Geophysical Research Abstracts
Vol. 20, EGU2018-13874-1, 2018
EGU General Assembly 2018

Estimating Correlations between Paleoclimate Time Series: Sources of Uncertainty and Potential Pitfalls

Jasper G. Franke (1,2) and Reik V. Donner (1)

(1) Potsdam Institute for Climate Impact Research, RD 4, Potsdam, Germany (jasper.franke@pik-potsdam.de), (2) Humboldt University of Berlin, Berlin, Germany

Estimating correlations between paleoclimate time series is complicated by a multiplicity of typical features, including irregular sampling, age model uncertainty (errors due to interpolation between radiocarbon sampling points) and time uncertainty (uncertainty in calibration), which - taken together - result in unequal observation times of the individual time series to be correlated. Specifically, the concept of correlations is based on the joint (bivariate) distribution of the observed data, and most existing estimators accordingly require concurrent observations. As a consequence, in the case of unequal sampling times, the key challenge is obtaining a feasible approximation of the joint distribution from the given data. In this context, a variety of methods have been proposed, most of which rely either on interpolation or temporal downsampling.

Here, we investigate and thoroughly compare the performance of some of the most popular existing correlation estimators using synthetic paleoclimate data. For this purpose, we construct pseudoproxies representing typical features of marine sediment records like irregular sampling, age model and time uncertainty. We employ a Bayesian approach, which has several advantages in comparison with classical point estimators. Using appropriate statistical models and prior distributions, we implicitly include multiple sources of uncertainty, most of which would have to be treated explicitly in a classical approach. In addition to estimating lag-zero correlations, we use this probabilistic approach to estimate the magnitude of a possible time shift between two series that maximizes their mutual correlation, representing the problem of inferring leads and/or lags between the variability recorded by different paleoclimate archives. Finally, we complement our pseudoproxy experiments by applying the same methodology to a pair of marine benthic d18O records from the Atlantic Ocean.

In general, we find that methods based upon interpolation commonly yield better results in terms of precision and accuracy than such which reduce the number of observations by downsampling. In all cases, the specific characteristics of the studied time series are, however, more important than the choice of a particular interpolation method. Relevant features include the number of observations, the persistence of each record, and the imposed coupling strength between the paired series.

It is important to note that persistence is necessary to be able to correlate paleoclimate time series, as its presence is necessary for obtaining meaningful approximations of the joint distributions from observed data. On the other hand, for the appropriate identification of mutually lagged statistical relationships, persistence causes additional challenges, since it increases the range of possible lags that are identified as yielding significant correlations and, thus, results in an increasing uncertainty of the estimate of the optimum lag. In order to quantify this effect, we apply different measures of dispersion to characterize the level of confidence one can put into a specific time lag estimate. In most of our pseudoproxy experiments, the errors resulting from time uncertainty are much smaller than that due to unequal sampling and age model uncertainty. Thus, it can be reasonable to rely on published time scales as long as calibration uncertainties are not known.