Geophysical Research Abstracts
Vol. 14, EGU2012-5737, 2012
EGU General Assembly 2012

Mutual information 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

For the automated, objective and joint analysis of time series, similarity measures are crucial. Used in the analysis of climate records, they allow for a complimentary, unbiased view onto sparse datasets. The irregular sampling of many of these time series, however, makes it necessary to either perform signal reconstruction (e.g. interpolation) or to develop and use adapted measures. Standard linear interpolation comes with an inevitable loss of information and bias effects. We have recently developed a Gaussian kernel-based correlation algorithm with which the interpolation error can be substantially lowered, but this would not work should the functional relationship in a bivariate setting be non-linear.

We therefore propose an algorithm to estimate lagged auto and cross mutual information from irregularly sampled time series. We have extended the standard and adaptive binning histogram estimators and use Gaussian distributed weights in the estimation of the (joint) probabilities. To test our method we have simulated linear and nonlinear auto-regressive processes with Gamma-distributed inter-sampling intervals. We have then performed a sensitivity analysis for the estimation of actual coupling length, the lag of coupling and the decorrelation time in the synthetic time series and contrast our results to the performance of a signal reconstruction scheme.

Finally we applied our estimator to speleothem records. We compare the estimated memory (or decorrelation time) to that from a least-squares estimator based on fitting an auto-regressive process of order 1. The calculated (cross) mutual information results are compared for the different estimators (standard or adaptive binning) and contrasted with results from signal reconstruction.

We find that the kernel-based estimator has a significantly lower root mean square error and less systematic sampling bias than the interpolation-based method. It is possible that these encouraging results could be further improved by using non-histogram mutual information estimators, like k-Nearest Neighbor or Kernel-Density estimators, but for short (<1000 points) and irregularly sampled datasets the proposed algorithm is already a great improvement.