DISCUSSION OF: A STATISTICAL ANALYSIS OF MULTIPLE TEMPERATURE PROXIES: ARE RECONSTRUCTIONS OF SURFACE TEMPERATURES OVER THE LAST 1000 YEARS RELIABLE?

By Lasse Holmström

Department of Mathematical Sciences, University of Oulu

POB 3000, 90014 University of Oulu, Finland

lasse.holmstrom@oulu.fi

This is an impressive paper. The authors present a thorough examination of the ability of various climate proxies to predict temperature. The prediction method is one much used in climate science literature and assumes a linear relationship between the proxies and the temperature. The idea is to use instrumental temperature data together with the corresponding proxy records to estimate a regression model to which historical proxy values are then input in order to produce a backcast of past temperature variation. The authors demonstrate convincingly that the data used in Mann et al. (2008) does not allow reliable temperature prediction using this approach and that purely random artificial proxy records in fact perform equally well or even better.

While this is certainly striking and thought-provoking, one should not be left under the impression that this is the standard approach to understanding past climate and that temperature reconstruction per se is impossible. In fact, some of the "proxies" used in the paper are themselves supposedly successful temperature reconstructions and therefore arguably of a more fundamental character than the predictor produced by the Lasso. There is for example a long and well established paleoecological tradition of quantitative environmental reconstruction based on diatoms, pollen, chironomids and other biological proxies that in some important aspects differs from the regression approach used in the present paper and that can offer better prediction accuracy (e.g. Birks (1995); Birks et al. (2010)). A typical temperature reconstruction in this tradition uses a sediment core from a selected lake together with training data from a number of other lakes to backcast temperatures hundreds or thousands of years in time. As a proxy one can use the relative abundances of various organisms (say, different diatom taxa) measured at various depths along the core. A model for the dependence between the temperature and the abundances is built using a training set that consists of the relative abundances of the same the same organisms in surface sediment samples from a large number of lakes located in the same general area as the core lake and their current temperatures, mean July temperature being a typical climate variable. The training lakes are selected to cover a wide range of environmental conditions to make possible temperature backcasting to times when the conditions at the core lake possibly were very different from what they are today. This is referred to as space-for-time substitution.

Compared to the approach studied in this paper, at least two differences seem clear: the relative locality of the analysis and the possibility in some cases to incorporate ecological information in the model. Instead of a global proxy network, this method

typically uses local proxy data to backcast local environmental conditions. Ecological information can enter the model for example through a hierarchical model component that explicitly incorporates the fact that different organisms have different optimal temperatures, that is, temperatures at which they fare particularly well. Various models, including methods based on parametric or nonparametric regression as well Bayesian approaches have been proposed. Such reconstructions appear quite successfully capture many large-scale Holocene climate features such as the Medieval Warm Period and the Little Ice Age. I wonder if the rather strict locality of the approach combined with at least some degree of ecological plausibility in modeling the proxy data generating process are factors in their apparent success?

Recently, Bayesian approaches have begun to enter the field. The first papers that used detailed Bayesian modeling for paleoclimate reconstruction were Vasko et al. (2000), Toivonen et al. (2001) and Korhola et al. (2002). More recently, Haslett et al. (2006) and Erästö and Holmström (2006) have taken similar approaches. As reconstruction in the Bayesian setting produces the posterior distribution of the past temperature variation, such important questions as joint (pathwise) credibility of past temperature features also discussed in the present paper can be easily answered.

Another question I would like to consider is the possible role of smoothing, briefly touched upon at the end of Section 3. As pointed out by the authors, one difficulty is the choice of a proper smoothing level. However, in the so-called scale space approach this difficulty is turned into an opportunity when, instead of just one, in some sense optimal smooth, one instead considers a whole family of smooths. Our paleoecologist collaborators have found quite useful a procedure where such multi-level smoothing is applied to the reconstructed temperature time series. Each smooth can then be interpreted to provide information about the underlying past temperature variation at a particular time scale, little smoothing showing the short time scale details and heavy smoothing leaving only the coarsest features, such as the over all trend. For a Bayesian reconstruction such scale space smoothing can easily be combined with a credibility analysis of these multi-scale features but similar analyses are possible also in a non-Bayesian setting (e.g. Erästö and Holmström (2005, 2006, 2007); Holmström and Erästö (2002); Godtliebsen at al. (2003); Rohling and Pälike H. (2005)).

It is also possible to combine a non-Bayesian reconstruction with Bayesian scale space analysis. This is useful as the vast majority of existing paleoreconstructions are non-Bayesian. Thus, suppose that $\mu(t_i)$ is the true past temperature at time t_i, $i = 1,\ldots,n$, and let $y_i = \mu(t_i) + \varepsilon_i$ be its reconstructed value. After specifying priors for μ and the reconstruction errors ε_i the posterior distribution $p(\mu' \mid y_1,\ldots,y_n)$ of the derivative of μ can then be obtained. The scale space analysis now consists of smoothing this posterior at different levels in order to find the credible features of past temperature variation in different scales. It is also possible to handle correlated errors as well as errors in the time points t_i. For details see Erästö and Holmström (2005, 2007).

An example of this idea is shown in Figure 1, where an analysis of a diatom-based temperature reconstruction in Northern Fennoscandia for the past 800 years is shown. In the upper panel, asterisks show the actual reconstructed temperatures together with three smooths (posterior means) reflecting variation in millennial, centennial and decadal timescales. The lower panel map summarizes the scale space inference. Following any horizontal line in this map we see when, at that particular smoothing

level, or time scale, the local temperature trend has been credibly positive (red) or negative (blue) or when no credible change has occurred (gray). Inference is joint over all time points. In this analysis, the little Ice Age shows as a credible broad minimum in several different scales, from century to millennial scales (the color changing from blue to red), whereas the overall millennial trend has been negative. The very recent warming after industrialization shows as red in many scales.

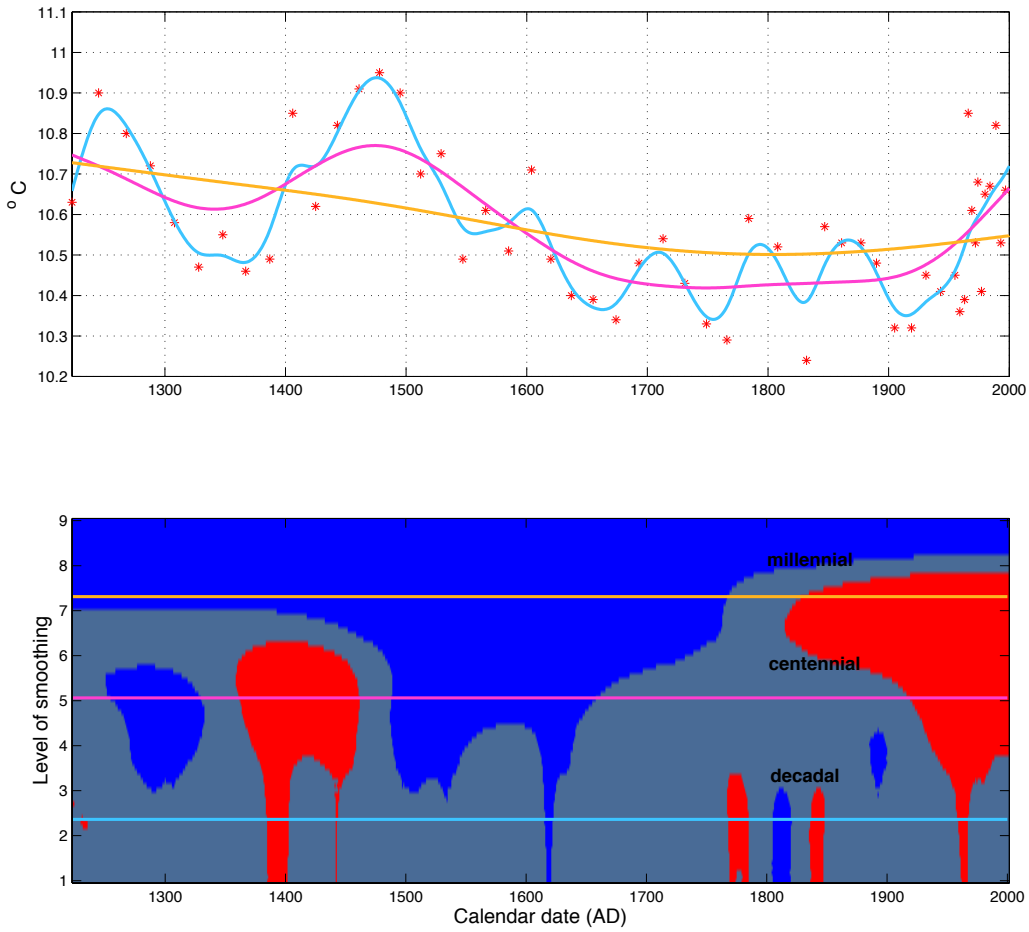

Figure 1. Scale space analysis of the salient features of a diatom-based temperature reconstruction in Northern Fennoscandia for the past 800 years. Upper panel: the reconstructed temperatures (asterisks) together with three smooths. Lower panel: scale space credibility map with the horizontal lines indicating the corresponding three levels of smoothing. For more details, see Weckström et al. (2006) and the text.

References

Birks H.J.B. (1995). Quantitative palaeoenvironmental reconstructions. In: Statistical Modelling of Quaternary Science Data. Technical Guide 5 (eds. Maddy D & Brew JS). Quaternary Research Association Cambridge, pp. 161-254.

Birks H.J.B. et al. (2010). Strengths and Weaknesses in Quantitative Climate Reconstructions based on Late-Quaternary Biological Proxies, unpublished.

Erästö P. and Holmström L. (2005). Bayesian multiscale smoothing for making inferences about features in scatter plots. *Journal of Computational and Graphical Statistics*, 14(3):569-589.

Erästö P. and Holmström L. (2006). Prior selection and multiscale analysis in Bayesian temperature reconstruction based on species assemblages. *Journal of Paleolimnology*, 36(1):69-80.

Erästö P. and Holmström L. (2007). Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors. *Journal of Statistical Computation and Simulation*, 77(5):421-431.

Godtliebsen F., Olsen L.R. and Winther J.G. (2003). Recent devel- opments in statistical time series analysis: examples of use in climate research. *Geophysical Research Letters* 30:1654. doi:10.1029/2003GL017229.

Haslett, J., M. Whiley, S. Bhattacharya, F. Mitchell, J. Allen, B. Huntley, S. Wilson, and M. Salter-Townshend (2006). Bayesian paleoclimate reconstruction. *Journal of the Royal Statistical Society*, Series A 169(3), 395–438.

Holmström L. and Erästö P. (2002). Making inferences about past environmental change using smoothing in multiple time scales. *Computational Statistics & Data Analysis*, 41(2):289-309.

Korhola A., Weckström J., Holmström L. and P. Erästö (2000). A quantitative Holocene climatic record from diatoms in northern Fennoscandia. *Quaternary Research*, 54:284-294.

Korhola, A., Vasko K., Toivonen H. T. T., and Olander H. (2002). Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling. *Qaternary Science Reviews* 21(16–17), 1841–1860.

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

Rohling E.J. and Pälike H. (2005). Centennial-scale climate cooling with a sudden cold event around 8,200 years ago. *Nature* 434:975–979.

Toivonen, H. T. T., Mannila H., Korhola A., and Olander H. (2001). Applying Bayesian statistics to organism-based environmental reconstruction. *Ecological Applications,* 11 (2), 618–630.

Vasko, K., Toivonen H. T., and Korhola A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *Journal of Paleolimnology*, 24(3), 243–250.

Weckström J., Korhola A., Erästö P., and Holmström L (2006). Temperature Patterns over the Past Eight Centuries in Northern Fennoscandia Inferred from Sedimentary Diatoms. *Quaternary Research,* 66:78-86.