



HAL
open science

Homogénéisation de séries climatiques

Olivier Mestre, Victor Venema

► **To cite this version:**

Olivier Mestre, Victor Venema. Homogénéisation de séries climatiques. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494708

HAL Id: inria-00494708

<https://hal.inria.fr/inria-00494708>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HOMOGENEISATION DE SERIES CLIMATIQUES

Olivier Mestre¹, Victor Venema², & groupe COST-ES0601 "HOME"³

¹Ecole Nationale de la Météorologie, 42, avenue Coriolis, 31057 Toulouse cedex, France

²University of Bonn, Auf dem Huegel 20, 53121 Bonn, Allemagne

³ groupe COST-ES0601 composé de: Enric Aguilar, Ingeborg Auer, José A. Guijarro, Petr Stepanek, Claude.N.Williams, Matthew Menne, Emilie Lebarbier, Franck Picard, Alexis Hannart, Peter Domonkos, Julien Viarre, Tamas Szentimrey, Monika Lakatos, Dubravka Rasol, Elke Rustemeyer, Gregor Vertacnik, Kostas Kolokythas, Tania Marinova, Fiorella Acquavotta, Lars Andersen, Tanja Likso, Matija Klancar, Michele Brunetti, Christine Gruber, Marc Prohom Duran, Theo Brandsma.

Mots clef : Environnement, processus qualité-fiabilité

Résumé

Dans un grand nombre d'endroits, on peut constituer des séries d'observations climatologiques relativement complètes, remontant au XIXème siècle. Cependant, les conditions de mesure ont été profondément modifiées au cours du temps. Les changements d'emplacement, d'instrumentation... se traduisent par autant de biais dans les séries de données. Or ces ruptures artificielles peuvent être du même ordre de grandeur que les phénomènes climatiques que l'on étudie. Leur détection et leur correction sont donc indispensables avant toute étude climatique : c'est ce que l'on appelle l'homogénéisation des séries. Le problème se traite en deux phases : détection des ruptures, et correction, pour lesquelles toute une variété de méthodes existent.

On présentera les résultats de l'Action Européenne COST ES0601 : intercomparaison des méthodes d'homogénéisation des séries de données, qui permet une synthèse des meilleurs aspects des différentes procédures.

Abstract

Long instrumental climate records are the basis of climate research. However, these series are usually affected by inhomogeneities (artificial shifts), due to changes in the measurement conditions (relocations, instrumentation and others). As the artificial shifts often have the same magnitude as the climate signal, such as long-term variations, trends or cycles, a direct analysis of the raw data series can lead to wrong conclusions about climate change. In order to deal with this crucial problem many statistical homogenisation procedures have been developed for detection and correction of these inhomogeneities.

We present here the results of European COST Action ES0601, whose objectives are to compare existing procedures and achieve a general method for homogenising climate and environmental datasets.

Dans un grand nombre d'endroits, on peut constituer des séries d'observations climatologiques relativement complètes, remontant au XIX^{ème} siècle. Cependant, les conditions de mesure ont été profondément modifiées au cours du temps. Les changements d'emplacement, d'instrumentation... se traduisent par autant de biais dans les séries de données. Or ces ruptures artificielles peuvent être du même ordre de grandeur que les phénomènes climatiques que l'on étudie. Leur détection et leur correction sont donc indispensables avant toute étude climatique : c'est ce que l'on appelle l'homogénéisation des séries.

Un grand nombre de méthodes ont été développées pour traiter ce problème. Ces méthodes reposent la plupart sur le principe d'homogénéité relative [1].

Dans les tests d'homogénéité classiques, on étudie les propriétés des séries de différences entre séries proches et bien corrélées, ce qui permet de supprimer l'essentiel du signal climatique: on met ainsi en évidence les changements artificiels dans les séries. Généralement, ces séries de comparaison ont les propriétés d'un échantillon gaussien présentant des ruptures au niveau de la moyenne.

Un grand nombre de procédures de segmentation existent, des plus simples, dérivées du test de student ou de tests non-paramétriques, et appliquées itérativement, aux plus récentes, reposant sur divers critères : BIC [2], BIC modifié [13], Caussinus & Lyazrhi [3], pénalisations adaptatives [8,9], MDL [5], etc...

De même, la correction des séries peut se faire simplement en utilisant les différences de moyennes, ou en employant des procédures multivariées.

Le but de l'Action Européenne COST ES0601 "HOME" est l'intercomparaison des méthodes utilisées pour homogénéiser les séries climatiques.

Les performances de 25 procédures de détection, largement employées en climat, économétrie, biostatistique, traitement du signal sont étudiées par simulations, en termes de puissance, niveau, robustesse (autocorrélation, dissymétrie). En fonction du rapport nombre de sauts sur taille des séries, les résultats peuvent être surprenants.

Les procédures d'homogénéisation les plus utilisées en climatologie sont également testées : SNHT [1], MASH [11], Rhtest [12], PRODIGE [4], etc...

Pour cela, des séries climatiques sont générées en reproduisant les corrélations spatiales et temporelles de séries réelles, puis des ruptures sont insérées.

Les participants à l'expérience (le groupe "HOME") ont homogénéisé "à l'aveugle" les séries perturbées, qui sont ensuite comparées aux séries originales.

On montre que si ces procédures permettent d'améliorer la qualité des données, certaines sont beaucoup plus efficaces que d'autres. On montrera les résultats sur divers scores classiques (RMSE) ou en termes de reconstitution de tendances à long terme.

L'analyse des résultats permet de proposer une méthode "universelle", synthèse des meilleurs aspects des différentes méthodes, qui sera présentée, ainsi que des recommandations d'ordre pratique.

Un problème subsiste toutefois, celui de l'automatisation des procédures, car une part d'expertise humaine reste nécessaire pour le moment.

Bibliographie

- [1] Alexandersson, H. (1986) *A homogeneity test applied to precipitation data*, Int. J. of climatol., 6, 661-675
- [2] Braun J.V., Braun R.K. and Müller H.G. (2000). *Multiple change-point fitting via quasilikelihood, with application to DNA sequence segmentation*. Biometrika 87, 301-314.
- [3] Caussinus H. and Lyazghi F. 1997. *Choosing a linear model with a random number of change-points and outliers*. Ann. Inst. Statist. Math. 49, n°4, 761-775.
- [4] Caussinus H. and Mestre O. 2004. *Detection and correction of artificial shifts in climate series*. Applied Statistics 53, part 3: 405-425.
- [5] Davis, R., Lee, T. C., Rodriguez-Yam, G (2008) *Break Detection for a Class of Nonlinear Time Series Models*. J. of Time Ser. Anal., 29, 5, 834-867
- [6] Easterling D. R. & Peterson D. C. (1995) *A new method for detecting undocumented discontinuities in climatological time series*. Int. J. of Climatol., Vol. 15, 369-377
- [7] Hawkins D.M. 2001. *Fitting multiple change-points to data*. Comput. Statist. Data Anal., 37: 323-341.
- [8] Lavielle, M. and Moulines, E. (2000). *Least squares estimation of an unknown number of shifts in a time series*. Journal of Time series analysis 21, 33-59.
- [9] Lebarbier, E. (2005). *Detecting multiple change-points in the mean of Gaussian process by model selection*. Signal Processing 85, 717-736.
- [10] Picard, F., Robin, S., Lavielle, M., Vaisse, C. and Daudin, J.-J. (2005). *A statistical approach for CGH microarray data analysis*. BMC Bioinformatics 6, 27.
- [11] Szentimrey T. (1996) *Statistical procedure for joint homogenization of climatic time series*. Proc. Seminar on Homogenization of surf. Instrum. Data, Budapest, 47-62
- [12] Wang, X. L., Wen Q. H., Wu Y. (2007) *Penalized maximal t test for detecting undocumented mean change in climate data series*. J. Appl. Meteor. Climatol., 46 (No. 6), 916-931.
- [13] Zhang, N. R. et Siegmund, D. O. (2006) *A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data*. Biometrics 63, 22-32.

<http://www.homogenisation.org/>

<http://www.meteo.uni-bonn.de/venema/themes/homogenisation/>