Understanding Sensitivities in Paleoclimatic Reconstructions

Richard L. Smith*

August 23, 2010

Abstract

Recent publications have seen the introduction of a number of new statistical methods for paleoclimatic reconstructions. When applied to archived multiproxy datasets, most of these methods produce reconstructed curves very similar to the "hockey stick" shape that was first observed by Mann, Bradley and Hughes. However, one recent reconstruction, by McShane and Wyner, produced a sharply different shape. Trying to understand the reasons for this leads to important insights for both statistical methodology and paleoclimatic datasets. The "divergence" phenomenon — that the relationship between temperature and some of the proxies may not be constant over time — has been extensively discussed in the paleoclimate literature, but mostly in the context of certain classes of northern hemisphere tree rings, which are not included among the proxies examined here. Closer scrutiny of the data suggests a new divergence phenomenon, associated with lake sediments. When these are removed from the data, the resulting reconstruction is much closer to the familiar hockey stick shape. This highlights the need both for careful scrutiny of the data, and for statistical methods that are robust against the divergence phenomenon.

*Richard L. Smith is the Director of the Statistical and Applied Mathematical Sciences Institute (SAMSI), Research Triangle Park, N.C., and Mark L. Reed III Distinguished Professor in the Department of Statistics and Operations Research, University of North Carolina, Chapel Hill. SAMSI is supported by the National Science Foundation, grant DMS 0635449. Contact email address: rls@email.unc.edu. The author would like to thank Dr. Michael Mann for answering a number of queries about the data and for some very helpful discussion of the divergence phenomenon.

1 Introduction

The problem of paleoclimatic reconstruction and the "hockey stick curve" — the assertion that global temperatures remained nearly constant for 900 years, and below those of the present day, before the well-documented rise of the 20th century — continues to fascinate scientists, statisticians, and the general public. Although arguably tangential to the central problem of climate change science, which I would characterize as the projection of future climate changes and their impacts under various scenarios of human activity, the evidence for a hockey stick curve has been widely interpreted as reinforcing the case for an anthropogenic influence on climate.

The hockey stick curve originated in two papers of Mann, Bradley and Hughes (1998, 1999). Their statistical methods were criticized by a number of authors, most notably McIntyre and McKitrick (2005) and Wegman, Scott and Said (2006). A report by the National Research Council (North *et al.*, 2006) reinforced some of the criticism but argued that the evidence for a hockey stick curve was compelling because of the wide variety of scientific studies, using different data and statistical methods, that had supported this conclusion. Papers by Ammann and Wahl (2007), Wahl and Ammann (2007) countered the objections of McIntyre and Wegman, in particular, arguing that the focus of the latter authors on a single leading principal component of the proxy data record was inappropriate. Meanwhile Li, Nychka and Ammann (2007) took a more comprehensive statistical approach to the problem, using generalized least squares time series regression combined with cross-validation and bootstrap resampling to generate "ensembles" from the full 1000 years of reconstruction; in this way, they were able to provide statistically rigorous answers to questions related to the probability that recent years or decades were the warmest of the millenium. A number of other papers, including Mann *et al.* (2008), have continued the discussion by introducing expanded datasets and statistical analyses.

Against this background, there have been a number of recent papers proposing new statistical techniques or analyses. Bayesian hierarchical models (BHMs) have been proposed by Li, Nychka and Ammann (2010), Tingley and Huybers (2010a, 2010b) and Brynjarsdóttir and Berliner (2010). Smith (2010), in discussion of Li, Nychka and Ammann (2010), argued that many questions of interest in this field of research could be obtained by more elementary statistical methods, in particular, a combination of principal component expansions and time series regression. As illustration, Smith showed how the hockey stick curve for a smoothed temperature signal could be reconstructed, with pointwise prediction intervals, based on a well studied dataset of North American tree rings. However, McShane and Wyner (2010) used similar though not identical methods, applied to a dataset of Mann *et al.* (2008), to obtain quite different conclusions. Their result indicated that the Medieval Warm Period (MWP), covering roughly the period 1000-1400, was substantially warmer than previous reconstructions have shown, and argued that this cast doubt on the hockey stick curve. Despite this negative conclusion, they also reported that by their Bayesian analysis, there was a 0.8 probability that the decade 1997–2006 was the warmest decade in 1000 years, and a 0.36 probability that 1998 the warmest year. In this respect, and taking into account that exact probabilities are almost impossible to estimate precisely in this sort of context, their conclusions are not substantially different from those of other authors who have examined similar issues.

In addition to that analysis, McShane and Wyner included a simulation, in which the true proxy records were replaced by "pseudoproxies" — random series uncorrelated with the true data — and argued that under certain circumstances, regressions could be obtained using the pseudoproxies that fitted the real observational data better than those based on the true proxies. They argued that this showed the proxies contained little useful information.

The present paper is largely motivated by the contrast between the two last papers — that of Smith (2010), that used principal component regression with time series errors to obtain results very similar to those that earlier authors had obtained using sometimes much more complicated methodology, and McShane and Wyner (2010), with a slightly different formulation of principal components regression with autocorrelated errors, who obtained completely different conclusions. Using the same methods as those of Smith (2010), applied to the data of McShane and Wyner (2010), I produce in one analysis a reconstructed curve that shows an even warmer MWP than theirs. However, I also show that this conclusion is sensitive to a number of features of the analysis, including the choice of starting and ending dates for the observational data fitting part of the analysis, and the inclusion or omission of individual principal components. With minor variations in the methodology, I produce results that are much more consistent with the hockey stick curves of other authors. Further examination of the data shows that a specific class of proxies — those associated with lake sediments — are largely responsible for these discrepancies. If the analyses are repeated without those proxies, the results are much more stable, though there remains evidence of instabilities which need to be studied more closely.

The remainder of the paper is organized as follows. In Section 2, I expand on the methods used and consider different approaches to the incorporation of autocorrelation into a PC regression. After discussing pseudo-proxies in Section 3, I consider the reconstruction of 1000 years of data, using the same datasets as McShane and Wyner (2010), in Section 4. Section 5 contains further discussion and conclusions.

Throughout the paper, I make use of the following data sources:

1. Michael Mann's archive of proxy data, available at

   http://www.meteo.psu.edu/ mann/supplements/MultiproxyMeans07

2. the observational temperature data from the Climate Research Unit of the University of East Anglia (HADCRUT3v series, mean annual NH anomalies) available at

   http://www.cru.uea.ac.uk/cru/data/temperature

Datasets and programs written in the course of the present analysis are archived on the author's webpage
http://www.stat.unc.edu/faculty/rs/reconstructions/

2 Methods

The statistical methods are based on those of Smith (2010). Suppose we have observed temperatures y_t for a given series of years indexed by t, and proxy series $\{x_{jt},\ j = 1, ..., 93\}$ for $t = 898, ..., 1998$. We perform a correlation-based (or "standardized") principal components analysis on the proxies to create PCs $\{u_{kt},\ k = 1, ..., 93\}$. We then fit the regression

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k u_{kt} + z_t \tag{1}$$

to the first K PCs, where K is some number to be determined, to the time period of observations.

With estimated regression coefficients $\hat{\beta}_j,\ j = 0, 1, ..., K$, we then compute for the historical reconstruction period the predicted temperature series

$$\hat{y}_t = \hat{\beta}_0 + \sum_{k=1}^{K} \hat{\beta}_k u_{kt}, \tag{2}$$

and also a "smoothed" and "differentiated" version, $\tilde{y}_t = \sum_{i=-12}^{12} w_i \hat{y}_{t-i}$ and $\tilde{y}'_t = \sum_{i=-49}^{49} w'_i \hat{y}_{t-i}$, where the weights w_i and w'_i are given by

$$w_i = (13-|i|)/169, \; i=-12,-11,...,12, \tag{3}$$

$$w'_i = \begin{cases} \frac{-50-i}{31250}, & i=-49,...,-25, \\ \frac{i}{31250}, & i=-24,...,24, \\ \frac{50-i}{31250}, & i=25,...,49, \end{cases} \tag{4}$$

which satisfy $\sum_i w_i = 0,\; \sum_i w'_i = 0,\; \sum_i i w'_i = 1$.

The smoothed version is simply meant to show a smooth trend which is easier to interpret than individual-year results. The differentiated version is meant as an approximation to the slope of the trend curve, though note that it is smoothed over the very long time period of 100 years (shorter smoothing periods were tried, but the results were so noisy as to be uninterpretable).

One significant issue with this analysis is that the errors (z_t in (1)) are easily shown to be autocorrelated, so the standard method of computing standard errors will not be valid. Here, I outline three methods of dealing with this issue, and discuss their relative merits.

1. Li *et al.* (2007) assumed the errors z_t form an autoregressive, moving average (ARMA) process (Brockwell and Davis 2003). They selected an AR(2) process after checking that the residuals from such an analysis were uncorrelated. Writing this in backward-time form (because the objective of paleoclimatology is to predict backwards), their model becomes

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k u_{kt} + z_t, \quad z_t = \phi_1 z_{t+1} + \phi_2 z_{t+2} + \epsilon_t, \tag{5}$$

   with ϵ_t uncorrelated (we assume they are independent $N(0,\sigma^2)$ for some σ^2). They then estimated the regression parameters by Generalized Least Squares (GLS).

2. McShane and Wyner (2010, page 31) assumed a regression equation of the form

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k u_{kt} + \beta_{K+1} y_{t+1} + \beta_{K+2} y_{t+2} + \epsilon_t, \; \epsilon_t \sim N[0,\sigma^2] \text{ (independent)}. \tag{6}$$

   Although their ultimate method was Bayesian, the simplest way to fit model (6) would be ordinary least squares (OLS) regression. The time series dependence is here incorporated through the terms y_{t+1} and y_{t+2} (again backward in time compared with traditional time series analysis, because of the intention to do backward prediction).

3. Smith (2010) proposed the simpler solution of fitting the model $y_t = \beta_0 + \sum_{k=1}^{K} \beta_k u_{kt} + z_t$ by OLS regression, but then correcting the standard errors for autocorrelation. This is assuming the same model as (5), but with OLS instead of GLS regression.

The first point to make is that (6) is *not* equivalent to (5). To see why, rewrite (5) in the form

$$
\begin{aligned}
\epsilon_t &= y_t - \beta_0 - \sum_{k=1}^{K} \beta_k u_{kt} - \phi_1\{y_{t+1} - \beta_0 - \sum_{k=1}^{K} \beta_k u_{k,t+1}\} - \phi_2\{y_{t+2} - \beta_0 - \sum_{k=1}^{K} \beta_k u_{k,t+2}\} \\
&= y_t - (1 + \phi_1 + \phi_2)\beta_0 - \sum_{k=1}^{K} \beta_k (u_{kt} - \phi_1 u_{k,t+1} - \phi_2 u_{k,t+1}).
\end{aligned}
$$

This is equivalent to (6) only if u_{kt} in (6) is replaced by u^*_{kt} say, where $u^*_{kt} = u_{kt} - \phi_1 u_{k,t+1} - \phi_2 u_{k,t+1}$. Conversely, if u^*_{kt} is the value of the kth PC in year t, as assumed by McShane and Wyner (2010), we must solve the last equation iteratively to derive u_{kt}, in order to rewrite the McShane and Wyner (2010) model in the form (5).

There is no clear-cut rationale for preferring one model to the other, but it seems to me that (5) is the more logical model in the present context. If there were indeed a close relationship between the temperature in year t and the proxies in year t, then we would want to regress y_t directly on the u_{kt} (for the same t) rather than some linear combination of $u_{kt'}$ across different years t'.

However, there is also a difficulty with model (5), if this is fitted by GLS regression. It is well known that the variance of the GLS regression coefficients is smaller than that of the OLS regression coefficients. However, in the event that the model is misspecified, it is entirely plausible that the bias of GLS could be larger than that of OLS. Figure 1 displays some evidence that this might be happening: the fitted regression curve by OLS regression follows the data quite closely, but the corresponding GLS curve does not. I have observed this behavior in a number of other cases, though not in all cases (sometimes the agreement between OLS and GLS is good).

It is possible that further research may cast light on this issue and result in a theoretical explanation. My experience with the limited number of analyses in this paper and in Smith (2010) is that OLS reconstructions of the historical temperature curve are more stable than GLS reconstructions, and I therefore prefer the OLS method.

There remains the question of how to specify the ARMA model. Li, Nychka and Ammann (2007) used an AR(2) model in (5) after examining the residuals from this model to check that there was no autocorrelation; McShane and Wyner (2010) made the same claim with respect to their

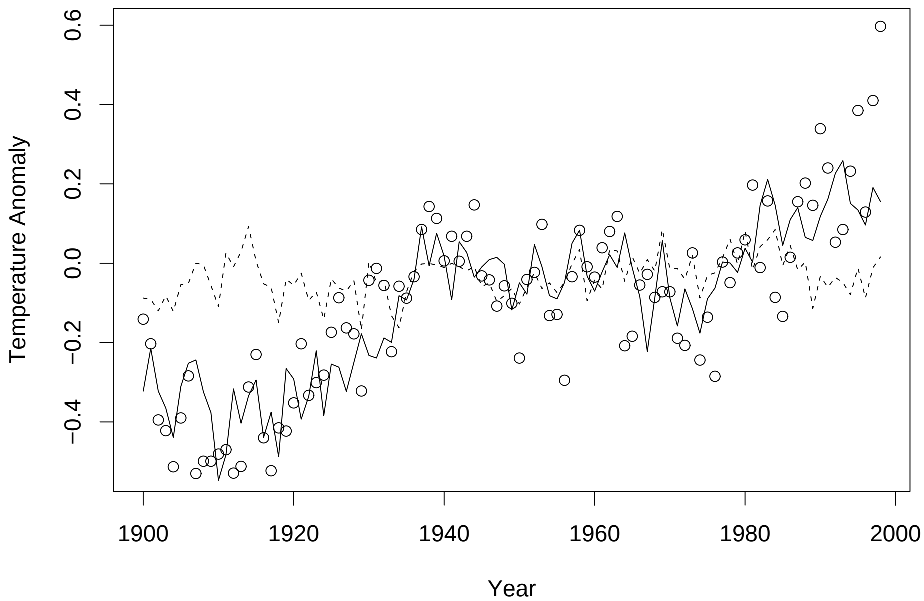

Figure 1: *NH temperature anomalies for 1900-1998, with the fitted curves of two regressions. Solid curve: OLS regression based on the first 6 PCs. Dashed curve: GLS regression based on the first 6 PCs.*

model (6). Smith (2010) examined the residuals from an OLS regression and used the statistical model selection criteria AIC, BIC and AICC to determine the order of the time series model to be fitted to the residuals; all three criteria pointed to AR(1). In limited checks with the present data, I have come to the same conclusion. However, for consistency with the papers of Li, Nychka and Ammann (2007, 2010) and McShane and Wyner (2010), I have used AR(2) in the remainder of this paper.

2.1 Technical Notes

If the regression is written in the form $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{z}$ then the covariance matrices of OLS and GLS regression are respectively $(X^TX)^{-1}X^TVX(X^TX)^{-1}$ and $(X^TV^{-1}X)^{-1}$ where V is the covariance matrix of $\mathbf{z}$. In the examples that used OLS regression, an ARMA model was fitted to the OLS regression residuals and V was estimated using the theoretical autocovariances of the fitted ARMA model.

For prediction, suppose $y_t = \mathbf{u}_t^T\boldsymbol{\beta} + z_t$ is the observation to be predicted, where $\mathbf{u}_t^T = (\ 1 \quad u_{1t} \quad \ldots \quad u_{Kt}\ )$ and $\boldsymbol{\beta}^T = (\ \beta_0 \quad \beta_1 \quad \ldots \quad \beta_K\ )$. We assume $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$ (either OLS or GLS) with covariance matrix V_β (either $(X^TX)^{-1}X^TVX(X^TX)^{-1}$ or $(X^TV^{-1}X)^{-1}$). We also assume $\hat{\boldsymbol{\beta}}$ and z_t are independent (not exactly true in the autocorrelated case, but true in practice because the autocorrelations of z_t are negligible at the time lags being used for the reconstruction).

Consider the error in predicting $\tilde{y}_t = \sum_i w_i y_{t-i}$ for given weights w_i. Defining $u_{0t} = 1$ and $\hat{y}_t = \mathbf{u}_t^T\hat{\boldsymbol{\beta}}$, we have

$$\begin{aligned}\sum_i w_i(y_{t-i} - \hat{y}_{t-i}) &= \sum_i w_i \sum_k u_{k,t-i}(\beta_k - \hat{\beta}_k) + \sum_i w_i z_{t-i} \\ &= \sum_k \tilde{u}_{kt}(\beta_k - \hat{\beta}_k) + \sum_i w_i z_{t-i}\end{aligned}$$

where $\tilde{u}_{kt} = \sum_i w_i u_{k,t-i}$.

Then

$$E\left[\left\{\sum_i w_i(y_{t-1} - \hat{y}_{t-i})\right\}^2\right] = \tilde{\mathbf{u}}_t^T V_{\boldsymbol{\beta}} \tilde{\mathbf{u}}_t + \sum_i \sum_{i'} w_i w_{i'} \gamma_{i-i'} \qquad (7)$$

where $\{\gamma_\cdot\}$ are the autocovariances of z_t. Equation (7) has been used throughout to calculate the prediction standard errors of the weighted reconstructions. This calculation does not account for the error in estimating the ARMA parameters.

3 Pseudoproxies

One of the arguments made by McShane and Wyner (2010) is that if the proxies are replaced by "pseudoproxies" generated as random "red noise" series, the fit of a regression model may actually be better than that due to the real proxies. They imply that this invalidates the temperature reconstruction process.

The argument is reminiscent of McIntyre and McKitrick (2005), who made a similar argument with respect to the original method of Mann, Bradley and Hughes (1998). However, the paper of McIntyre and McKitrick (2005) has been generally interpreted as a specific criticism of the MBH centering technique, not as a critique of paleoclimatology methods more generally. The critique of McShane and Wyner (2010) is based on the "lasso" method of regression analysis, which is widely accepted as a statistically valid method of performing regression analysis when the number of covariates is large in comparison to the number of observations. Nevertheless, to my knowledge the lasso method has not actually been used in paleoclimatology.

A side comment is that McShane and Wyner's definition of red noise is different from that of McIntyre and McKitrick (2005). The latter interpreted red noise as meaning a fractionally differenced process, for which the autocorrelations decay polynomially with lag. McShane and Wyner used an AR(1) process with autoregressive parameter ϕ, for which the kth autocorrelation is proportion to ϕ^k, an exponential decay even if ϕ is close to 1. However they also considered the nonstationary "Brownian motion" case for which $\phi = 1$.

As an independent test of McShane and Wyner's assertion, the following simulation experiment was performed.

The new simulation is based on the data and statistical methodology of Smith (2010). That paper used the NOAMER tree ring dataset, which consists of 70 temperature series constructed from tree rings for 581 years (1400-1980). As training data, I use mean global annual anomalies for 1901-1980, exactly as in Smith (2010) except that I extended the time period by one year to make it a round number 80 years. I then performed the following cross-validation exercise:

1. The period 1901-1980 was divided into four 20-year lengths (1901-1920, 1921-1940, 1941-1960, 1961-1980).

2. Each of the four periods, in turn, was omitted. A regression model based on the first two PCs

from a correlation-based (standardized) PCA of the tree ring dataset was fitted by ordinary least squares (OLS) based on the remaining 60 years of data.

3. The fitted model was applied to the 20 years of omitted data and compared with the true values to compute a total mean squared prediction error (MSPE) for that 20 years.

4. The sum of MSPE for each of the 20-year periods was computed as a total cross-validated MSPE.

The result of this exercise was a MSPE of 3.53.

The exercise was now repeated with 1000 simulations of pseudoproxies, computed as follows. First, 70 realizations of a stationary AR(1) process of length 581 with autoregressive parameter ϕ were generated. These 70 random series were assembled into a 581×70 matrix and a PCA performed, exactly as for the real tree ring data. The first two PCs were taken, and the above sequence of steps was repeated to calculate a cross-validated MSPE for each set of pseudoproxy data. Finally, the result was averaged over the 1000 simulations to produce a simulated MSPE for each ϕ. The results are displayed in Table 1.

ϕ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99	0.999
MSPE	4.03	4.01	4.00	3.99	3.94	3.93	3.86	3.76	3.54	3.13	2.60	2.32

Table 1: *MSPE for the pseudoproxy experiment*

As can be seen, the MSPE declines monotonically with ϕ being smaller than the MSPE for the real proxy data for $\phi > 0.8$.

McShane and Wyner (2010) claimed that the pseudoproxies performed worse than the real proxies when $\phi = 0.25$ or $\phi = 0.4$, but better when $\phi = 0$ (the white noise case) or in the limiting case $\phi = 1$ (Brownian motion).

The experiment performed here does not back up the assertion of McShane and Wyner (2010) in the white noise case and suggests that, perhaps, what happened in their experiment was an artifact of the lasso procedure. However, the result in the limiting case $\phi \to 1$ is consistent with the conclusion of McShane and Wyner.

The following possible explanation suggests itself. It is well known that a simple linear regression fitted to 20th century temperature data produces a highly significant result, even when the standard error of the regression coefficient is corrected to allow for autocorrelation. It is entirely conceivable that such a linear trend will produce a better fit to the data than any combination of proxies computed over a 581-year time period. A limiting red noise process will behave rather like a simple linear trend. The result from Table 1 essentially confirms confirm that a linear trend does, indeed, fit the data better than two 581-year-based PCs of the proxy data.

However, it seems to me that this result does not tell us anything meaningful about paleoclimate reconstruction. Nobody seriously believes that the linear trend has existed for 1000 years, and the experiment does confirm (by comparison with the white noise case) that the proxies do contain some information about 20th century temperatures.

I believe the result of McShane and Wyner (2010) is essentially correct, at least in the limiting cases where the random input signal is nonstationary, but it is hard to see the relevance of this result for the problem of paleoclimate reconstruction, since nobody would try to do a paleoclimatic reconstruction based on a simple linear regression.

4 Analysis of proxy data 998-1998

In this section, I perform some alternative analyses of the 93 proxies contained in the dataset of McShane and Wyner (2010), for the time period 998-1998, using the PC regression methods described in Section 2.

The observational data were taken to be 161 years (1850-2010) of mean annual anomalies from the HADCRUT3v northern hemisphere datasets. However, the hemispheric averages computed before 1900 are based on a substantially smaller coverage of the earth's surface and are therefore widely believed to be less reliable than 20th-century data. In this analysis, we use observational data only from 1900 onwards.

All analyses here are for the OLS regression method, with standard errors corrected for autocorrelation assuming an AR(2) model for the residuals.

The initial analyses are based on observational data fitted to the period 1900-1978. Later, they are extended to 1998 (the end of the time period of the proxy data, that was also used by McShane and Wyner). The reason for doing the analysis in this order will become clear as we proceed.

Figure 2 shows the reconstructed temperature curve based on the first K PCs for six values of

K. Also shown are the smoothed projections $\tilde{y}_t$, and 90% pointwide prediction intervals for those smoothed projections, computed using equation (7). This is similar to Figure 5 of Smith (2010), except that in that paper, the initial analyses were presented without correction for autocorrelation, whereas here, autocorrelation is taken into account from the beginning. The impression created by Figure 2 is that the curves are stable for $K \geq 6$.

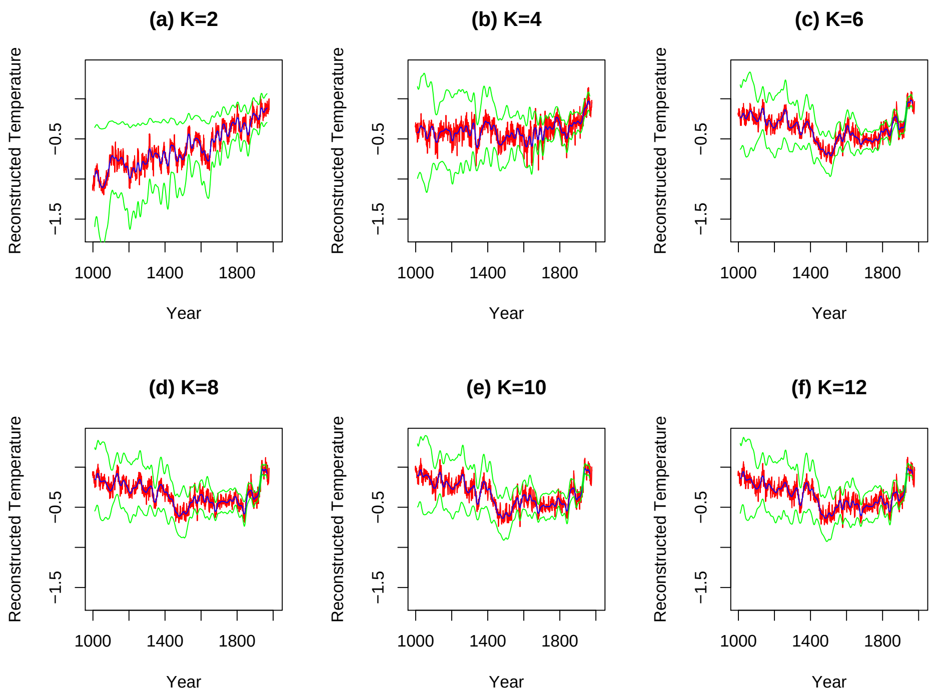

Figure 2: *Reconstructions of historical temperature anomalies, with smoothed trends and pointwise 90% prediction intervals on the smoothed trends, for 6 values of K; data fitting period 1900-1978.*

Figure 3 shows the same curve for $K = 6$ as given in Figure 2(c), but with the observations and the same smoother applied to the observations (blue dashed curve). This is of traditional "hockey stick" shape, with recent observations well above any of the historical reconstructions, though the prediction interval bands for the reconstructions are quite wide in the period 1000-1400.

On the other hand, Figure 4 shows the reconstruction computed by the same method but using observational data from 1900 to 1998. This curve departs from the hockey stick to an even greater extent than the reconstructions in McShane and Wyner (2010). In particular, if this curve is to be believed, temperatures around AD 1100 were substantially warmer than the present day.

As one comment on this, however, we do note an instability between OLS and GLS regression,

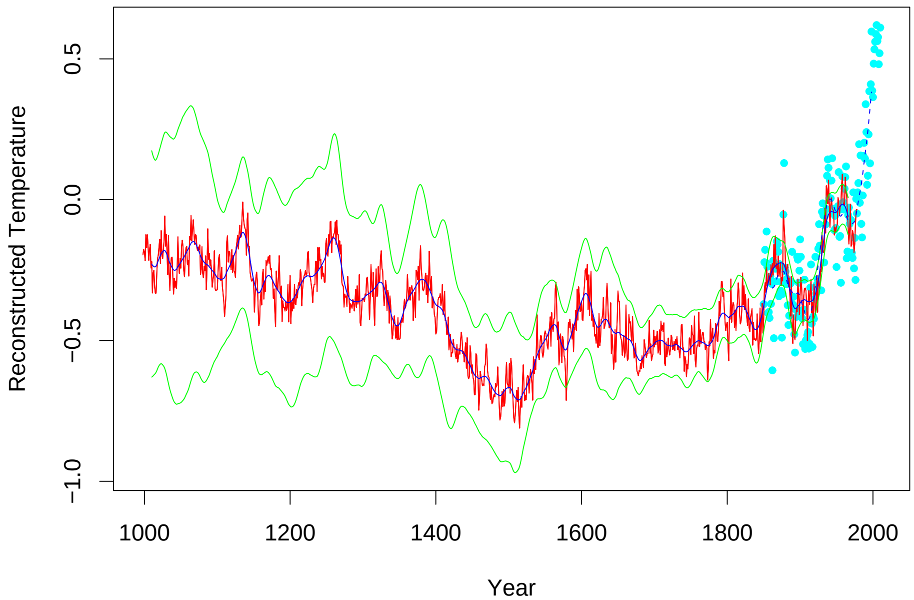

Figure 3: *Same as Figure 2(c) (K = 6), with observational data added and the same smoothed trend applied to the observational data (dashed curve).*

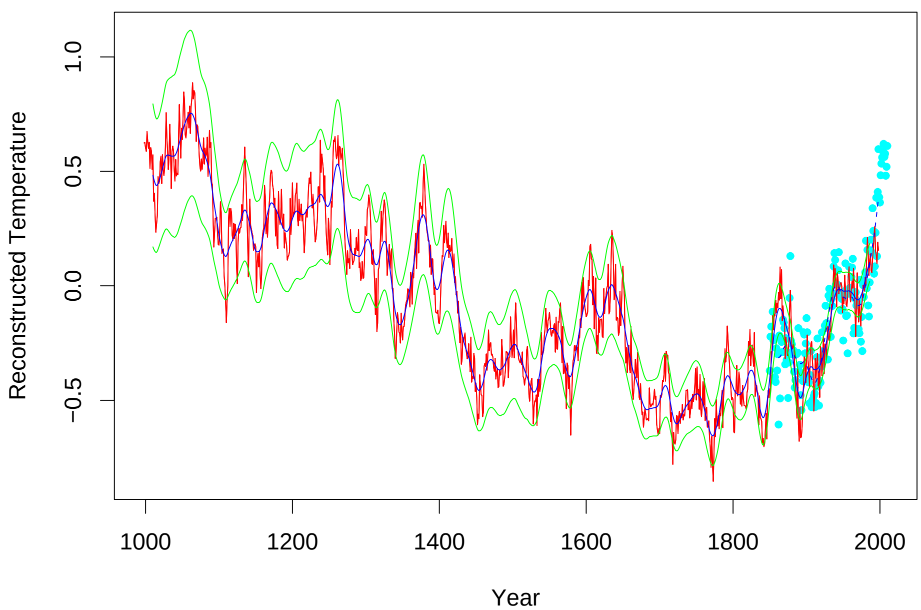

Figure 4: *Same as Figure 3, but with data fitted to 1900-1998.*

already noted in Section 2. Figure 6 shows the same curve constructed by GLS regression, based on the model of equation (5), which looks quite different from Figure 4. This is disturbing in itself, since by classical regression theory, both OLS and GLS lead to unbiased estimators of the regression coefficients. The discrepancy therefore add further evidence that the model is misspecified. (In contrast, the GLS regression corresponding to Figure 3, not shown here, has a much closer though not exact agreement with the OLS regression.)

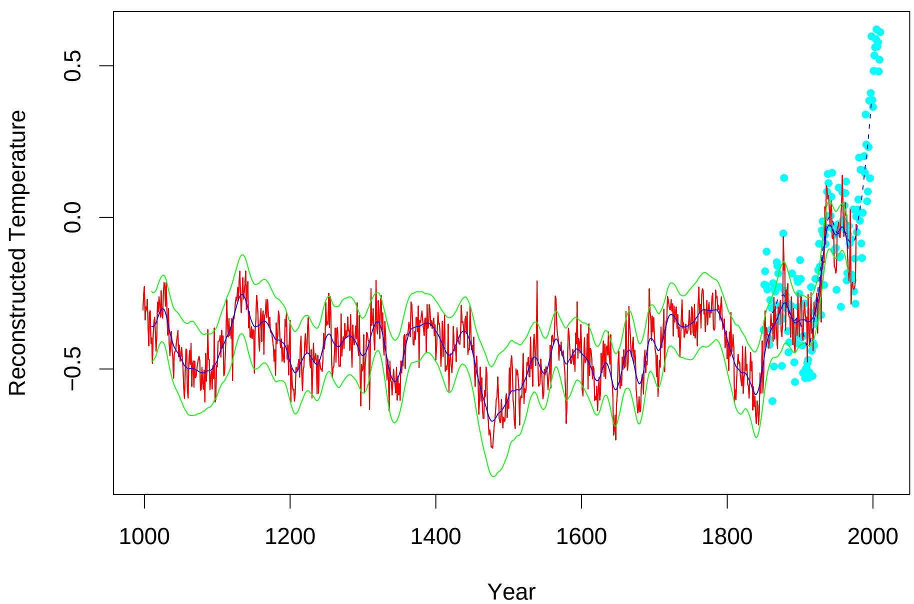

Figure 5: *Same as Figure 4, but based on GLS regression.*

From the analysis so far, it is clear that there is something discrepant about the data for 1979-1998. In fact, the climate literature contains a fair amount of discussion of the "divergence" phenomenon with tree rings (Section 5 explains this more fully) and this is what originally suggested removing the last twenty years of data. However, as the following discussion will make clear, the explanation for what has been observed has nothing to do with trees.

It should be pointed out in passing that McShane and Wyner also highlighted the divergence in paleoclimatic reconstructions over the last 20 years; see, in particular, their Figure 18. However, they did not explore the reasons for this in greater depth.

One approach that was tried was to return to the original 93 proxies, to look at their correlations (over 1900-1998) with the observed temperature record, and to remove those with low correlation. This turns out to be much too crude a method for identifying which proxies might be causing a problem.

An alternative method is to look at the influence of individual PCs. Figure 6 is calculated the same way as Figure 4, but with each of the first six PCs omitted in turn. As can be seen, if we remove PC2 from the analysis (but not any of the other PCs), the reconstructed curve again returns to the hockey stick shape. This suggests that proxies that contribute heavily to PC2 might be behaving differently as a group.

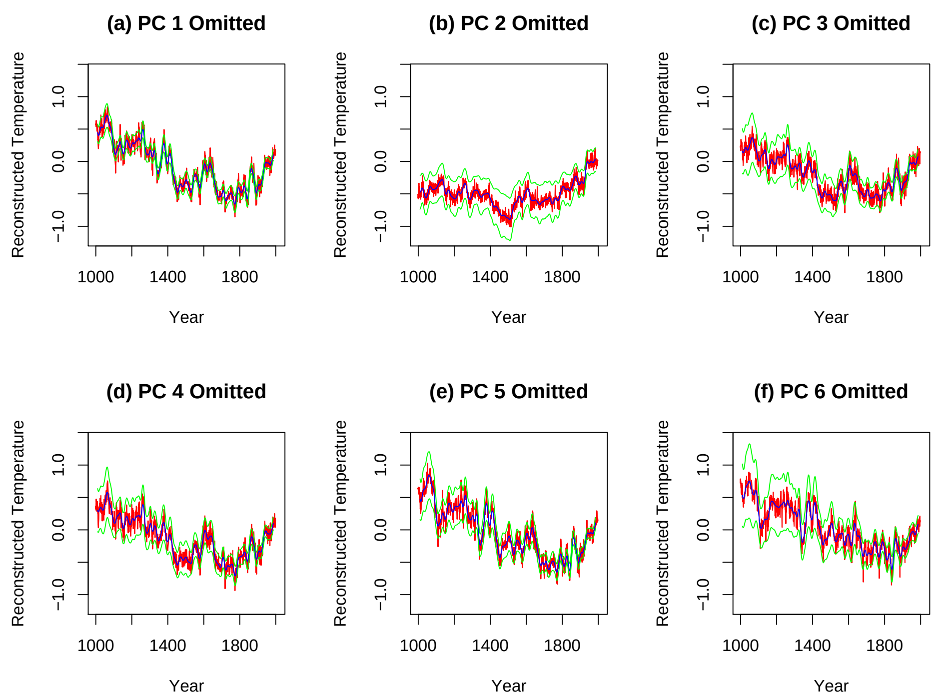

Figure 6: *The analysis of Figure 4 repeated with each of the six PCs omitted in turn, while the model is fitted to the other five.*

In fact, closer examination of the data shows that several of the proxies which contribute heavily to PC2 are of the "lake sediment" type (data codes 4000, 4001). Sediment deposits in lakes (also known as varves) are often used as a proxy for temperature over time scales as much as several thousand years, though it is also known that they are affected by variations in precipitation and by atmospheric pollution (North *et al.* 2006, pp. 60–61). Of the 93 proxies being used in the current analysis, 12 are of lake sediment type. This suggests repeating the analyses with those proxies removed.

Hence a “reduced proxy” dataset was constructed, consisting of 81 proxies, removing all those connected with lakes. I repeated many steps of the foregoing analysis. Once again, it appeared that an acceptable reconstruction was obtained with 6 PCs. The reconstruction using observational data from 1900-1978 is given in Figure 7, and that based on 1900-1998 in Figure 8.

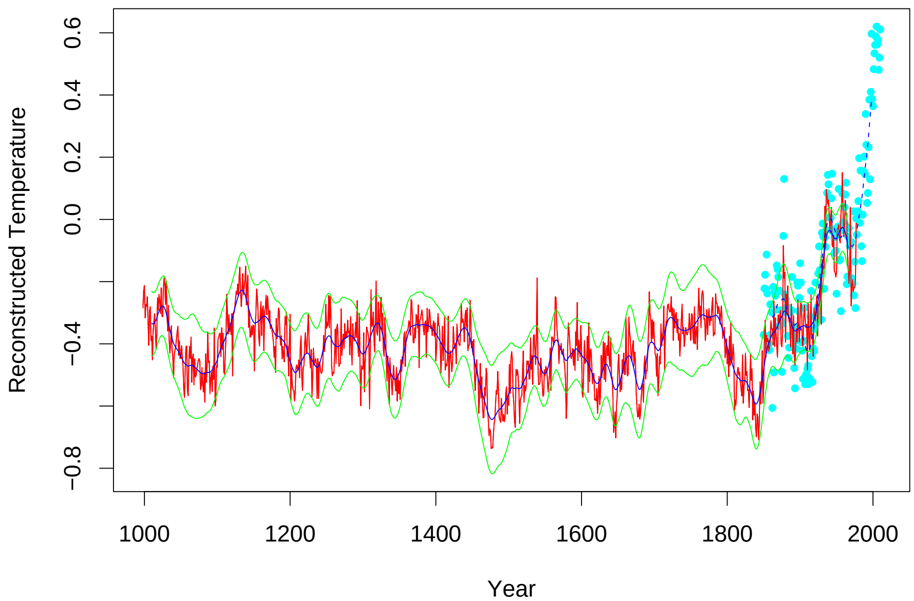

Figure 7: *Same as Figure 3, but fitted to the reduced proxy dataset.*

Both figures are essentially of hockey stick shape, though on closer examination, there are still non-trivial discrepancies. For instance, since the observational data from 1850-1899 were not used in the reconstruction, those data themselves could be considered a test of the quality of the reconstruction. Close examination of Figure 8 shows that the observation-based smooth trend (dashed curve) is sometimes outside the confidence bands from the reconstruction during this period; the agreement in Figure 7 is much better. So even after removing all the lake sediment proxies, it still seems there are some problems with the reconstruction that uses observational data up to 1998, but the discrepancy between the reconstruction based on 1900-1978 observational data and the reconstruction based on 1900-1998 observational data is nowhere near as great as in the original analysis.

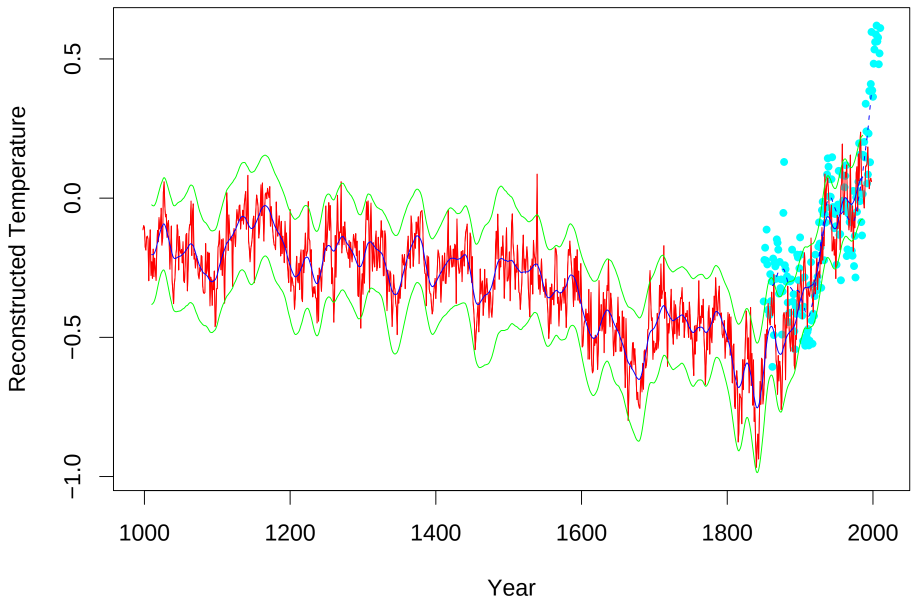

Figure 8: *Same as Figure 4, but fitted to the reduced proxy dataset.*

I also computed a figure analogous to Figure 6, omitting each of PCs 1-6 in turn. This figure (not shown here) showed a generally consistent shape of the reconstruction, and did not suggest that any single PC is dominating the overall reconstruction.

Finally, we look at the estimated slope of the temperature trend. It is often stated that, regardless of the exact temperatures that were observed in the distant past, the rapid pace of recent changes is much greater than anything in known history. This hypothesis can be examined by using the predicted values of $\tilde{y}'_t$ as an approximation to the slope of the temperature trend. These values, together with pointwise 90% prediction intervals, are shown in Figure 9. The data fitting period in this case was 1900-1978: the corresponding plot based on 1900-1998 is of similar overall form but with less good agreement to 19th-century observational data (as has already been noted).

The result in this case shows that the slope of the temperature trend in the twentieth century was about 0.5 K/century, but that the slope has approached this several times in the past, in fact appears to have done so at regular intervals with a period of about 150 years. In contrast, recent NH temperatures since 1980 have been rising at a rate of about 2 K/century, though since this is

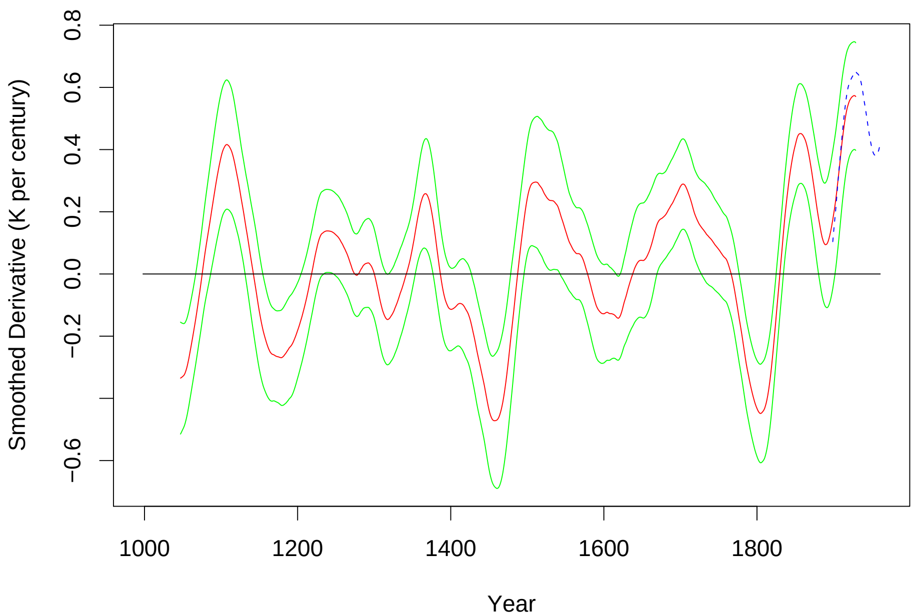

Figure 9: *Estimated slope of the trend curve, $\tilde{y}'_t$, with pointwise 90% confidence bands. The analysis is based on the same data and statistical model as Figure 7. Blue dashed curve: the same estimated slope applied to the observational data.*

based on a much shorter averaging time, this result is not directly comparable with the values in Figure 9.

5 Discussion and Conclusions

All paleoclimate reconstruction relies on an implicit assumption of stationarity: that is, that the relationships between proxies and true temperatures do not change across time (North *et al.* 2006, Chapter 9). Such an assumption should never be taken for granted but needs critical examination for each proposed new reconstruction. Paleoclimatologists have coined the term "divergence" to describe cases in which the stationarity assumption appears to be breaking down within the timescale of observational data.

The best known example of divergence concerns trees; see for example Briffa *et al.* 1998 or pages 48-52 of North *et al.* (2006). However, the problem does not (so far as is known) apply uniformly to all tree-ring proxies; the specific class of proxies for which it is known to be a problem are tree-ring latewood density records. However, most of the known records of this type go back no further than AD 1400; in particular, none of them are among the 93 proxies used in the present analysis (Dr. Michael Mann, personal communication). Therefore, it appears that the known divergence problem with tree rings is not responsible for the results in the present paper.

However, it has also been suggested that the divergence problem is not confined to tree rings and may affect other proxies as well (Mann *et al.* 2008, page 13254). To the best of my knowledge, no previous study has explicitly identified lake sediment records as subject to this problem, though with the benefit of hindsight, it seems obvious that lake sediment deposits in the late 20th century would be affected by anthropogenic activity other than increasing CO_2.

With the elimination of lake sediment proxies, the agreement between the 1900-1978 and 1900-1998 reconstructions is much better, and both show a clear hockey stick shape. However, we have observed that there are still non-trivial differences between the two: in particular, the reconstruction from 1900-1998 fails to reproduce the observed signal over part of the nineteenth century, which is within the timescale of observational data but not used in the reconstruction itself.

The paper by McShane and Wyner (2010) and the follow-up analyses presented here have highlighted the sensitivity of paleoclimatic reconstructions to the time period of observational data and to the selection of proxies. I am not suggesting that the elimination of lake sediment proxies is the complete solution to the problem, but it does substantially reduce the discrepancies observed in the earlier analysis. I believe the whole analysis highlights the dangers of applying automated statistical methods to large datasets without carefully considering the structure of the data.

The recent series of papers on paleoclimatic reconstruction have also highlighted a variety of different statistical methods. The papers by Li, Nychka and Ammann (2007), Smith (2010), McShane and Wyner (2010) and the present contribution have used more or less direct regression methods, regressing observed temperatures on proxies to develop a regression relationship and then projecting backward in time, with slightly different approaches to the question of autocorrelation. As noted in Section 1, there has been a recent trend towards the use of Bayesian Hierarchical Models, which may ultimately be the most comprehensive solution though they are also the most computationally intensive. A third trend in paleoclimate reconstruction methods is the use of Errors in Variables (EIV) techniques that allow for randomness in the regressors as well as the observations, see e.g. Schneider (2001), Hegerl *et al.* (2006) and a recent statistical contribution by Ammann, Genton and Li (2010). Mann *et al.* (2008) remark that EIV methods seem less sensitive to the divergence problem than classical regression techniques; given that McShane and Wyner (2010) and the present paper have highlighted the importance of the divergence problem, the fact that some statistical methods appear to be more robust than others at dealing with it would seem to highlight the need for more systematic study of these issues.

In summary, recent papers published in this field have highlighted the sensitivity of paleoclimatic data reconstructions to choices in both data selection and statistical methodology. The fact that statistical methods that seem at first sight extremely logical, applied to well-documented datasets, can produce results totally at odds with previous literature, is an important warning against the automated use of statistical methods without consideration of the data to which they are being applied.

These new analyses do not settle the question of whether the hockey stick relationship, for northern hemisphere temperatures over the last 1000 years, is true. However, given that the majority of recently published analyses support that relationship, and in the cases of those that do not, the reasons for the discrepancy are easily understood and explained, it seem likely that the consensus of the scientific community will continue to support the hockey stick relationship.

6 References

Ammann, C.M., Genton, M.G. and Li, B. (2010), Correcting for signal attenuation from noisy proxy data in climate reconstruction. *Climate of the Past* **6**, 273–279.

Ammann, C.M. and Wahl, E.R. (2007), The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climatic Change* **85**, 71–88

Briffa, K.R., Schweingruber, F.H., Harris, I.C., Jones, P.D., Shiyatov, S.G. and Vaganov, E.A. (1998), Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* **391**, 678–682.

Brockwell, P.J. and Davis, R.A. (2003), *Introduction to Time Series and Forecasting.* Second edition: Springer.

Brynjarsdóttir, J. and Berliner, L.M. (2010), Beysian hierarchical modeling for paleoclimate reconstruction from geothermal data. Preprint, Ohio State University.

Hegerl, G.C., Crowley, T.J., Hyde, W.T. and Frame, D.J. (2008), Climate sensitivity constrained by temperature reconstruction over the past six centuries. *Nature* **440**, 1029–1032.

Li, B., Nychka, D.W. and Ammann, C.M. (2007), The 'hockey stick' and the 1990s: a statistical perspective on reconstructing hemispheric temperatures. *Tellus* **59A**, 591–598.

Li, B., Nychka, D.W. and Ammann, C.M. (2010), The value of multi-proxy reconstruction of past climate. *Journal of the American Statistical Association*, to appear.

Mann, M.E., Bradley, R.S. and Hughes, M.K. (1998), Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787.

Mann, M.E., Bradley, R.S. and Hughes, M.K. (1999), Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters* **26**, 759–762.

Mann, M.E., Zhang, Z., Hughes, M.K., Bradley, R.S., Miller, S.K., Rutherford, S. and Ni, F. (2008), Proxy-based reconstruction of hemispheric and global surface temperature variations over the past two millenia. *Proceedings of the National Academy of Sciences* **105**, 36.

McIntyre, S. and McKitrick, R. (2005), Hockey sticks, principal components, and spurious significance. *Geophysical Research Letters* **32**, L03710, doi:10.1029/2004GL021750.

McShane, B.B and Wyner, A.J. (2010), A statistical analysis of multiple temperature proxies:

Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*, to appear. Available online at
http://www.e-publications.org/ims/submission/index.php/AOAS/user/submissionFile/6695?confirm=63ebfddf

North, G.R., Biondi, F., Bloomfield, P., Christy, J.R., Cuffey, K.M., Dickinson, R.E., Druffel, E.R.M., Nychka, D., Otto-Bliesner, B., Roberts, N., Turekian, K.K. and Wallace, J.M. (2006), *Surface Temperature Reconstructions for the Last 2,000 Years.* National Research Council ISBN: 0-309-66144-7, available online from http://www.nap.edu/catalog/11676.html

Schneider, T. (2001), Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate* **14**, 853–871.

Smith, R.L. (2010), Elementary Reconstruction of the Hockey Stick Curve: Discussion of Paper by Li, Nychka and Ammann. *Journal of the American Statistical Association*, to appear. Available online at
http://www.stat.unc.edu/postscript/rs/PaleoASA.pdf

Tingley, M.P. and Huybers, P. (2010a), A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *Journal of Climate* **23**, 2759–2781.

Tingley, M.P. and Huybers, P. (2010b), A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: Comparison with the regularized expectation-maximization algorithm. *Journal of Climate* **23**, 2782–2800.

Wahl, E.R. and Ammann, C.M. (2007), 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* **85**, 33–69.

Wegman, E.J., Scott, D.W. and Said, Y.H. (2006), *Ad Hoc Committee Report on the 'Hockey Stick' Global Climate Reconstruction.* Report presented to the Committee on Energy and Commerce and the Subcommittee on Oversight and Investigations, U.S. House of Representatives.