The Annals of Applied Statistics
2011, Vol. 5, No. 1, 71–75
DOI: 10.1214/10-AOAS398H
Main article DOI: 10.1214/10-AOAS398
© Institute of Mathematical Statistics, 2011

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

BY LASSE HOLMSTRÖM

University of Oulu

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

Received September 2010; revised September 2010.