Geophysical Research Abstracts,
Vol. 11, EGU2009-10930, 2009
EGU General Assembly 2009

Bayesian multiple change-points and segmentation: application to homogenization of climatic series homogenization of climatic series

A. Hannart (1) and P. Naveau (2)

(1) LOCEAN, Institut Pierre-Simon Laplace, CNRS/IRD, France (alexis.hannart@locean-ipsl.upmc.fr), (2) LSCE, Institut Pierre-Simon Laplace, CNRS/CEA, France (philippe.naveau@lsce.ipsl.fr)

We describe a new multiple change-point detection technique based on segmenting the time series under study into subsequences. These segments correspond to the episodes that are likely to contain at most a unique jump. They are found by applying Bayesian decision theory through the minimization of simple cost functions. To perform this task, we recall that a change-point is universally defined as an *abrupt* shift. In our view, this characterization makes change-point a *local* concept and problem. Abrupt changes are only apparent with respect to their immediate surroundings. By contrast, values that are remote from the change-points are irrelevant to the detection process. In this context, our main assumption is that the whole time series does not have to be treated globally but can be segmented into shorter subsequences that may contain one change-point and can be treated with a single change-point algorithm. Concerning the computational cost, we no longer need to work with a complex multiple change-point model and a high number of interrelated parameters, leading to inextricable inference procedures – especially in the Bayesian context. Rather, two tools are required: a criterion capable of quickly quantifying the amount of evidence in favour of the existence of a single change-point in a particular subsequence and a fast and single change-point model to infer change-point characteristics in each subsequence. This plan can be implemented because basic Bayesian single change-points with explicit solutions are already available (Lee and Heghinian, 1977) and can be modified in a decision and cost minimization framework. As a result of this simplifying scheme, calculations can be performed explicitly, without falling back on MCMC methods and resulting in particularly light implementation. Through prior distributions derived from a stochastic renewal process description of jump occurrences, past knowledge on jump amplitude and frequency is also introduced in our decision process. Results on simulated series lead to improved speeds and inferences compared to past penalized likelihood methods. At a low computational cost, the method therefore benefits from the strengths of the Bayesian framework, which mainly consist in introducing expert knowledge through prior distributions, and in quantifying the uncertainty on change-points characteristics through posterior distributions. In applications to homogenization, those benefits could be leveraged in two foreseeable ways. Posteriors of jumps position can be found useful to help objectivize a decision on the existence of jumps when they are detected simultaneously on multiple series of pairwise station comparison, a process which is currently performed visually. Also, joint posteriors of jumps position and amplitude can be used to derive confidence intervals on the corrected series, and to quantify the uncertainty introduced by homogenization in climatic trends further obtained.