GL014554

Mann ME, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J. Clim. 18: 4097-4107

Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstructions. J. Geophys. Res. 112: D12109, doi:10.10292006JD008272

Makse, HA, Havlin, S, Schwartz, M, Stanley, HE (1996) Method for generating long-range correlations for large systems. Phys. Rev. E 53: 5445-5499

Moberg, A, Sonechkin DM, Holmgren K, Datsenko N,and Kart'en W (2005) High variable Northern Hemisphere temperatures reconstructed from low and high resolution proxy data. Nature 433: 613-617

Min S-K, Legutke S, Hense A, and Kwon W-T (2005) Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G II. El Niño Southern Oscillation and North Atlantic Oscillation. Tellus 57A: 622-640

National Research Council (2006) Surface temperature reconstructions for the last 2,000 years, National Academies Press, Washington, DC, 141 pp

Osborn TJ and Briffa KR (2000) Revisiting timescale-dependent reconstruction of climate from tree-ring chronologies. Dendrochronologia 18: 9-26

Osborn TJ, Raper SCB, and Briffa KR (2006) Simulated climate change during the last 1000 years: comparing the ECHO-G general circulation model with the MAGGIC simple climate model. Climate Dynamics 27: 185-197

Roeckner, E., Bengtsson L, Feichter J, Lelieveld J, and Rodhe H (1999) Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle, J. Clim. 12: 3004-3032

Rutherford, S., Mann ME, Osborn TJ, Bradly RS, Briffa KR, Hughes MK, and Jones PD (2005) Proxy-based reconstruction Northern Hemisphere surface temperature reconstructions: sensitivity to method, prediction network, target season and target domain. J. Climate 18: 2308-2329

Schneider T (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. J. Climate 14: 854-871

Scheffer, M., Brovkin, V., and Cox M (2006) Positive feedback between global warming and atmospheric $C0_2$ concentrations inferred from past climate change. Geophy. Res. Lett., 33: L10702 doi 10.1029/2005/GL025044

Storch H, Zorita E, Jones J , Dimitriev Y, González-Rouco JF, and Tett SFB (2004) Reconstructing past climate from noise data. Science 306: 679-682

Storch H, Zorita E, Jones J, González-Rouco JF, and Tett SFB (2006) Reply to ” Comment on reconstructing past climate from noisy data” by Wahl et al. Science 312: 529b-529b

Treydte KS, Schleser GH, Helle G,Frank DC, Winiger M, Haug GH, and Esper J (2006) Nature 440: 1179-1182

Vaganov EA, Hughes ME, Kirdyanov AV, Schwiengruber FH and Silkin PP (1999) Influence of snowfall and melt timing on tree growth in subarctic Euroasia. Nature 400: 149-151

Wahl ER, and Ammann CM (2007) Robustness of the Mann, Bradly,Hughes reconstructions of northern hemisphere surface temperatures: examination of criticism based on the nature and processing of proxy climate evidence, Climate Change 85:71-88

Yule GU (1926) Why do we sometimes get nonsense correlation between timeseries?. J. Royal Stat. Soc. 89: 1-69

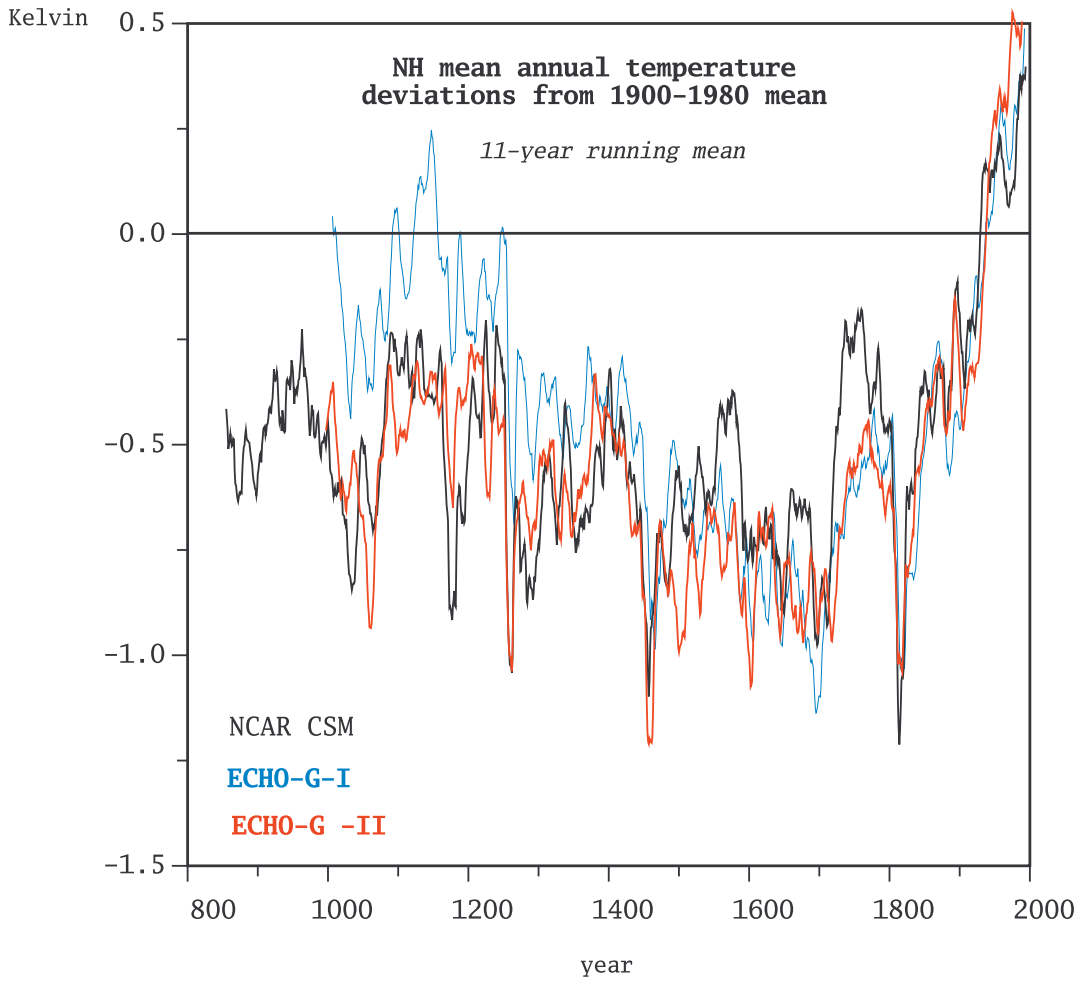

Figure 1: Simulated mean annual Northern Hemisphere temperature (North of Equator) in three simulations of the past centuries with General Circulation Models: one simulation with the Climate System Model (CSM, Mann et al 2005) and two simulations with the model ECHO-G (ref González-Rouco et al (2006)). Curves represent deviations from the 1900-1980 mean.

.

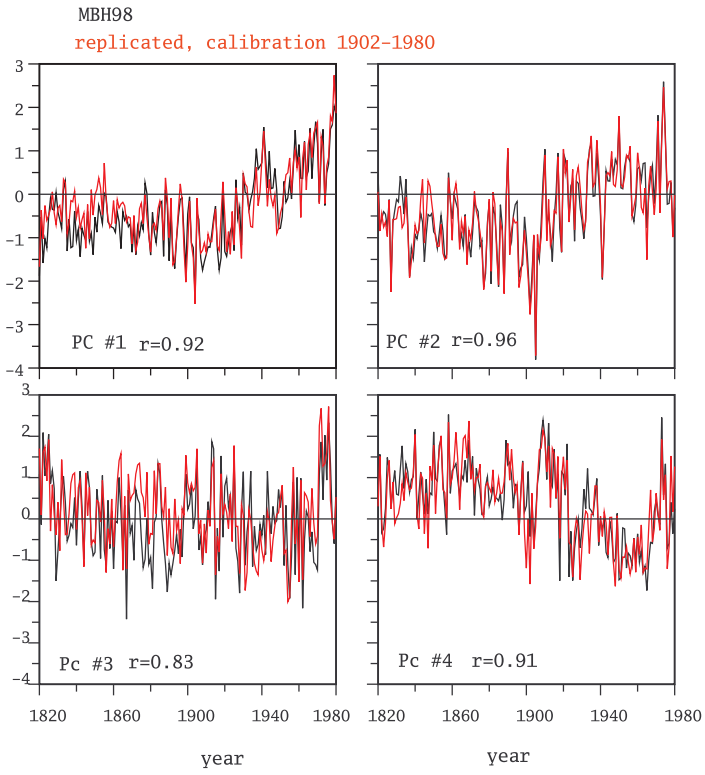

Figure 2: Emulation of the reconstruction algorithm of Mann et al (1998). The proxy-based reconstructed four leading Principal Components of the Northern Hemisphere annual temperature field in the period 1820-1980 from Mann et al (1998) (black) and the result of our emulation of the algorithm using the same calibration data (red) (Mann et al. 2004).

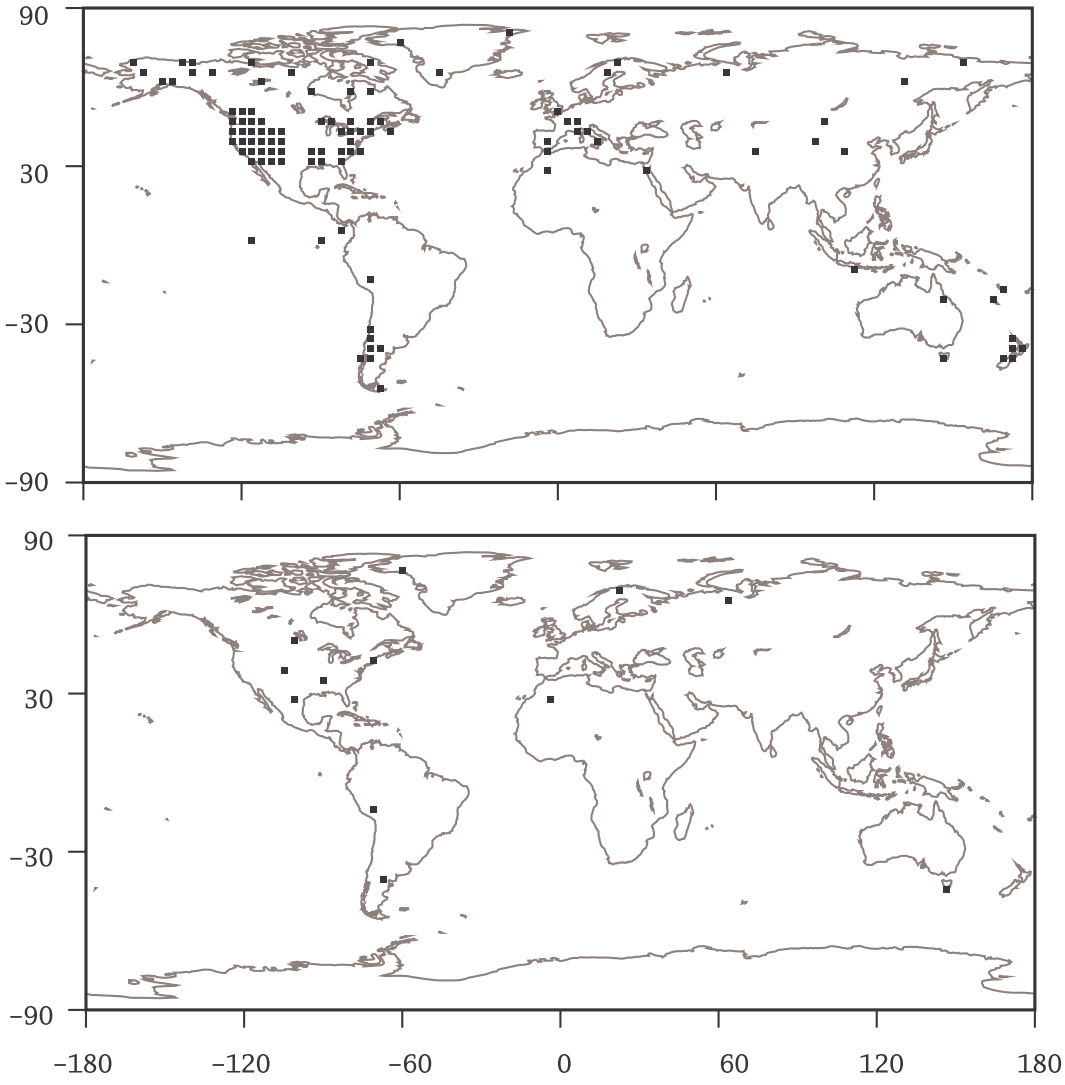

Figure 3: Network of pseudo-proxies in the ECHO-G grid. Upper panel depicts the full network which emulates the full complete network of real proxy indicators in Mann et al. (1998); lower, the small pseudo-proxy network emulating the minimum network of real proxies in Mann et al. (1999).

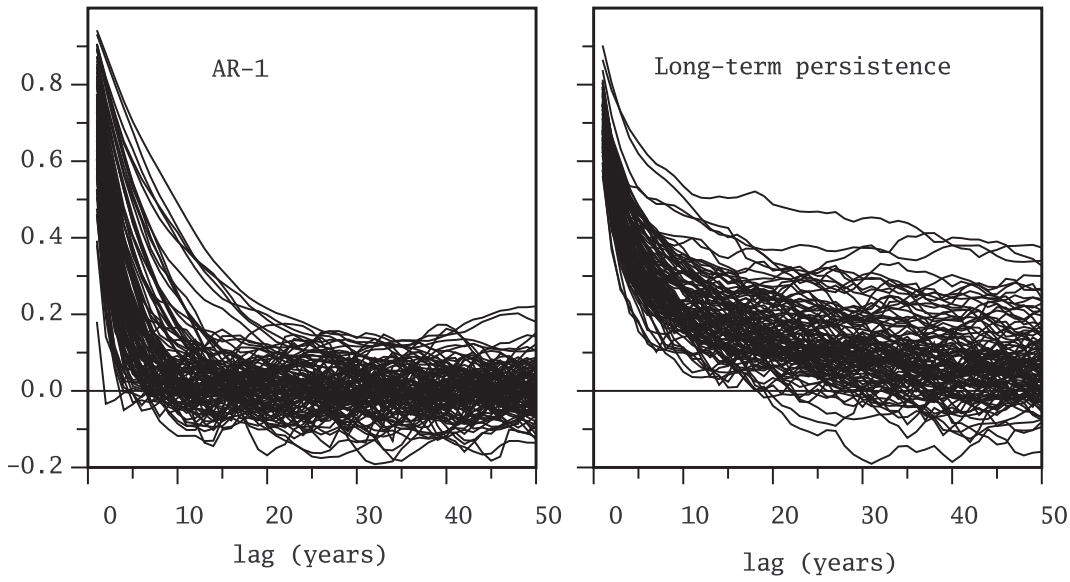

Figure 4: Sample autocorrelation estimated from random timeseries generated by autoregressive process of order 1 with lag-1 autocorrelation drawn from a beta distribution with mean 0.7 and standard deviation 0.42 (left) and estimated from long-term-persistence noise timeseries generated with a lag-1 autocorrelation of 0.7 and a γ parameter (describing the power-law decay of the autocorrelation function, see text) drawn from a beta distribution with mean 0.5 and standard deviation 0.18 .

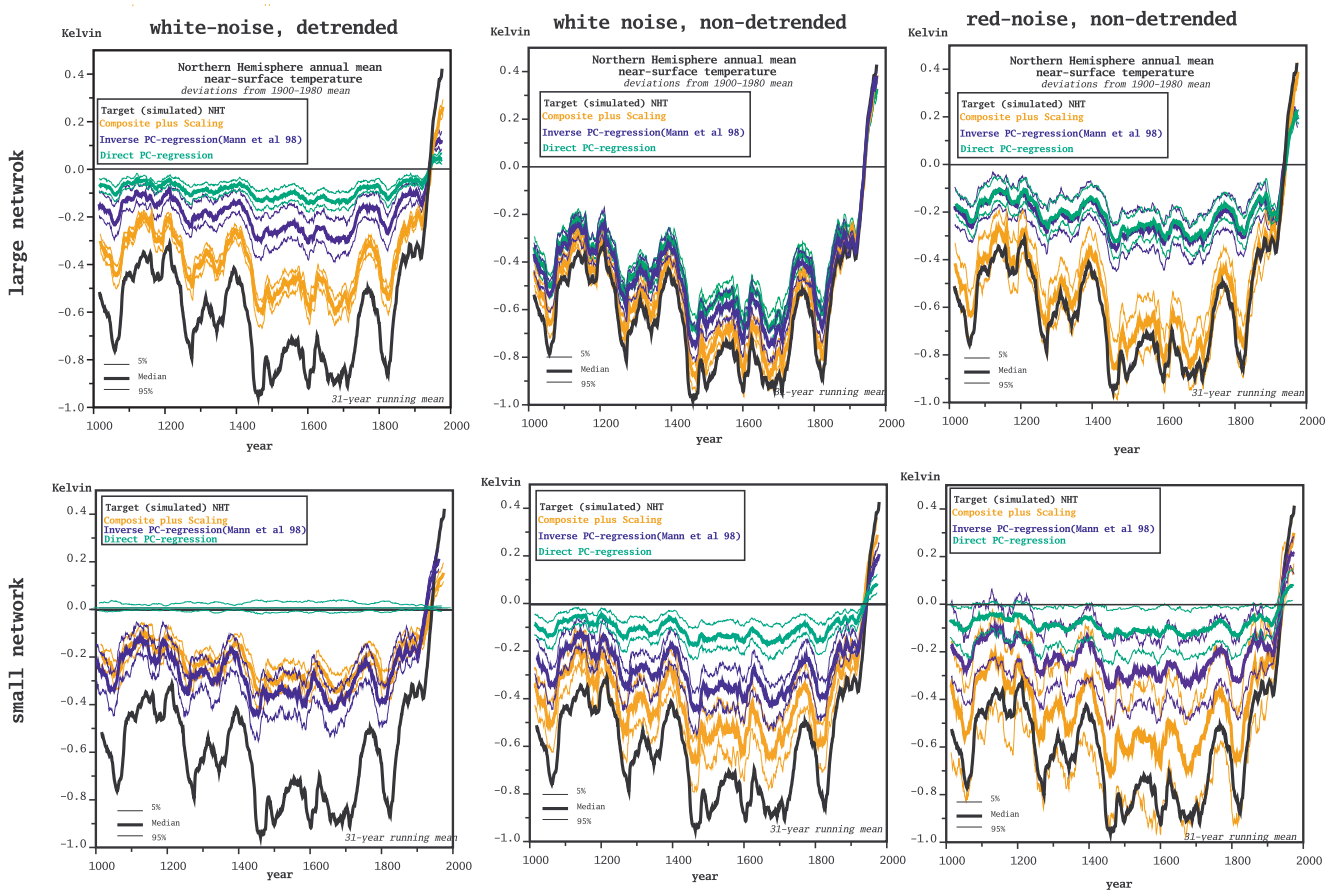

Figure 5: Pseudo-reconstructions of the annual mean Northern Hemisphere temperature in the model ECHO-G for different setups: upper and lower rows, large and small network, respectively; left column: white-noise pseudo-proxies and detrended calibration, middle column: white-noise pseudo-proxies and non-detrended calibration; right column, red-noise pseudo-proxies and non-detrended calibration. The interannual correlation between pseudo-proxies and local temperature is 0.5. Curves represent deviations from the mean of the calibration period 1900-1980. Thick and thin lines represent the median and the 5%-95% range resulting from 100 Monte Carlo realisations of the noise in each running 31-year period.

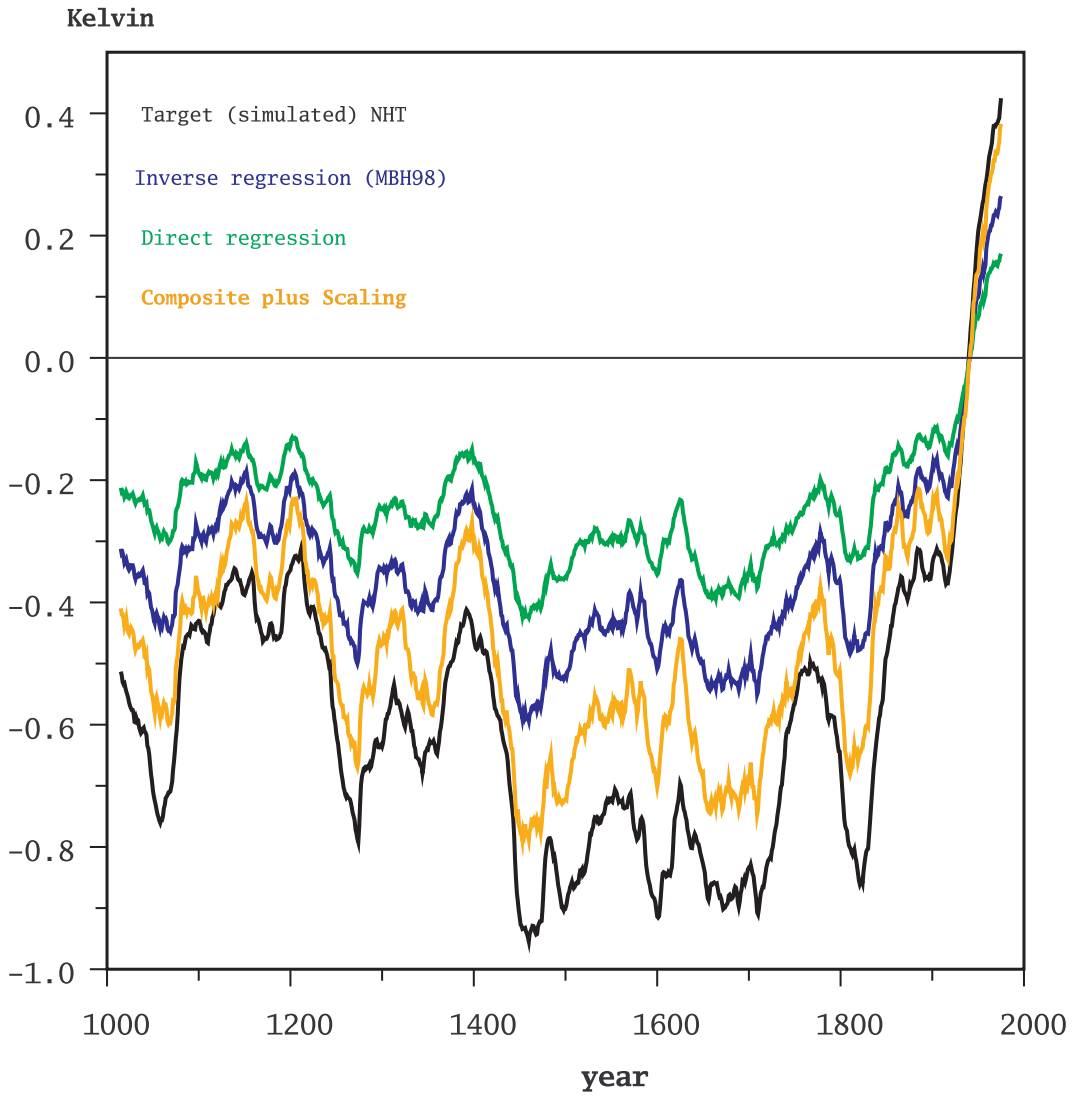

Figure 6: Pseudo-reconstructions of the annual mean Northern Hemisphere temperature in the model ECHO-G with perfect pseudo-proxies and small proxy-network. Curves show deviations from the mean of the calibration period 1900-1980. The calibration variant is non-detrended calibration.

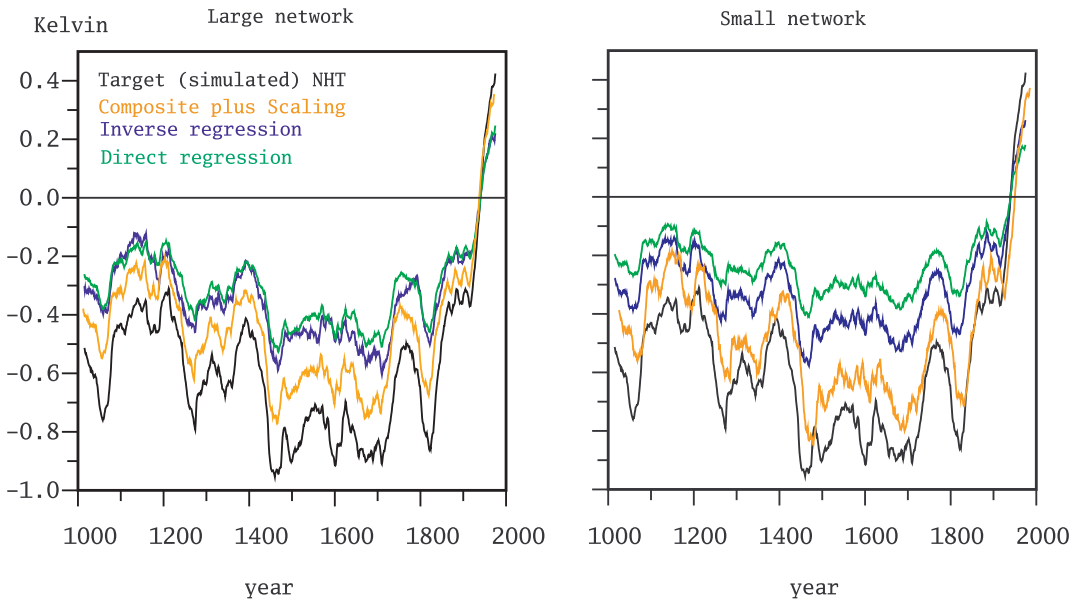

Figure 7: Pseudo-reconstructions of the annual mean Northern Hemisphere temperature in the model ECHO-G with the large and small proxy networks and with pseudo-proxies constructed contaminating the grid-point temperature with simulated precipitation. The amount of non-temperature variance is 50% of the total pseudo-proxy variance of each pseudo-proxy. The calibration variant is non-detrended calibration.

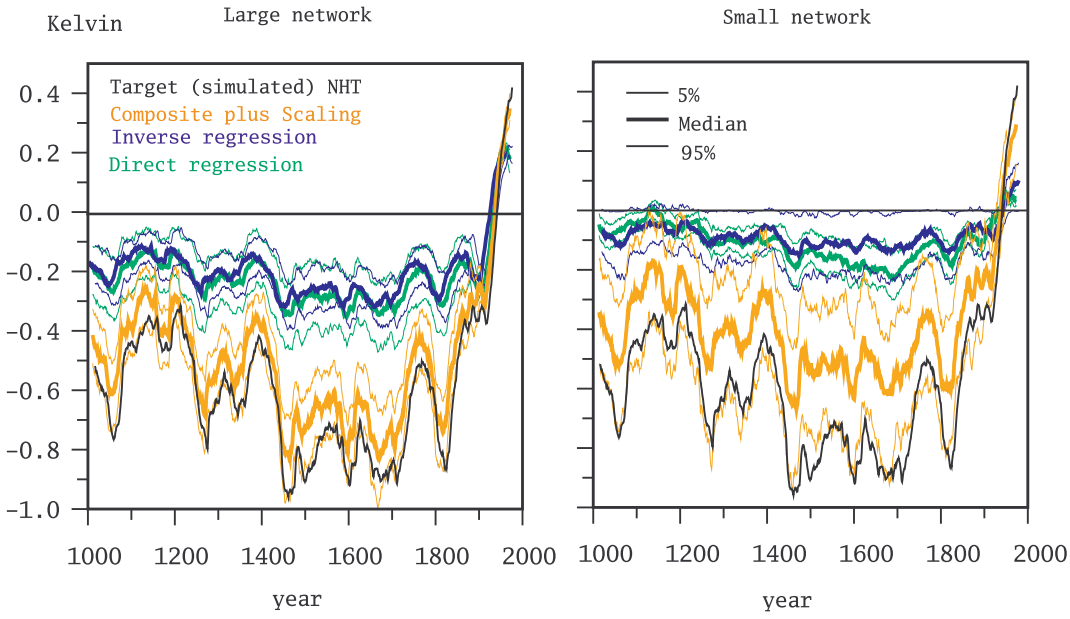

Figure 8: As figure 7 with pseudo-proxies constructed contaminating the grid-point temperature with spatially heterogenous autoregressive process of order 1. The (population) lag-1 noise autocorrelation is drawn from a beta distribution in the interval (0,1) with mean 0.7 and standard deviation 0.42. The amount of non-temperature variance at interannnual timescales is 75% of the total variance of each pseudo-proxy.

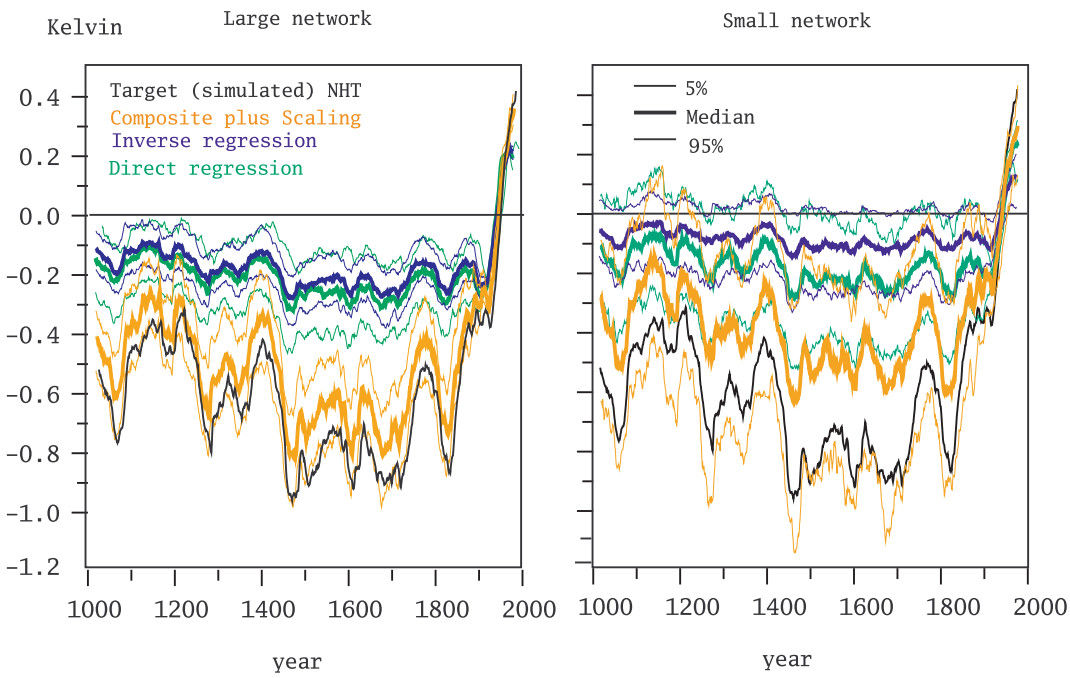

Figure 9: As figure 7 with pseudo-proxies constructed contaminating the grid-point temperature with spatially heterogeneous long-term-persistence noise. The (population) lag-1 noise autocorrelation is 0.7 and the exponent describing the power-law decay of the autocorrelation function is drawn from a beta distribution in the interval (0,1) with mean 0.5 and standard deviation 0.18. The amount of non-temperature variance at interannnual timescales is 75% of the total variance of each pseudo-proxy.

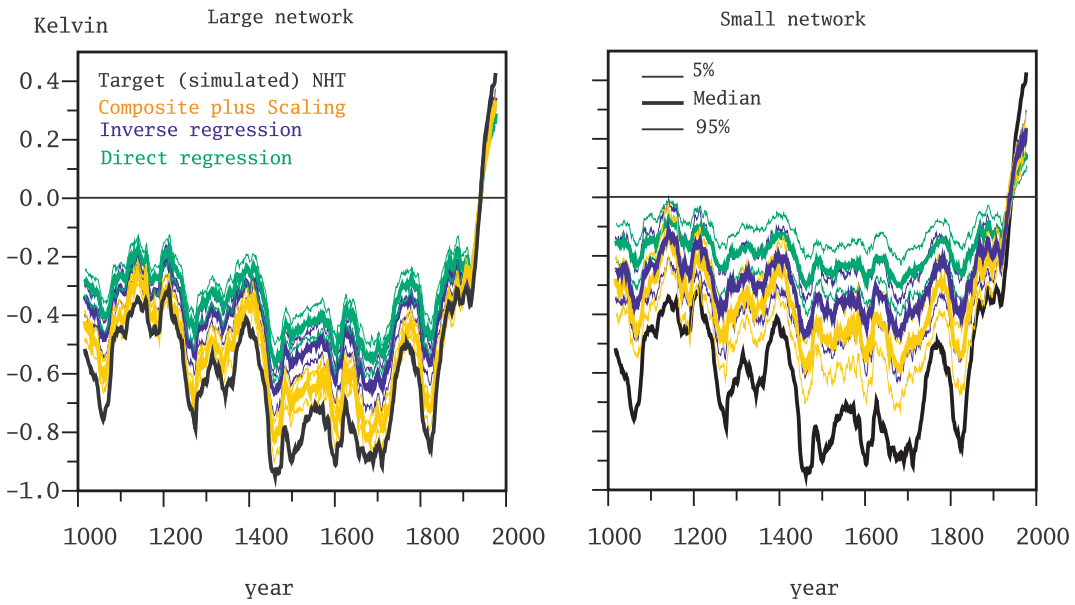

Figure 10: Pseudo reconstructions of the mean annual Northern Hemisphere temperature obtained from *summer* white-noise pseudoproxies. The local correlation between summer temperature and summer pseudoproxy is 0.5 for all pseudoproxies. For the direct PC regression and MBH98 methods the spatial temperature EOF patterns are calculated from the annual mean temperature field.

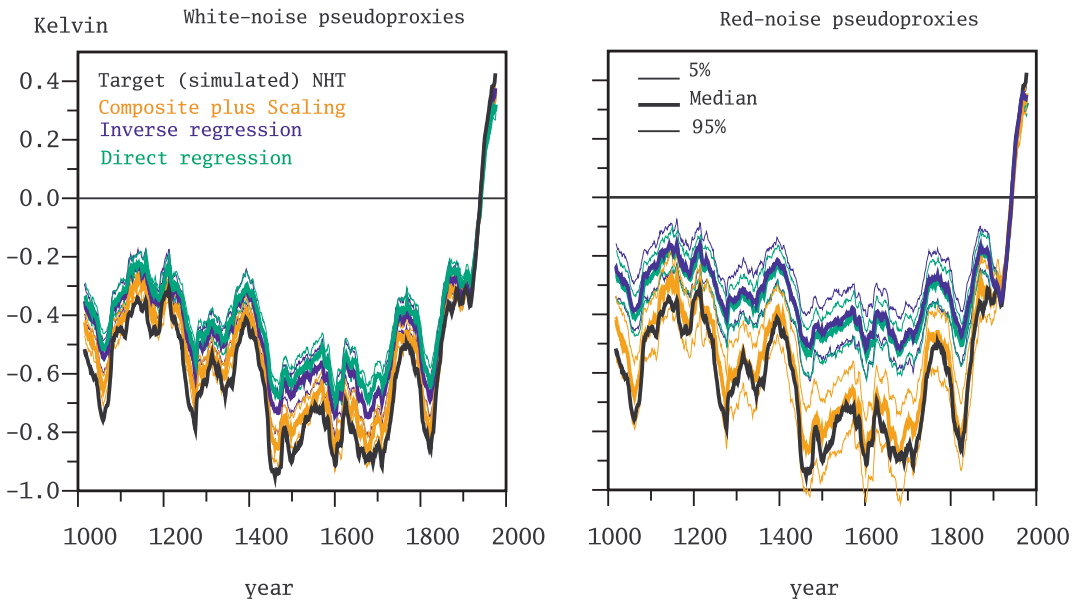

Figure 11: Pseudo reconstructions of the mean annual Northern Hemisphere temperature obtained after transferring the proxy locations of the full network (Fig 3, upper panel) in the geographical box (30N-55N;120W-60W) to (30N-55N;60E-120E), and vice-versa, by swapping east and west longitudes. The pseudo-proxies contain white-noise with local correlations to temperature of 0.5 .