



A comparative analysis of clustering algorithms to identify the homogeneous rainfall gauge stations of Bangladesh

Sangita Paul*

University of Dhaka, Dhaka, Bangladesh - spaul@isrt.ac.bd

Mohammad Samsul Alam

University of Dhaka, Dhaka, Bangladesh - msalam@isrt.ac.bd

Abstract

Bangladesh has an agro-based economy which mainly depends on rain. Dealing with the rainfall at each of the stations separately is time consuming as well as subject to more errors. It seems more advantageous and flexible to deal with a group of homogeneous stations rather than individual stations. Different regions of Bangladesh experience different precipitations every year where some regions receive very much similar precipitations. To identify the homogeneous stations clustering algorithms of multivariate techniques are applied in this study and a comparison between these algorithms is made. Annual and seasonal (pre monsoon, monsoon and post monsoon) precipitation data of 30 stations from 1977 to 2012 recorded by Bangladesh Meteorological Department are used in this study. Fuzzy C-means, agglomerative hierarchical and K-means clustering methods are applied to classify the precipitation series and identify the hydrologically homogeneous groups. The optimal numbers of clusters are four, three, three and four chosen for annual, pre monsoon, monsoon and post monsoon rainfalls respectively using Gap statistic. The spatial distributions of stations in the clusters identified by each of the clustering methods are obtained for annual precipitations and for precipitations during pre monsoon, monsoon and post monsoon. The regional homogeneity test based on L-moments showed that the clusters identified by Fuzzy C-means method are sufficiently homogeneous compared to that by hard clustering methods, hierarchical and K-means. Thus it is recommended to prefer Fuzzy C-means method to classify the precipitation series and for identifying hydrologically homogeneous regions to those hard clustering methods.

Keywords: gap statistic; hierarchical; fuzzy c-means; L-moments.