Geophysical Research Abstracts,
Vol. 11, EGU2009-3081, 2009
EGU General Assembly 2009

Time series segmentation: a new approach based on Genetic Algorithm and Hidden Markov Model

A. Toreti (1,2,3), F.G. Kuglitsch (1,3), E. Xoplaki (1,3,4), J. Luterbacher (1,3,5)

(1) Oeschger Centre for Climate Change Research (OCCR), University of Bern, Bern, Switzerland, (2) Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Rome, Italy, (3) Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland, (4) The Cyprus Institute, EEWRC, Nicosia, Cyprus, (5) Institute of Geography, Climatology and Climate Change Research Group, University of Giessen, Germany

The subdivision of a time series into homogeneous segments has been performed using various methods applied to different disciplines. In climatology, for example, it is accompanied by the well-known homogenization problem and the detection of artificial change points.
In this context, we present a new method (GAMM) based on Hidden Markov Model (HMM) and Genetic Algorithm (GA), applicable to series of independent observations (and easily adaptable to autoregressive processes). A left-to-right hidden Markov model, estimating the parameters and the best-state sequence, respectively, with the Baum-Welch and Viterbi algorithms, was applied. In order to avoid the well-known dependence of the Baum-Welch algorithm on the initial condition, a Genetic Algorithm was developed. This algorithm is characterized by mutation, elitism and a crossover procedure implemented with some restrictive rules. Moreover the function to be minimized was derived following the approach of Kehagias (2004), i.e. it is the so-called complete log-likelihood. The number of states was determined applying a two-fold cross-validation procedure (Celeux and Durand, 2008). Being aware that the last issue is complex, and it influences all the analysis, a Multi Response Permutation Procedure (MRPP; Mielke et al., 1981) was inserted. It tests the model with K+1 states (where K is the state number of the best model) if its likelihood is close to K-state model.
Finally, an evaluation of the GAMM performances, applied as a break detection method in the field of climate time series homogenization, is shown.

1. G. Celeux and J.B. Durand, Comput Stat 2008.
2. A. Kehagias, Stoch Envir Res 2004.
3. P.W. Mielke, K.J. Berry, G.W. Brier, Monthly Wea Rev 1981.