Geophysical Research Abstracts
Vol. 14, EGU2012-5752, 2012
EGU General Assembly 2012

Pearson correlation estimation for irregularly sampled time series

K. Rehfeld (1,2), N. Marwan (1), J. Heitzig (1), J. Kurths (1,2)

(1) Potsdam-Institute for Climate Impact Research, Transdisciplinary Concepts & Methods, Potsdam, Germany (rehfeld@pik-potsdam.de), (2) Department of Physics, Humboldt University Berlin, Berlin, Germany

Many applications in the geosciences call for the joint and objective analysis of irregular time series. For automated processing, robust measures of linear and nonlinear association are needed. Up to now, the standard approach would have been to reconstruct the time series on a regular grid, using linear or spline interpolation. Interpolation, however, comes with systematic side-effects, as it increases the auto-correlation in the time series.

We have searched for the best method to estimate Pearson correlation for irregular time series, i.e. the one with the lowest estimation bias and variance. We adapted a kernel-based approach, using Gaussian weights. Pearson correlation is calculated, in principle, as a mean over products of previously centralized observations. In the regularly sampled case, observations in both time series were observed at the same time and thus the allocation of measurement values into pairs of products is straightforward. In the irregularly sampled case, however, measurements were not necessarily observed at the same time. Now, the key idea of the kernel-based method is to calculate weighted means of products, with the weight depending on the time separation between the observations. If the lagged correlation function is desired, the weights depend on the absolute difference between observation time separation and the estimation lag.

To assess the applicability of the approach we used extensive simulations to determine the extent of interpolation side-effects with increasing irregularity of time series. We compared different approaches, based on (linear) interpolation, the Lomb-Scargle Fourier Transform, the sinc kernel and the Gaussian kernel. We investigated the role of kernel bandwidth and signal-to-noise ratio in the simulations. We found that the Gaussian kernel approach offers significant advantages and low Root-Mean Square Errors for regular, slightly irregular and very irregular time series. We therefore conclude that it is a good (linear) similarity measure that is appropriate for irregular time series with skewed inter-sampling time distributions.