Using Hierarchical Clustering in Order to Increase Efficiency of Self-Organizing Feature Map to Identify Hydrological Homogeneous Regions for Flood Estimation

F. Farsadnia[1*] - B. Ghahreman[2]
Received: 04-17-2014
Accepted: 08-19-2015

Introduction: Hydrologic homogeneous group identification is considered both fundamental and applied research in hydrology. Clustering methods are among conventional methods to assess the hydrological homogeneous regions. Recently, Self Organizing feature Map (SOM) method has been applied in some studies. However, the main problem of this method is the interpretation on the output map of this approach. Therefore, SOM is used as input to other clustering algorithms. The aim of this study is to apply a two-level Self-Organizing feature map and Ward hierarchical clustering method to determine the hydrologic homogenous regions in North and Razavi Khorasan provinces.

Materials and Methods: SOM approximates the probability density function of input data through an unsupervised learning algorithm, and is not only an effective method for clustering, but also for the visualization and abstraction of complex data. The algorithm has properties of neighborhood preservation and local resolution of the input space proportional to the data distribution. A SOM consists of two layers: an input layer formed by a set of nodes and an output layer formed by nodes arranged in a two-dimensional grid. In this study we used SOM for visualization and clustering of watersheds based on physiographical data in North and Razavi Khorasan provinces. In the next step, SOM weight vectors were used to classify the units by Ward's Agglomerative hierarchical clustering (Ward) methods. Ward's algorithm is a frequently used technique for regionalization studies in hydrology and climatology. It is based on the assumption that if two clusters are merged, the resulting loss of information, or change in the value of objective function, will depend only on the relationship between the two merged clusters and not on the relationships with any other clusters. After the formation of clusters by SOM and Ward, the most frequently applied tests of regional homogeneity based on the theory of L-moments are used to compare and modify the clusters which are formed by clustering algorithms and find the best clustering method to achieve hydrologically homogeneous regions. Two statistical measures are used to form a homogeneous region, (i) discordancy measure and (ii) heterogeneity measure. The discordancy measure, D_i, is used to find out unusual sites from the pooling group (i.e., the sites whose at-site sample L moments are markedly different from the other sites). Generally, any site with $D_i > 3$ is considered as discordant. The homogeneity of the region is evaluated using homogeneity measures which are based on sample L-moments (LCv, LCs and LCk), respectively. The homogeneity measures are based on the simulation of 500 homogeneous regions with population parameters equal to the regional average sample l-moment ratios. The value of the H-statistic indicates that the region under consideration is acceptably homogeneous when H<1, possibly heterogeneous when $1 \leq H < 2$, and definitely heterogeneous when $H \geq 2$.

Results and Discussion Conclusions: At first by principal component analysis we reduced SOM input matrix dimension, then the SOM was used to form a two-dimensional features map. Then to determine homogeneous regions for flood frequency analysis, SOM output nodes were used as input into the Ward method. The regions identified by the clustering algorithms are, in general, not statistically homogeneous. Consequently, they have to be adjusted to improve their homogeneity. The sites that were flagged discordant by the discordancy measure were first identified. Secondly, the heterogeneity measures of the adjusted region were examined as they change with exclusion of each site from the region. Thirdly, the discordant site, whose exclusion reduces the heterogeneity measures of the region by a significant amount, was identified and removed. After adjustment of homogeneity regions by L-moment tests, five hydrologic homogeneous regions were identified. Finally adjusted regions are created by a two-level SOM and then the best regional distribution function and associated parameters are selected by the L-moment approach. The main results of this study are briefly mentioned:

The results showed that the combination of self-organizing maps and Ward hierarchical clustering by principal components as input is more effective than the hierarchical method, by principal components or

---

1,2 - PhD Student of Irrigation and Drainage and Professor, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad
(*-Corresponding Author Email: farhadfarsadnia@stu.um.ac.ir)

standardized inputs to achieve hydrologic homogeneous regions.

SOM is a useful method to achieve homogeneous regions, because SOM has shown a high performance for visualization and abstraction of attributes, and displayed a distribution of each component. It is found that Ward's algorithm is an easy way to cluster SOM units because Ward's algorithm does not need to determine the optimum number of clusters before calculations.