EMS Annual Meeting Abstracts
Vol. 15, EMS2018-346, 2018

New data quality control tools for operational use in ProClimDB software

Petr Štěpánek (1,2), Pavel Zahradníček (1,2), and Jan Meitner (1)

(1) Global Change Research Institute CAS, Brno, Czech Republic (zahradnicek.p@czechglobe.cz), (2) Czech Hydrometeorological Institute, Brno, Czech Republic

There is a lack of a generally accepted methodology for data quality control (contrary to homogenization for which such methods are developed). This is due to a fact that a huge variety of reasons exists leading to flawed data. However, without outliers being treated properly, homogenization and analysis may produce misleading results.

The ProClimDB software (www.climahom.eu) has been tested and applied in various regions throughout the world in the past 10 years, solving problems not only in the most common meteorological elements (air temperature and precipitation), but also other characteristics like wind speed, relative humidity and sunshine duration. In order to apply ProClimDB to a large (for example pan-European) dataset efficiently and to be able to share this software with other users, the important goal was set to re-program the original method from ProClimDB software (based on database language Visual FoxPro), into a modern and open software (R).

One of the changes is in a new selection of reference series. Proper selection of reference stations for QC is a crucial issue. The software chooses the neighbour stations automatically, based on the criteria set (correlation coefficient, distance, difference in altitude, length of time series, number of neighbours etc.) but a user is still able to further manually manage (edit) the selection. Other changes are in improved error detection statistics. This corresponds to the experiences with testing large number of different datasets that had different types of errors (duplicated stations, repeated values, zeros instead missing value and typical errors for each element). Tables with statistics are accompanied with various plots to help a user quickly evaluate found errors and take correspondent decision. Further QC outputs are given also in a form of interpolated maps, where the plotted residuals help to better conceive spatial relationships.

This method was evaluated within Copernicus project C3S.311a.Lot.4 on real and surrogate datasets of the maximum and minimum temperature with known errors. The software gave good results, when it showed great success in right detections with a small percentage of falls alarms.