Geophysical Research Abstracts
Vol. 20, EGU2018-1324, 2018
EGU General Assembly 2018

Filling gaps in streamflow data with the Direct Sampling technique

Moctar Dembélé, Grégoire Mariéthoz, and Bettina Schaefli
Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Lausanne, Switzerland (moctar.dembele@unil.ch)

In many regions around the world, observed streamflow time series show too many missing data for a direct use in the context of water resources management, engineering applications or hydrological modelling. Here, we propose the use of the Direct Sampling (DS) method, which was originally designed for the generation of random fields, but sees an increasing number of applications focused on time series. The principle of DS is that it generates new simulated values based on a conditional resampling of a provided training set. In the case of reconstructing gaps in time series, the training set consists of historical data in the same time series during periods not affected by gaps. The method can be extended to multisite time series by considering patterns that span across several neighboring locations. Most importantly, the approach is stochastic and considers short as well as long-range variability of the considered variable. Compared to other methods, one advantage of the DS is that it is almost entirely data-driven, can fit any data structure and simulate the outcome of a complex natural process without assuming a specific statistical model.

This study aims at using the DS technique to fill gaps in daily streamflow data in the Volta river basin (VRB) in West Africa. The area is known to be a data scarce region due to the poor density of gauging stations on the hydrographic network. Fifteen years of daily streamflow data have been collected for the VRB with some stations and years more complete than others. In a first step, different scenarios of missing data with irregular gap sizes, ranging from days to months, are used to assess the performance of the DS technique in reproducing the data patterns.

DS performed very well and even better than linear regression model with high Pearson's correlation coefficient ($r>0.86$) for all scenarios. In scenario 3, the reference and auxiliary stations were poorly correlated ($r=0.43$), and the reference station's time series contained 65% of gaps while the auxiliary station contained 30%. However, DS showed good performance even with low r and gaps in conditioning data. Moreover, there was a good similarity of probability distributions among the simulated and the reference values, which become better at monthly scale and with less variation in the simulated values.

Given the simplicity of the DS technique, the method is readily transferable to any other variable and is thus very promising for environmental modeling in general. The next step is the comparison of DS to other data-driven methods. Once validated, the method can be used to generate continuous streamflow data that will be used for hydrological modelling and water accounting in the VRB.