Geophysical Research Abstracts
Vol. 19, EGU2017-9021, 2017
EGU General Assembly 2017

Infilling and quality checking of discharge, precipitation and temperature data using a copula based approach

Faizan Anwar, András Bárdossy, and Jochen Seidel

Institute for Modelling Hydraulic and Environmental Systems - Dept. of Hydrology and Geohydrology, University of Stuttgart, Germany (faizan.anwar@iws.uni-stuttgart.de)

Estimating missing values in a time series of a hydrological variable is an everyday task for a hydrologist. Existing methods such as inverse distance weighting, multivariate regression, and kriging, though simple to apply, provide no indication of the quality of the estimated value and depend mainly on the values of neighboring stations at a given step in the time series.
Copulas have the advantage of representing the pure dependence structure between two or more variables (given the relationship between them is monotonic). They rid us of questions such as transforming the data before use or calculating functions that model the relationship between the considered variables. A copula-based approach is suggested to infill discharge, precipitation, and temperature data. As a first step the normal copula is used, subsequently, the necessity to use non-normal / non-symmetrical dependence is investigated.
Discharge and temperature are treated as regular continuous variables and can be used without processing for infilling and quality checking. Due to the mixed distribution of precipitation values, it has to be treated differently. This is done by assigning a discrete probability to the zeros and treating the rest as a continuous distribution.
Building on the work of others, along with infilling, the normal copula is also utilized to identify values in a time series that might be erroneous. This is done by treating the available value as missing, infilling it using the normal copula and checking if it lies within a confidence band (5 to 95% in our case) of the obtained conditional distribution.
Hydrological data from two catchments Upper Neckar River (Germany) and Santa River (Peru) are used to demonstrate the application for datasets with different data quality.
The Python code used here is also made available on GitHub. The required input is the time series of a given variable at different stations.