Homogenization in mean and standard deviation (MASHv4.01)

Beatrix Izsák and Tamás Szentimrey

HungaroMet, Climate Research Unit, Budapest, Hungary (izsak.b@met.hu)
Varimax Limited Partnership, Budapest, Hungary

In essence the theme of homogenization can be divided into two subgroups, such as monthly and daily data series homogenization. These subjects are in strong connection with each other of course, for example the monthly results can be used for the homogenization of daily data. The earlier versions of our method MASH (Multiple Analysis of Series for Homogenization; Szentimrey) were developed for homogenization of the daily and monthly data series in the mean i.e. the first order moment. The software MASH was developed as an interactive automatic, artificial intelligence (AI) system that simulates the human intelligence and mimics the human analysis on the basis of advanced mathematics. The new version MASHv4.01 is able to homogenize also the standard deviation i.e. the second order moment. We remark if the data are normally distributed (e.g. mean temperature) then the homogenization of mean and standard deviation is sufficient since in case of normal distribution if the first two moments are homogenous then the higher order moments are also homogeneous.

In our presentation, we present the application of the MASH4 software on real climate data. From 1 January 1901 to 31 December 2023, the daily mean temperature data series of 27 stations in Hungary were homogenized in both first and second moments. We present our results based on verification statistics. The developed automatic verification procedure of MASH makes it possible to test the homogenized time series systematically in order to clear up the uncertainty. The basic concept of the verification procedure is that confidence in the homogenized series may be increased by the joint comparative examination of the original and the homogenized series systems. The comparison is based on such adequate questions, which can also be formulated mathematically, consequently, a programmed statistical test procedure can be obtained to evaluate the quality of the homogenized series. The questions are related to the estimated inhomogeneity of the series before and after homogenization the comparison of the modification measure of the series with the estimated inhomogeneity of the original series, the representativity of the given station network in respect of the homogenization, the relation between the estimated inhomogeneities and the metadata.

The research presented in the article was carried out within the framework of the Széchenyi Plan Plus program with the support of the RRF 2.3.1 21 2022 00008 project.