Dealing with incomplete and inhomogeneous data in deriving climatological normals for the 1981-2010 period in Croatia

Irena Nimac, Melita Perčec Tadić , and Dubravka Rasol
Meteorological and Hydrological Service of Croatia, Zagreb, Croatia

Calculating climatological normals is one of the most important tasks for national meteorological services. Due to requirements regarding homogeneity and completeness of data record, "data cleaning" is necessary since missing or inhomogeneous data can significantly alter the estimated statistical properties of the whole series. Lately, various techniques and methods are used for dealing with problem of inhomogeneity. One of such tools for homogenization and interpolation is package Climatol developed in the framework of the R (r-project.org). Data used in this study are mean monthly temperature data series from 38 meteorological stations in Croatia divided into three different climate regions (continental, mountain and coastal). Missing data problem can be diminished or overcome if they are replaced with some statistically justified estimates. Therefore, gaps were filled here by random selection from corresponding distribution determined from observed data records so multiannual average remains the same. Later, Climatol was applied to the original data and new homogenized dataset is obtained. In order to validate selected method, obtained break-points are compared to the ones determined using ACMANT software. Homogenization process using Climatol results in increase in correlation coefficient amongst stations in certain region and decrease in its variability, i.e. smoothness of the regions is increased. Also, principal component analysis gave increase in fraction of variance explained by first component for every region and all area. Finally, changes in statistical characteristics of temperature field, such as annual and seasonal mean and decadal trend, are determined. Significance of change in mean is tested using Student's test, while for trend Mann-Kendall's test is used. Significant changes in mean are observed only at one station, while trend analysis showed to be more sensitive to inhomogeneities. Not only change in trend rate has occurred but also few stations experienced change in trend significance at annual and seasonal scale.