CORRESPONDENCE

Comments on “Reconstruction of the Extratropical NH Mean Temperature over the Last Millennium with a Method That Preserves Low-Frequency Variability”

ANDERS MOBERG

Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden

(Manuscript received 22 July 2011, in final form 29 February 2012)

ABSTRACT

Christiansen and Ljungqvist have presented an extratropical NH temperature reconstruction using a method (LOC) that they claim “preserves” low-frequency variability, at the expense of exaggerated high-frequency variability. Using theoretical arguments and a pseudoproxy experiment, it is demonstrated here that the LOC method is not guaranteed to preserve variability at any frequency. Rather, LOC reconstructions will have more variance than true large-scale temperature averages at all frequencies. This variance inflation, however, can be negligible at those frequencies where the noise variance in individual proxies is small enough to be effectively cancelled when computing an average over the available proxies. Because the proxy noise variance at low frequencies cannot be directly estimated, and thus has to be regarded as unknown, it is safer to regard a reconstruction with the LOC method as providing an estimate of the upper bound of the large-scale low-frequency temperature variability rather than one with a correct estimate of this variance.

1. Introduction

Christiansen and Ljungqvist (2011, hereafter CL) have presented a reconstruction of extratropical Northern Hemisphere (NH) temperatures for the last millennium using the LOC method, which is explained in Christiansen (2011). According to Christiansen (2011) and CL the LOC method is designed to avoid underestimation of low-frequency climate variability, although this occurs at the expense of exaggerated high-frequency variability. These efforts should be viewed in light of a vigorous debate held during the last decade, where a range of statistical approaches for reconstructing past large-scale temperature variations from proxy data have been proposed and where their relative performance has been evaluated in more than 40 investigations based on pseudoproxy experiments [see Smerdon (2012) for a review]. In those experiments, output from climate model simulations is used to benchmark reconstruction methods in situations where the “true” climate variability is known, as defined by the simulations.

Christiansen et al. (2009) observed, in one such multi-method intercomparison, that all methods they investigated tend to underestimate low-frequency NH temperature variability. On the contrary, some other methods that were not analyzed by Christiansen et al. (2009) have performed well in pseudoproxy tests: notably, the two closely related methods of Hegerl et al. (2007) and Ammann et al. (2010) have not shown any significant loss of low-frequency variability.

It is beyond my scope here to try and summarize the many lessons learned from all pseudoproxy experiments undertaken so far, but I wish to point out that there are good reasons to assume that the LOC method not only avoids underestimation of low-frequency variance, but also has an opposite tendency, such that the new CL reconstruction may exaggerate temperature variability at the time scale for which it is presented, which is 50-yr averages. I will demonstrate why such an assumption is reasonable by first providing some thoughts from a theoretical viewpoint and then by undertaking a new pseudoproxy experiment.

2. Some thoughts on the LOC method

The basic idea underlying the LOC method is simple and defines the two essential components: First, we calibrate each local noisy proxy record such that its

Corresponding author address: Anders Moberg, Department of Physical Geography and Quaternary Geology, Stockholm University, SE-10691 Stockholm, Sweden.
E-mail: anders.moberg@natgeo.su.se

temperature signal component is correctly estimated, and then take an average of a sufficiently large number of proxy sites such that the supposedly uncorrelated noise terms effectively cancel and the average of the correctly estimated temperature signals dominate.

The method used in the first step of LOC is known in the statistical literature as “classical calibration” (e.g., Osborne 1991) or “controlled calibration” (e.g., Brown 1993). In our context, this means that one calibrates each proxy against local temperatures by regressing the proxy on the instrumental data and then inverts the regression slope to predict the local temperatures at each proxy location. In the second step, an average is taken over the calibrated local proxy series to obtain an estimate of temperatures averaged over the region that the collection of proxies represents together. This approach should guarantee that underestimation of variability is avoided, because it will actually lead to a climate reconstruction that has more variance than what would have been obtained if the instrumental observations could have been used for the entire period.

Why will this approach lead to too much variance? Let us first, for simplicity, assume that the instrumental observations are perfect (noise free) recorders of the true temperature variability at each proxy site. Also, assume that each proxy record consists of a temperature signal plus some noise (which is uncorrelated with temperature) and, further, that there is a linear relation between the temperature component in the proxy record and the true local temperature. Classical calibration then yields an unbiased estimation of the regression slopes that translate each proxy record to local temperature. This property, in turn, provides noise-contaminated local temperature reconstructions where the *temperature signal component* is estimated with its correct amplitude. This is the fundamental justification for the LOC method, as discussed in appendix A in Christiansen (2011). However, the local temperature reconstructions will have more variance than the corresponding instrumental records, because each local reconstruction will consist of a correctly estimated temperature signal *plus* noise. It is reasonable to assume that the noise components in the different proxy sites are uncorrelated with each other. Thus, when taking averages over a large number of calibrated local proxy series, the noise components will tend to partly cancel each other and the average of the correctly estimated temperature signals will tend to dominate over the noise. But any average over a finite number of proxy series must contain *some* noise superimposed on the temperature signal. Their average would be the same as the average of the corresponding instrumental temperatures only in the hypothetical situation of an infinite number of calibrated proxy series, and thus no overestimation of variance would occur.

According to Brown (1993) the idea of classical (or controlled) calibration goes back at least to Eisenhart (1939), who contended that, for prediction of the explanatory variable in a linear regression model, one should minimize the errors in the direction in which they occur. Temperature is the explanatory variable in the temperature reconstruction case. Thus, when assuming that the instrumental data are noise-free, the error (noise) only occurs in the proxy data. Hence, the errors should be minimized in the direction of the proxy; and this is what the LOC method does. However, when noise occurs also in the instrumental observation data, an adjustment of the calibration is needed to obtain an unbiased estimate of the regression slope. Without such an adjustment, using errors-in-variables (EIV) methods (also called measurement error methods; e.g., Fuller 1987), the approach explained above will lead to an exaggerated variance of the temperature signal component. Several authors (e.g., Hegerl et al. 2007; Ammann et al. 2010; Christiansen 2011; Kutzbach et al. 2011; Moberg and Brattström 2011) have recently discussed the theory of EIV methods in a paleoclimate reconstruction context.

It is of interest to note that the two EIV-based temperature reconstruction methods proposed by Hegerl et al. (2007) and Ammann et al. (2010) share a central idea with the LOC method, namely, that they seek to obtain an unbiased relation between the true temperature and the proxy. However, Hegerl et al. (2007) and Ammann et al. (2010), unlike Christiansen (2011) and CL, do not calibrate individual proxies against local temperatures but rather use the large-scale temperature average directly as the calibration target. Ammann et al. (2010, p. 277) state that their method results in an increased variance, which is most concentrated at the interannual scale but decadal smoothing essentially compensates for this. Importantly, they further state that with red noise in the proxy data, variance increase will also appear over longer time scales. It seems natural to assume also for the LOC method that variance can be overestimated at *all* frequencies, not just at the high frequencies. I will now more closely discuss why this may affect the CL reconstruction.

Let us, for simplicity, ignore the problem of how to estimate temperatures representing the average over the entire NH or, for example, the part of NH north of 30°N, from a set of temperature series from a small number of sites. Instead, consider an arithmetic average of (noise free) temperature data from these sites as the “truth,” which we want to predict from the calibrated proxies. If the noise in the individual proxy records has variance at all frequencies, then so will their average. (We assume noise across sites is uncorrelated at all frequencies.) However, if the noise variance varies with frequency in a systematic way (e.g., with less noise variance

at low frequencies than at high frequencies), then this character will be reflected also in the proxy average. Thus, the question of whether the LOC method “preserves low-frequency variance” is a question of how much low-frequency noise there is in the individual proxy series and whether the number of proxy series is sufficiently large for the low-frequency proxy noise at the individual sites to cancel out when taking their average.

Is the low-frequency noise variance in the individual proxies used by CL small enough, and is the number of them large enough, that the low-frequency variance in their reconstruction is (nearly) correctly estimated? Their main reconstruction is based on 23 proxy series that pass screening against local temperature data by means of a significance test of correlation. The accepted proxies have correlations with local temperature data typically around 0.4. In other words, these proxy series contain a considerable amount of noise; typically about 80% of their variances (integrated across all frequencies that can be analyzed using data for the calibration period) is noise. It is relevant to ask how much of this noise variance that occurs at low frequencies, in particular corresponding to time scales longer than the moving 50-yr averages used in the figures. Noise variance at such low frequencies can hardly be estimated simply by comparison with the rather short instrumental records, and thus it has to be considered as an unknown quantity. Moreover, there is certainly no guarantee that proxy records that passed the screening are not affected by substantial low-frequency distortions affecting their long-term temporal evolution in the prescreening period.

Some information on the potential presence of low-frequency distortions can be obtained by examining the 50-yr smoothed records for the 23 proxies that passed the screening (see Fig. 8 in CL). In some cases, pairs or groups of geographically neighboring proxies exist. In these cases one should expect significant positive correlations among the neighbors when the smoothed data are analyzed. There are, however, some cases where the correlation is insignificant or even negative. I have observed this after having first downloaded data from ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/christiansen2011/christiansen2011.txt, then processed the proxy records much as CL did. After linear interpolation to annual resolution (when needed) and then application of a 50-yr moving average filter, I found negative correlations between their two records from Alaska (records 2 and 18, $r = -0.45$) and between the two Greenland ice-core records GISP2 and GRIP (records 16 and 17, $r = -0.03$). Small positive but statistically insignificant correlations are found between the two tree-ring records from Sweden (records 23 and 37, $r = 0.09$) and between two Chinese records (records 12 and 14, $r = 0.14$).

Furthermore, visual inspection of CL's Fig. 5 reveals that some proxy records that passed the screening (e.g., records 8 and 25) have trends that deviate from the corresponding local instrumental temperature records over periods spanning more than a decade.

I interpret these observations as evidence of potentially substantial low-frequency noise in at least some of the proxies used. Although the reasons for any low-frequency distortions are unique for each affected proxy record, one may point out some general plausible reasons, as discussed, for example, by Jones and Mann (2004) and Jones et al. (2009). For example, dating uncertainties in sedimentary records may be on the order of several decades or more. Together with potential stratigraphic mixing (bioturbation), or just by itself, this may cause distortions on multidecadal and longer scales in reconstructed temperature series derived. Another example is the necessary removal of biological growth trends, by means of some curve-fitting technique, that is applied to individual tree-ring series when tree-ring chronologies are constructed. Depending on how this is done, the multidecadal variability in the resulting temperature reconstructions is affected to varying extent (Melvin and Briffa 2008; Briffa and Melvin 2011). Also, temporally varying multivariate influences (i.e., not temperature alone) may influence several types of proxy records including trees, lake sediments, speleothems, and ice cores (Jones and Mann 2004; Jones et al. 2009). Such different types of potential low-frequency distortions cannot be collectively modeled by a single simple stochastic noise model. But we can at least make some simple assumptions and then analyze what kind of potential problems we are facing. I will do so in the next section.

3. A pseudoproxy experiment

I will here describe a pseudoproxy experiment that mimics essential details in the LOC reconstruction by CL. Temperature data from the 991-yr-long Erik-II simulation with the coupled ECHAM4 and the global Hamburg Ocean Primitive Equation (ECHO-G) model (González-Rouco et al. 2006) are used to define the pseudoworld within which the experiment is made. This simulation includes solar, volcanic, and greenhouse gas forcing for the period AD 1000–1990 and has been used in previous pseudoproxy studies (e.g., Mann et al. 2007; Moberg et al. 2008; Riedwyl et al. 2009). From its 2-m temperature field, I select annual mean data for the grid points closest to each of the 23 proxy sites used by CL. The arithmetic mean of these 23 temperature series is then taken as the truth, which I try to predict from pseudoproxy data for the same sites.

These pseudoproxies are constructed by adding noise to the temperature series, where the noise variance is

chosen such that the expected correlation between each pseudoproxy and its corresponding temperature series is the same as the correlations found by CL (as given in their Fig. 5). With this way of constructing the pseudoproxies, no calibration is needed because classical calibration would in this situation simply mean multiplication with the factor one. The crucial property is that each pseudoproxy contains the true temperature signal with its correct amplitude plus the same amount of noise as in the CL proxies. For simplicity, I will ignore the fact that some of the CL proxies have decadal resolution. I will instead consider all as being annually resolved, as the experiment is intended to point out a principal problem that has nothing to do with the temporal resolution of the proxies.

It is not obvious how to choose the structure of the noise. What I want to demonstrate is how the level of noise at different time scales will affect the variance of the final reconstruction at different time scales. This implies that the variance spectra of the noise and the reconstructions are the interesting aspects to consider. To this end, it is sufficient to choose two different types of noise with notably different variance spectra. In previous pseudoproxy studies, a first-order autoregressive process [AR(1)] has sometimes been used as a model for the noise. Such a process is determined by its lag-1 autocorrelation parameter ρ . A value of $\rho = 0$ corresponds to white noise, which has a power spectrum that is constant for all frequencies. However, its variance spectrum, which is the power spectrum multiplied by frequency, monotonously increases from low to high frequencies (Fig. 1a). Such a process is not realistic in a pseudoproxy study as it has almost no variance at the lowest frequencies. Nevertheless, it has been the most frequently used type of noise (Smerdon 2012). Mann et al. (2007) argued that a value of $\rho = 0.32$ is more reasonable for annually resolved proxies and this value has been used in some pseudoproxy studies (e.g., Mann et al. 2007; Lee et al. 2008; Moberg et al. 2008; Riedwyl et al. 2009). The substantially larger value of $\rho = 0.70$ or $\rho = 0.71$ is also sometimes used (von Storch et al. 2006; Mann et al. 2007; Moberg et al. 2008; Riedwyl et al. 2009), although Mann et al. (2007) argued that this value is unrealistically large for many annually resolved proxies. However, it gives considerably more variance at low frequencies than with $\rho = 0.32$ and is thus a relevant choice to demonstrate the effect of different noise variance spectra. In normalized noise realizations (Fig. 1a), there is more variance with $\rho = 0.71$ than with $\rho = 0.32$ at all low frequencies corresponding to time periods longer than about a decade. At higher frequencies, there is more variance with $\rho = 0.32$.

I choose AR(1) noise with $\rho = 0.32$ and $\rho = 0.71$ for my pseudoproxy experiments, where I add a random

FIG. 1. Wavelet variance spectra for (a) realizations of three types of noise with unit variance and (b) LOC pseudoproxy reconstructions. (b) The thick black line shows the spectrum for the simulated area-weighted temperature north of 30°N , with 95% confidence intervals defined by the dashed lines. The thick green line shows the spectrum for the average of 23 locations. This is the "true" target for the LOC reconstructions, whose spectra are identified by colors defined in the graph. Noise spectra in (a) and reconstruction spectra in (b) are represented by ensemble medians (thick lines) and 2.5 and 97.5 percentiles (thin lines). (c) The ratio between the reconstruction ensemble median variances and the variance of the 23-site "truth."

realization of each type of noise 1000 times to each local temperature series. This makes it possible to analyze variance spectra in ensembles of LOC reconstructions, with 1000 ensemble members for each type of noise.

FIG. 2. (a)–(d) Time sequences for 10 arbitrarily selected members of LOC pseudoproxy temperature reconstructions, for each of four cases explained in the text (thin lines, different colors). The thick black line in each panel shows the time sequence for the simulated area-weighted temperature north of 30°N. The thick green line shows the average over the 23 pseudoproxy locations. All data are shown as moving 50-yr averages and as anomalies from their long-term means.

Figure 1b shows the estimated variance spectra for the reconstructions and their target. All variance spectra are computed with wavelet analysis tools, using the same method (Percival and Walden 2000) as in the pseudoproxy study by Moberg et al. (2008). Note that the wavelet variance spectrum is defined only for a discrete set of frequencies corresponding to the respective wavelet scales. For the specific wavelet used here [called LA(8); see Percival and Walden 2000], the smallest scale corresponds to Fourier periods of about 2.8 years and the scales are separated by a power of 2. The largest scale is dictated by the length of the time series being analyzed; here this scale corresponds to about 720-yr periods.

The thick black line in Fig. 1b is the spectrum for the area-weighted mean over all grid points north of 30°N, which corresponds to the extratropical reconstruction target in CL. The dashed black lines show its 95% confidence interval (Percival and Walden 2000). The thick green line is the spectrum for the arithmetic average over the 23 local temperature series (i.e., the “truth” with which the reconstructions should be compared).

This green line lies close to the spectrum for the extratropical mean at the lowest frequencies corresponding to periods longer than about 180 years, whereas at higher frequencies there is more variance in the 23-series mean (even outside the black 95% confidence interval for periods shorter than about 45 years). This property, with too much variance at high frequencies but about the right variance at low frequencies, is a consequence of fewer spatial degrees of freedom in the temperature field at low frequencies than at high frequencies (e.g., Jones et al. 1997). Thus, at the lowest frequencies it is sufficient (in Erik-II) to take an average over the 23 sites to obtain a reconstruction with about the right variance, whereas at high frequencies data would be needed from more sites. This is, however, not the problem I am focusing on here: the matter of interest is to compare the variance spectra of the LOC reconstructions with the spectrum for the true 23-series mean.

The red lines in Fig. 1b show the spectrum for reconstructions with noise having $\rho = 0.71$, while dark blue lines show the spectrum for the case with $\rho = 0.32$.

Spectra for individual reconstructions are not shown; instead, for each wavelet scale, the median (thick lines) of the 1000 members is plotted together with the 2.5 and 97.5 percentiles (thin lines). To make it easier to visualize how much these spectra deviate from that of the 23-series truth, the ratio between the median variance for each ensemble and the variance of the truth is shown with red and dark blue lines, respectively, in Fig. 1c, separately for each scale. As expected from the theoretical considerations, the reconstructions' spectra have on average (as represented by the ensemble medians) more variance than the spectrum of the truth (i.e., the variance ratio is >1) at all scales. *This confirms my main suspicion, namely, that the LOC method may exaggerate variance at all time scales.*

It is of particular interest to study the variance spectra and ratios for the periods that are longer than the 50-yr mean used by CL in their time series plots (i.e., at the roughly 90-, 180-, 360- and 720-yr scales computed in the wavelet spectra). The reconstructions with $\rho = 0.32$ have on average about 30% too much variance at the 90-yr scale, but the variance inflation decreases to about 16% at 180 years, 8% at 360 years, and less than 1% at 720 years. This decreased variance inflation toward the longer periods reflects the fact that the variance spectrum for noise with $\rho = 0.32$ decreases toward the longer scales (Fig. 1a). At periods shorter than about 90 years, however, there is considerably too much variance, with a maximum of 3.5 times too much variance at the 5.6-yr scale.

For reconstructions contaminated with noise having $\rho = 0.71$, there is on average nearly twice as much variance at the 90-yr scale compared to the truth. As with the less red noise case, the variance inflation decreases toward the longest scale, with a median inflation by about 50% at 180 years, 30% at 360 years and 3% at the longest scale. The maximum variance inflation, by a factor of 3.7, occurs at the 22-yr scale.

Both types of proxy noise used here have rather little variance at the very lowest frequencies (Fig. 1a). If another type of noise, with more long-term persistence (as discussed, e.g., in Rybski et al. 2008) and thus relatively more variance at the lowest frequencies, had been used, then we would have seen more overestimation of variance at the lowest frequencies than we can see in this experiment. However, even if the two types of noise chosen here may not be realistic representations of the noise in the CL proxies, the main message is clearly demonstrated:

The LOC method can provide a reconstruction with a reasonably correctly estimated variance only at those frequencies where the noise variance in individual proxies is sufficiently small to be effectively cancelled when taking the average over the proxies used.

This message can be further illuminated by showing results for a hypothetical situation where the proxy sample size is increased. Increasing the sample size means that averages are taken over a larger number of proxy series, where each is calibrated to have the temperature signal component correctly estimated. This implies that the noise variance in the resulting reconstruction would decrease compared to when averaging over a smaller sample, if noise across proxies is uncorrelated. I have undertaken two such experiments, where I simply produced first 3 and then 10 replicates of pseudoproxies with noise having $\rho = 0.32$ for each of the 23 proxy sites. Thus, I obtain two situations with 69 and 230 pseudoproxies, respectively. For each replicate, ensembles with 1000 noise realizations are produced. Medians and percentiles for variance spectra of the resulting LOC reconstructions and the corresponding variance ratios are shown as light blue and gray lines in Figs. 1b and 1c. An increase of the proxy sample size clearly improves the performance of the LOC reconstruction method. With 69 pseudoproxies, the median variance inflation at the 90-yr scale is reduced to about 11% and with 230 proxies the corresponding inflation is less than 4%. The inflation at the longer scales is even smaller. These numbers should not be regarded as any recommendation for how many proxies that would be needed to yield a LOC reconstruction with negligible variance inflation. But the experiment demonstrates how the variance inflation in a LOC reconstruction—for a given noise type and local signal-to-noise ratio in the proxies—can be diminished by increasing the proxy sample size. Figure 2 helps to visualize these characteristics also in the time domain, by showing 50-yr moving averages (i.e., as in CL) for 10 arbitrary members from each of the four ensembles.

4. Conclusions

CL have presented an extratropical NH temperature reconstruction for the last millennium using the LOC method, which they claim “preserves” low-frequency variability, at the expense of an exaggerated high-frequency variability. Using theoretical arguments and a pseudo-proxy experiment, I have demonstrated that the LOC method is not guaranteed to preserve variability at any frequency. In fact, it may lead to exaggerated variance at all frequencies. This variance inflation, though, can be negligible for those frequencies where the noise variance is small enough to be effectively cancelled when computing an average over the available proxies. Thus, for the CL reconstruction, we would need to know the noise variance for each proxy at time scales longer than the 50-yr means they use in their main figures. This noise

variance, however, cannot be estimated from comparison with the rather short instrumental record. Therefore, we do not know if an average over their 23 proxies will be significantly contaminated by low-frequency noise. A safe way to view the CL reconstruction is therefore to regard it as providing an estimate of the upper bound of the variance of 50-yr mean extratropical temperatures. The same thinking would hold for any other reconstruction produced with the LOC method.

Acknowledgments. This work was funded by the Swedish Research Council (VR) Grants 90751501 and B0334901. Eduardo Zorita kindly provided the ECHO-G data.

REFERENCES

- Ammann, C. M., M. G. Genton, and B. Li, 2010: Correcting for signal attenuation from noisy proxy data in climate reconstructions. *Climate Past*, **6**, 273–279.
- Briffa, K. R., and T. Melvin, 2011: A closer look at regional curve standardization of tree-ring records: Justification of the need, a warning of some pitfalls, and suggested improvements in its application. *Dendroclimatology: Progress and Prospects*, M. K. Hughes, T. W. Swetnam, and H. F. Diaz, Eds., Springer, 113–146.
- Brown, P. J., 1993: *Measurement, Regression, and Calibration*. Oxford University Press, 201 pp.
- Christiansen, B., 2011: Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided? *J. Climate*, **24**, 674–692.
- , and F. C. Ljungqvist, 2011: Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. *J. Climate*, **24**, 6013–6034.
- , T. Schmith, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, **22**, 951–976.
- Eisenhart, C., 1939: The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Stat.*, **10**, 162–186.
- Fuller, W. A., 1987: *Measurement Error Models*. Wiley, 440 pp.
- González-Rouco, J. F., H. Beltrami, E. Zorita, and H. von Storch, 2006: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling. *Geophys. Res. Lett.*, **33**, L01703, doi:10.1029/2005GL024693.
- Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new, validated 1500-year temperature reconstruction. *J. Climate*, **20**, 650–666.
- Jones, P. D., and M. E. Mann, 2004: Climate over past millennia. *Rev. Geophys.*, **42**, RG2002, doi:10.1029/2003RG000143.
- , T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate*, **10**, 2548–2568.
- , and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *Holocene*, **19**, 3–49, doi:10.1177/0959683608098952.
- Kutzbach, L., B. Thees, and M. Wilmking, 2011: Identification of linear relationships from noisy data using errors-in-variables models—Relevance for reconstruction of past climate from tree-ring and other proxy information. *Climatic Change*, **105**, 155–177, doi:10.1007/s10584-010-9877-7.
- Lee, T. C. K., F. W. Zwiers, and M. Tsao, 2008: Evaluation of proxy-based millennial reconstruction methods. *Climate Dyn.*, **31**, 263–281, doi:10.1007/s00382-007-0351-9.
- Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.
- Melvin, T. M., and K. R. Briffa, 2008: A “signal-free” approach to dendroclimatic standardisation. *Dendrochronologia*, **26**, 71–86, doi:10.1016/j.dendro.2007.12.001.
- Moberg, A., and G. Brattström, 2011: Prediction intervals for climate reconstructions with autocorrelated noise—An analysis of ordinary least squares and measurement error methods. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **308**, 313–329, doi:10.1016/j.palaeo.2011.05.035.
- , R. Mohammad, and T. Mauritsen, 2008: Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction. *Climate Dyn.*, **31**, 957–971, doi:10.1007/s00382-008-0392-8.
- Osborne, C., 1991: Statistical calibration: A review. *Int. Stat. Rev.*, **59**, 309–336.
- Percival, D. B., and A. T. Walden, 2000: *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 4, Cambridge University Press, 594 pp.
- Riedwyl, N., M. Küttel, J. Luterbacher, and H. Wanner, 2009: Comparison of climate field reconstruction techniques: Application to Europe. *Climate Dyn.*, **32**, 381–395, doi:10.1007/s00382-008-0395-5.
- Rybski, D., A. Bunde, and H. von Storch, 2008: Long-term memory in 1000-year simulated temperature records. *J. Geophys. Res.*, **113**, D02106, doi:10.1029/2007JD008568.
- Smerdon, J., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *WIREs Climate Change*, **3**, 63–77, doi:10.1002/wcc.149.
- von Storch, H., E. Zorita, J. M. Jones, J. F. González-Rouco, and S. F. B. Tett, 2006: Response to comment on “Reconstructing past climate from noisy data.” *Science*, **312**, 529, doi:10.1126/science.1121571.