Testing of homogenisation methods: the use of simulated data sets with varied statistical characteristics of inhomogeneities

P. Domonkos

dpeterfree@freemail.hu / Phone: +36-209911774

Effectiveness of seventeen versions of 11 objective homogenisation methods are under examination in this study. All the methods have commonly been used in climatology (t-test, Buishand-test, Standard Normal Homogeneity Test, Wilcoxon Rank Sum Test, Multiple Linear Regression, Bayesian test, Pettitt test, Mann-Kendall test, method of Mestre, Easterling-Peterson test and Multiple Analysis of Series for Homogenisation). Simulated data sets are used for characterising and evaluating the differences between detectable and factual inhomogeneities. Before the starting of homogenisation, relative time series were created applying Alexandersson's guides. In processing the first simulated data set a principle was to approach the statistical properties to those of the relative time series from an observational data base, as closely, as it is possible. The compared properties include the statistical characteristics of the detected inhomogeneities obtained from the application of the 17 homogenisation methods. The aim of this step is to ensure a kind of compatibility between factual and simulated data sets, since otherwise the test-results might lead to false conclusions concerning the practical applications in real time series of observations. The used observational database comprises 215 temperature and 112 precipitation time series, they are originated mostly from the observing network of the Hungarian Meteorological Service.

After the first simulated data set was created, further data sets are simulated with predefined differences in the statistical properties, e.g. mean frequency or mean magnitude of change-points, frequency of short-term deviations, etc. In this way the representation of the test-results is also assessable.

A distinct problem is to find a proper measure for evaluating efficiencies of different homogenisation methods. The recognition of change-points, and a satisfactorily accurate determination of their position (timing) and magnitude are generally expected, but other expectations may also realistic: rate of false detection must be low, and real trend- and variability-characteristics of the time series must be well producible from the homogenised series. I introduce two measures for evaluating efficiencies, one is for evaluating the reliability of change-point recognition only, while the other one is a more complex index.

The presentation of the study will show results and conclusions those can be drawn from the examination of the first few simulated data sets.