Investigation of Heterogeneity Measure using various Nonlinear approaches by deriving characteristics through SWAT model

Shiyamalagowri G, Ganapathy G.P

Centre for Disaster Mitigation and Management, Vellore Institute of Technology (VIT), Vellore-632014, India. Email: shiyamalagowri1947@gmail.com

Introduction

Flooding ranks as the most devastating natural disaster causing high economic loss and large number of fatalities. The main objective of frequency analysis of hydrologic data is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions. Globally many linear clustering techniques are employed to categorize the watershed which are ineffective when dealing with noise and outliers. The present study overcomes this by proposing a relatively new Non-Linear Clustering Algorithm based on Hierarchical estimation of Densities (NLCAHD) for the Cauvery basin, where the Homogeneity test (H) is enforced to identify the group of stations with same populations.

Study Area

- Daily discharge data
- 1971-2017 (CWC)
- 28 gauging stations
- Average annual rainfall -1250 mm
- Mean annual discharge -21.36 BCM
- Drainage area
- Elevation
- Baseflow index
- Mean annual precipitation
- Annual maximum streamflow

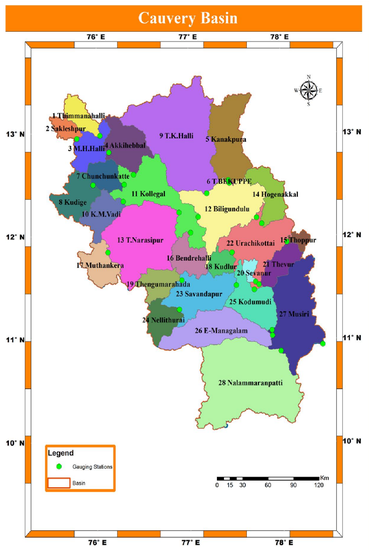

Fig.1 Cauvery Basin map with locations of gauging stations

Heterogeneity Test

A homogeneity test (H) is conducted to measure the degree of uniformity among all the gauging stations belonging to a cluster.

- H<1 = purely homogeneous
- 1≤ H < 2 = possibly heterogeneous
- H>2 = absolutely heterogeneous.

$$H = \frac{(V - \mu_V)}{\sigma_V}$$

Methodology

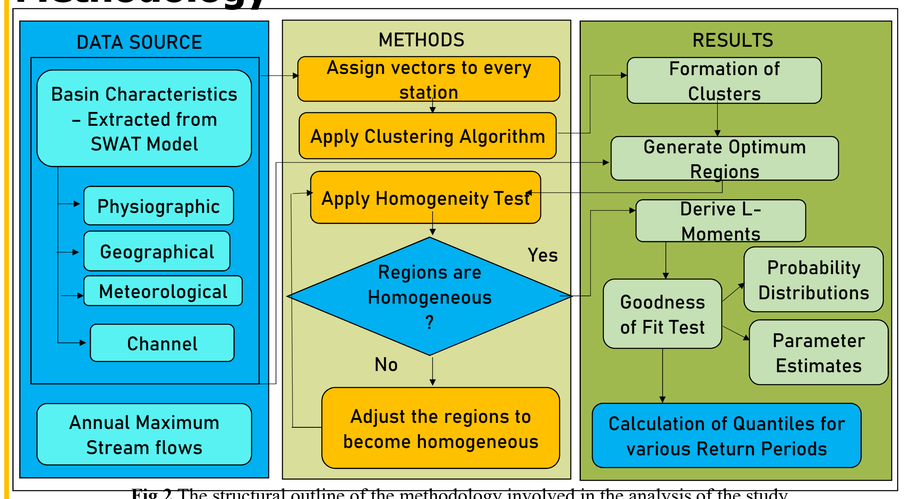

Fig.2 The structural outline of the methodology involved in the analysis of the study

1. Calculate the kernel distance w.r.t l_{dps} for all the gauging stations denoted by $d_{kernel}(g_{si})$, which is the distance from ith point to the l_{dps} nearest neighbor. The gauging station $g_{si} \in G_s$ is said to be a kernel station, if $d_{kernel}(g_{si}) \leq \beta$
2. Estimate the shared attainable distance between the two stations g_{si} and g_{sj} in accordance with l_{dps} given by
   $d_{sattain}(g_{si},g_{sj}) = \max\{d_{kernel}(g_{si}),d_{kernel}(g_{sj}),d(g_{si},g_{sj})\}$
3. Design the Minimum Spanning Tree (MST) for the graph G with l_{dps} in order to obtain the shared attainable graph which is a complete graph G, where stations are represented as vertices and value for each edge is the shared attainable distance w.r.t l_{dps} between the pair of stations.
4. Add a self edge to each station with the kernel distance of the corresponding station as value in order to obtain MSText.
5. From the dendrogram generated from MSText isolate the hierarchy of density clusters.
   i. Same name is assigned to all the stations at the root of the tree. Delete the edges in descending order w.r.t weights.
   ii. Once the edges are deleted, allot labels to the connected components, which contains the last node of the deleted edges, in order to obtain the succeeding level of hierarchy.

References

- Basu B, Srinivas V V. (2016) Regional Flood Frequency Analysis Using Entropy-Based Clustering Approach. J Hydrol Eng 21:04016020.
- Burn DH (1990) Evaluation of Regional Flood Frequency Analysis With a Region of Influence Approach. Water Resour Res 26:2257–2265
- Hosking JRM, Wallis JR (1997) Regional Frequency Analysis: An approach based on L-moments, Cambridge University Press, New York.

Results and Conclusion

Table 1. Clusters formed by NLCAHD approach

Regions	Number of sites	H value	Homogeneity Status
1	7	0.8093	Homogeneity
2	4	0.5787	Homogeneity
3	5	0.8165	Homogeneity
4	6	0.1171	Homogeneity
5	3	-0.5557	Homogeneity
6	3	-0.7257	Homogeneity

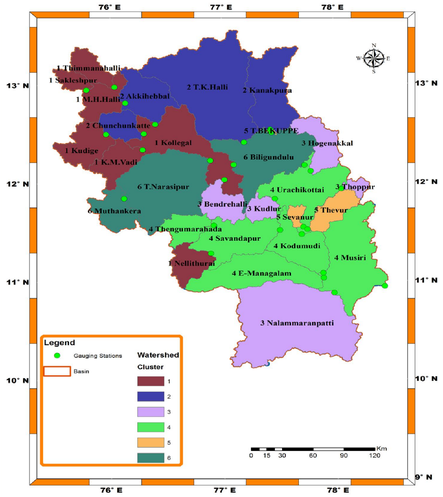

Fig.3 Cauvery Basin – Classification of 6 Clusters

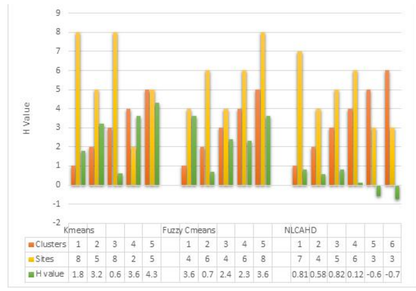

Fig.4 Comparison of NLCAHD with K means and C means

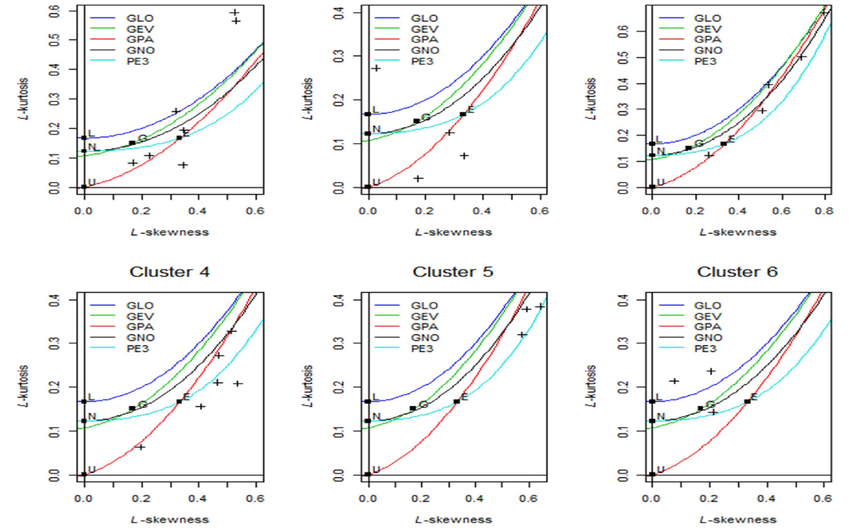

Fig.5 L-moment ratio diagrams for the 6 clusters

The approach automatically classifies the region into 6 homogeneous clusters, where the number of clusters were neither known nor mentioned a prior, whereas in the case of other clustering algorithms like K-means and C-means the number of clusters should be initialized before the partitioning of datasets takes place. The algorithm measures the density over the stations based on the kernel distance of 0.4, which serves as the radius of the neighborhood in order to form the clusters.