EGU21-6680
https://doi.org/10.5194/egusphere-egu21-6680
EGU General Assembly 2021

A validation scheme for homogenization techniques on a Swedish temperature network using artificial inhomogeneities (1950-2005)

Giulio Nils Caroletti[1], Tommaso Caloiero[2], Magnus Joelsson[3], and Roberto Coscarelli[4]
[1]MIUR, Italy (iuliusscaevola@yahoo.it)
[2]Institute for Agricultural and Forest Systems in the Mediterranean, National Research Council of Italy, Rende (CS), Italy (tommaso.caloiero@isafom.cnr.it)
[3]SMHI, Information och statistik, Norrköping, Sweden (magnus.joelsson@smhi.se)
[4]Research Institute for Geo-hydrological Protection , National Research Council of Italy, Rende (CS), Italy (roberto.coscarelli@irpi.cnr.it)

Homogenization techniques and missing value reconstruction have grown in importance in climatology given their relevance in establishing coherent data records over which climate signals can be correctly attributed, discarding apparent changes depending on instrument inhomogeneities, e.g., change in instrumentation, location, time of measurement.

However, it is not generally possible to assess homogenized results directly, as true data values are not known. Thus, to validate homogenization techniques, artificially inhomogeneous datasets, also called benchmark datasets, are created from known homogeneous datasets. Results from their homogenization can be assessed and used to rank, evaluate and/or validate techniques used.

Considering temperature data, the aims of this work are: i) to determine which metrics (bias, absolute error, factor of exceedance, root mean squared error, and Pearson's correlation coefficient) can be meaningfully used to validate the best-performing homogenization technique in a region; ii) to evaluate through a Pearson correlation analysis if homogenization techniques' performance depends on physical features of a station (i.e., latitude, altitude and distance from the sea) or on the nature of the inhomogeneities (i.e., the number of break points and missing data).

With this aims, a southern Sweden temperature database with homogeneous, maximum and minimum temperature data from 100 ground stations over the period 1950-2005 has been used. Starting from these data, inhomogeneous datasets were created introducing up to 7 artificial breaks for each ground station and an average of 107 missing data. Then, 3 homogenization techniques were applied, ACMANT (Adapted Caussinus-Mestre Algorithm for Networks of Temperature series), and two versions of HOMER (HOMogenization software in R): the standard, automated setup mode (Standard-HOMER) and a manual setup developed and performed at the Swedish Meteorological and Hydrological Institute (SMHI-HOMER).

Results showed that root mean square error, absolute bias and factor of exceedance were the most useful metrics to evaluate improvements in the homogenized datasets: for instance, RMSE for both variables was reduced from an average of 0.71-0.89K (corrupted dataset) to 0.50-0.60K (Standard-HOMER), 0.51-0.61K (SMHI-HOMER) and 0.46-0.53K (ACMANT), respectively.

Globally, HOMER performed better regarding the factor of exceedance, while ACMANT outperformed it with regard to root mean square error and absolute error. Regardless of the technique used, the homogenization quality anti-correlated meaningfully to the number of breaks. Missing data did not seem to have an impact on HOMER, while it negatively affected ACMANT, because this method does not fill-in missing data in the same drastic way.

In general, the nature of the datasets had a more important role in yielding good homogenization results than associated physical parameters: only for minimum temperature, distance from the sea and altitude showed a weak but significant correlation with the factor of exceedance and the root mean square error.

This study has been performed within the INDECIS Project, that is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).