Geophysical Research Abstracts
Vol. 20, EGU2018-14438-1, 2018
EGU General Assembly 2018

Input variable selection for hydrological predictions in ungauged catchments: with or without clustering?

Nilay Dogulu (1), Inci Batmaz (2), and Elcin Kentel (3)

(1) Middle East Technical University (METU), Civil Engineering Dept., Water Resources Laboratory, Ankara, Turkey (ndogulu@metu.edu.tr), (2) Middle East Technical University (METU), Statistics Dept., Ankara, Turkey (ibatmaz@metu.edu.tr), (3) Middle East Technical University (METU), Civil Engineering Dept., Water Resources Laboratory, Ankara, Turkey (ekentel@metu.edu.tr)

A key step in data-driven environmental modelling, including for hydrological purposes, is input variable selection (IVS) to ensure that the least number of variables with minimum redundancy are used to characterize the inherent relationship between inputs and outputs. Hydrological predictions in ungauged catchments is one such area where the information on influential predictors of runoff signatures guides in understanding dominant controls of meaningful information transfer from gauged to ungauged locations (i.e. regionalization). This understanding is valuable especially for the analysis of hydrological similarity among ungauged catchments, e.g. to identify reference (donor) catchment(s). Large-sample hydrology can help to gain useful insights on how these significant predictors change over different runoff signatures representing particular hydrological conditions as well as within different groups of similar catchments. This study explores the added value of clustering for input variable selection in the case of catchments across continental USA using the CAMELS dataset (Addor et al., 2017). We employ the method of k-means clustering on the input space consisting of topography, soil, geology, vegetation and climate attributes as a way to deal with heterogeneity and complexity in hydrological processes, and thus, aiming to identify similar groups of catchments. The input variables (for predicting selected hydrological attributes) are determined by three IVS filter algorithms (partial mutual information, partial correlation input selection, and iterative input selection), then evaluated and compared for among different clusters and when no clustering method is applied. We present and discuss the results for three hydrological attributes – 95% flow percentile (low flows), mean daily discharge (medium flows), and 5% flow percentile (high flows) – in order to account for variability in hydrological conditions, and to investigate the dominant catchment and/or climate characteristics controlling low/medium/high flow predictability at ungauged locations. The findings of our study have implications for reference (donor) catchment selection and hydrological model independent regionalization (aka hydrostatistical or data-driven methods), and are particularly relevant to understanding hydrological similarity and catchment classification in the absence of local runoff observations.

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017) The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences, 21(10), 5293-5313.