Geophysical Research Abstracts
Vol. 14, EGU2012-2546, 2012
EGU General Assembly 2012

Multiproxy Paleoclimate Reconstruction Using Gaussian Graphical Models

D. Guillot (1), B. Rajaratnam (2), and J. Emile-Geay (3)

(1) University of Southern California and Stanford University, United States (dguillot@usc.edu), (2) Stanford University, United States (brajarat@stanford.edu), (3) University of Southern California, United States (julieneg@usc.edu)

Understanding centennial-scale climate variability hinges on the availability of datasets that are accurate, long, continuous, and of broad spatial coverage. Since instrumental measurements are generally only available after 1850, temperature fields must be reconstructed by using information embedded in proxy records spanning all or part of the Common Era. Various climate field reconstructions (CFR) methods have heretofore been proposed to infer past climate from multiproxy networks. A popular one is the regularized EM algorithm (RegEM) with several regularization schemes like truncated total least squares (TTLS) and individual ridge regression (iRIDGE). Like RegEM, most CFR methods are based on multivariate regression. The cornerstone of such methods is to obtain an accurate estimate of the covariance matrix Σ of the data. In the paleoclimate context, this is a challenging problem because the number of variables (number of temperature grid points, p) vastly exceeds the number of observations (number of years of instrumental data, n). The matrix Σ is often estimated by the sample covariance matrix of the dataset, but this estimator is known to be very poor when $p > n$. Significant improvements can be made by estimating Σ using a Gaussian graphical model (GGM). A GGM (a.k.a. Markov random field) is a multivariate normal model where variables display certain conditional independence relations encoded by a graph. Such relations are inherent to climate fields and can be exploited to yield better temperature reconstructions.

In this work, we propose a new CFR method based on GGMs. By modeling the conditional independence structure of climate fields, the number of parameters to estimate can be greatly reduced. Precise and well-conditioned estimates of Σ can then be computed, even when the sample size is much smaller than the number of variables. Those estimates can then be used in various reconstruction procedures like the EM algorithm. I will discuss efficient algorithms to discover the conditional independence structure of climate data and show how GraphEM (GGM + EM algorithm) can be used to reconstruct past climate variability. I will then compare the performance of GraphEM to RegEM TTLS using pseudoproxy experiments and real datasets. Our results show that, in many cases, GraphEM outperforms RegEM TTLS. This increase in performance is directly tied to the sparsity of the covariance model, confirming the usefulness of ℓ^1 model selection prior to ℓ^2 regularization.