

EGU2020-5365

<https://doi.org/10.5194/egusphere-egu2020-5365>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Uncertainty of Climatol adjustment algorithm for daily time series of additive climate variables

Oleg Skrynyk^{1,2}, Enric Aguilar¹, José A. Guijarro³, and Sergiy Bubin⁴

¹Universitat Rovira i Virgili, Center for Climate Change (C3), Vila-seca, Spain (oleg.skrynyk@urv.cat, enric.aguilar@urv.cat)

²Ukrainian Hydrometeorological Institute, Kyiv, Ukraine

³AEMET (State Meteorological Agency), Palma de Mallorca office, Spain (jguijarro@aemet.es)

⁴Department of Physics, Nazarbayev University, Nur-Sultan, Kazakhstan (sergiy.bubin@nu.edu.kz)

Before using climatological time series in research studies, it is necessary to perform their quality control and homogenization in order to remove possible artefacts (inhomogeneities) usually present in the raw data sets. In the vast majority of cases, the homogenization procedure allows to improve the consistency of the data, which then can be verified by means of the statistical comparison of the raw and homogenized time series. However, a new question then arises: how far are the homogenized data from the true climate signal or, in other words, what errors could still be present in homogenized data?

The main objective of our work is to estimate the uncertainty produced by the adjustment algorithm of the widely used Climatol homogenization software when homogenizing daily time series of the additive climate variables. We focused our efforts on the minimum and maximum air temperature. In order to achieve our goal we used a benchmark data set created by the INDECIS* project. The benchmark contains clean data, extracted from an output of the Royal Netherlands Meteorological Institute Regional Atmospheric Climate Model (version 2) driven by Hadley Global Environment Model 2 - Earth System, and inhomogeneous data, created by introducing realistic breaks and errors.

The statistical evaluation of discrepancies between the homogenized (by means of Climatol with predefined break points) and clean data sets was performed using both a set of standard parameters and a metrics introduced in our work. All metrics used clearly identifies the main features of errors (systematic and random) present in the homogenized time series. We calculated the metrics for every time series (only over adjusted segments) as well as their averaged values as measures of uncertainties in the whole data set.

In order to determine how the two key parameters of the raw data collection, namely the length of time series and station density, influence the calculated measures of the adjustment error we gradually decreased the length of the period and number of stations in the area under study. The total number of cases considered was 56, including 7 time periods (1950-2005, 1954-2005, ..., 1974-2005) and 8 different quantities of stations (100, 90, ..., 30). Additionally, in order to find out how stable are the calculated metrics for each of the 56 cases and determine their confidence

intervals we performed 100 random permutations in the introduced inhomogeneity time series and repeated our calculations. With that the total number of homogenization exercises performed was 5600 for each of two climate variables.

Lastly, the calculated metrics were compared with the corresponding values, obtained for raw time series. The comparison showed some substantial improvement of the metric values after homogenization in each of the 56 cases considered (for the both variables).

*INDECIS is a 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). The work has been partially supported by the Ministry of Education and Science of Kazakhstan (Grant BR05236454) and Nazarbayev University (Grant 090118FD5345).