Climatic Change (2007) 85:71–88
DOI 10.1007/s10584-007-9276-x

# The importance of the geophysical context in statistical evaluations of climate reconstruction procedures

**Caspar M. Ammann · Eugene R. Wahl**

Received: 22 August 2000 / Accepted: 13 June 2007 / Published online: 24 August 2007

**Abstract** A portion of the debate about climate reconstructions of the past millennium, and in particular about the well-known Mann-Bradley-Hughes ("MBH" 1998, 1999) reconstructions, has become disconnected from the goal of understanding natural climate variability. Here, we reflect on what can be learned from recent scientific exchanges and identify important challenges that remain to be addressed openly and productively by the community. One challenge arises from the real, underlying trend in temperatures during the instrumental period. This trend can affect regression-based reconstruction performance in cases where the calibration period does not appropriately cover the range of conditions encountered during the reconstruction. However, because it is tied to a unique spatial pattern driven by change in radiative balance, the trend cannot simply be removed in the method of climate field reconstruction used by MBH on the statistical argument of preserving degrees of freedom. More appropriately, the influence from the trend can be taken into account in some methods of significance testing. We illustrate these considerations as they apply to the MBH reconstruction and show that it remains robust back to AD 1450, and given other empirical information also back to AD 1000. However, there is now a need to move beyond hemispheric average temperatures and to focus instead on resolving climate variability at the socially more relevant regional scale.

## 1 Introduction

According to all commonly used instrumental climate data series, ten (NASA-GISS) or eleven (HadCRU and NOAA) out of the twelve most recent years have been among the twelve warmest on record for the Northern Hemisphere. Although this warming is generally in good agreement with climate model-based projections (e.g., Tett et al. 1999; Folland et al. 2001;

---

C. M. Ammann (✉)
Climate and Global Dynamics Division, National Center for Atmospheric Research,
1850 Table Mesa Drive, Boulder, CO 80305, USA
e-mail: ammann@ucar.edu

E. R. Wahl
Environmental Studies and Geology Division, Alfred University,
One Saxon Drive, Alfred, NY 14802, USA

Meehl et al. 2004; Hansen et al. 2005), the precise separation of the anthropogenic signal from natural variations is difficult. The temporary cooling due to the Pinatubo eruption exemplifies this issue through superposition of natural and anthropogenic forcing signals along with inherent internal variability of the climate system (an El Niño was going on at the time). A significant portion of the debate concerning the nature and intensity of anthropogenic climate change has recently shifted its attention to natural climate variability prior to significant human alterations of atmospheric composition (cf. National Research Council (NRC) Report 2006). How much do we know about the hemispheric climate of the past Millennium? What has caused these fluctuations? The most critical arguments in this context concern the limited information we have about the past, and more generally, the fact that climate has always been varying and will continue to change naturally. Such reasoning challenges instrumental- and proxy-based disciplines alike to come up with (and test) plausible explanations of observed changes by putting the available pieces of evidence into a self-consistent geophysical framework. Over the last several years, process-oriented interpretations of this kind have become increasingly comprehensive and robust. A great deal of information concerning the nature and mechanisms of climate change no longer hinges on simple assumptions or correlations, but now fits into a multi-disciplinary, spatio-temporal geophysical systems perspective held together by both theory and observations. This self-consistent framework has allowed the warming over the past few decades to be formally *detected and attributed* to the anthropogenic influence on the atmosphere (Folland et al. 2001; Hegerl et al. 2003, 2007; Santer et al. 2003; Mears and Wentz 2005; Stott et al. 2006).

Our knowledge about the climate of the pre-industrial past, however, is significantly more uncertain. In part, this situation is not surprising because the record is obviously less complete and most information has to be inferred from proxy data. Additionally, as we consider in more detail here, the combined facts that the instrumental record is rather short (only about 100 to 150 years long) and at the same time coincides with the period over which humans have steadily been increasing their influence on the system, makes it harder to isolate natural aspects of variability. Nevertheless, the underlying question is very much the same, independent of being applied to the past or present: Can we *understand* the general causes for the observed or reconstructed climate, despite uncertainty and the noise arising from internal variations? For an explanation to be scientifically acceptable, it has to be *consistent* for both the past and the present.

The ultimate goal of climate reconstructions is to provide basic information for the study of cause and effect in the climate system, and thus, despite clear restrictions and uncertainties, high resolution climate reconstructions *fundamentally represent an attempt to extend the instrumental record back in time*. Depending on what questions are addressed, one can exploit the proxy information with different techniques, estimating local, regional, continental and even hemispheric aspects of climate and its variability (global patterns are still difficult to reconstruct due to extreme limitation of Southern Hemisphere records). Most challenging are reconstructions that attempt to resolve the underlying spatio-temporal structures in the evolving climate fields. These climate field reconstructions (CFRs) not only provide information over regions where direct proxy data are available, but they can make use of the fact that the climate system anomalies can be projected onto a relatively small set of principal modes of variability (Kaplan et al. 1997; Mann et al. 1998; Corti et al. 1999; Palmer 1999; Stephenson et al. 2004). This ability allows exploitation of systematic teleconnection relationships between proxies and variations in certain modes. Although the underlying concept is relatively simple, such approaches can be complex to implement. The benefit, however, of obtaining spatially resolved climate information is substantial. It serves as the essential ingredient in the scientific process of extending the short instrumental record and, thus, for building a coherent geophysical understanding of how the atmosphere, oceans, and other components in the climate system vary by themselves, and how they respond jointly to external forcing. In this light, it becomes clearer that averages of hemispheric or global climate are more of a secondary product, rather than the fundamental goal, of CFRs. Also from a "user" perspective (be it the broad public or decision makers), there is a clear need for techniques that help identify and explain climate variations on the most relevant and accessible of spatial scales, the local and regional levels (Luterbacher et al. 2004; Xoplaki et al. 2005; Casty et al. 2005a). Such information would be particularly useful if it could also go beyond annual average temperature and resolve seasonal variations, and even more so if hydrologic and circulation information could be captured.

Throughout such efforts, it is also important to be aware of and to define various levels of uncertainty that are inherent in the reconstruction process. It is in the interest of the scientific community to identify uncertainties in data (in individual series as well as networks through their sampling density and spatial coverage), reconstruction procedures (reconstruction options, properties of residuals), and interpretations (distinguishing between interannual and longer term, i.e. climatic, skill) very carefully and to cross-verify findings in order to limit spurious or non-robust results. Using climate models, several of these uncertainties can be studied and quantified (see below; von Storch et al. 2004, 2006; Bürger et al. 2006; Mann et al. 2007; Küttel et al. 2007). Recently, part of the scientific debate about uncertainties in the available climate reconstructions has become rather heated and counter-productive. If we are to advance our knowledge of how the climate system really works, then we have to be able to scale down the underlying tone in the debate and be ready to discuss the various issues openly and with a constructive intent. Rather than approaching this science as if it were a "make-or-break" business deal, all parties involved need to omit inflammatory language (and editors should require removal of such language) and be more constructive. We should focus on key problems, such as: (a) the remaining issues of how well the various proxies retain and store climate signals at various temporal and spatial scales; and (b) how best to circumvent the burden put on reconstructions and their verification procedures resulting from the short instrumental record, which additionally comes with an "inconvenient" trend. In this process, we should not lose sight of the primary goal of generating information with spatial resolution, rather than focusing on hemispheric temperatures alone. Only by understanding the spatial – and thus dynamical – structures of past climate will we be able to address and answer a number of important questions about climate changes (Casty et al. 2005b; Xoplaki et al. 2005). In the following sections, we discuss what we have learned from the recent debate around the Mann-Bradley-Hughes (MBH) reconstruction and its "hockey stick" Northern Hemisphere temperature output (Mann et al. 1998, 1999, subsequently called MBH98 and MBH99), and put key issues into a larger perspective. One overarching conclusion we find, which needs to be highlighted from the outset, is that despite a clear need for careful consideration of procedural details and minutia, in the end the big picture should not hinge on them. And as noted, the results need to fit into a logical, geophysical framework that is self-consistent. These requirements alone constitute an important, and general, measure of robustness.

Additional information and illustrations beyond the present text are provided in an electronic supplement and on our WEB site (http://www.cgd.ucar.edu/ccr/ammann/millennium/MBH-reevaluation.html).

## 2 Proxy PCs and bristlecone pines issues

The MBH reconstruction has been evaluated on various grounds. One focus has been on details of mathematical data processing steps and the data themselves (McIntyre and

McKitrick 2003, 2005a,b; referred to as "MM" with a date designation). Wahl and Ammann (2007, this issue; referred to as "WA") systematically evaluate and discuss these issues. It is found that the original results from MBH98, given the underlying assumptions and input data, are reproducible, and despite some reasonable corrections, hold up to scrutiny.

The primary issue concerning the MBH climate reconstruction arose from a data pre-processing step concerning how to generate principal component (PC) summaries for a set of tree ring series, which subsequently are put into the reconstruction algorithm instead of the individual data series themselves. From first principles, it should be apparent that – independent of the actual procedures used to calculate and determine the number of "representative" PCs (i.e., independent of centering conventions, whether the data are normalized or not, whether the variance–covariance matrix or the correlation matrix is used in PC calculation, and how one identifies the number of PCs used to represent the full data) – all appropriate procedures should lead to a very similar climate reconstruction because the underlying information does not change. The only difference in process should relate to how the signal is separated from noise across the rank ordering of the PCs. In the end, all important climate information should be extractable and useable.

Although applicable for all reconstruction subsets, MM have focused on the 15th century segments of the MBH98 reconstruction, for which they suggested that the fundamental conclusions of MBH98 would not hold if PC calculations were done correctly (cf. WA for a full discussion of this claim). One can easily test this proposition in a few iterative steps: for each PC calculation method the climate reconstruction is performed using successively larger numbers of proxy PCs until convergence of the reconstructions occurs (see Figure S1, supplementary electronic material). Based on this analysis, warm early 15th century conditions that approach late 20th century temperatures (MM05b) can be demonstrated to be an artifact of data truncation, which is reflected in the lack of convergence in the reconstructions and poor validation performance of the non-converged results (WA). In all "flavors" of calculations without data loss, independent of the centering, scaling or PC calculation method applied, all climate reconstructions in fact converge towards a relatively cooler hemispheric temperature series that resembles the original MBH result (Figure S1).

MM05a have, however, correctly identified a systematic bias arising from the short-centering PC calculation convention used in MBH (centering and normalizing relative to the mean and standard deviation of the 1902–1980 calibration period rather than over the full length of the proxy data series). This methodological bias can easily be corrected (WA), and results show that there are no significant effects on the reconstruction itself. Rather, the impact is very small, a fact that can be recognized through comparison of the resulting principal components (see electronic supplement, Figure S2). Application of the simple convergence test to the resulting climate reconstructions then confirms that the centering choice does not introduce a different type of reconstruction outcome, only that the offset slightly inflates the amplitude of the original MBH reconstruction, particularly so during the early 15th century, where this effect reaches approximately 0.05° (WA). While it is very important to be cautious about data (pre-) processing, one also has to be careful not to truncate information when representing a large number of individual records through a subset of their PCs – a danger that exists in examination of the methods used by MBH if one focuses exclusively on the first PC (NRC 2006; Wegman et al. 2006) or an a priori fixed number of PCs (MM05a,b). Convergence in the resulting climate reconstruction is a straightforward empirical test to guard against inappropriate data truncation. In the end, it is elucidation of the signal in the proxy data that matters, which – with proper attention – is transparent to details of how the signal is retained across PCs.

With regard to the proxies themselves, MM have emphasized (as MBH99 had noted already) that the much discussed bristlecone/foxtail pine (BC) series play an important role in MBH. What is often overlooked, however, is that the BC influence is only crucial over the first 50 years in MBH98 (cf. WA, and in particular WA-Figure 5). The whole issue can be summarized in a simple way (see electronic supplement, Figure S3): *if BC series are included*, then the shape of the full 1400–1980 reconstruction of MBH is essentially unchanged, although there is an option for a small amplitude correction in the early 15th century (WA); s*hould BC series be excluded*, then the MBH reconstruction would have to start in 1450. Thus, the sensitivity to the BC series – and in fact all of the important MM arguments about MBH98 – only concerns the 1400–1449 segment of MBH98. A detailed discussion for or against inclusion of BC series is provided in WA, where it is concluded that the evidence indicates the BC data are not introducing spurious information in this early segment.

The debate around the bristlecone pine issue, along with that of proxy PC calculation methods, provide good examples of how details can sometimes unnecessarily obscure the bigger picture, where it should be clear that different approaches concerning how exactly to separate signal from noise simply *cannot lead to greatly different results* (as some outcomes have been interpreted), as long as information does not get truncated or otherwise altered in significant ways. However, the important point remains that pre-processing of data needs to be performed with significant caution in order to avoid introduction of systematic biases, even if these may be quite small.

## 3 The 20th century trend

Climate reconstructions have to deal with the limited length and underlying trend of the instrumental data. In the literature, this issue has – explicitly or implicitly – been approached from different angles. While the brevity of the record is something scientists must simply accept, options exist to limit the influence of the trend from unduly dominating calibration algorithms. In fitting procedures, one would ideally first remove any trends and then establish the link between predictor (proxy) and predictand (some instrumental time series). Further justifying a trend removal, von Storch et al. (2004; personal comm.) have made the argument that it should not matter on what time scale the link is established between a proxy record and the (local) temperature. Thus, not being fully aware of the details of the MBH algorithm (von Storch et al. 2006), the authors removed the underlying trend in an attempt to optimize the fitting procedure, arguing that at least part of the 20th century trend could be due to non-climatic changes in the proxies (von Storch et al. 2004; a concern that cannot be entirely refuted). Bürger and Cubasch (2005) and Bürger et al. (2006) went a step further and evaluated a host of possible procedural choices that have to be made in an MBH-type reconstruction. They considered the influence of different regression approaches and combined them with detrending as an option for their reconstruction matrix, and it is this latter component that is of concern here.

While the intent of guarding against overfitting is important, detrending does not work in the MBH reconstruction framework because of its underlying geophysical and mathematical concept. The MBH approach estimates the temporal weights of the dominant patterns of variability (called empirical orthogonal functions, or EOFs) to reconstruct climate back in time. It is important to recall that at least the leading EOF patterns have some (more or less direct) relationship with the large-scale geophysical processes that combine to make up the variability and trends in the spatio-temporally evolving surface temperature field. By removing the underlying trend in calibration, which is the first-order time trajectory of MBH's instrumental EOF #1, the assumption is implicitly made that the pattern related to the trend can still be achieved in the reconstruction through a recombination of the other retained patterns. Because the MBH reconstruction is based on truncated EOFs (meaning that it does not attempt to reconstruct all, but rather a subset, of the dominant temperature patterns that can be estimated with reasonable accuracy back in time; cf. MBH98 and MBH99), what is left to capture this primary variance are only modes that resemble strong interannual to decadal variability, such as ENSO, or high latitude modes related to the AO/NAO. While these internal modes of variability might contain some information in terms of the 20th century trend (Hurrell et al. 2004), they alone do not represent the principal geophysical process that dominates global/hemispheric climate on the longer time scales: a simple change in energy balance (Crowley 2000; Wahl et al. 2006). If no separate energy balance-related pattern exists (because it has been effectively removed), then the signal from the rather uniform warming currently occurring over the globe (with a few dynamically expected non-uniformities) cannot be fully "recognized" as such by the limited, highly-varying interannual to decadal modes. In fact, the trend information would then project onto any of the remaining signal and residual noise patterns that carry minor components of the energy balance effect. Given the truncated-EOF method, some, and perhaps a substantial, amount of the structure representing the overall energy balance will thus get lost.

In short, the detrending procedure used in von Storch et al. (2004) and numerous Bürger and Cubasch (2005) and Bürger et al. (2006) variants systematically removes a good deal of what, in the end, is expected to dominate the mean hemispheric temperature variation back in time (Crowley 2000; Bürger et al. 2006). The fact that the model-based exercises are still capable of capturing the correct overall shape (relative evolution in time) of the target series shows that part of the energy balance signal is present in the first few high-variance patterns, but the detrending in combination with the EOF truncation has caused a significant reduction of its full magnitude. This fact can easily be recognized in the poor calibration performance (strong loss of amplitude) of such reconstructions (cf. Wahl et al. 2006). Unless a way can be found to properly restore the energy balance information contained in the underlying trend at the reconstruction level (e.g., through separate calibration of low- and high-frequency components, one form of which can be found in Rutherford et al. 2005 and Mann et al. 2007), one needs to include it in the MBH reconstruction procedure as an important geophysical structure.

## 4 Validation thresholds and measures of merit

While removal of the 20th century trend in the MBH truncated EOF framework is not an appropriate solution from a geophysical standpoint, the intent of guarding against the loss of degrees of freedom over the calibration period that arises from an underlying trend is clearly a consideration that requires attention. A more suitable alternative to achieve this goal can be found in more extensive tests of reconstruction significance. Cross-validation in an independent verification period with carefully chosen significance thresholds can help to explicitly guard against spurious "success" in reconstructing the 20th century trend. This issue has been discussed before in MBH98 (cf. WA), and then by McIntyre and McKitrick (2005a,c). Particularly, MM05c (cf. Huybers 2005) have evaluated the extent to which random red-noise pseudoproxy series can generate spurious verification significance when propagated through the MBH reconstruction algorithm. It is important to examine this issue further and compare it against other methods of benchmarking verification significance.

MBH and WA argue for use of the of Reduction of Error (RE) metric as the most appropriate validation measure of the reconstructed Northern Hemisphere temperature within the MBH framework, because of its balance of evaluating both interannual and long-term mean reconstruction performance and its ability thereby to avoid false negative (Type II) errors based on interannual-focused measures (WA; see also below). A threshold of zero was used in these studies, above which a hemispheric reconstruction was regarded as possessing at least some skill in relation to the calibration period climatology (MBH98, Huybers 2005; WA). Standard practice in climatology uses the red-noise persistence of the target series (here hemispheric temperature) in the calibration period to establish a null-model threshold for reconstruction skill in the independent verification period, which is the methodology used by MBH in a Monte Carlo framework to establish a verification RE threshold of zero at the >99% significance level.

Rather than examining a null model based on hemispheric temperatures, MM05a,c report a Monte Carlo RE threshold analysis that employs random red-noise series modeled on the persistence structure present in the *proxy* data (note, noise here is meant in the sense of the ‘signal’ itself, rather than as an addition to the signal). They argue that random iterations of even a single red-noise series based on the first PC of the North American tree-ring proxy set in MBH can systematically generate a mean offset in the resulting “reconstructed” 20th century calibration period temperatures compared to the immediately preceding verification period. Because the MBH method employs unitless proxy series in a regression against the primary PCs of instrumental surface temperatures, the calibration will not “care” if such random mean differences in the (red-noise) proxy series are positive or negative (MM05a). Rather than averaging to zero, all differences in mean will be treated as an inherent structure of the data and they will be used as if ‘oriented’ in the same way, and thus the verification significance thresholds at the 95 and 99% levels can be expected to be well above zero. It is important to note that such a situation, however, does not occur in the case when the predictors are direct temperature series (with associated sign) that are compared against another temperature series (e.g., hemispheric mean values), and equally not when composites of proxy time series are scaled against a hemispheric (or other) temperature series in “Composite Plus Scale” (CPS) methods. While pointing in the right direction of the need for a higher bar due to the trend in the instrumental (calibration) period, this approach suggested by MM05c contains a series of significant drawbacks, which require the application of caution and correction, as follows.

- To generate “random” noise series, MM05c apply the full autoregressive structure of the real world proxy series. In this way, they in fact train their stochastic engine with significant (if not dominant) low frequency *climate signal* rather than purely non-climatic noise and its persistence. The resulting verification RE significance thresholds for rejecting the likelihood of spurious, random “success” derived from reconstructions using these pseudo-proxies are therefore too large by construction, precisely because they contain at least some climate signal along with purely stochastic processes. Such thresholds thus enhance the danger of committing Type II errors (inappropriate failure to reject a null hypothesis of *no* climatic information for a reconstruction).
- Furthermore, the MM05c proxy-based threshold analysis only evaluates the verification-period RE scores, ignoring the associated calibration-period performance. However, any successful real-world verification should always be based on the presumption that the associated calibration has been meaningful as well (in this context defined as RE >0), *and* that the verification skill is generally not greatly

larger than the skill in calibration. When the verification threshold analysis is modified to include this real-world screening and generalized to include *all* proxies in each of the MBH reconstruction segments – even under the overly-conservative conditions discussed above – previous MBH/WA results can still be regarded as valid, contrary to MM05c. Ten of the eleven MBH98 reconstruction segments emulated in WA are significant above the 95% level (probability of Type I error below 5%) when using a conservative minimum calibration/verification RE ratio of 0.75, i.e. accepting poorer relative calibration performance than the lowest seen in the WA reconstructions (0.82 for the MBH 1400-network). Only the MBH 1600-network is significant at a slightly lower level (89%), and the much discussed 1400- and 1450-networks are significant at the 99% and 96% levels, respectively. (See electronic supplement for further discussion and details, including code and tables with established thresholds for a variety of calibration/verification RE ratios and for the other WA scenarios examined.)

- A final, and important, shortcoming of the MM05c procedure is that RE performance thresholds established using this proxy-based approach have the disadvantage of not being uniformly applicable; rather, they need to be established individually for each proxy network. Individually-established thresholds are not necessary for verification significance evaluations based directly on the characteristics of a target temperature series, such as Northern Hemisphere average surface temperature. For these cases, an RE threshold of zero can still be considered appropriate.

It is interesting to compare the verification significance threshold derived from the Northern Hemisphere temperature series itself with the thresholds derived from the proxy-based analysis implemented here. Overall, the reconstructions that perform with high scores – and thus can be considered validated – in both approaches can be recognized as a clear group (the WA/MBH emulations as discussed above, WA scenario 5a for both the 1400- and 1450-networks, and WA scenarios 5b–d and 6a–c for the 1450-network). Equally clear is the identification of networks that lead to failing reconstructions in both approaches (WA scenarios 5d and 6a–c for the 1400-network). The networks approaching significance limits in both cases (WA scenarios 5b–c for the 1400-network) make up a third group in which one could argue which kind of measure should be given most weight for a decision of significance (cf. electronic supplement, WA).

MM have stressed that RE should only be one of various measures of merit and that other tests, such as Pearson's $r^2$ and the sign test, should be used as well. WA lay out in detail why $r^2$ should be used with caution in the reconstruction framework, not least because the MBH/WA verification period (1854–1901) has a mean significantly different from that of the calibration period (1902–1980), a crucial factor against which $r^2$ applied to the isolated subperiods is insensitive. More fundamentally, there is a simple (but mostly overlooked) geophysical reason for not using $r^2$ with any of the MBH networks prior to approximately 1700. By construction, $r^2$ focuses on a reconstruction's ability to estimate interannual tracking with the instrumental record. Clearly, such a focus is misdirected for reconstruction networks that do not include most of the important and dominant interannual modes of variability as part of the retained instrumental EOFs. For example, the MBH 1400-network exclusively estimates the first PC of instrumental temperatures, largely representing the hemispheric mean (Wahl et al. 2006). This PC, as discussed above, is closely related to the processes that translate the underlying change in energy balance into a surface temperature response. Therefore, if $r^2$ were to be applied to the 1400-network reconstructed temperature field, then one is actually testing if the estimate of energy balance on an interannual basis (based on the relatively data-poor 1400-network) would provide good interannual reconstruction information within the time span of 1854–1901. Over these short 48 years, very little change in energy balance actually happened (with exception of a few volcanic eruptions); rather, the dominant variance information for Northern Hemisphere surface temperature was in terms of interannual variability, such as from ENSO, NAO/AO and other modes. The first EOF pattern based on instrumental data from 1902–1980 will contain little information to estimate much of this interannual variability with recognizable skill. Therefore, $r^2$ is almost certain not to perform well for such a case because the reconstruction does not contain much explicit interannual information in the first place. The *climate* reconstruction, however, should be capable of recognizing that the longer-term mean had changed, as could be expected because the underlying global energy balance was altered between 1854–1901 and 1902–1980. This skill in terms of identifying a change in climatic mean from that of the calibration period is rewarded in verification through use of the RE metric, while the lack of performance in $r^2$ on the hemispheric level can simply be expected and thus is not an adverse result. Putting weight on the use of $r^2$ in this situation would significantly increase Type II Error (false negatives), an issue WA explicitly attempt to address (coupled with joint reduction of Type I Error-false positives) by using RE. (cf. Mann et al. 2007, who find that $r^2$ is particularly prone to both kinds of error in experimental tests.)

## 5 Amplitude issues

Several recent climate reconstructions have determined hemispheric temperature variations in which amplitudes are larger than those indicated by MBH – although the overall shapes of the temperature evolution are similar (e.g., Esper et al. 2005). The offsets could arise from differences in latitudinal extent of the proxy temperature information used, in the land and ocean points included, in the representation of the seasonal cycle, in the different preservation of climate signal in the employed proxy records themselves (cf. Rutherford et al. 2005), or through differences in the calibration procedures (Osborn and Briffa 2006). Von Storch et al. (2004, 2006) as well as Bürger et al. (2006) pointed towards – and our own assessment confirmed (Wahl et al. 2006) – a potential for amplitude loss in the MBH truncated-EOF method. To illustrate this issue, we have applied our verified MBH algorithm (WA) to climate model output.

Because in the MBH truncated-EOF CFR each year is treated independently, one can selectively, or even randomly, reorder years of climate model output for specific analytical purposes with no spurious effects on the resulting model-based reconstructions. This technique allows isolation of the impact of the climatological range present in the years used for calibration on the reconstruction itself (note, however, this approach might not be appropriate for examination of low-frequency spatial characteristics). Figure 1 shows 30-year smoothed output from the NCAR CSM 1.4 (Ammann et al. 2007) in such a temporal rearrangement. In this example, we have sorted the model output into a sequence that exactly mirrors the rank ordering sequence found in the reconstruction of MBH98. The model data were then masked to contain information only on the grid cells present in the calibration instrumental network of MBH. Using the MBH 1820 proxy locations, we sub-sampled the relevant model fields (temperature and precipitation at seasonal resolution) to isolate the 112 individual time series used in this network, and then added a large amount of white-noise (with a conservative, individual signal-to-noise variance ratio of 1-to-4). Based on these pseudo-instrumental and pseudo-proxy time series, we then emulated the MBH

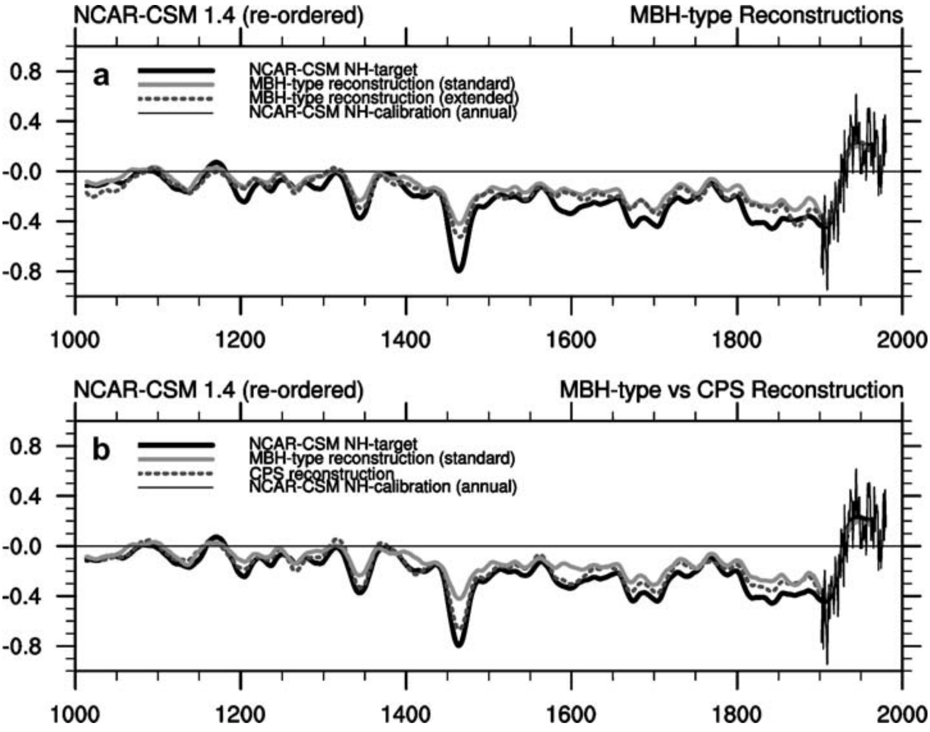

**Fig. 1** Using climate model output as a laboratory to test reconstruction performance. **a** NCAR CSM 1.4 model output (Ammann et al. 2007) of the experiment with traditional ("medium") solar irradiance scaling, reordered to match the rank-structure of MBH99 (*black line*). The model Northern Hemisphere "target" (*black line*, MBH calibration grid cells) is shown with MBH-type reconstructions based on the "standard" 1902–1980 calibration using the full calibration grid (*solid grey line*) or the "extended" verification-grid-only calibration over 1854–1980 (*dashed grey line*). **b** Same as in **a**, but the model Northern Hemisphere mean (*black*) and "standard" case (*grey*) are compared with Composite Plus Scale (CPS) reconstructions based on the MBH 1820-grid of 112 pseudo-proxies using only temperature information at all proxy grid points (*dashed grey line*). The CPS results were obtained using a 1902–1980 calibration. All series are smoothed with a 30-year Gaussian weighted running filter, the calibration period annual average Northern Hemisphere temperature from the model is shown at annual resolution. (A color version of this figure is available with the electronic supplement)

reconstruction. Figure 1a compares the model Northern Hemisphere temperature series (black line) as the target series with a reconstruction based on the "standard" 1902–1980 calibration period using the full calibration grid (solid grey line), and an "extended" reconstruction calibrated on the entire 1854–1980 "instrumental" time series restricted to the spatially-limited MBH verification grid (dashed grey line). Figure 1b shows the model hemispheric temperatures (black) with the 1902–1980 calibration results (grey, same as in Fig. 1a), and compares them with a version of the 1820 grid information used in a CPS method (calibrated over 1902–1980), where the model's annual average temperature was sampled at all 112 proxy locations (grey dashed).

The results of the MBH-type reconstructions based on the reordered model output shown in Fig. 1a indicate excellent overall reconstruction skill. The calibration RE-scores are high (0.85) in both cases, and the verification period RE-score is 0.81 for the "standard" case (no MBH-equivalent verification is possible in the "extended" case), all highly significant. Because of the use of the detailed 1820 proxy network combined with maximum retention of "instrumental" patterns of variability in the MBH-framework, $r^2$-values are also high (0.86 and 0.51 in the "standard" case for calibration and verification respectively, and 0.85 for calibration in the "extended" case), despite large noise in the pseudo-proxies, confirming the good visual interannual tracking. There is, however, an indication for some amplitude loss. Qualitatively, the "standard" reconstruction agrees with von Storch et al. (2006) (version *without* trend removal). The average difference over 1000–1980 is 0.072°C in comparison to the model output masked to the calibration grid, and less than half that amount for the "extended" case (0.031°C in comparison to the model output masked to the corresponding verification grid).

Although these results are generally consistent with observations of some systematic amplitude loss by von Storch et al. (2004, 2006) and Bürger and Cubasch (2005) (cf. Wahl et al. 2006), Fig. 1a shows that the most obvious continuous deviations are concentrated in a few periods (particularly around the 1450–70, 1580–1610, 1670–1700, and first-half of the 19th century cooling episodes). As found in Mann et al. (2005b), these periods are associated with significant volcanic perturbations. There are actually also a few limited periods with amplitude gain in the "extended" case. Overall, there is clear loss of amplitude, but the extent is comparatively small in absolute terms. In relative terms the loss averages to ~35% in the "standard" case and ~25% in the "extended" case, which is between the 'perfect pseudoproxies' and 'white noise' exercises in the updated model-based tests reported by Zorita and von Storch (2005) (version *without* trend removal), despite a much larger signal-to-noise ratio of 1:4 used here compared to about 1:2 in Zorita and von Storch (2005).

An important source for this amplitude uncertainty is likely connected to the range of climate available for calibration (Bürger et al. 2006). With respect to the range of Northern Hemisphere temperatures, the "standard" calibration period covers the warm end of the distribution very well (the late 20th century is far warmer than any other decade), but does not fully cover the cool part. In the "extended" (though spatially more limited) experiment, the overall representation of the cool years is improved (simply through more samples) without degradation at the warm end. This example illustrates the importance of the range of the climate information present in calibration in the MBH truncated-EOF reconstruction algorithm. Interestingly, the available instrumental *and* proxy data used in MBH span the range quite well, with the late 19th–early 20th century being close to the coldest continuous period (MBH99), while no period can rival the late 20th century on the warm end (cf. Juckes et al. 2006). Therefore, while some amplitude loss is still possible, the problem appears to remain limited in our model-based tests and does not affect the overall interpretation of the original MBH reconstruction. The problem, however, will be more accentuated in situations when pre-calibration climates deviate significantly from the calibration range (von Storch et al. 2004, 2006; cf. Juckes et al. 2006 and Wahl et al. 2006).

The results for the CPS method shown in Fig. 1b also indicate some, but even smaller, average amplitude loss (0.041° compared to the calibration-grid Northern Hemisphere average) in relation to the MBH CFR. Thus, for the specific goal of reconstructing the Northern Hemisphere mean temperature evolution (in particular its coldest excursions), our results suggest that the CPS method based on temperature information might be favored over the MBH CFR (cf. Mann et al. 2005b, where similar results are reported) until the exact source of the amplitude issue in MBH (and possibly other variants of truncated-EOF techniques) can be isolated. Alternatively, within the MBH CFR-framework, our results suggest that calibration with an extended, lower-resolution grid (equivalent to the verification network resolution) leads to average results roughly equivalent to the CPS outcome, though it still captures less amplitude in the strongest cooling episodes.

It is interesting to note that model-based reconstructions of Northern Hemisphere average temperatures (similarly emulating the MBH proxy data) done with the newer regularized expectation maximization (RegEM) field reconstruction method (Schneider 2001) show no tendency for low frequency amplitude loss across the range of noise levels (both white and red), calibration-period lengths, and proxy richness values examined (Mann et al. 2007). In this light, the fact that application of RegEM to the MBH real-world proxy data set results in a reconstruction very much like the original (Mann et al. 2007) suggests that the *potential* for amplitude loss in the truncated-EOF method noted here had little effect on the MBH outcome in actuality. Simulated reconstructions of the regional-scale NINO3 index done with RegEM, however, do exhibit systematic amplitude loss when calibrated over 1900–1980, which is ameliorated by lengthening the calibration period to 1856–1980 (Mann et al. 2007), similar to the results presented here for Northern Hemisphere temperatures in the MBH truncated-EOF method.

## 6 Conclusions

The MBH reconstruction is a robust climate field reconstruction (CFR) that separates itself from non-field oriented reconstructions through its attempt to capture additional spatial information beyond individual contributions of the proxies themselves (i.e., the joint information is larger than the sum of the individual pieces). It achieves this goal through use of the spatially coherent signal related to a few primary patterns (EOFs) of climate variability, and then estimates the annual weights of these patterns that jointly make up the past temperature fields. In this operation, the climate indicators are not restricted to direct temperature proxies; rather, *any* climatically relevant combination (temperature, precipitation, etc.) that is related to one or more of the primary patterns of variability can contribute to the reconstruction. This is the case because large-scale temperature patterns are often dynamically linked to other climatic fields (strength of pressure fields, shifts in storm tracks, etc.). The interesting aspect of such an approach is that it offers spatially resolved climate information at the level of individual grid cells, rather than an end result of exclusively large-scale means. As such, the reconstructions can serve as a valuable, albeit somewhat crude, extension of the instrumental record, which can be very helpful in geophysical analyses and interpretations of past climate. This advantage has proven particularly interesting for the study of variations in some of the climate modes with global significance, such as ENSO (Mann et al. 2000) or the NAO (Cook et al. 2002; Luterbacher et al. 2002). The MBH reconstruction has thus stimulated research into understanding geophysical processes that translate global or hemispheric radiative forcing into a regional response on various time scales (Adams et al. 2003; Mann et al. 2005a; Graham et al. 2007; Fischer et al. 2007; Küttel et al. 2007).

There are a number of issues with the MBH (and other) CFR methods that need further attention. There is much we still have to learn about: the robustness of some of the regional results (particularly early in the record); the stationarity of the primary patterns used in the calibration period and the sensitivity of the reconstructions (particularly the spatial gradients) to whether or not such stationarity exists; the general issue of amplitude fidelity (at least in the MBH truncated-EOF method); and a more comprehensive way of evaluating uncertainty due to procedural choices (Bürger et al. 2006) or unresolved variance in the data (Li et al. in press). [Note that in both Bürger et al. (2006) and Li et al. (in press) no independent verification-screening step of reconstruction performance is utilized.] Compared to the heated debate concerning some of the details in data processing, relatively less attention in large-scale field reconstructions has been given to the proxies themselves (but cf. MM05b). Clearly, a climate reconstruction can only be as good as its input, and particularly retention of more slowly-developing climatic change signals might be more difficult to achieve in practice (Briffa et al. 2001; Esper et al. 2005; Moberg et al. 2005). However, because of their exploitation of spatially coherent signals, one can expect the CFR methods to be less sensitive to erroneous individual series on both interannual as well as longer time-scales (i.e., red noise would not be expected to be coherent in space and time). There is need to further explore the potential of CFRs to combine information from proxies with different time resolutions, where one would hope that some of the lower-resolution records might more reliably preserve the low-frequency climate signal (Moberg et al. 2005). First attempts in this direction have been done with the RegEM technique (Rutherford et al. 2005; Mann et al. 2007). Successful integration of lower- and higher-frequency proxy information will also help resolve some discomfort that exists with the MBH method's sensitivity to the calibration information and its related potential for amplitude loss. As indicated here, we believe this problem to be small within the original proxy framework, and our examination may actually overstate its extent (as the reported RegEM results suggest); nevertheless, it needs a more complete resolution.

Since the algorithm itself cannot explain most of the amplitude difference between MBH CFR-temperatures and other reconstructions, it is important to ask what could be affecting the hemispheric mean across the reconstructions. Geographical coverage has been identified before as influencing the range of climate over the past millennium (Rutherford et al. 2005). A directly associated factor concerns latitudinal extent. In this regard, one issue has recently been recognized as particularly challenging for our understanding from a more dynamical viewpoint: numerous records from within and around the Pacific Basin suggest that during the time span often referred to as "Medieval Warm Period" the tropical Eastern Pacific was probably generally cooler than present. Subsequently, during the hemispheric cool periods of the "Little Ice Age," the Eastern Equatorial Pacific seems to have warmed (Cobb et al. 2003; Graham et al. 2007). One could explain these trends through a shift in mean state of the zonal gradients at the equator on top of which El Niño variations occur – although there are a few records that don't fit well into this possibly over-simplified picture (see discussion in Graham et al. 2007). How climate system dynamics can explain this structure is currently being debated (cf. Mann et al. 2005a; Graham et al. 2007) and has to be left open at this time. But it is clear that a cool Eastern Equatorial Pacific would affect the climate system through a dampening of a generally warm Northern Hemisphere during medieval times, and equally, a warm eastern Equatorial Pacific would dampen the hemispheric cool signal from the "Little Ice Age". CPS reconstructions that are mostly based on mid- to high-latitude land data cannot be expected to pick up this spatial variation. Borehole temperatures, which generally point towards the higher end of the proposed range of past temperature amplitudes, are equally restricted to continental areas. Fundamentally, this kind of pattern variability is what CFR methods can help resolve.

Finally, where do these considerations leave us with regard to the MBH climate reconstruction? After correction for a very small methodological bias, and despite some amplitude issues, which seem to be reasonably in check – at least given the available instrumental and proxy data in the original publications – there currently is no objective ground on which one would have to dismiss the reconstruction. The RegEM/MBH comparison results reported by Mann et al. (2007) strongly support this conclusion. Figure 2 shows our emulation of the full MBH99 reconstruction from 1000 to 1980, which contains corrections for the centering convention in North American proxy PCs as well as elimination of the Gaspé-series during 1400–1449. The overall correction is minimal and

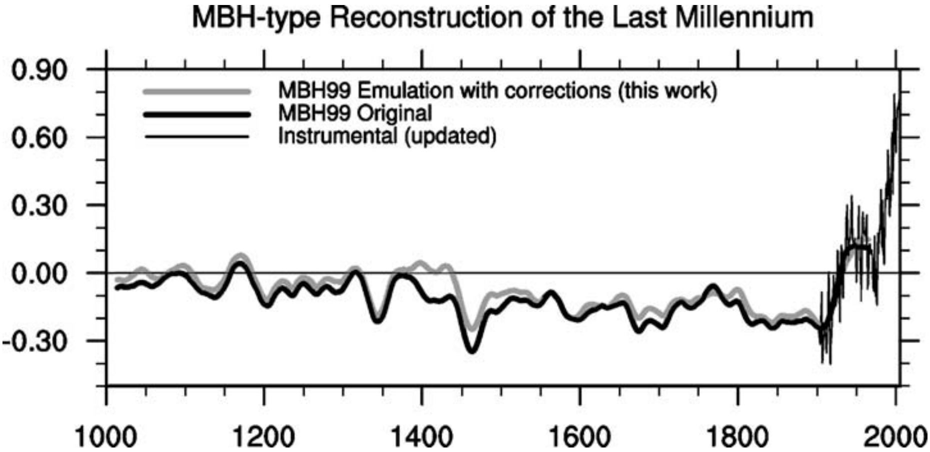

**Fig. 2** Correction of MBH99: our emulation of the real world proxy-based MBH99 reconstruction containing full-period proxy PC-centering corrections and omission of the Gaspé-series during 1400–1449 (*solid grey line*) is compared to the original MBH99 reconstruction (*black line*)

averages to roughly 0.04°C over the millennium, but is somewhat larger during 1400–1449 because of the Gaspé-series removal during this time interval. The new reconstruction of the 1000–1399 segment passes validation (calibration RE of 0.39 and verification RE of 0.22, the latter is significant at the 94% level using the very conservative proxy-based significance analysis described above).[1]

We initially asked the question if climate science could *understand* the general causes for the observed or reconstructed climate, despite the multi-facetted uncertainties (as well as noise from internal climate variability). It has previously been found that changes to the Earth's energy balance are predominantly responsible for variations in the *mean climate* at the global and hemispheric scale (Tett et al. 1999; Crowley 2000; Goosse et al. 2005; Hegerl et al. 2007; Ammann et al. 2007). This issue has been summarized most convincingly in IPCC 2007 (Jansen et al. 2007, p. 466 ff). The MBH (and Mann et al. 2007) reconstruction results are largely consistent with other climate reconstructions, particularly so when issues of proxy geographical coverage are kept in mind and when the reconstructions are evaluated using relative variations that mask some of the amplitude differences (Esper et al. 2005; Büntgen et al. 2005; and see also Osborn and Briffa 2006). If not crippled by removal of the geophysically and statistically relevant energy balance signal embedded in the 20th century trend, the MBH CFR approach can make use of a diverse set of climate proxy time series (again, note that strict local temperature representation is not required!) to identify the large-scale structure of climate variability.

---

[1] This significance value is a very conservative estimate because it was determined using the reconstruction of the full-length-centered, yet not $CO_2$-corrected, PCs for the 28 North American tree rings (see MBH99). Any $CO_2$-correction would reduce the redness of the proxies, which would lower the RE significance threshold (and thus increase the significance of the result). Due to the change in how the "hockey-stick"-like signal in the underlying data is separated into two PCs by full-length centering (rather than appearing in the first PC, similar to the discussion in WA), a $CO_2$-correction that would be directly comparable to MBH99 is not possible. Therefore, the effect of $CO_2$-correction was estimated using the original MBH99 proxies (with and without $CO_2$-correction) and this impact was added to the reconstruction based on full-period centering where the PCs have not been corrected for $CO_2$ changes. The corresponding significance level for the 1000–1399-segment using the original MBH PCs is 97%.

Nevertheless, MBH's primary benefits are not in the area for which it has most commonly been used, the Northern Hemisphere mean temperature, but much more in its information content on regional scales. It is this information which should be exploited more thoroughly. Together with the RegEM technique (Schneider 2001; Rutherford et al. 2003, 2005; Mann et al. 2005b, 2007), truncated-EOF climate field reconstructions (Kaplan et al. 1997; MBH98/99; Evans et al. 2002; Luterbacher et al. 2004; Xoplaki et al. 2005; Casty et al. 2005b; Pauling et al. 2006) will only gain in importance in coming years as the focus of climate change research moves from questions concerning whether the current warming is human-made or not to how the warming will express itself in space. For this, we need a more complete picture of how regional climate patterns have responded to radiative forcing in the past (Fischer et al. 2007). Reconstructed large-scale climate fields help to verify such cause-and-effect relationships (with their associated mechanisms) over different time scales by providing the necessary extension to the short instrumental record. In this process, it will be important to further refine the reconstructions and to resolve seasonal rather than only annual variations (Mann et al. 2000; Luterbacher et al. 2004; Xoplaki et al. 2005), and as the data density increases one can even approach expansion to non-temperature fields (Casty et al. 2005a,b; Pauling et al. 2006). Thus, CFRs provide a crucial ingredient for our overall geophysical understanding of the Earth's climate system. The more we can explain past climate variations, rather than simply describing them, the better and stronger becomes our geophysical framework within which we can evaluate what the causes of past climate impacts might have been (Xoplaki et al. 2005) and what future climate changes might look like, particularly on the most relevant, regional scale. The Mann-Bradley-Hughes reconstruction has opened the way for such analyses, and we need to further pursue the geophysical promises embedded in CFRs while objectively and constructively evaluating uncertainties and potential biases.

**Acknowledgements** The authors are grateful for careful and detailed guidance by three anonymous reviewers that have significantly helped to improve and clarify the manuscript. Linda Mearns and the Weather and Climate Impacts Assessment Science Program at the National Center for Atmospheric Research supported this work. Technical and methodological advice from Claudia Tebaldi and Doug Nychka were invaluable, and helpful discussions with Hee-Seok Oh are acknowledged. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

## References

Adams JB, Mann ME, Ammann CM (2003) Proxy evidence for an El Niño-like response to volcanic forcing. Nature 426:274–278

Ammann CM, Joos F, Schimel D, Otto-Bliesner BL, Tomas R (2007) Solar influence on climate during the past millennium: results from transient simulations with the NCAR Climate System Model. Proc Nat Acad Sci 104(10):3713–3718

Briffa KR, Osborn TJ, Schweingruber FH, Harris IC, Jones PD, Shiyatov SG, Vaganov EA (2001) Low-frequency temperature variations from a northern tree ring density network. J Geophys Res 106 (D3):2929–2941

Büntgen U, Esper J, Frank DC, Nicolussi K, Schmidhalter M (2005) A 1052-year tree-ring proxy for Alpine summer temperatures. Clim Dyn 25:141–153

Bürger G, Cubasch U (2005) Are multiproxy climate reconstructions robust? Geophys Res Lett 32:L23711, doi:10.1029/2005GL024155

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

Casty C, Wanner H, Luterbacher J, Esper J, Boehm R (2005a) Temperature and precipitation variability in the European Alps since 1500. Int J Climatol 25:1855–1880

Casty C, Handorf D, Sempf M (2005b) Combined winter climate regimes over the North Atlantic/European sector 1766–2000. Geophys Res Lett 32:L12801

Cobb KM, Charles CD, Cheng H, Edwards RL (2003) El Niño/Southern Oscillation and tropical Pacific climate during the last millennium. Nature 424:271–276

Cook ER, D'Arrigo RD, Mann ME (2002) A well-verified, multiproxy reconstruction of the Winter North Atlantic Oscillation Index since A.D. 1400. J Climate 15:1754–1764

Corti S, Molteni F, Palmer TN (1999) Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature 398:799–802

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

Esper J, Wilson RJS, Frank DC, Moberg A, Wanner H, Luterbacher J (2005) Climate: past ranges and future changes. Quat Science Rev 24:2164–2166

Evans MN, Kaplan A, Cane MA (2002) Pacific sea surface temperature field reconstruction from coral d$^{18}$O data using reduced space objective analysis. Paleoceanography 17(7):1–7,13

Fischer EM, Luterbacher J, Zorita E, Tett SFB, Casty C, Wanner H (2007) European climate response to tropical volcanic eruptions over the last half millennium. Geophys Res Lett 34:L05707, doi:10.1029/2006GL027992

Folland CK, Karl TR, Christy JR, Clarke RA, Gruza GV, Jouzel J, Mann ME, Oerlemans J, Salinger MJ, Wang S-W (2001) Observed climate variability and change. In Houghton JT et al. (eds.) Climate Change 2001: The scientific basis. Cambridge Univ. Press, Cambridge, pp 99–181

Goosse H, Rensen H, Timmermann A, Bradley RS (2005) Internal and forced climate variability during the last millennium: a model-data comparison using ensemble simulations. Quat Sci Rev 24:1345–1360

Graham NE, Hughes MK, Ammann CM, Cobb KM, Hoerling MP, Kennett DJ, Kennett JP, Rein B, Stott L, Wigand PE, Xu T (2007) Tropical Pacific – mid-latitude teleconnections in medieval times. Climatic Change doi:10.1007/s10584-007-9239-2 (in press)

Hansen et al. (2005) Earth's energy imbalance: confirmation and implications. Science 308:1431–1435

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

Hegerl GC, Crowley TJ, Baum SK, Kim K-Y, Hyde WT (2003) Detection of volcanic, solar and greenhouse gas signals in paleo-reconstructions of Nothern Hemispheric temperature. Geophys Res Lett 30(5):1242 doi:10.1029/2002GL016271

Hurrell JW, Hoerling MP, Phillips AS, Xu T (2004) Twentieth century North Atlantic climate change. Part I: assessing determinism. Clim Dyn 23:371–389

Huybers P (2005) Comment on "Hockey sticks, principal components, and spurious significance" by McIntyre and McKitrick. Geophys Res Lett 32:L20705, doi:10.1029/2005GL023395

Jansen E, Overpeck J, Briffa KR, Duplessy J-C, Joos F, Masson-Delmotte V, Olago D, Otto-Bliesner B, Peltier WR, Rahmstorf S, Ramesh R, Raynaud D, Rind D, Solomina O, Villalba R, Zhang D (2007) Paleoclimate. In: Solomon S., Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Juckes MN, Allen MR, Briffa KR, Esper J, Hegerl GC, Moberg A, Osborn TJ, Weber SL, Zorita E (2006) Millennial temperature reconstruction intercomparison and evaluation. Clim Past Discuss 2:1001–1049

Kaplan A, Kushnir Y, Cane MA, Blumenthal MB (1997) Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. J Geophys Res 102(C13):27,835–27,860

Küttel M, Luterbacher J, Zorita E, Xoplaki E, Riedwyl N, Wanner H (2007) Testing a European winter surface temperature reconstruction in a surrogate climate. Geophys Res Lett 45:L07710

Li B, Nychka DW, Ammann CM (2007) The "Hockey Stick" and the 1990s: a statistical perspective on reconstructing hemispheric temperatures. Tellus A, TeA-07-03-0034 (in press)

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H (2004) European seasonal and annual temperature variability, trends and extremes since1500. Science 303:1499–1503

Luterbacher J, Xoplaki E, Dietrich D, Jones PD, Davies TD, Portis D, Gonzalez-Rouco JF, von Storch H, Gyalistras D, Casty C, Wanner H (2002) Extending North Atlantic Oscillation Reconstructions back to 1500. Atmos Sci Lett 2:114–124

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

Mann ME, Bradley RS, Hughes MK (1999) Northern Hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. Geophys Res Lett 26:759–762

Mann ME, Bradley RS, Hughes MK (2000) Long-term variability in the El Niño Southern Oscillation and associated teleconnections. In Diaz HF, Markgraf V (eds) El Niño and the Southern Oscillation:

Multiscale Variability and its Impacts on Natural Ecosystems and Society, Cambridge University Press, Cambridge, UK, pp 357–412

Mann ME, Cane MA, Zebiak SE, Clement A (2005a) Volcanic and solar forcing of the tropical Pacific over the past 1000 years. J Climate 18:447–456

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

Mann ME, Rutherford S, Wahl ER, Ammann CM (2007) Robustness of proxy-based climate field reconstruction methods. J Geophys Res (in press)

McIntyre S, McKitrick R (2003) Corrections to the Mann et al (1998) proxy data based and Northern Hemispheric average temperature series. Energy and Environment 14:751–771

McIntyre S, McKitrick R (2005a) Hockey sticks, principal components, and spurious significance. Geophys Res Lett 32:L03710, doi:10.1029/2004GL021750

McIntyre S, McKitrick R (2005b) The M&M critique of the MBH98 Northern Hemisphere climate index: update and implications. Energy and Environment 16:69–100

McIntyre S, McKitrick R (2005c) Reply to comment by Huybers. Geophys Res Lett 32:L20713, doi:10.1029/2005GL023586

Mears CA, Wentz FJ (2005) The effect of diurnal correction on satellite-derived lower tropospheric temperature. Science 309:5740, doi:10.1126/science.1114772

Meehl GA, Washington WM, Ammann CM, Arblaster JM, Wigley TML, Tebaldi C (2004) Combinations of natural and anthropogenic forcings in twentieth-century climate. J Climate 17:3721–3727

Moberg A, Sonechkin D, Holmgren K, Datsenko N, Karlén W (2005) Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. Nature 433:613–617

National Research Council Report (2006) Surface temperature reconstructions for the last 2,000 years. Committee on surface temperature reconstructions for the last 2,000 years, Board on Atmospheric Sciences and Climate, Division of Earth and Life Studies. The National Academies Press

Osborn TJ, Briffa KR (2006) The spatial extent of 20th-century warmth in the context of the past 1200 years. Science 311:841–844

Palmer TN (1999) A nonlinear perspective on climate change. J Clim 12:575–591

Pauling A, Luterbacher J, Casty C, Wanner H (2006) 500 years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. Clim Dynam 26:387–405

Rutherford S, Mann ME, Delworth TL, Stouffer RJ (2003) Climate field reconstruction under stationary and nonstationary forcing. J Climate 16:462–479

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

Santer BD, Wehner MF, Wigley TML, Sausen R, Meehl GA, Taylor KE, Ammann CM, Arblaster J, Washington WM, Boyle JS, Brüggemann W (2003) Contributions of anthropogenic and natural forcing to recent tropopause height changes. Science 301:479–483

Schneider T (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. J Climate 14:853–887

Stephenson DB, Hannachi A, O'Neill A (2004) On the existence of multiple climate regimes. QJR Meteorol Soc 130:583–606

Stott PA, Jones GS, Lowe JA, Thorne P, Durman C, Johns TC, Thelen J-C (2006) Transient climate simulations with the HadGEM1 climate model: causes of past warming and future climate change. J Climate 19(12):2763–2782

Tett SFB, Stott PA, Allen MR, Ingram WJ, Mitchell JFB (1999) Causes of twentieth-century temperature change near the Earth's surface. Nature 399:569–572

Von Storch H, Zorita E, Jones JM, Gonzalez-Rouco F, Tett SFB (2006) Response to Comment on "Reconstructing past climate from noisy data. Science 312:529

Von Storch H, Zorita E, Jones JM, Dimitriev Y, Gonzalez-Rouco F, Tett SFB (2004) Reconstructing past climate from noisy data. Science 306:679–682

Wahl ER, Ritson D, Ammann CM (2006) Comment on 'Reconstructing Past Climate from Noisy Data. Science 312, doi:10.1126/science.1120866, 2006

Wahl ER, Ammann CM (this volume) Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence. Climatic Change, (in press)

Wegman EJ, Scott DW, Said YH (2006) Ad hoc committee report on the "Hockey Stick" global climate reconstruction. Report delivered to the Chairman of the United States House of Representatives Committee on Energy and Commerce, as well as the Chairman of the Subcommittee on Oversight and

Investigations , July 14, 2006. (non peer-reviewed). Available: http://energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf

Xoplaki E, Luterbacher J, Paeth H, Dietrich D, Steiner N, Grosjean M, Wanner H (2005) European spring and autumn temperature variability and change of extremes over the last half millennium. Geophys Res Lett 32:L15713

Zorita E, von Storch H (2005) Methodical aspects of reconstructing non–local historical temperatures. Mem SA It 76:794. http://sait.oat.ts.astro.it/MSAIt760405/PDF/2005MmSAI..76..794Z.pdf