

A self-organizing map exploratory analysis of the flow duration curve in the United States

*Geoffrey Fouad¹, André Skupin²

1. Monmouth Univ, 2. San Diego State Univ

The flow duration curve (FDC) describes the full range of streamflow magnitude observed at a site, and is strongly influenced by upstream conditions of the basin. Upstream conditions are quantified using basin characteristics, such as mean elevation and annual precipitation. A large variety of data now exists to characterize basins in the United States (US). However, a greater understanding of how this data relates to the FDC is critical, considering basin characteristics are typically the basis for predicting the FDC of ungauged basins. The present study performs an exploratory analysis of the FDC and characteristics of 918 basins in the US using a neural network technique called the self-organizing map (SOM). The SOM is applied for its ability to cluster and visualize fine-scale variation in large datasets. Both of these exploratory frameworks (i.e. clustering and visualization) are used to compare individual flows of the FDC to basin characteristics. Clusters based on common basin characteristics poorly agree with those of the FDC (36% agreement), which is less than prior work in smaller study areas, such as Italy. This is an important point because clusters based on basin characteristics are used to deploy models for predicting the FDC. Basin characteristics primarily cluster basins into geographic regions, whereas the FDC generates clusters of basins distributed throughout the US. Geographic proximity therefore may not be an indicator of similarity in the FDC between basins. Variation of the FDC is also unrelated to some common basin characteristics, such as topographic variables, as indicated through SOM data visualizations. This may partially explain the disagreement between the two sets of clusters. The disagreement may also be because basin characteristics are only associated with certain parts of the FDC, but not the overall FDC. For instance, aridity, an index of precipitation lost to evapotranspiration, suppresses high flows possibly due to lower antecedent moisture conditions that moderate storm flows. High flows are also related to spring snowmelt represented using the percent of precipitation delivered as snow. Another association to a part of the FDC is that average to low flows vary with groundwater contributions (i.e. baseflow). Basin characteristics describing surface runoff are more related to high flows, whereas subsurface drainage has more influence on average to low flows. The processes that generate different flows should be accounted for in the clusters used to predict the FDC, and future research should evaluate if the tradition of using a single set of basin characteristics to cluster basins for predicting the FDC should be revised to select different basin characteristics depending on the flow targeted for prediction.

Keywords: basin hydrology, flow duration curve, self-organizing map, cluster analysis, data visualization, United States