

Clustering patterns in streamflow to produce regionally or anthropogenically similar groups

Karen R. Ryberg^{1,2,*}

1. U.S. Geological Survey

2. North Dakota State University

*Contact author: kryberg@usgs.gov

Keywords: streamflow, environmetrics, clustering, trends, hydrology

Trends in streamflow characteristics (flood peaks, critical low flows, annual-mean flows) are often discussed in various climate change scenarios, whether that climate change is human induced or the result of natural climatic cycles. Increasing or decreasing trends are important for emergency management, agricultural, and municipal supply, as well as other industrial uses. Trends are often summarized graphically with a few sites highlighted in a paper, for example, or summarized in tables of trend values. However, when looking at a large number of sites with varying drainage basin sizes over a large geographic region, the number of sites involved makes these graphical and tabular methods difficult to comprehend and summarize.

R is used here to calculate a number of streamflow-related variables and extract features from the daily streamflow time series of approximately 500 long-term, mostly unregulated, streamgage sites operated by the U.S. Geological Survey. Cluster analysis is then performed to identify regionally similar areas and to identify sites that may be geographically distant but experiencing similar trends, such as a downward trend in streamflow because of large-scale irrigation which depletes groundwater and baseflow to the streams. A number of cluster variables and techniques (including techniques in **cluster** and **pyclust**) are explored and the results compared to existing knowledge of hydrologic trend and regional differences. The clustering is then visualized geospatially (using **maptools** and **rgdal**) with other hydrologic and geographic data. The clustering and visualization then allows researchers to focus on key sites in the regions (sites with the longest periods of record, largest basins, critical sites where flooding has economic impacts, and so on), as well as identify outliers. Outliers are of interest because they could be sites with unique conditions such as extensive land use change, urbanization, or regulation not indicated in the source database.