The 'hockey stick' and the 1990s: a statistical perspective on reconstructing hemispheric temperatures

Chris Cabanski

April 25, 2009

Introduction Over the past 10 years, the quality of regional and hemispheric temperature reconstruction for the past millennium has increased substantially, due to both methodological development as well as to improved data availability. The recent debate about the hockey stick temperature reconstruction has highlighted the importance of these past temperature estimates. Since there is only instrumental temperature data from (approximately) 1850 onward, different techniques have been used to estimate temperatures before 1850. However, it is widely acknowledged that these reconstructions may contain substantial uncertainty that is difficult to quantify.

Traditionally, most large-scale temperature reconstructions are presented as a single time series without error bars or confidence ranges. Although it is recognized that the proxies contain noise, the reconstruction outcomes are often reported as 'unique'. This, however, does not take into account that the noise component in all proxies is but a single realization of many possible noise realizations. Hypothetically, should the exact same climate occur again, the noise realization would be different. Therefore, it would be more appropriate to recognize each proxy-based climate reconstruction as an individual member of a family of possible realizations.

In this paper, the authors develop a statistical method to reconstruct past temperatures together with its confidence ranges by keeping track of different sources of uncertainty. They then illustrate their approach with multivariate reconstruction of hemispheric mean temperatures using a limited proxy set.

Data and Methods For the example explored in this paper, the authors restrict their network of proxy records to the 14 series originally used in MBH99 for the period back to the year 1000. They also assume the linear and stationary relationship between these 14 proxies and the temperature evolution following MBH98, MBH99.

Let T_t and $\mathbf{p}_t$ denote the temperature and proxies at time t, respectively. The basic statistical model is

$$T_t = \mathbf{p}_t'\beta + e_t \tag{1}$$

where β is a vector of regression coefficients, and the vector of errors $\mathbf{e} = (e_1, \ldots, e_t)' \sim \text{Normal}(0, \Sigma)$. This model takes uncertainties from both the instrumental data and the proxies into account.

The authors are interested in studying the uncertainties underlying this linear regression reconstruction. Accodingly, they address three types of uncertainties:

1. Effects of possible autocorrelation in the errors of the linear model, owing to the assumption of independent errors for the ordinary linear model.

2. Potential prediction errors due to the overfitting from the calibration period. This type of uncertainty is probably relatively small since the series is so long.

3. Uncertainty about the proxies themselves and how they represent the climatic conditions over multiple timescales.

The full time period from 1850 to 1980 was chosen as the calibration period for model (1). The authors then applied the estimated model to the proxies from 1000 to 1849 to reconstruct the temperature for that time period. After fitting an ordinary least squares model, it was seen that the errors exhibited a clear temporal correlation. An AR(2) model was found sufficient to capture the temporal correlation in the residuals. An AR(2) process is defined as

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \epsilon_t,\ \epsilon_t \sim \text{iid Normal}(0, \sigma^2) \tag{2}$$

where ϕ_1 and ϕ_2 are coefficients governing the correlation of time lag 1 and time lag 2.

Motivated by the temporal correlated errors, model (1) is fit using generalized least-squares, which allows the errors to be correlated. The parameters involved in this model are $\theta = (\beta',\ \sigma^2,\ \phi_1,\ \phi_2)'$. Maximum likelihood estimates are chosen in the GLS fitting to obtain the parameter estimates $\hat{\theta} = (\hat{\beta}',\ \hat{\sigma}^2,\ \hat{\phi}_1, \hat{\phi}_2)'$ simulatenously. Being concerned with the possible overfitting problem during the calibration period, the authors use 10-fold cross-validation to quantify how much overfitting there is, if any. By applying 10-fold cross-validation and fixing $\hat{\phi}_1$ and $\hat{\phi}_2$ but leaving the other parameters varying, they find that an inflation factor of 1.30 is needed.

To account for the uncertainty from parameter estimates, $\hat{\theta}$, the authors employ a paramteric bootstrap to determine the sample distribution of $\hat{\theta}$. This begins with generating ensembles of temperature based on model (1). Particular to each ensemble, they first generate an AR(2) error, $\mathbf{e}_0$, as defined in (2) taking $\hat{\sigma}^2$, $\hat{\phi}_1$, $\hat{\phi}_2$ as true parameters. Then letting $\tilde{\mathbf{T}} = (\tilde{T}_{1850}, \ldots, \tilde{T}_{1980})'$ denote an ensemble of temperature and $\mathbf{P}$ denote the known proxy matrix containing rows $\mathbf{p}_{1850}, \ldots, \mathbf{p}_{1980}$, $\tilde{\mathbf{T}} = \mathbf{P}\hat{\beta} + \mathbf{e}_0$ produces one valid ensemble. These temperature ensembles have the same mean function $\mathbf{P}\hat{\beta}$, but each ensemble has its own noise, which makes any individual ensemble feature differently from the others. For each ensemble, they repeat the GLS model fitting procedure to get the parameter estimates, denoted by $\tilde{\theta} = (\tilde{\beta}',\ \tilde{\sigma}^2,\ \tilde{\phi}_1, \tilde{\phi}_2)'$, which are considered to have the same distribution as $\hat{\theta}$ estimated from the real data. They generate 1000 temperature ensembles and thus obtain 1000 $\tilde{\theta}$. The sampling distribution of these $\tilde{\theta}$ is a valid estimate of the distribution of $\hat{\theta}$.

Now let $\tilde{\mathbf{T}} = (\tilde{T}_{1000}, \ldots, \tilde{T}_{1849})'$ and $\mathbf{P}$ be the proxy matrix containing rows $\mathbf{p}_{1000}, \ldots, \mathbf{p}_{1849}$. An ensemble is given by $\tilde{\mathbf{T}} = \mathbf{P}\tilde{\beta} + (e_{1000}, \ldots, e_{1849})' \mid (e_{1850}, \ldots, e_{1980})'$. The conditional error term ensures that the ensemble is temporally correlated with the instrumental temperature. Each $\tilde{\theta}$ is picked in turn from the 1000 members of $\tilde{\theta}$ generated above, and $\tilde{\sigma}^2$ is inflated by the inflation coefficient to get $\tilde{\tilde{\sigma}}^2 = 1.30\tilde{\sigma}^2$. The selected $\tilde{\theta}$, with $\tilde{\tilde{\sigma}}^2$ used in place of $\tilde{\sigma}^2$, are taken as true parameters to generate ensembles. This brings all the components of uncertainty into the ensembles. In this way, the 1000 temperatures ensembles that are generated correspond to the 1000 individual $\tilde{\theta}$.

Temperature Reconstruction and Results The authors calculate the running decadal average using 10 years as the moving window. They then identify the maximum decadal average and record its corresponding year. They do this for each of the 1000 temperature ensembles generated in the previous section, and thus obtain 1000 decadal maxima with their corresponding years. The result enables the authors to compute the probability of each year corresponding to the maximum decadal average across all the ensembles.

All the temperature ensembles are possible realizations with equal chance of occurrence given the same setting of the proxy and instrumental data. Hence, the statistical inference obtained from each ensemble is an equally valid estimate of that inference, and moreover, the variability of inferences from different ensembles shapes the distribution of that inference.

Figure 4 shows a summary of the 1000 temperature ensembles. Up to the 1980s, instrumental temperatures are not significantly higher than the reconstructed maxima of before 1850 (the 95% confidence interval of the reconstructed decadal maxima, shown by the black dashed line in Figure 4, is 0.22°C, and instrumental temperatures before 1980 fall below this level). However, with the 1990s, observed temperatures (shown by the red asterisks in Figure 4) begin to rise clearly above the maxima from the ensembles. For example, the 1990s is 0.08° warmer than the upper bound of the 95% confidence interval, and the most recent decade (1997-2006) is yet another 0.22° warmer. This example seems to confirm the 'hockey stick' phenomenon that temperatures over the past few decades are much warmer than would be expected compared to the reconstructed temperature ensembles.

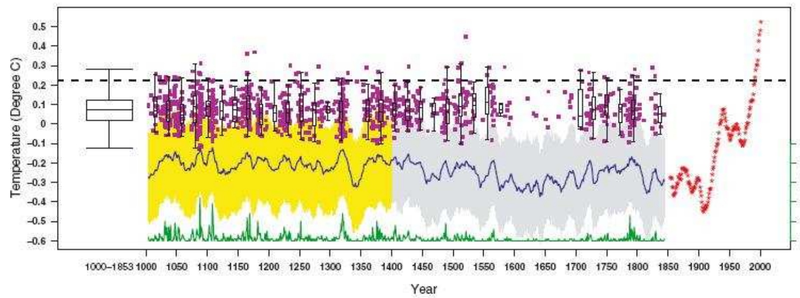

Fig. 4. Summary of 1000 temperature ensembles: the decadal average of mean temperature over the 1000 ensembles (blue curve) with its 95% confidence region (yellow and grey band); decadal maxima from 1000 individual ensembles (purple dots) and the chance of each year corresponding to the decadal maximum (green curve scaled by the green label); the upper bound of the 95% confidence interval of the decadal maxima (dashed line) and decadal instrumental temperatures (red asterisks). The small box plots overlapped with purple dots show the distribution of decadal maxima in small groups with each group containing about 20 yr, and the leftmost big box plot shows the distribution of all the decadal maxima. The decadal average of mean temperature before 1400 (the blue curve embedded in the yellow band) and its 95% confidence region (yellow band) can be compared to the corresponding section of Fig. 3(a) in MBH99.

Conclusions A benefit of introducing the notion of ensemble reconstruction is that ensembles make it easy to draw more sophisticated inferences about past temperture evolution such as the decadal maximum. There are several important differences in this analysis from previous work. Adjustment was made for temporal correlation in errors, cross-validation was used to adjust for overfitting, and bootstrapping was used to determine uncertainty in the estimated parameters.

References

[1] 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.