

PARIS, September 28, 2010

Scientific Report PARIS EM (AgroParisTech)

COST Action ES0601 : “Advances in homogenisation methods of climate series: an integrated approach (HOME)”

The purpose of this meeting is to study applicability of Mixed Linear Models to segmentation of climate and GPS series. Olivier Mestre presents the climate problem at hand, and shows the two options to be tested for detection in the future software:

- pairwise comparisons, that requires an additional attribution step. A clue there would be an improved version of Menne & Williams algorithm, modified according to Hannart bayesian probabilities.
- Perform a joint segmentation of climate series, using DP algorithms recently published by Picard, Lebarbier & Robin.

Participants then discuss similarities between DNA and climate segmentation. It appears that both problems are very close. Franck Picard, Emilie Lebarbier and Stéphane Robin have developed a new package R, CGHseg, that allows perform joint detection in multiple series. This package has not been released yet, but they agree to provide a beta version, for experimentation on the benchmark by Olivier Mestre. A modified Zhang and Siegmund criterion is used to choose model dimensions; as is, this technique requires no input from the user, a highly desirable feature.

According to Stéphane Robin and Emilie Lebarbier, climate effect modelled as a fixed effect (regression) and as a random effect (MLM) are fairly equivalent, since MLM can be seen as a regularized regression technique (ridge regression). CGHseg handles the fixed effects case, and Emilie Lebarbier agrees to provide additional sources for MLM so that both techniques can be compared.

Further more, this package allows for covariates. The “UBRIS” correction model, designed for urban effect, may then be directly implemented using this package. But the ACMANT ideas of joint segmentation of seasonal series are not directly implementable in CGHclust.

Possibility to include spatial dependency in the model is then discussed. According to Stéphane Robin and Emilie Lebarbier, DP algorithms should be able to handle this extra complexity, but this feature requires more development, as well as model the spatial dependency, in order to reduce the number of involved parameters (a suggestion already made during July EM in Toulouse).

This afternoon session is dedicated to segmentation on GPS data series, in connexion with COST ES0701. Unfortunately, Simon Williams (ES0701 WG leader) could not attend, due to late organization of this EM.

GPS series, used to measure the post glacial rebound, also face instrumental problems, directly influencing the observations. But unlike climate series, observed noises in those series have the characteristics of a mixture of white noise and flicker noise, which requires adaptation of techniques.

Xavier Collilleux presents an overview of characteristics of GPS series, as well as standard techniques used in this field: maximum likelihood estimation, and iterative least squares.

Discussion follows around shape of GPS series, that look like linear trends with constant coefficient, affected by abrupt changes. The most problematic phenomenae are the autoregressive properties of noise, that theoretically destroys optimality properties of DP algorithm. Robustness to autocorrelation will be first tested using CGHclust package, on data series provided by Xavier Collilleux.

Emilie Lebarbier, Stéphane Robin, Olivier Mestre, Xavier Collilleux.