Geophysical Research Abstracts
Vol. 18, EGU2016-87, 2016
EGU General Assembly 2016

Independent variable complexity for regional regression of the flow duration curve in ungauged basins

Geoffrey Fouad, André Skupin, and Allen Hope
Department of Geography, San Diego State University, San Diego, United States (geoffrey.fouad@gmail.com)

The flow duration curve (FDC) is one of the most widely used tools to quantify streamflow. Its percentile flows are often required for water resource applications, but these values must be predicted for ungauged basins with insufficient or no streamflow data. Regional regression is a commonly used approach for predicting percentile flows that involves identifying hydrologic regions and calibrating regression models to each region. The independent variables used to describe the physiographic and climatic setting of the basins are a critical component of regional regression, yet few studies have investigated their effect on resulting predictions.

In this study, the complexity of the independent variables needed for regional regression is investigated. Different levels of variable complexity are applied for a regional regression consisting of 918 basins in the US. Both the hydrologic regions and regression models are determined according to the different sets of variables, and the accuracy of resulting predictions is assessed. The different sets of variables include (1) a simple set of three variables strongly tied to the FDC (mean annual precipitation, potential evapotranspiration, and baseflow index), (2) a traditional set of variables describing the average physiographic and climatic conditions of the basins, and (3) a more complex set of variables extending the traditional variables to include statistics describing the distribution of physiographic data and temporal components of climatic data. The latter set of variables is not typically used in regional regression, and is evaluated for its potential to predict percentile flows.

The simplest set of only three variables performed similarly to the other more complex sets of variables. Traditional variables used to describe climate, topography, and soil offered little more to the predictions, and the experimental set of variables describing the distribution of basin data in more detail did not improve predictions. These results are largely reflective of cross-correlation existing in hydrologic datasets, and highlight the limited predictive power of many traditionally used variables for regional regression. A parsimonious approach including fewer variables chosen based on their connection to streamflow may be more efficient than a data mining approach including many different variables. Future regional regression studies may benefit from having a hydrologic rationale for including different variables and attempting to create new variables related to streamflow.