

## A blind test of monthly homogenisation algorithms

V.K.C. Venema (1), O. Mestre (2), and the COST-HOME Team

(1) University of Bonn, Meteorological institute, Bonn, Germany (victor.venema@uni-bonn.de), (2) Météo France, Ecole Nationale de la Météorologie, Toulouse, France (olivier.mestre@meteo.fr)

As part of the COST Action HOME (Advances in homogenisation methods of climate series: an integrated approach) a dataset was generated that serves as a benchmark for homogenisation algorithms. This talk will shortly describe this benchmark dataset and focus on the results and lessons learned.

Based upon a survey among homogenisation experts we chose to work with monthly values for temperature and precipitation. Temperature and precipitation were selected because most participants consider these elements the most relevant for their studies. Furthermore, they represent two important types of statistics (additive and multiplicative).

The benchmark has three different types of datasets: real data, surrogate data and synthetic data. The real datasets allow comparing the different homogenisation methods with the most realistic type of data and inhomogeneities. Thus this part of the benchmark is important for a faithful comparison of algorithms with each other. However, as in this case the truth is not known, it is not possible to quantify the improvements due to homogenisation. Therefore, the benchmark also has two datasets with artificial data to which we inserted known inhomogeneities: surrogate and synthetic data. By comparing the statistical properties of the detected inhomogeneities in the real dataset and in the two artificial ones, we can also study how realistic the inserted inhomogeneities are.

The aim of surrogate data is to reproduce the structure of measured data accurately enough that it can be used as substitute for measurements. The surrogate climate networks have the spatial and temporal auto- and cross-correlation functions of real homogenised networks as well as the exact (non-Gaussian) distribution for each station.

The idealised synthetic data is based on the surrogate networks. The change is that the difference between the stations has been modelled as uncorrelated Gaussian white noise. The idealised dataset is valuable because its statistical characteristics are assumed in most homogenisation algorithms and Gaussian white noise is the signal most used for testing the algorithms. The presentation will focus on the results of the more realistic surrogate data.

The surrogate and synthetic data represent homogeneous climate data. To this data inhomogeneities are added: outliers, as well as breaks and local trends. Breaks are either introduced randomly or simultaneously in a fraction of the stations. Furthermore, missing data values are simulated and a random global (network wide) trend is added.

The participants have returned 25 blind contributions, as well as 22 further contributions, which were submitted after revealing the truth.

The quality of the homogenised data is assessed by a number of metrics: the root mean square error, the error in (linear and nonlinear) trend estimates and contingency scores. The metrics are computed on the station data and the network average regional climate signal, as well as on monthly data and yearly data, for both temperature and precipitation. The performance of the contributions depends significantly on the error metric considered. Still a group of better algorithms can be found, which includes Craddock, PRODIGE, MASH,

ACMANT and USHCN. Clearly algorithms developed for solving the multiple breakpoint problem with an inhomogeneous reference perform best; splitting and semi-hierarchical algorithms should be considered outdated. The results suggest that the correction algorithms are currently an important weakness of many methods.

For more information on the COST Action on homogenisation see:

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