Clustering approaches for analysing similarity in ungauged catchments: input variable selection for hydrological predictions

Nilay Dogulu¹, Inci Batmaz², and Elcin Kentel¹

1 Middle East Technical University, Department of Civil Engineering, Water Resources Laboratory, Ankara, Turkey
2 Middle East Technical University, Department of Statistics, Ankara, Turkey

**Background**

Catchments are hydrological units that exhibit unique but distinct features that greatly contribute to heterogeneity and complexity of rainfall-runoff processes. While the type of understanding such diversity has provided a focus of many research efforts in hydrology, including predictions in ungauged basins, there is still room for improving our ability to benefit from this diversity in the context of data-driven hydrologic regionalization.

An outstanding issue in this line of research concerns enhanced in identifying similar groups of catchments in ungauged basins. This is a critical step in moving towards the development of catchment-based predictive models for hydrological attributes and over different runoff attributes representing particular hydrological conditions.

**Research Objectives:**

To explore the potential value of different clustering methods in identifying similar groups of catchments
To determine input variables that control streamflow predictability within such group of catchments and over different runoff attributes representing particular hydrological conditions.

**Methodology**

CAMELS

CATCHMENT Attributes and MEeTeorology for Large-sample Studies

Adar et al., 2017, HESS

671 watersheds across continental USA

(impacted / less impacted by anthropogenic changes)

**Input Variable Selection (IVS)**

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

**Data**

Clustering approaches for analysing similarity in ungauged catchments:

**Summary**

Clustering of catchments using available topography, soil, geology, vegetation and climate attributes

**Contact**

Supervisor Co-Supervisor

Nilay Dogulu

E-Mail: nilaydogulu@metu.edu.tr

Blog: blog.metu.edu.tr/e149313

Supervisor: Nilay Dogulu

Co-Supervisor: Nilay Dogulu

**References**


**Research Contribution:**

The effect of clustering method choice needs to be carefully explored for assessing clustering similarity.
Clustering approaches for analysing similarity in ungauged catchments: input variable selection for hydrological predictions

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)

Catchments are hydrological units that exhibit unique but distinct features that greatly contribute to heterogeneity and complexity of rainfall-runoff processes. While the lure of understanding such diversity has underpinned the focus of many research efforts in hydrology, including predictions in ungauged basins, there is still room for improving our ability to benefit from this diversity in the context of data-driven hydrologic regionalization. An outstanding issue in this line of research concerns enhanced utilization of knowledge on dominant factors affecting catchments’ hydrologic response behaviour under different types of streamflow. Our study addresses this issue by grouping similar catchments across continental USA using the CAMELS dataset (Addor et al., 2017) for the purpose of determining input variables that control streamflow predictability within each group of catchments.

To this aim, we explore the performance of different clustering methods in identifying similar catchments based on available topography, soil, geology, vegetation and climate attributes, and then evaluate the set of variables which characterize hydrological attribute of interest (95% flow percentile for low flows, mean daily discharge for medium flows, and 5% flow percentile for high flows) using iterative input variable selection method (Galelli and Castelletti, 2013). We compare three clustering approaches that belong to different family of methods: partitional clustering algorithm (k-means clustering), density-based clustering algorithm, and spectral clustering algorithm.

We discuss the results from the perspective of underlying assumptions and capabilities of these methods, and provide insights into effects of clustering method choice in analysing variability of catchment similarity with respect to high, medium and low flows.
