



Exploratory insights into hydrological applications of data clustering

Nilay Dogulu

Middle East Technical University (METU), Civil Engineering Dept., Water Resources Laboratory, Ankara, Turkey
(ndogulu@metu.edu.tr)

Clustering of hydrological data offers rich insights that allow researchers to understand diverse concepts and relations that underline heterogeneity and complexity inherent in catchment systems both spatially and temporally. In hydrology, this unsupervised data mining technique for multivariate data is often used as a part of the full methodology (i.e. intermediate step). Furthermore, the issues and challenges associated with clustering algorithms are usually overlooked. Nevertheless, despite finding a range of applications within hydrological studies, as reflected in the rapid growth of the literature, there has been little research into the implications of such growth. There are few studies examining different uses of alternative clustering methods in hydrology while none of these provide a roadmap to assist in selecting the most appropriate method for a particular purpose. This review provides a comprehensive analysis and synthesis of selected journal articles from the hydrology literature in the last two decades. The growth of the literature is considered quantitatively with respect to: (i) the purpose of clustering, (ii) the method(s) of clustering used, (iii) justification behind clustering method selection, (iv) the number of clusters, (v) how the optimal number of clusters is determined, (vi) reported performance of the methods used, and (vii) the case study (e.g. dataset, location). Based on the analysis, current trends in the literature are summarized. Also, future challenges and opportunities for international hydrological research agenda are highlighted.