



Journal of Experimental Biology and Agricultural Sciences

<http://www.jebas.org>

ISSN No. 2320 – 8694

IDENTIFICATION OF HOMOGENEOUS RAINFALL STATIONS IN HARYANA

Mohit Nain^{*}, B.K. Hooda

Department of Mathematics and Statistics, College of Basic sciences and Humanities, CCS Haryana Agriculture University, Hisar-125004 (Haryana), India

Received – July 10, 2019; Revision – August 21, 2019; Accepted – 05 September 12, 2019

Available Online – October **, 2019

DOI: [http://dx.doi.org/10.18006/2019.7\(5\).***.***](http://dx.doi.org/10.18006/2019.7(5).***.***)

KEYWORDS

Rain gauge stations

Cluster analysis

Ward's method

Common principal components

ABSTRACT

Rainfall analysis is vital phenomenon for farming creation as well as for regulatory purposes and plays an important role in designing water harvesting structures as well as crop planning. In present study, monthly rainfall data of 42 years (1970-2011), covering 27 rain gauge stations of Haryana was used for the classification and identification of homogeneous rainfall stations in Haryana. Clustering of rainfall stations for monsoon period was done by utilizing Ward's method on the common principal components scores (CPCs). The results of present study showed that there are four clusters of rain gauge stations having similar monsoon rainfall spread over Haryana. Cluster I consisted of six stations; Cluster II consisted of eight stations; Cluster III consisted of ten stations and Cluster IV consisted of three stations. Cluster analysis of mean monthly rainfall was also performed by using Ward's method. On the basis of mean monthly rainfall we observed that there are again four clusters of rain gauge stations. Cluster I consisted of five stations; Cluster II consisted of eight stations; Cluster III consisted of ten stations while Cluster IV consisted of four stations. The two analyses gave the patterns in close agreement and it was found that Haryana can be grouped into four clusters based on Monsoon rainfall.

* Corresponding author

E-mail: nainbir93@gmail.com (Mohit Nain)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI]
(<http://www.horizonpublisherindia.in/>).
All rights reserved.

All the articles published by [Journal of Experimental Biology and Agricultural Sciences](#) are licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#) Based on a work at www.jebas.org.



1 Introduction

Rainfall is one of the most variable climatic characters and its variability varies with both spatially and temporally. India is a tropical country and its agricultural planning and water utilization mainly depends on monsoon rainfall. More than 75% of the India rainfall occurs during the monsoon season. The Agriculture of the state depend on the rainfall received and rainfall characteristics like magnitude, frequency and intensity vary from spatially and temporally both. The random nature of rainfall occurrence suggests need for its statistical analysis and logical interpretation. In particular, the monthly rainfall of a region is very helpful for farmers in deciding when and where to sow and reap for successful cultivation with proper utilization of available water and irrigation facilities. The Eastern agro climatic zone of Haryana has high rainfall (>400mm) whereas the Western agro climatic zone has lesser amount of rainfall (200-400mm) and maximum rainfall reaches to 800mm in Northern districts of Panchkula, Ambala, Yamunanagar, Kurukshetra etc. (Agro climatic Atlas of Haryana, Technical Bulletin No. 15, 2010). There are two main cropping patterns in Haryana i.e. Rabi and Kharif. Wheat is the main crop of Rabi season and second main crop season is Kharif coinciding with hot weather and south-west monsoon season. In Kharif season main crops are rice (Eastern agro climatic zone) and cotton (Western agro climatic zone). Due to frequent abnormalities in the magnitude and distribution of rainfall make the cropping more risky.

Multivariate techniques are very useful tools to find hydro logically homogeneous regions and to classify regions based meteorological data such as rainfall. Gadgil & Iyengar (1980) applied principal component analysis to derive patterns of temporal variation of the rainfall at fifty-three stations in peninsular India and eight clusters was found. Further, Kulkarni & Rao (2000) used Common Principal Components (CPC) approach for classification of the 20 districts of Andhra Pradesh based on monthly rainfall data. Similarly, Kulkarni & Reddy (1994) used average linkage method to group the districts of Andhra Pradesh and found that districts were classified into 5 to 7 clusters which depend on the season. Further, Munoz-Diaz & Rodrigo (2004) used Ward's clustering methods and principal component analysis technique to find out climatically homogeneous zones, based on seasonal rainfall for 32 Spanish localities and found that cluster analysis technique to more suitable than principal component analysis. Similarly, cluster analysis technique was used by various researchers in various regions of India for indentifying homogeneous rainfall regimes, among these some most popular works are Venkatesh & Jose, 2007 (Western Ghats region of Karnataka); Yashwant & Sananse, 2015 (Marathwada region in Maharashtra) and Shirin & Thomas, 2016 (Kerala). Oliveira-Júnior

et al. (2017) identified three homogeneous rainfall regions in Tocantins State, Brazil using Ward's algorithm of cluster analysis. Similarly, Terassi & Galvani (2017) also identified the homogeneous rainfall regions in the eastern watersheds of the State of Paraná, Brazil. Recently, Siraj-Ud-Doulah & Islam (2019) analyzed monthly rainfall data from 34 climate stations of Bangladesh using five agglomerative hierarchical clustering measures and found that Ward method based on Euclidean distance, K-means, Fuzzy were the most suitable methods in this particular case. They found seven different climate zones in Bangladesh. Similarly, Gonçalves et al. (2018) used annual mean precipitation and found six homogeneous regions through cluster analysis using Ward's agglomeration method, applied to a historical series of 31 years (1960-1990) at 413 satellite monitoring points in the state of Pará, in the Amazon where the selected years occurred during an El Niño or a La Niña event. The aim of this study was to identify homogeneous regions (rain-gauge stations) in Haryana using cluster analysis and common principal component analysis techniques. For the study monthly rainfall data of 42 years (1970-2011), covering 27 rain gauge stations of Haryana was used for the identification of homogeneous rainfall stations in Haryana.

2 Material and Methods

2.1. Location of study and Data

The state Haryana is located in north western India and occupies 1.3 per cent geographical area of the country. The latitude and longitude coverage of the state extends between 27°39' to 30°55'N and 74°27' to 77°36'E respectively. The data for this study include monthly rainfall data obtained from Indian Meteorological Department (IMD) Pune covering 27 rain gauge stations scattered in all the districts of Haryana state. Depending on the availability, 42 years' data (1970-2011) were obtained for rain gauge stations Fatehabad, Gurgaon, Sohana, Jind, Narwana, Ferozpur Jhirka, Nuh, Panipat, Rohtak, Sonapat, Sirsa, Hisar, Bawal, Karnal, Ambala and Kaithal. For the stations Tohana, Jhajjar, Dujana, Kalka, Dadupur and Jagadhari the data for the 2006 was missing. Also for stations Dadri, Ballabgarh, Thanesar, Hassanpur and Narnaul the data were available for 36 years (data from 1984-1990 was not available).

2.2. Ward's Cluster Method

Cluster analysis (CA) is a convenient method for identifying homogenous groups of objects called clusters. There are number of methods that can be used to carry out cluster analysis and in this study, Ward's (1963) method of cluster analysis was used which is also known as "minimum variance method". This method is different from other hierarchical clustering methods because it uses an analysis of variance approach to evaluate the distances between clusters. In Ward's method the within- cluster sum of squares is minimized and clusters with minimum between-cluster distance

are merged. Let we have two clusters C_k and C_l which are merged to form a new cluster C_m , then the Euclidean distance between the new cluster and another cluster C_j is given by the formula:

$$d_{j,m} = \frac{(n_j + n_k)d_{jk} + (n_j + n_l)d_{jl} - n_j d_{kl}}{n_j + n_m}$$

Where n_j, n_k, n_l and n_m are the number of objects in clusters j, k, l and m , respectively and d_{jk}, d_{jl} and d_{kl} represent the distances between the observations in clusters j and k , between j and l , and between k and l , respectively (Ramos, 2001). The, Ward's algorithm can be implemented by updating a stored Euclidean distance between cluster centroids. Although clustering results may be sensitive to the chosen method, Blashfield (1976) found that the Ward's method provides the most accurate solutions among the hierarchical methods

2.3.1. Clustering under Multiple Sampling

Common principal components approach (Kulkarni & Rao, 2000) was used and described as: let we have 'n' objects which are to be classified into $k (< n)$ homogeneous groups. Suppose that the j -th object has observations ($j = 1, \dots, n$) which are recorded by drawing a random sample of size N_j from it. Let X be the random vector consisting of p variables, then X_{ij} represents the i -th observation vector on the j -th object ($i = 1, \dots, N_j, j = 1, \dots, n$). Thus on the basis of the observation vector X_{ij} the n objects are to be classified into $k (< n)$ distinct groups.

This approach involves determining a vector subspace which represents the vector subspaces of all the objects as closely as possible. Several developments have taken place in this field (Flury, 1988). Suppose that principal component analysis has been carried out for each of the 'n' objects. Furthermore, the first $q (< p)$ principal components are adequate for summarizing the total variance of each of the covariance matrices. Let $L_t (q \times p)$ be the matrix of these vectors corresponding to the t -th object ($t = 1, \dots, n$) whose rows are the Eigen vectors of the p -principal components. Let Σ is the covariance matrix and $H (p \times p) = \sum L_t' L_t$ be a matrix whose first $q (< k)$ principal components represent the "common principal components".

For obtaining Common Principal Components, each of the covariance matrices of the rain gauge stations were subjected to principal component analysis. It was observed that the first 3 PCs accounted for at least 85 per cent of the sample variance and so adequately summarized the total variance of the 4 rainfall variables in all the 27 stations. Hence the matrix L_t ($t = 1, 2, \dots, 27$) was defined on the basis of the first 3 components. Using these CPCs, component scores for the stations (based on mean rainfall) are obtained and clustering was carried out on the basis of these scores.

3 Results and Discussion

For clustering various rainfall stations data for monsoon-period (June-September), common principal components approach was used. It was observed that the first 3 PCs accounted for at least 86 per cent of the sample variance and so adequately summarized the total variance of the 4 rainfall variables (June-September) in all the 27 stations. Hence the matrix L_t ($t=1, \dots, 27$) was defined on the basis of the first 3 components. The results of the principal component analysis of the matrix $H = \sum L_t' L_t$ given below which give the common principal components are presented in Table 1.

$$H = \begin{bmatrix} 23.13 & -0.18 & 0.53 & -0.12 \\ -0.18 & 20.86 & 0.06 & 4.06 \\ 0.53 & 0.06 & 18.99 & 5.25 \\ -0.12 & 4.06 & 5.25 & 18.02 \end{bmatrix}$$

Table 1 Common Principal Components of monsoon rainfall for various rain gauges stations in Haryana

Variable	Common Principal Components			
	1	2	3	4
June rainfall	0.038	0.988	-0.144	-0.030
July rainfall	0.554	-0.120	-0.750	0.342
August rainfall	0.519	0.089	0.630	0.571
September rainfall	0.650	-0.026	0.144	-0.746
Latent Root	25.7	23.2	20	12.1

It can be observed that the latent roots of H , which represent the measure of similarity between the CPCs and vector subspaces of all the 27 stations, were almost similar corresponding to the first 3 components, ($\lambda_1=25.7, \lambda_2=23.2, \lambda_3=20$) whereas it was considerably low in the fourth component ($\lambda_4=12.1$). The results thus indicate that all the stations were close together along the first 3 CPCs (i.e., the first three components of H).

The vector subspaces of the common principal components indicated that the vector subspace of the first common principal components is heavily loaded on September rainfall (loading = 0.650) while rainfall of June (loading = 0.988) and July (loading = -0.750) were found to be respectively in second and third components. This behavior was exhibited for all the districts in their three-dimensional subspaces (i.e. three principal components). These results indicated that only the first three common principal components can be considered common to all the vector subspaces of the districts. These three components also

revealed the common cause for variation in the rainfall of the stations viz. rainfall of June, July, and September. Using these CPCs, Components scores for the stations (based on mean rainfall) were obtained and are given in Table 2. Cluster analysis of scores carried out using Ward's method.

Table 2 Common Principal Components Scores

Stations	CPC1	CPC2	CPC3	CPC4
Sirsa	137.4	33.4	-37.4	42.2
Fatehabad	120.0	40.3	-5.6	47.9
Hisar	174.7	47.1	-27.2	54.5
Tohana	154.9	28.1	-20.1	44.0
Gurgaon	314.8	54.2	-18.5	129.0
Sohana	233.4	26.1	-2.9	99.1
Thanesar	227.7	77.5	-30.7	95.9
Panipat	215.0	53.1	-11.9	69.1
Rohtak	223.6	45.0	-2.5	100.3
Ballabgarh	281.2	51.5	-3.2	98.0
Hassanpur	176.4	14.9	20.7	65.4
Bawal	240.8	56.8	-0.2	86.2
Karnal	290.4	82.2	-18.4	104.9
Ambala	317.4	90.5	-71.8	101.7
Dadupur	394.8	112.5	-51.1	152.8
Jagadhari	374.2	98.6	-24.6	145.4
Dujana	202.5	34.6	-28.0	73.9
Jhajjar	251.8	47.8	-20.1	81.8
Jind	196.1	43.5	-21.1	46.8
Narwana	185.8	44.9	-11.4	48.8
Firozpur Jhirka	282.9	60.8	-4.6	107.1
Nuh	267.8	49.7	1.0	85.3
Kaithal	224.4	48.9	-15.3	78.2
Sonipat	276.4	58.9	-17.6	103.5
Dadri	206.0	42.6	-11.5	80.6
Kalka	461.9	101.7	-36.5	167.7
Narnaul	181.7	38.6	6.3	16.7

Dendrogram based on Common Principal Components Scores is shown in Figure 1. The dendrogram revealed that there are four clusters of rain gauge stations having similar monsoon rainfall spread over Haryana. Cluster I consisted of six stations i.e. Ballabgarh, Gurgaon, Ambala, Karnal, Firozpur Jhirka and Sonipat while Cluster II made of eight stations i.e. Hassanpur, Fatehabad,

Tohana, Sirsa, Hisar, Jind, Narwana and Narnaul; Cluster III comprised of 10 stations i.e. Sohana, Thanesar, Panipat, Rohtak, Bawal, Dujana, Jhajjar, Nuh, Kaithal and Dadri and Cluster IV has three stations i.e. Kalka, Dadupur and Jagadhari. Thus, Haryana can be divided into four rainfall zones based on common principal component scores.

3.2 Hierarchical Clustering Analysis (Ward's Method)

The Ward's method of Hierarchical clustering was also applied for classifying the 27 rain gauge stations of Haryana based on average monthly rainfall for the period 1970-2011. Three seasons viz., Monsoon (June-September); Pre-monsoon (March-May) and Overall period (June-May) were considered for the present study. Post-monsoon (October-December) and winter-period (January-February) were not considered for classification, as in most of the years the rainfall during these months was low (mainly November it was near about zero). Dendrogram based on pre-monsoon, monsoon and overall period are presented in the Figures 2, 3 and 4 respectively. From the analysis of these dendrogram, it has been concluded that 4-clusters solutions are appropriate for grouping of stations. Further, the stations classified under a cluster need not be from the same region.

About 80 per cent of annual rainfall comes from the south-west monsoon in the month of June to September in Haryana; hence we are interested in monsoon rainfall. In monsoon rainfall there are 4-clusters as suggested by dendrogram (Figure 3). Cluster I (C1) consisted of five stations i.e. Ballabgarh, Gurgaon, Karnal, Firozpur Jhirka and Sonipat. Cluster II (C2) having eight stations i.e. Hassanpur, Fatehabad, Tohana, Sirsa, Hisar, Jind, Narwana and Narnaul. Cluster III (C3) comprised of 10 stations i.e. Sohana, Thanesar, Panipat, Rohtak, Bawal, Dujana, Jhajjar, Nuh, Kaithal and Dadri. Cluster IV (C4) consisted of four stations i.e. Ambala, Kalka, Dadupur and Jagadhari.

Cluster analyses of rain gauge stations based on monsoon rainfall in Haryana are given in Table 3 and distances between the cluster centroids are given in Table 4. It was interesting to note that the results of the Ward's method of Hierarchical clustering either based on common principal component score or based on mean monthly rainfall data of Monsoon period are almost in agreement similar as depicted in Figure 1 and Figure 3.

The cluster profile of rain gauge stations based on monsoon rainfall period is presented in Table 3 and it showed that the mean monsoon rainfall was minimum for cluster-II (333.6) and maximum for Cluster-IV (839.2) showing the maximum variation among the characteristics of these two clusters. But the minimum difference between mean monsoon rainfall was found for cluster-III (461.7) and Cluster-I (582.8) showing the similarity between the characteristics of these clusters. Recently, Swaminathan &

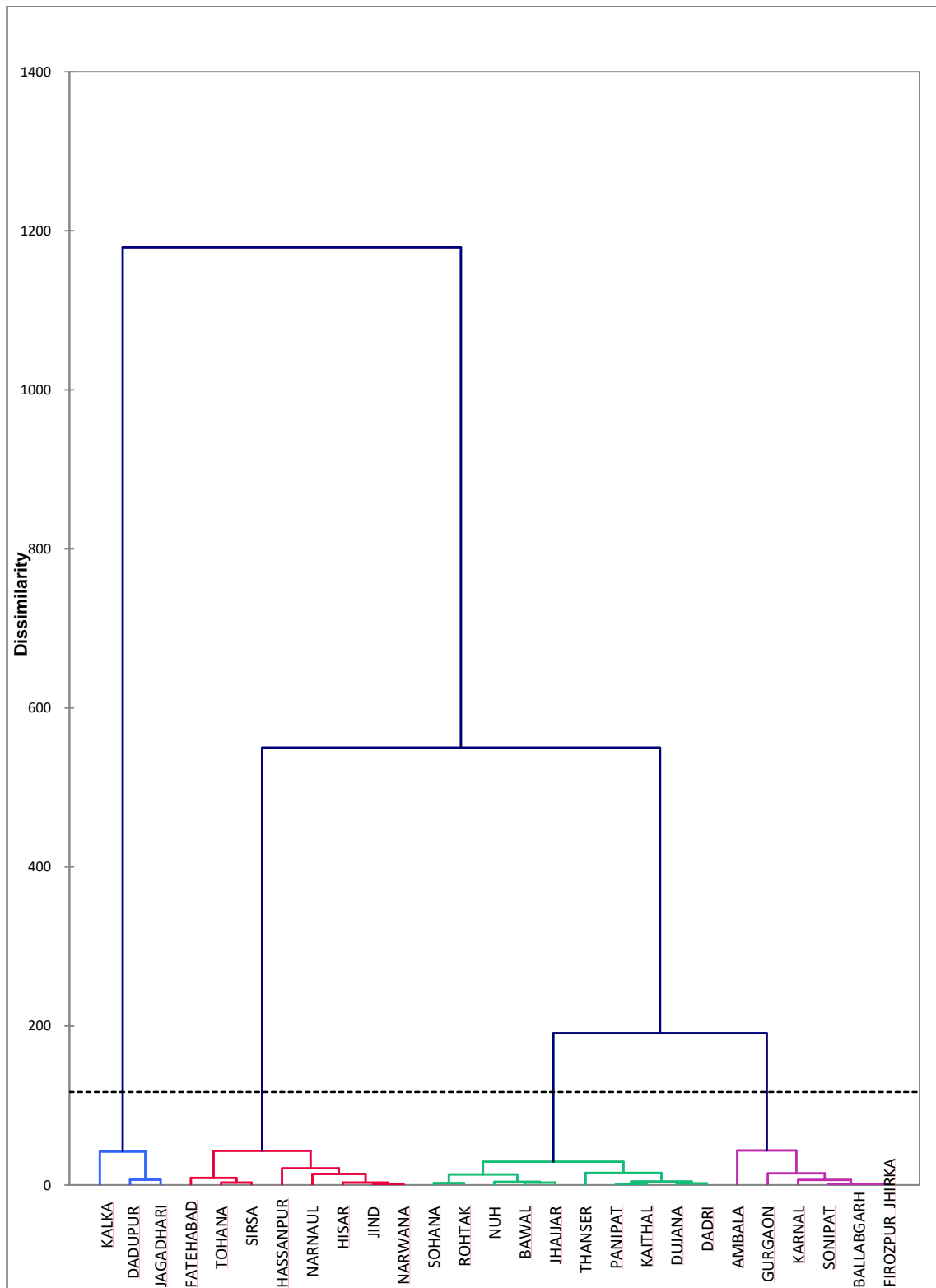


Figure 1 Dendrogram for rain gauge stations in Haryana based on Common Principal Components Scores for Monsoon rainfall

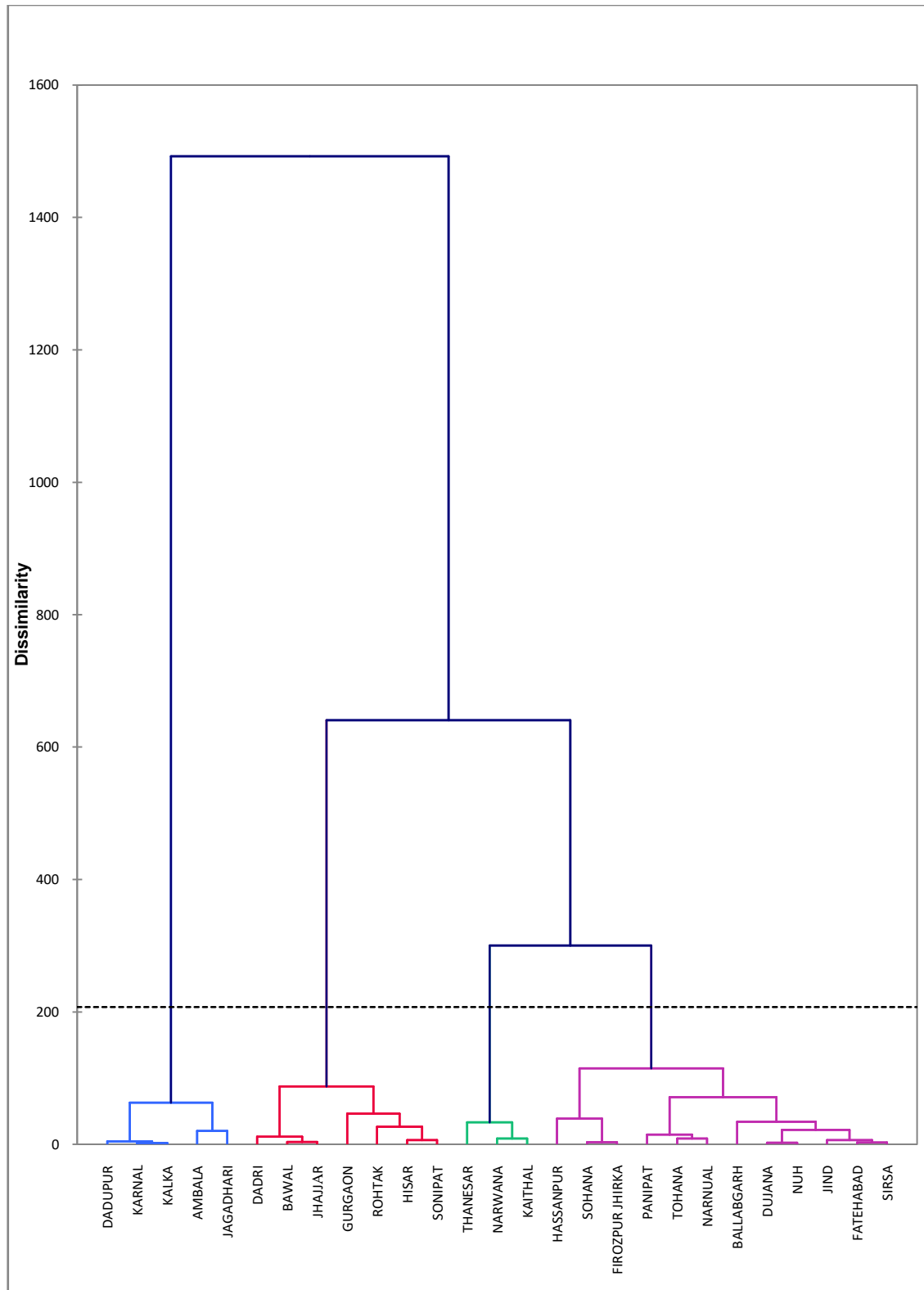


Figure 2 Dendrogram for rain gauge stations in Haryana based on pre-monsoon (March-May)

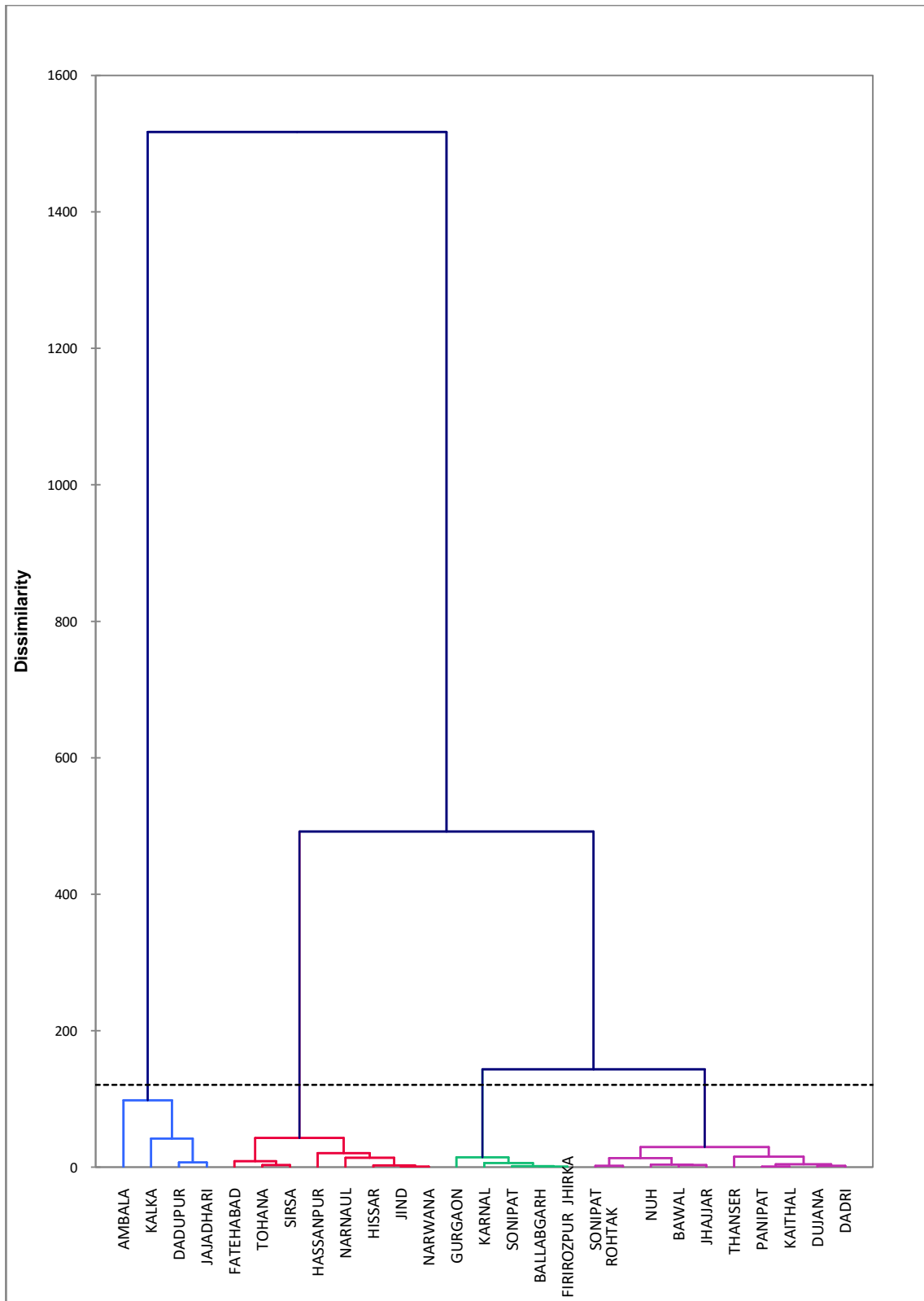


Figure 3 Dendrogram for rain gauge stations in Haryana based on monsoon period (June- Sept.)

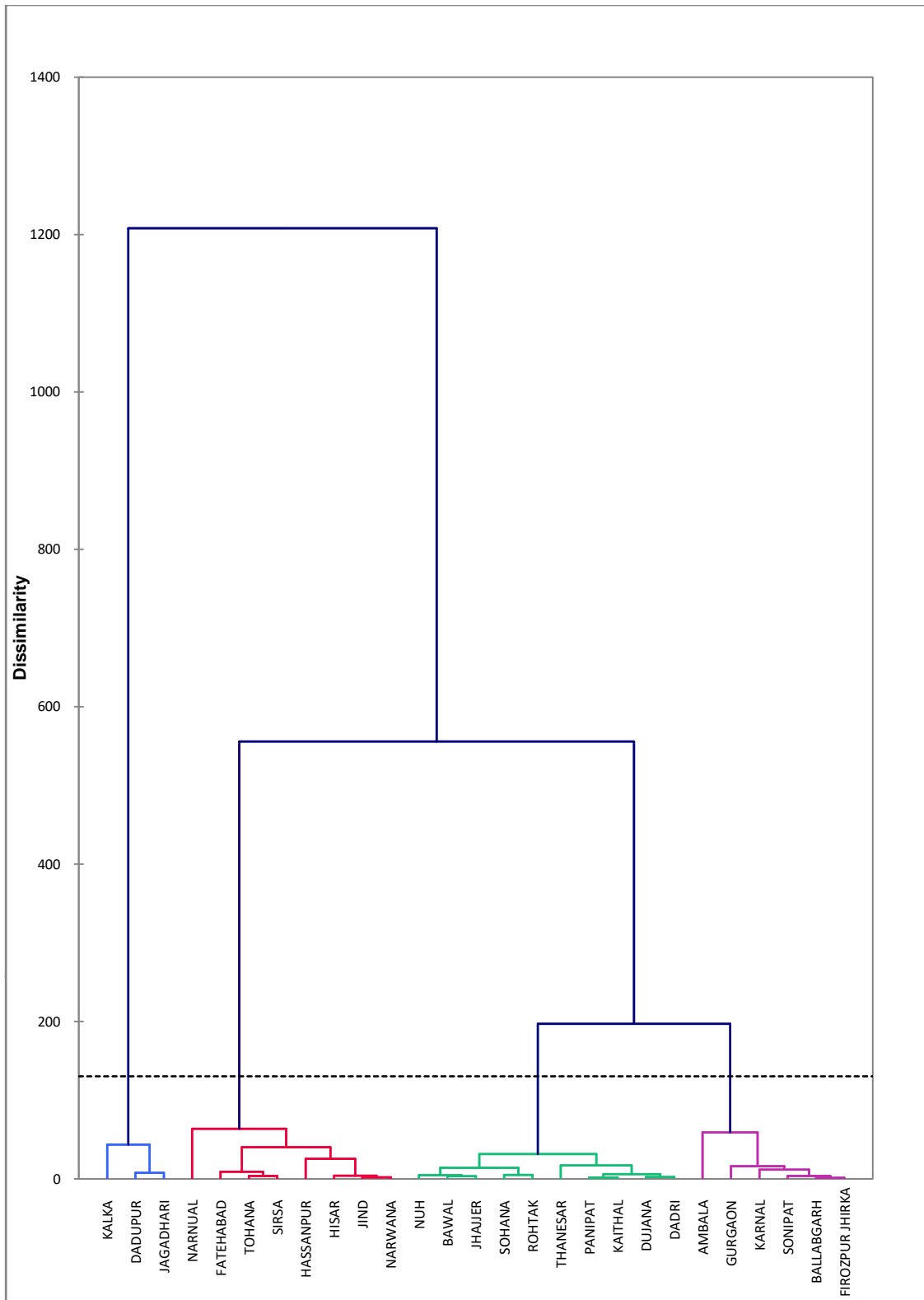


Figure 4 Dendrogram for rain gauge stations in Haryana based on overall period (June-May)

Table 3 Cluster analysis of rain gauge stations based on monsoon rainfall in Haryana

Clusters	Stations	June	July	Aug.	Sept.	Mean Monsoon rainfall	Avg. distance to Centroid
C1	Ballabgarh, Gurgaon, Karnal FirozpurJhirka, Sonipat	70.3	199.2	209.7	103.6	582.8	20.02
C2	Hassanpur, Fatehabad,Tohana, Sirsa, Hisar, Jind, Narwana, Narnaul	42.5	112.1	108.0	71.0	333.6	32.57
C3	Sohana, Thanesar, Bawal, Dujana, Jhajjar, Panipat, Rohtak, Nuh, Kaithal, Dadri	55.5	159.4	164.2	82.6	461.7	25.81
C4	Ambala, Kalka, Dadupur, Jagadhari	121	307.0	262.0	149.2	839.2	56.21

Table 4 Distances between the cluster centroids

Clusters	1	2	3	4
1	0	140.57	65.63	137.93
2	140.57	0	75.53	272.01
3	65.63	75.53	0	200.18
4	137.93	272.01	200.18	0

Meganathan (2018) employed EM method and K Means on rainfall data of a Thanjavur region in Tamil Nadu for 100 years and found EM method was accurate than K Means. Singh et al. (2010) described that eastern agro climatic zone of Haryana has high amount of rainfall (>400 mm) and maxima of rainfall reaches to > 800 mm in the northern districts of Panchkula, Ambala, Yamunagar, Kurukshetra etc. Eastern agro climatic zone of Haryana is also called wet zone. The western agro climatic zone has lesser amount of rainfall (200-400 mm). These results are mostly in agreement with the present study.

The distance between the cluster centroid presented in Table 4 indicate that maximum distance of 272.01 has been observed between cluster 2 and 4 showing that these clusters were most different characteristics to each other and minimum distance of 65.63 between clusters 1 and 3 showing that these were the most similar clusters. An objective of this study was to classify the entire state of Haryana into relative number of homogeneous zones based on monthly rainfall. Clustering is the process of dividing the area under consideration to a limited number of climatologically homogeneous zones, based on any hydrologic parameter. It was found that Haryana can be grouped into four clusters based on Monsoon rainfall.

Conclusion

For clustering the rainfall stations for monsoon-period (June-September), common principal components (CPCs) approach was

used. Using these CPCs, Components scores for the stations (based on mean rainfall) were obtained. Clustering was carried out on these scores using ward's method of clustering and dendrogram was prepared. These results indicated that there are four clusters of rain gauge stations having similar monsoon rainfall spread over Haryana. Cluster I consisted of six stations i.e. Ballabgarh, Gurgaon, Ambala, Karnal, Firozpur Jhirka and Sonipat; Cluster II consisted of eight stations i.e. Hassanpur, Fatehabad, Tohana, Sirsa, Hisar, jind, Narwana and Narnaul; Cluster III consisted of 10 stations i.e. Sohana, Th0anesar, Panipat, Rohtak, Bawal, Dujana, Jhajjar, Nuh, Kaithal and Dadri while Cluster IV consisted of three stations i.e. Kalka, Dadupur and Jagadhari. Thus, Haryana can be divided into four rainfall zones based on common principal component scores.

The alternate approach of clustering the rainfall stations for monsoon-period was based on mean monthly rainfall using Ward's method. Again the results indicated 4-clusters. Cluster I consisted of five stations i.e. Ballabgarh, Gurgaon, Karnal, Firozpur Jhirka and Sonipat; Cluster II consisted of eight stations i.e. Hassanpur, Fatehabad, Tohana, Sirsa, Hisar, Jind, Narwana and Narnaul; Cluster III consisted of 10 stations i.e. Sohana, Thanesar, Panipat, Rohtak, Bawal, Dujana, Jhajjar, Nuh, Kaithal and Dadri while Cluster IV consisted of four stations i.e. Ambala, Kalka, Dadupur and Jagadhari. The Ward's method of Hierarchical clustering either based on common principal component score or based on mean monthly rainfall data of Monsoon period gave almost similar result.

Conflict of Interest

Authors would hereby like to declare that there is no conflict of interests that could possibly arise.

References

- Blashfield RK (1976) Mixture Model tests of Cluster Analysis: Accuracy of four Agglomerative Hierarchical Methods. *Psychological Bulletin* 83: 377–388.
- Flury B (1988) *Common Principal Components and Related Multivariate Models*. John Wiley and Sons, New York, USA.
- Gonçalves M F, Blanco CJC, Santos VC dos, Oliveira LL dos S (2018) Homogenous regions and rainfall probability models considering El Niño and La Niña in the State of Pará in the Amazon. *Acta Scientiarum Technology* 40: 1:10.
- Gadgil S, Iyengar RN (1980) Cluster Analysis of Rainfall Stations of the Indian Peninsula. *Quarterly Journal of the Royal Meteorological Society* 106: 873-886
- Kulkarni BS, Rao GN (2000) The Common Principal Components a for Clustering Under Multiple Sampling. *Journal of Indian Society for Agricultural Statistics* 53: 1-11.
- Kulkarni BS, Reddy DD (1994) The Cluster Analysis Approach for Classification of Andhra Pradesh on the Basis of Rainfall. *Mausam* 45: 325-332.
- Muñoz-Díaz D, Rodrigo FS (2004) Spatio-temporal patterns of Seasonal Rainfall in Spain (1912-2000) Using Cluster and Principal Component Analysis: Comparison. *Annals of Geophysics* 22:1435-1448.
- Oliveira-Júnior JF de, Xavier FMG, Teodoro Pe,Gois D de , Delgado RC (2017) Cluster Analysis Identified Rainfall Homogeneous Regions In Tocantins State, Brazil. *Bioscience Journal* 33: 333-340.
- Ramos MC (2001) Divisive and Hierarchical Clustering Techniques to Analyze Variability of Rainfall Distribution Patterns in a Mediterranean Region. *Atmospheric Research* 57: 123–138.
- Singh D, Singh R, Anuragh SC, Rao VUM, Singh S (2010) *Agro Climatic Atlas of Haryana*. Technical Bulletin No. 15, Department of Agricultural Meteorology, CCS HAU, Hisar.Pp 80.
- Shirin AHS, Thomas R (2016) Regionalization of Rainfall in Kerala State. *Procedia Technology* 24: 15-22.
- Swaminathan S, Meganathan S (2018) Identifying efficient clustering techniques for classifying rainfall data. *Asian Journal of Microbiology, Biotechnology and Environmental Sciences* 20:566-568
- Siral-Ud-Doulah M, Islam MN (2019) Defining Homogenous climatic Zones of Bangladesh using Cluster Analysis. *International Journal of Statistics and Mathematics* 6:119-129.
- Terassi PMB, Galvani E (2017) Identification of Homogeneous Rainfall Regions in the Eastern Watersheds of the State of Paraná, Brazil. *Climate* 53: 1-13.
- Venkatesh B, Jose MK (2007) Identification of Homogeneous Rainfall Regimes in Parts of Western Ghat Regions of Karnataka. *Journal of Earth System Science* 116:321-329.
- Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association* 58: 236–244.
- Yashwant S, Sananse SL (2015) Comparisons of Different Methods of Cluster Analysis with Application to Rainfall Data. *International Journal of innovative Research in Science, Engineering and Technology* 4: 10861-10872.