Received: 2 November 2017 | Revised: 3 August 2018 | Accepted: 24 September 2018 | Published on: 16 January 2019

DOI: 10.1002/met.1763

Meteorological Applications RMetS

RESEARCH ARTICLE

Clustering of rainfall stations and distinguishing influential factors using PCA and HCA techniques over the western dry region of India

Deepesh Machiwal[1] | Sanjay Kumar[2*] | Hari M. Meena[3] | Priyabrata Santra[4] | Ranjay K. Singh[4] | Dharam V. Singh[3]

[1]Regional Research Station, ICAR-Central Arid Zone Research Institute, Bhuj, Gujarat, India

[2]Krishi Vigyan Kendra, ICAR-CAZRI, Bhuj, India

[3]Division of Natural Resources, ICAR-Central Arid Zone Research Institute, Jodhpur, Rajasthan, India

[4]Division of Agricultural Engineering and Renewable Energy, ICAR-Central Arid Zone Research Institute, Jodhpur, Rajasthan, India

Correspondence
Deepesh Machiwal, Regional Research Station, ICAR-Central Arid Zone Research Institute, Kukma-Bhuj, Gujarat, India 370105.
Email: dmachiwal@rediffmail.com

[*]**Present address**
College of Forestry, Banda University of Agriculture and Technology, Banda-210001, Uttar Pradesh, India.

This study used hierarchical cluster analysis (HCA) to delineate the spatial patterns of monthly, seasonal and annual rainfall by clustering 62 stations in the western arid region of India based on a 55 year (1957–2011) data set. The statistical properties of clusters were computed and box–whisker plots plotted. Furthermore, the relative influence of three geographical factors (longitude, latitude and altitude) and five statistical parameters (the mean, standard deviation (SD), co-efficient of variation (CV), and maximum and minimum rainfall) on mean rainfall was investigated using principal component analysis (PCA). The use of HCA resulted in four rainfall clusters geographically located at a distinct position. Cluster I, characterized by the lowest mean rainfall and highest CV, was located in the western portion, whereas mean rainfall was the highest for cluster IV situated in the eastern portion. Box–whisker plots revealed a slight skewness, although the monsoon and annual rainfall followed a normal distribution. The PCA results indicted two to three significant principal components (PCs) with eigenvalues > 1. In four clusters, two PCs explained the major variance, ranging from 69.41% (June) to 91.83% (August) in monthly rainfall, from 63.62% (monsoon) to 93.30% (post-monsoon) in seasonal rainfall, and from 71.48% to 90.73% in annual rainfall. In monthly and seasonal rainfall, first PC 1 is termed the "mean rainfall component", which has strong to moderate associations with longitude, and is equally opposed by the CV. These findings are vital for planners and decision-makers to formulate strategies to manage unusual rainwater quantities.

KEYWORDS

arid region, geographical factors, hierarchical cluster analysis, principal component analysis, rainfall, statistical parameters

1 | INTRODUCTION

Rainfall, which is the primary water source for global agriculture, is highly variable over space and time, and is an indicator for climate variability/change (Frich *et al*., 2002). Thus, knowledge about the spatial patterns of rainfall is essential for sustainable agricultural water management and to detect climate variability. The importance of spatial pattern recognition is high for low rainfall areas, especially arid and semi-arid regions encompassing about 35% (> 61 million km^2) of global lands (Mares, 1999). The arid region of India, which encompasses > 38.7 million ha, experiences both hot and cold climates. Of total hot–arid lands, a major part (28.57 million ha) exists in the northwest area of the country, with low, uncertain and highly variable rainfall (Kar *et al*., 2009). Therefore, it is imperative to understand the spatial rainfall patterns in the region.

Several methods have been used in the literature to investigate spatial rainfall patterns, for example, L-moments (Guttman, 1993), harmonic analysis (Suhaila and Jemain, 2009), multivariate regression (Sabziparvar *et al*., 2015), spatial correlation functions (Şen and Habib, 2001), spatial interpolation (Gupta *et al*., 2017), *k*-means clustering (Machiwal *et al*., 2017), empirical orthogonal functions (Jebari *et al*., 2009), Pearson's correlation (Haines and Olley, 2017), regional frequency analysis (Medina-Cobo *et al*., 2017), and support vector machines (Lin *et al*., 2017). Spatial pattern recognition by cluster analysis (CA) and principal component analysis (PCA) are other methods used to find homogenous rainfall clusters (Modarres and Sarhadi, 2011). Amissah-Arthur and Jagtap (1999) applied both PCA and CA to seasonal rainfall in Nigeria in order to cluster 23 stations into six groups. Decreasing rainfall trends in four delineated groups were identified. Golian *et al*. (2010) compared natural break and revised fuzzy *c*-means methods by clustering 25 stations using rainfall of northern Iran for the period 1975–2008 and showed coherence between outcomes of both methods. Again in Iran, regional rainfall patterns were quantified using CA and L-moments together using annual rainfall of 137 stations for the period 1952–2003 (Modarres and Sarhadi, 2011). In other study, Nnaji *et al*. (2016) grouped 24 rainfall stations into five clusters in Nigeria based on the co-efficient of variation (CV) through hierarchical cluster analysis (HCA). Furthermore, linkages between mean annual rainfall and the CV resulted in linear, power law and logarithmic relationships. Recently, Machiwal *et al*. (2017) classified 32 stations in hot and cold Indian arid regions into entirely and partially arid clusters based on 102 year (1901–2002) rainfall using *k*-means clustering. The results indicated prominent increasing trends in annual and wet season rainfall for the entirely arid cluster. In the literature, studies combining PCA and HCA to delineate rainfall clusters and identify major factors associated with clusters are rare (e.g. Darand and Daneshvar, 2014).

The present study was undertaken with two objectives: (1) to delineate rainfall clusters by HCA and to understand their spatial patterns at monthly, seasonal and annual scales in the northwest hot–arid lands of India; and (2) to identify the influential geographical and statistical factors by PCA. The study is the first of its kind to find significant influencing factors for spatial rainfall patterns by PCA in HCA-delineated clusters.

2 | STUDY AREA AND DATA DESCRIPTION

The western dry region of India (the study area), lying west of the Aravalli Hills, is located between latitudes 24 ° 37 ' 00 " and 30 ° 10 ' 48 " N and longitudes 69 ° 29 ' 00 " and 76 ° 05 ' 33 " E. The study area comprises 12 districts of the state of Rajasthan, namely, Barmer, Bikaner, Churu, Hanumangarh, Jaisalmer, Jalor, Jhunjhunu, Jodhpur, Nagaur, Pali, Sikar and Sriganganagar (Figure 1). Despite the adversity of the climate,

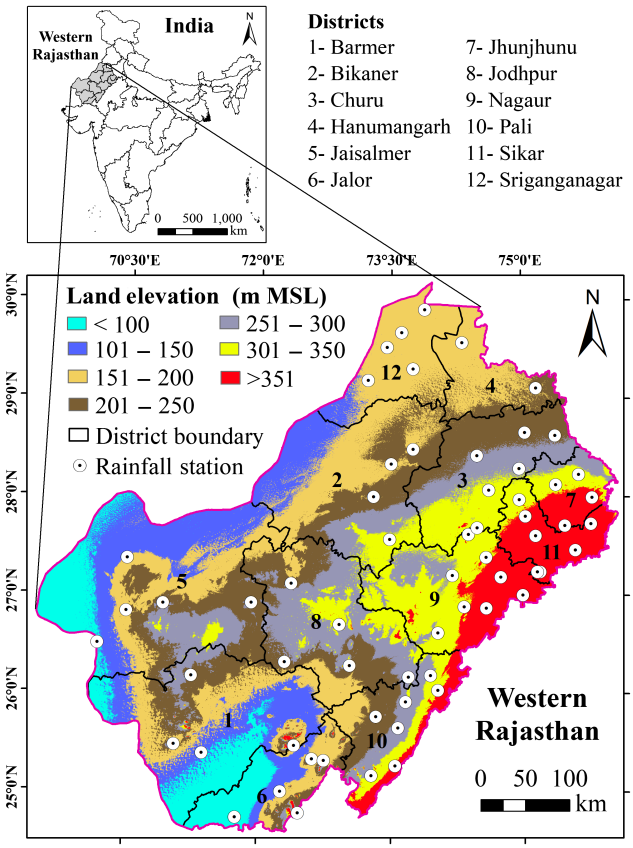

FIGURE 1 Study area in (western Rajasthan) in India, and sites of the rainfall stations [Colour figure can be viewed at wileyonlinelibrary.com]

a major portion of India's arid lands fall within the study area (62% of hot–arid lands) (Kar *et al*., 2009). The climate is characterized by low annual rainfall (100–500 mm) and a high potential evapotranspiration (1400–2000 mm). In summer, the mean maximum temperature ranges from 40 to 42°C, with May being the hottest month. In winter, the mean monthly maximum temperature ranges from 22 to 29°C, and the minimum temperature from 4 to 14°C (Moharana *et al*., 2016).

Daily rainfall data for 55 years (1957–2011) from 62 stations (acquired from the Water Resources Department of the Government of Rajasthan) were used to prepare monthly, seasonal and annual series. This study considered four monsoon months, June, July, August and September, and four seasons, monsoon (June–September), post-monsoon (October–December), winter (January–February) and pre-monsoon (March–May). The longitudes and latitudes of the stations are shown in a map of the study area (Figure 1). Missing data for some stations in a few years were estimated by the inverse distance-weighting method.

3 | MATERIALS AND METHODS

3.1 | HCA of rainfall series

The CA helps to group variables into clusters according to the high similarity of their features, such as geographical,

physical, statistical or stochastic properties (Modarres and Sarhadi, 2011). The CA is of two types: hierarchical and non-hierarchical (Otto, 1998). The HCA is the common approach where clusters of variables are sequentially formed, where each cluster depicts the least variance (or smallest dissimilarity) of variables. The present study considered Ward's agglomerative hierarchical algorithm (Ward, 1963) as a dissimilarity measure using Euclidean distance, as follows (Everitt and Dunn, 1991):

$$d_e = \left[\sum_{i=1}^{n} \left(P_{p,i} - P_{q,i}\right)\right]^{1/2} \quad (1)$$

where d_e is Euclidean distance; and $P_{p,i}$ and $P_{q,i}$ are quantitative variables i of individuals p and q, respectively.

The HCA was performed for monthly, seasonal and annual rainfall by using STATISTICA software (StatSoft Inc., 2004).

3.2 | Computing the statistical properties of delineated rainfall clusters

Salient statistics, the mean, SD, CV, skewness and kurtosis, were computed for rainfall clusters of monthly, seasonal and annual series. Furthermore, box–whisker plots were drawn to explore a cluster-wise summary of five key statistics along with outliers and extremes (USEPA, 1998).

3.3 | Exploring the relationship of geographical and statistical factors with rainfall

The study investigated the linear relationship between mean rainfall and geographical and statistical parameters by computing the correlation co-efficient (R) for two pairs before identifying the major factors dominating each cluster. Three geographical variables (longitude, latitude and altitude) and four statistical factors (the SD, CV, and minimum and maximum rainfall) were considered to find the extent of the relationship between these variables and the mean rainfall of monthly, seasonal and annual series.

3.4 | Applying PCA to identify the dominating features/components of rainfall clusters

The PCA was used to reduce dimensionality to obtain an economic description of the rainfall profile, where differences between various profiles may be clearly evidenced (Davis, 2002). The use of the PCA helps to recognize patterns by explaining the variance of a large set of intercorrelated variables, and it transforms them into a smaller set of independent factors (PCs) (Dillon and Goldstein, 1984). Significant PCs were chosen by the Kaiser (1958) criterion (eigenvalue > 1), which best describes the variance of the analysed variables with reasonable interpretation (Harman, 1960).

Rainfall $R(i,j)$ at the i-th station in the j-th year is expressed as the sum of the products of the co-efficients $A_n(i)$ varying over space, and associated with a temporal pattern or eigenvectors $B_n(j)$, as follows (Gadgil and Iyengar, 1980):

$$R(i,j) = \sum_{k=1}^{n} [A_n(i)] \times [B_n(j)] \quad (2)$$

where $i = 1, \ldots, N$; $j = 1, \ldots, n$; and $B_n(j)$ is the eigenvectors of correlation matrix.

The study applied the PCA to the geographical and statistical parameters of stations grouped under clusters. Stations' altitudes were taken from the digital elevation model of the Shuttle Radar Topographic Mission. The PCA aimed to find the relative influence of each variable explaining the variance of the system in each separate cluster for monthly, seasonal and annual series.

4 | RESULTS AND DISCUSSION

4.1 | Clustering of rainfall stations

The HCA delineated four clusters of 62 stations for monthly, seasonal and annual series, which were geographically located over space (Figure 2). It can be seen that stations exist in close proximity to each other in every cluster. Likewise, clustering resulted in the geographical contiguity of stations in some of the delineated groups in Nigeria (Amissah-Arthur and Jagtap, 1999). Also, clusters depict a definite geographical pattern justifying clustering. The spatial pattern of clusters slightly varies over months, seasons and year along with variation in the number of stations in every cluster. A careful examination revealed that stations of cluster I are mainly situated towards the western portion of the area in most months, seasons and year (Figure 2). On the contrary, stations of cluster IV are mainly located in the eastern portion. Likewise, cluster III exists in the northern and northeast portions, whereas cluster II is seen towards the southern portion.

4.2 | Statistical properties of rainfall clusters

Cluster-wise statistics of the mean rainfall over months, seasons and year are presented in Table 1. The mean rainfall remains the lowest in cluster I and the highest in cluster IV, except in August–September and the pre-monsoon. In general, the SD of rainfall is the least for cluster I and the most for cluster II. Both skewness and kurtosis values for all clusters of annual rainfall are within the permissible limits (±2.0) for a normal distribution. Similarly, skewness and kurtosis values are within normality limits for July–September and the monsoon, except for a single cluster over these periods. However, the mean rainfall deviates from normality in all clusters during June and the post-monsoon. A positive skewness indicates that rainfall is slightly right skewed in all clusters in all periods. It is further revealed that both the mean monthly rainfall and SD are the highest in August in all clusters, although the CV is

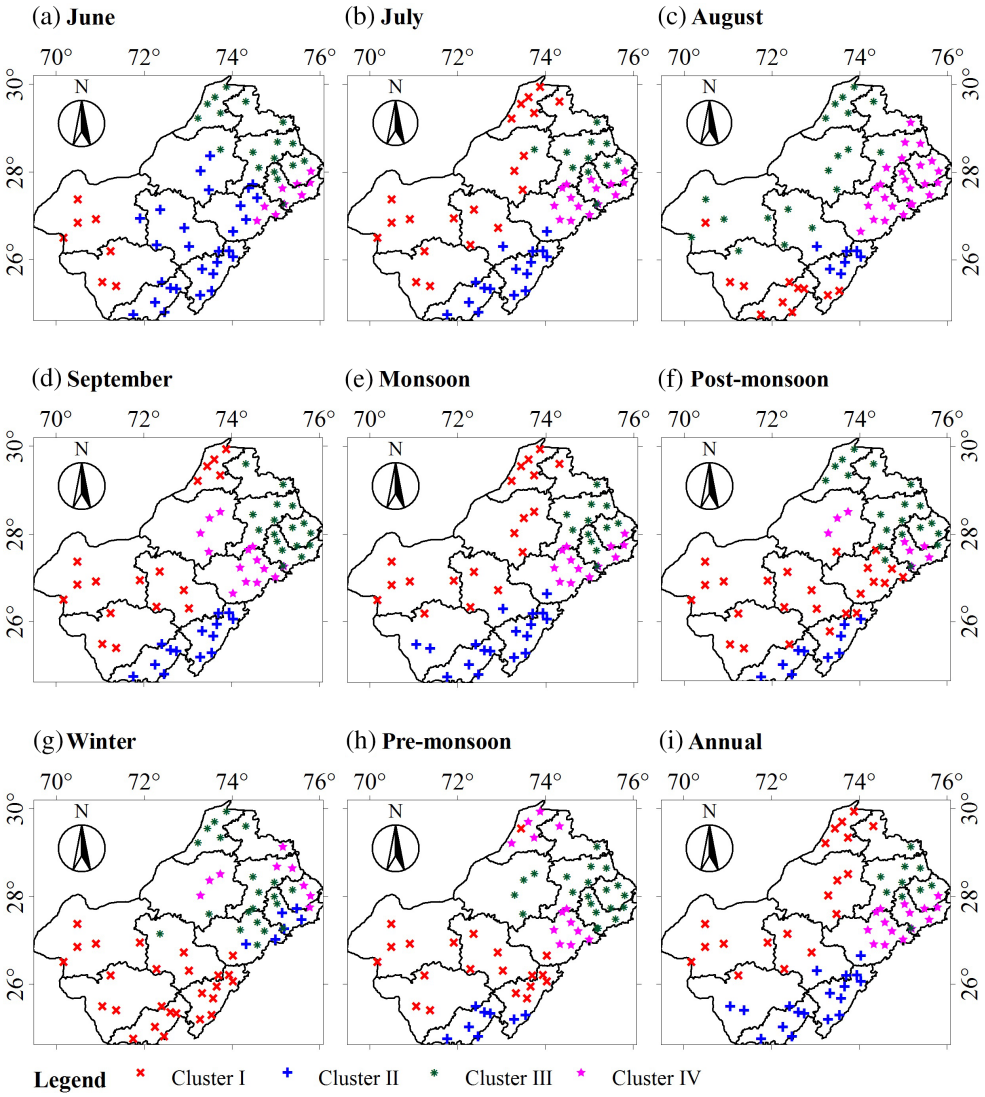

FIGURE 2 Clustered rainfall stations classified into four groups for monthly (a–d), seasonal (e–h) and annual (i) time-series [Colour figure can be viewed at wileyonlinelibrary.com]

comparatively low in July, indicating least temporal variations. In the monsoon, the mean and SD are the highest with the least amount of rainfall variations (CV < 45%) in all clusters in comparison with the other three drier seasons with contrasting statistics. The CV remains the lowest (35–42%) in annual rainfall in comparison with that in months and seasons. This indicates that rainfall variations are relatively high over the shorter periods of months and seasons containing scanty rainfall.

The box–whisker plots of four clusters for monthly, seasonal and annual scales are depicted in Figures 3a–c. The median rainfalls are the lowest for cluster I in June–September and for cluster III in August. An important finding is the larger length of the upper whisker compared with the lower whisker, and the presence of outliers and extremes towards the upper whisker. This indicates that the distribution of monthly rainfall is right skewed, which is in agreement with the results of the skewness values. Among seasons, rainfall distribution is more uniform in the pre-monsoon and winter, as revealed from fewer outliers/extremes in comparison with the post-monsoon (Figure 3b). Among all periods, the monsoon and annual rainfall are more uniformly distributed over time, and follow normality (Figure 3c). It is evident that cluster-wise patterns are quite comparable for the monsoon and annual rainfall (Figure 3c). The lowest median rainfalls are in cluster I, while the medians of clusters II and IV are relatively high, although rainfall distribution in cluster II is more uniform as outliers/extremes are absent (Figure 3c).

4.3 | Rainfall correlations with geographical factors and statistical parameters

Correlation co-efficients (R) are presented in Table 2. The minimum rainfall is mostly zero in monthly and seasonal rainfalls, and was not analysed. The mean rainfall for June is strongly correlated ($R > 0.70$) with longitude, altitude and the SD, and is moderately correlated ($R > 0.50$) with maximum rainfall. This finding suggests that monsoon rains are quite good at higher altitudes near the upper

TABLE 1 Statistics for cluster-wise mean rainfall for the monthly, seasonal and annual series

Period	Cluster	Stations	Range (mm)	Mean (mm)	Standard deviation (mm)	Co-efficient of variation (%)	Skewness	Kurtosis
June	I	7	0–262.6	**24.4**[a]	38.9	**160**[b]	4.6[b]	26.5[b]
	II	28	0.7–180.4	40.9	**31.6**[a]	**77**[a]	2.1[b]	6.4[b]
	III	18	1.2–146.4	39.8	32.1	81	1.6	2.6[b]
	IV	9	1.2–413.4	**58.5**[b]	**64.0**[b]	109	3.7[b]	18.0[b]
July	I	20	0.8–181.1	**73.5**[a]	**39.7**[a]	**54**[a]	0.8	0.2
	II	16	2.1–367.5	155.4	**89.2**[b]	57	0.6	−0.4
	III	11	9.5–566.2	124.4	80.5	**65**[b]	3.1[b]	16.0[b]
	IV	15	7.9–409.5	**155.5**[b]	88.3	57	1.1	1.1
August	I	11	3.1–626.3	127.9	**115.0**[b]	**90**[b]	2.5[b]	8.1[b]
	II	7	7.9–415.6	**136.4**[b]	97.4	71	1.0	0.6
	III	19	1.9–192.5	**63.1**[a]	**41.4**[a]	66	0.9	0.8
	IV	25	9.2–284.6	121.9	68.4	**56**[a]	0.6	−0.3
September	I	17	0.5–125.2	**28.9**[a]	**33.1**[a]	**115**[b]	1.5	1.3
	II	14	0–318.1	**64.2**[b]	**73.8**[b]	115	1.7	2.4[b]
	III	17	0.4–202.4	52.8	47.9	**91**[a]	1.3	1.3
	IV	14	0–210.2	42.7	41.9	98	1.9	4.9[b]
Monsoon	I	19	41–406.8	**192.3**[a]	**76.5**[a]	**40**[a]	0.5	0.3
	II	18	99.6–801.8	382.0	**170.4**[b]	**45**[b]	0.6	−0.3
	III	12	113.6–768.3	343.0	142.7	42	0.9	0.6
	IV	13	114.6–933.9	**389.7**[b]	155.3	40	1.2	2.4[b]
Post-monsoon	I	24	0–71.2	**10.8**[a]	**15.4**[a]	**143**[a]	2.5[b]	6.9[b]
	II	10	0–136.3	17.6	29.6	168	2.4[b]	5.9[b]
	III	20	0–96.4	12.7	19.5	154	2.7[b]	7.5[b]
	IV	8	0–231.3	**18.9**[b]	**36.3**[b]	**192**[b]	4.3[b]	22.0[b]
Winter	I	26	0–39.4	**5.1**[a]	**7.0**[a]	**138**[b]	2.6[b]	9.4[b]
	II	6	0–87.7	14.1	**18.2**[b]	129	1.8	4.0[b]
	III	21	0–45	11.6	12.2	**105**[a]	1.2	0.5
	IV	9	0–72.8	**15.8**[b]	16.9	106	1.7	3.0[b]
Pre-monsoon	I	20	0.7–62.6	16.0	**16.3**[a]	101	1.5	1.4
	II	8	0–133.3	**15.5**[a]	**27.8**[b]	**179**[b]	3.4[b]	11.6[b]
	III	21	0.5–116.3	**30.0**[b]	**27.8**[b]	93	1.7	2.5[b]
	IV	13	2.5–101.6	26.2	23.2	**89**[a]	1.7	2.7[b]
Annual	I	19	67.8–432.6	**231.8**[a]	**81.8**[a]	**35**[a]	0.3	−0.3
	II	18	128.3–845.2	418.6	**177.2**[b]	**42**[b]	0.5	−0.5
	III	10	182.7–822.1	390.4	144.9	37	0.7	0.3
	IV	15	157.8–961.7	**448.3**[b]	161.6	36	1.0	1.5

Notes: Entries shown in bold indicate the minimum/maximum among four clusters.
[a] Minimum (maximum) among four clusters of a given month/season.
[b] Skewness/kurtosis outside the limit of ±2.0.

portion of the Aravalli range. Similarly, rainfall in July has a very strong correlation ($R > 0.80$) with the SD, a strong correlation with maximum rainfall and a moderate correlation with longitude, altitude and the CV. August rainfall is strongly related to the SD, and moderately related to altitude and maximum rainfall. *R*-values in September indicate very strong linkages between the mean and both the SD and maximum rainfall. Among seasons, the monsoon and post-monsoon rainfall revealed very strong correlations with the SD, and moderate correlations with longitude and altitude. In addition, maximum rainfall has strong and moderate relations with means in the monsoon and post-monsoon, respectively. Winter rainfall is very strongly correlated with longitude and the SD, strongly correlated with latitude and the CV, and moderately correlated with maximum rainfall. However, pre-monsoon mean rainfall is strongly related to longitude, latitude, the SD and CV, and moderately related to maximum rainfall. In annual rainfall, the *R*-value suggests very strong relationships with the SD and minimum rainfall, a strong relationship with maximum rainfall, and a moderate relationship with longitude and altitude. It is inferred that the SD has very strong to strong relationships with the mean rainfall of every month, season and year. However, rainfall linkages with the CV and latitude are strongly related only with winter and the pre-monsoon.

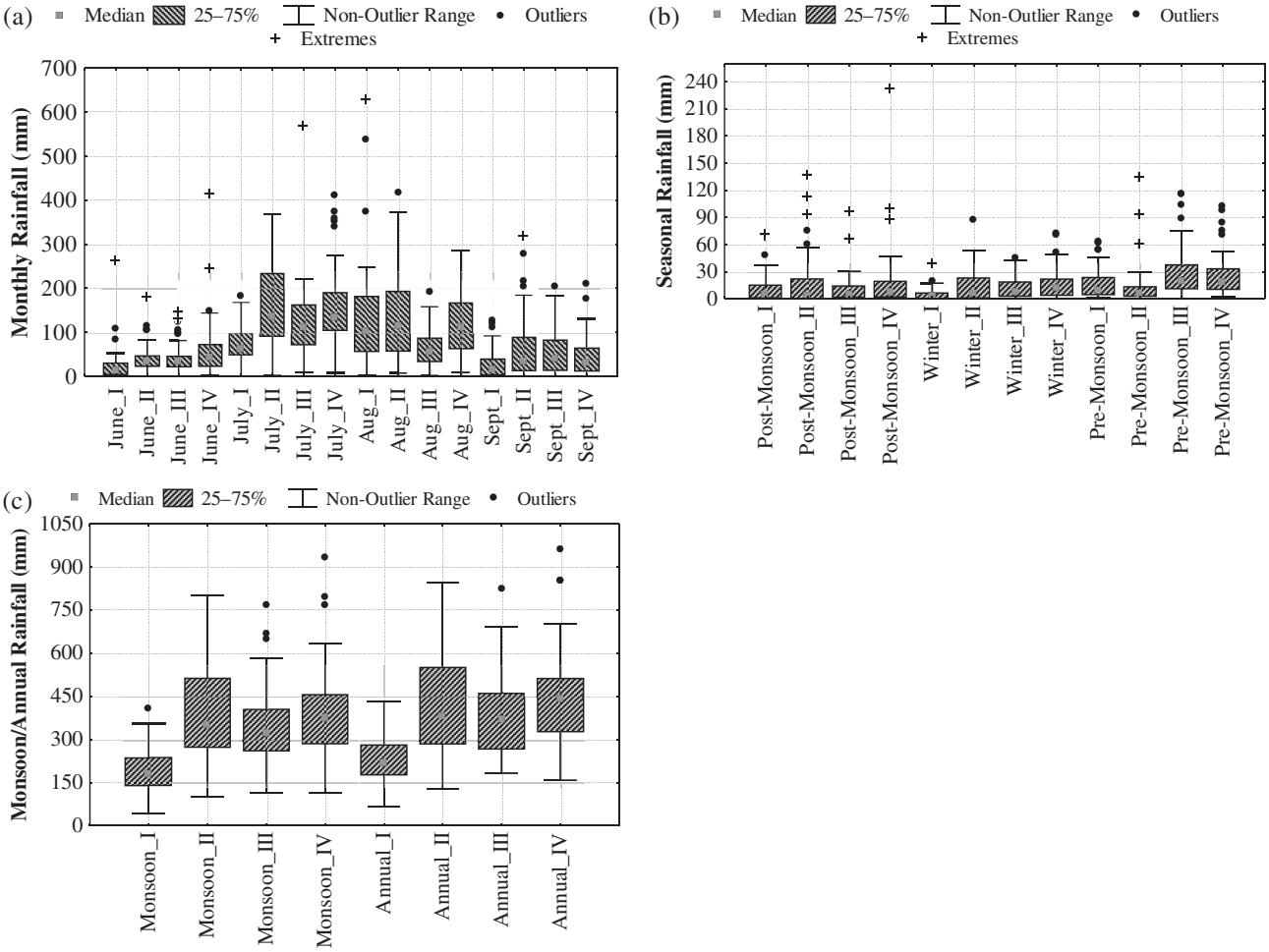

FIGURE 3 Box–whisker plots of cluster-wise rainfall for (a) monthly, (b) seasonal and (c) monsoon/annual rainfall series

TABLE 2 Correlation co-efficient (R) depicting the relationships between mean rainfall and the geographical and statistical parameters

	Geographical factors			Statistical parameters			
Period	**Longitude**	**Latitude**	**Altitude**	**Standard deviation**	**Co-efficient of variation**	**Minimum rainfall**	**Maximum rainfall**
June	**0.73**[a]	0.00	**0.78**[a]	**0.71**[a]	−0.47	n.a.	**0.51**[b]
July	**0.57**[b]	−0.36	**0.63**[b]	**0.90**[c]	**−0.62**[b]	n.a.	**0.75**[a]
August	0.46	−0.46	**0.58**[b]	**0.72**[a]	−0.42	n.a.	**0.53**[b]
September	0.41	−0.44	0.45	**0.90**[c]	−0.47	n.a.	**0.80**[c]
Monsoon	**0.55**[b]	−0.38	**0.63**[b]	**0.85**[c]	−0.49	n.a.	**0.78**[a]
Post-monsoon	**0.51**[b]	−0.23	**0.54**[b]	**0.87**[c]	−0.33	n.a.	**0.68**[b]
Winter	**0.83**[c]	**0.78**[a]	0.46	**0.91**[c]	**−0.70**[a]	n.a.	**0.60**[b]
Pre-monsoon	**0.78**[a]	**0.70**[a]	0.47	**0.75**[a]	**−0.75**[ba]	n.a.	**0.52**[b]
Annual	**0.63**[b]	−0.29	**0.67**[b]	**0.82**[c]	−0.48	**0.85**[c]	**0.77**[a]

Notes: Entries shown in bold indicate $R > 0.5$. n.a.: not applicable.
[a] Strong relationship ($0.8 > R \geq 0.7$).
[b] Moderate relationship ($0.7 > R \geq 0.5$).
[c] Very strong relationship ($R \geq 0.8$).

Similar findings are reported for Nigeria, where monthly rainfall has a power-law relationship with the CV (Nnaji *et al*., 2016). Looking at moderate to very strong linkages for geographical and statistical parameters with the mean, this study subsequently used the PCA to identify the influential geographical and statistical parameters dominating each cluster.

Before applying the PCA, a Gleason–Staelin redundancy measure (Gleason and Staelin, 1975) and Bartlett's test of sphericity (Bartlett, 1950) were used to evaluate the appropriateness of the variables to be used as inputs into the PCA. The redundancy measures for four clusters varied from 0.43 to 0.67 for monthly, from 0.39 to 0.71 for seasonal and from 0.44 to 0.60 for annual series. This indicates a moderate to

good correlation among the variables. Furthermore, the results of Bartlett's chi-squared test statistics were found to be statistically significant ($p < 0.05$) for all clusters for monthly, seasonal and annual series. Hence, the null hypothesis that the correlation matrix is an identity matrix was rejected, which suggests that the variables are appropriate to perform a PCA.

4.4 | Significant principal components (PCs) for rainfall clusters in individual months

The PCA resulted in six to eight PCs in clusters, although only the first two or three PCs were found to be significant following Kaiser's criterion of an eigenvalue > 1. A major portion of the variance was mostly explained by the first two PCs for monthly (75.36–88.44% for June, 69.41–83.53% for July, 74.60–91.83% for August and 73.70–86.58% for September), seasonal (63.62–90.39% for monsoon, 81.82–93.30% for post-monsoon, 75.74–88.60% for winter and 64.31–77.77% for pre-monsoon) and annual series (71.48–90.73%) (Figures 4–6). Thus, factor co-ordinates for the variables for the first two PCs were subsequently plotted against each other on the unit circle plots in order to examine their associations with two significant PCs (Figure 4) for four monthly clusters.

To scrutinize variables associated with two significant PCs, their co-ordinates were classified into strong (PC co-ordinate ≥ 0.75) and moderate ($0.75 >$ PC co-ordinate ≥ 0.50) associations following the criteria of Liu *et al.* (2003) (Table 3). In most monthly clusters, the mean (Avg) has either a strong or a moderate association with PC 1, which is steadier than that with PC 2. Hence, PC 1 is termed the "mean rainfall component" (Table 3). In June, cluster I is characterized by a strong positive association with the mean rainfall and longitude and a strong negative association with latitude and the CV, whereas cluster II is characterized by a strong negative association with longitude, altitude, the mean and CV, and a moderate negative association with latitude, the mean and maximum rainfall. In cluster III, a strong positive association with latitude, a strong negative association with longitude, altitude, the mean, SD and maximum rainfall, and a moderate positive association with the CV characterize PC 1. However, PC 1 of cluster IV is characterized by a strong positive linkage for longitude, latitude, altitude and the mean, a strong negative linkage for the CV, and a moderate negative linkage for the SD and maximum rainfall. In July, PC 1 of cluster I has a strong positive association with the mean, a negative association with the CV, and a moderate positive association with longitude, altitude and the SD, while cluster II represents strong positive associations with longitude and altitude, a negative association with the CV, and a moderate positive association with the mean. PC 2 of cluster III in July depicts moderate negative associations with altitude, the mean, SD and maximum rainfall, whereas cluster IV reveals strong positive linkages with longitude, altitude and the mean, a strong negative linkage with the CV, a moderate positive latitude and a negative maximum rainfall. For August, cluster I represents a strong positive association with longitude and the mean, a strong negative association with latitude and the CV, and a moderate negative maximum rainfall. Cluster II has a strong positive latitude, and a strong negative mean, SD, CV and maximum rainfall. In cluster III, PC 2 corresponds to strong negative co-ordinates for altitude and the mean, and a moderate negative co-ordinate for the SD, while PC 1 of cluster IV is represented by strong negative co-ordinates for altitude, the mean, SD and maximum rainfall. In September, PC 1 of cluster I is characterized by a moderate positive relationship with latitude, and strong negative relations of the mean, SD and maximum rainfall. However, PC 2 of cluster II is characterized by a strong positive co-ordinate for altitude, moderate positive co-ordinates for longitude, the mean and SD. Similarly, PC 2 of cluster III is characterized by strong positive relations of longitude, altitude and the mean, a moderate positive co-ordinate for the SD, a strong negative co-ordinate for latitude, and a moderate negative co-ordinate for the CV. However, PC 1 of cluster IV is characterized by strong positive co-ordinates for longitude, altitude, the mean and SD, a moderate positive co-ordinate for maximum rainfall, and strong negative co-ordinates for latitude and the CV.

Overall, it is apparent that maximum rainfall and the SD of monthly rainfall mostly remain close to each other. This indicates that the SD should be higher for a month with higher maximum rainfall, and *vice versa.* Further, the CV on the unit circle plot is always opposed by the mean, suggesting that months with low rainfall exhibit large variations.

4.5 | Dominating factors associated with clusters of seasonal and annual rainfall

The unit circle plots were first drawn between two significant PCs for seasonal (Figure 5) and annual (Figure 6) rainfall. For two PCs, the strong and moderate associations of variables for seasonal rainfall are summarized in Table 4. Similar to months, the mean rainfall over seasons has a strong association with PC 1 in most clusters and seasons, and thus PC 1 is termed the "mean rainfall component". It can be seen that cluster I of monsoon rainfall has strong associations with longitude (negative) and the CV (positive), and a moderate association with the mean (negative), SD (positive) and maximum rainfall (negative). However, cluster II has a strong coherence for longitude, altitude, the mean and maximum, and is strongly opposed by the CV. Cluster III has strong linkages for altitude and the SD, and a moderate linkage for the mean, which is opposed by a moderate linkage for latitude, while cluster IV is strongly linked to longitude, altitude, the mean and maximum rainfall, moderately linked to latitude, and strongly opposed by the CV. Likewise, for the post-monsoon, cluster I is strongly related to longitude, the mean and SD, moderately linked to altitude and maximum rainfall, and moderately opposed by the CV. Cluster II has strong linkages with the mean, SD, CV and maximum rainfall, and moderately linked to altitude, whereas

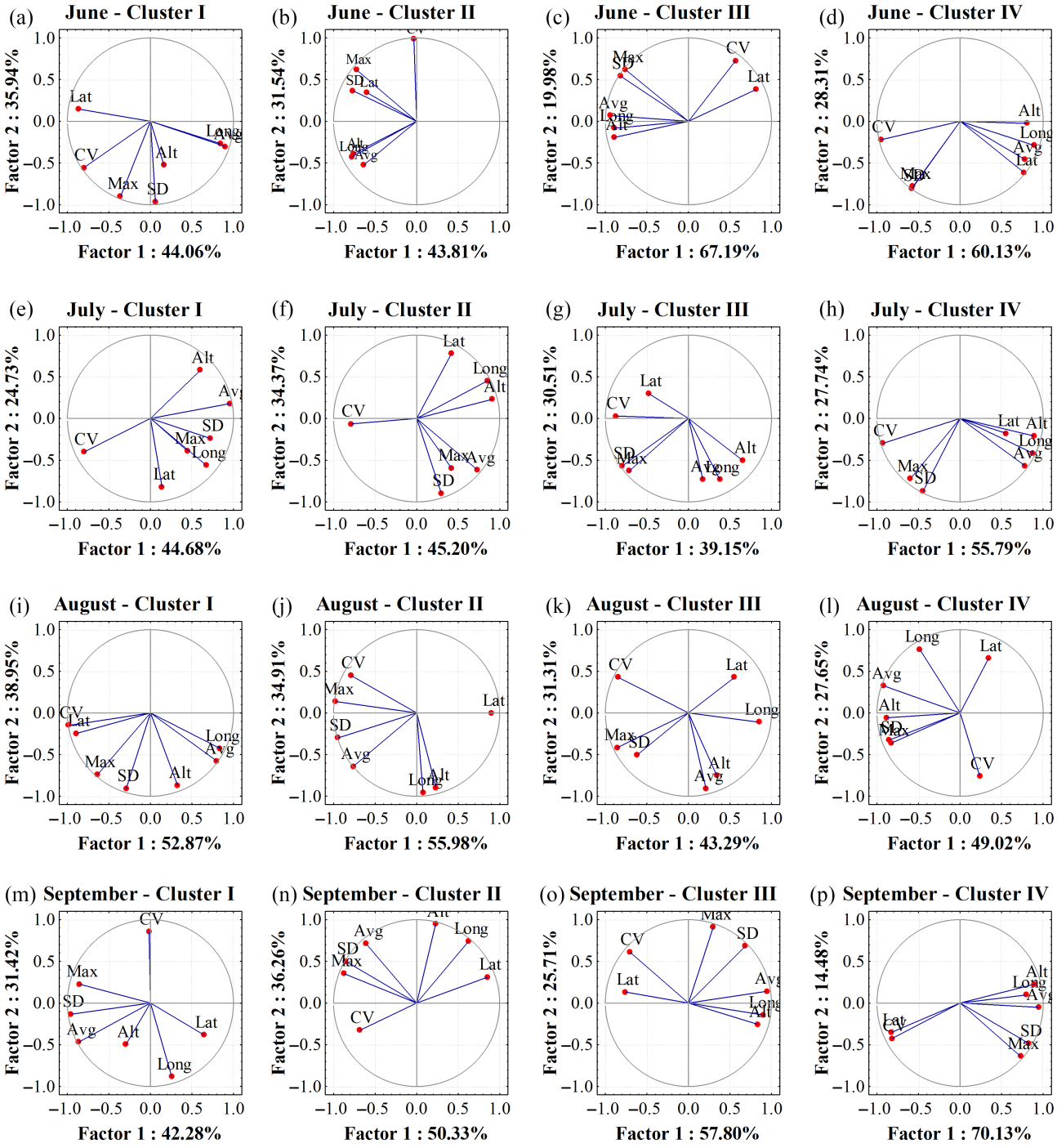

FIGURE 4 Unit circle plots of the two most significant principal components showing the association of seven influential variables with average monthly rainfall: Alt: altitude; Avg: average; CV: co-efficient of variation; Lat: latitude; Long: longitude; Max: maximum; and SD: standard deviation [Colour figure can be viewed at wileyonlinelibrary.com]

cluster III is strongly linked to longitude, altitude and the mean, which is strongly opposed to latitude and moderately to the CV. However, cluster IV is strongly associated with longitude, altitude, the mean and SD, and moderately to maximum rainfall, which is opposed by a strong linkage for latitude and a moderate linkage for the CV. In winter, cluster I is strongly related to longitude, altitude and the mean, and moderately related to the SD. In cluster II, PC 1 has a strong association with longitude, latitude, altitude and the mean, which is strongly opposed by the CV. PC 1 in cluster III has strong positive co-ordinates for latitude and the mean, and moderate co-ordinates for the SD and maximum rainfall, which is opposed by moderate co-ordinates for latitude and the CV. Cluster IV has strong linkages for longitude and the mean, and a moderate linkage for altitude, which is strongly opposed by the CV. In pre-monsoon, PC 1 of cluster I is represented by strong co-ordinates for the mean and SD, and moderate co-ordinates for longitude, altitude and maximum rainfall, which is moderately opposed by the CV, whereas cluster II is strongly favoured by longitude,

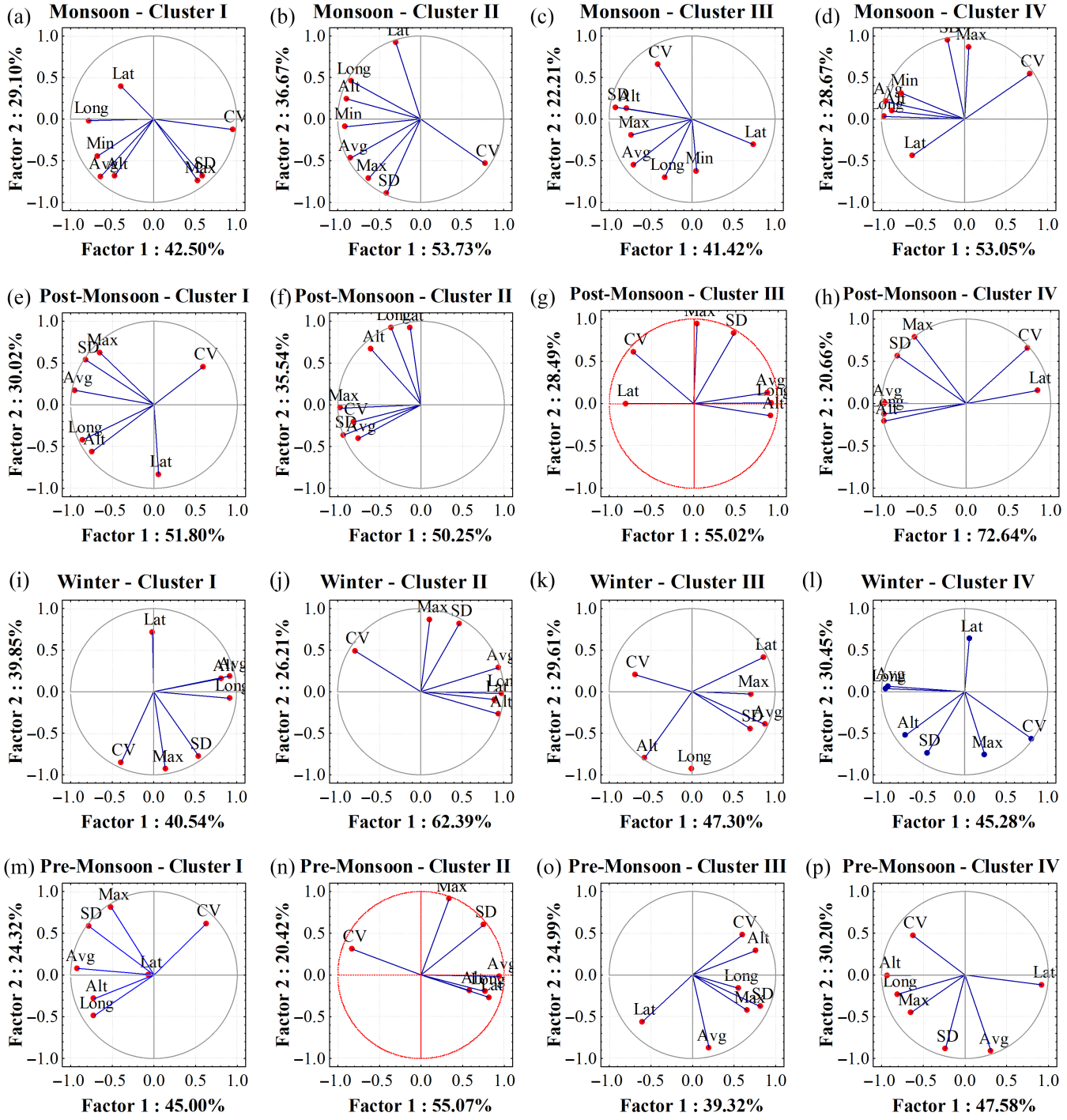

FIGURE 5 Unit circle plots of the two most significant principal components showing the association of eight influential variables with average seasonal rainfall: Alt: altitude; Avg: average; CV: co-efficient of variation; Lat: latitude; Long: longitude; Max: maximum; Min: minimum; and SD: standard deviation [Colour figure can be viewed at wileyonlinelibrary.com]

latitude, the mean and SD, moderately favoured by altitude, and strongly opposed by the CV. However, PC 2 of cluster III is characterized by a strong negative co-ordinate for the mean and a moderate negative co-ordinate for latitude, while cluster IV has strong associations with the mean and SD. It is evident that the maximum rainfall and SD for all clusters in the seasons remain close to each other, which is in agreement with the finding of monthly rainfall. Another prominent feature is that the CV, distinctly located to other variables, is mostly opposed by latitude and the mean.

In annual rainfall (Figure 6), PC 1 of cluster I has strong positive associations with longitude, the mean and maximum rainfall, a moderate association with latitude, and is strongly opposed by the CV. For cluster II, PC 1 has strong negative linkages for longitude, altitude, the mean and minimum, and a moderate negative linkage for maximum rainfall, which is strongly opposed by the CV (Table 4). In cluster III, PC 2 is characterized by strong negative co-ordinates for longitude and the mean, and is moderately negative for minimum and maximum rainfall. On the other hand, cluster IV has PC

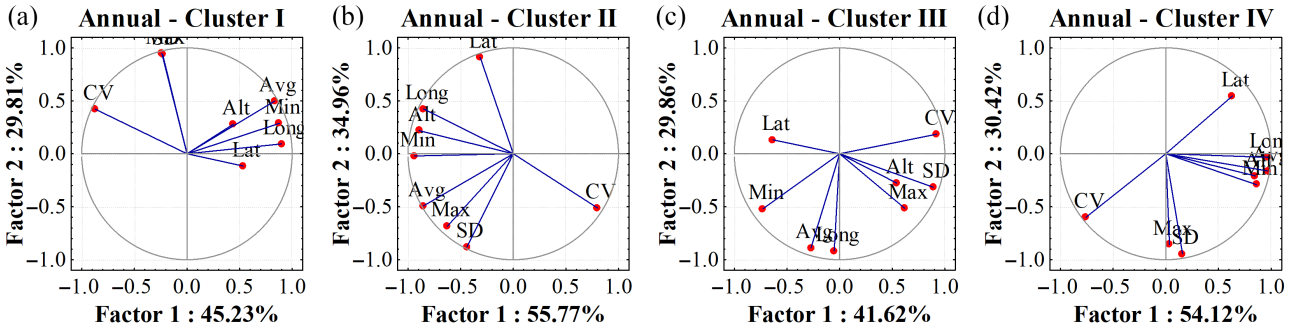

FIGURE 6 Unit circle plots of the two most significant principal components showing the association of eight influential variables with average monsoon and annual rainfall: Alt: altitude; Avg: average; CV: co-efficient of variation; Lat: latitude; Long: longitude; Max: maximum; Min: minimum; and SD: standard deviation [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Co-ordinates for the two most significant principal components (PCs) for monthly rainfall

		Cluster I		Cluster II		Cluster III		Cluster IV	
Month	**Variable**	**PC 1**	**PC 2**	**PC 1**	**PC 2**	**PC 1**	**PC 2**	**PC 1**	**PC 2**
June	Longitude	**0.8407**[a]	−0.2662	**−0.7835**[a]	−0.4288	**−0.8947**[a]	−0.0834	**0.8938**[a]	−0.2903
	Latitude	**−0.8682**[a]	0.1473	**−0.5954**[b]	0.3425	**0.8160**[a]	0.3831	**0.7725**[a]	**−0.6166**[b]
	Altitude	0.1621	**−0.5270**[b]	**−0.7621**[a]	−0.3879	**−0.8935**[a]	−0.1932	**0.8051**[a]	−0.0251
	Average	**0.9058**[a]	−0.3040	**−0.6366**[b]	**−0.5228**[b]	**−0.9376**[a]	0.0693	**0.7821**[a]	−0.4593
	SD	0.0603	**−0.9653**[a]	**−0.7666**[a]	0.3660	**−0.8151**[a]	**0.5404**[b]	**−0.5766**[b]	**−0.8034**[a]
	CV	**−0.8024**[a]	**−0.5565**[b]	−0.0331	**0.9834**[a]	**0.5683**[b]	**0.7270**[b]	**−0.9483**[a]	−0.2198
	Maximum	−0.3599	**−0.9008**[a]	**−0.7238**[b]	**0.6181**[b]	**−0.7564**[a]	**0.6183**[b]	**−0.5671**[b]	**−0.7822**[a]
July	Longitude	**0.6712**[b]	**−0.5609**[b]	**0.8518**[a]	0.4512	0.3783	−0.7347	**0.8773**[a]	−0.4165
	Latitude	0.1391	**−0.8219** [a]	0.4186	**0.7764**[a]	−0.4813	0.3006	**0.5562**[b]	−0.1785
	Altitude	**0.6004**[b]	**0.5778**[b]	**0.9130**[a]	0.2291	**0.6549**[b]	**−0.5049**[b]	**0.8917**[a]	−0.2102
	Average	**0.9624**[a]	0.1790	**0.7362**[b]	**−0.6206**[b]	0.1725	**−0.7267**[b]	**0.7831**[a]	**−0.5686**[b]
	SD	**0.7261**[b]	−0.2409	0.2956	**−0.8967**[a]	**−0.7982**[a]	**−0.5703**[b]	−0.4447	**−0.8696**[a]
	CV	**−0.8016**[a]	−0.4029	**−0.7913**[a]	−0.0654	**−0.8730**[a]	0.0289	**−0.9277**[a]	−0.2979
	Maximum	0.4486	−0.3932	0.4178	**−0.5945**[b]	**−0.7125**[b]	**−0.6297**[b]	**−0.5999**[b]	**−0.7238**[b]
August	Longitude	**0.8372**[a]	−0.4250	0.0758	**−0.9572**[a]	**0.8525**[a]	−0.1113	−0.4868	**0.7650**[a]
	Latitude	**−0.8906**[a]	−0.2456	**0.8999**[a]	−0.0001	**0.5534**[b]	0.4259	0.3416	**0.6539**[b]
	Altitude	0.3271	**−0.8694**[a]	0.2332	**−0.8973**[a]	0.3422	**−0.7533**[a]	**−0.8796**[a]	−0.0579
	Average	**0.8004**[a]	**−0.5787**[b]	**−0.7629**[a]	**−0.6411**[b]	0.2106	**−0.9056**[a]	**−0.9164**[a]	0.3240
	SD	−0.2925	**−0.9083**[a]	**−0.9444**[a]	−0.2993	**−0.6172**[b]	**−0.5043**[b]	**−0.8509**[a]	−0.3250
	CV	**−0.9838**[a]	−0.1499	**−0.7925**[a]	0.4506	**−0.8464**[a]	0.4273	0.2371	**−0.7605**[a]
	Maximum	**−0.6371**[b]	**−0.7400**[b]	**−0.9731**[a]	0.1365	**−0.8592**[a]	−0.4168	**−0.8269**[a]	−0.3613
September	Longitude	0.2568	**−0.8855**[a]	**0.6264**[b]	**0.7424**[b]	**0.9035**[a]	−0.1432	**0.8021**[a]	0.0973
	Latitude	**0.6462**[b]	−0.3832	**0.8573**[a]	0.3119	**−0.7596**[a]	0.1326	**−0.8291**[a]	−0.3528
	Altitude	−0.2974	−0.4995	0.2288	**0.9475**[a]	**0.8360**[a]	−0.2558	**0.8940**[a]	0.2252
	Average	**−0.8647**[a]	−0.4691	**−0.6109**[b]	**0.7134**[b]	**0.9453**[a]	0.1413	**0.9444**[a]	−0.0516
	SD	**−0.9557**[a]	−0.1337	**−0.8565**[a]	**0.5012**[b]	**0.6869**[b]	**0.6828**[b]	**0.8288**[a]	−0.4858
	CV	−0.0169	**0.8545**[a]	**−0.6857**[b]	−0.3276	**−0.7061**[b]	**0.6125**[b]	**−0.8178**[a]	−0.4241
	Maximum	**−0.8521**[a]	0.2258	**−0.8752**[a]	0.3526	0.2998	**0.9135**[a]	**0.7294**[b]	**−0.6408**[b]

Notes: Entries shown in bold indicate significant principal components. CV: co-efficient of variation; SD: standard deviation.
[a] Strongly significant.
[b] Moderately significant.

1 characterized by a strongly positive longitude, altitude, mean and minimum rainfall, and moderately positive latitude, which is strongly opposed by the CV.

It is inferred that in most clusters and periods (months, seasons and year), longitude has strong to moderate linkages with the mean. However, these linkages are opposed mainly by the CV. This finding suggests that mean rainfall increases when moving towards higher longitudes, and at the same time, the CV decreases, indicating less rainfall variations. This is true as rainfall in the area is lowest towards the western portion, and it increases towards the eastern portion, while the CV has a gradient from west to east. In some cases, an association with

TABLE 4 Co-ordinates for the two most significant principal components (PCs) for seasonal and annual rainfall

		Cluster I		Cluster II		Cluster III		Cluster IV	
Season/year	**Variable**	**PC 1**	**PC 2**	**PC 1**	**PC 2**	**PC 1**	**PC 2**	**PC 1**	**PC 2**
Monsoon	Longitude	**−0.7826**[a]	−0.0182	**−0.8365**[a]	0.4552	−0.3288	**−0.6995**[b]	**−0.9717**[a]	0.0317
	Latitude	−0.3943	0.3956	−0.2993	**0.9198**[a]	**0.7370**[b]	−0.3046	**−0.6259**[b]	−0.4375
	Altitude	−0.4710	**−0.6867**[b]	**−0.8951**[a]	0.2437	**−0.7857**[a]	0.1252	**−0.8715**[a]	0.0957
	Average	**−0.6354**[b]	**−0.6886**[b]	**−0.8496**[a]	−0.4710	**−0.7051**[b]	**−0.5517**[b]	**−0.9534**[a]	0.2118
	SD	**0.5938**[b]	**−0.6840**[b]	−0.4077	**−0.8927**[a]	**−0.9164**[a]	0.1422	−0.2035	**0.9503**[a]
	CV	**0.9572**[a]	−0.1254	**0.7778**[a]	**−0.5371**[b]	−0.4127	**0.6554**[b]	**0.7888**[a]	**0.5424**[b]
	Maximum	**−0.6743**[b]	−0.4444	**−0.9077**[a]	−0.0922	0.0561	**−0.6230**[b]	**−0.7574**[a]	0.3125
Post-monsoon	Longitude	**−0.8577**[a]	−0.4218	−0.3597	**0.9255**[a]	**0.9216**[a]	0.0097	**−0.9697**[a]	−0.1222
	Latitude	0.0551	**−0.8312**[a]	−0.1298	**0.9219**[a]	**−0.8058**[a]	−0.0003	**0.8481**[a]	0.1557
	Altitude	**−0.7450**[b]	**−0.5631**[b]	**−0.5976**[b]	**0.6639**[b]	**0.9148**[a]	−0.1440	**−0.9735**[a]	−0.2088
	Average	**−0.9474**[a]	0.1734	**−0.7515**[a]	−0.4048	**0.8793**[a]	0.1304	**−0.9643**[a]	0.0036
	SD	**−0.8133**[a]	**0.5342**[b]	**−0.9284**[a]	−0.3629	0.4820	**0.8329**[a]	**−0.8084**[a]	**0.5625**[b]
	CV	**0.5923**[b]	0.4561	**−0.8114**[a]	−0.2077	**−0.7130**[b]	**0.6067**[b]	**0.7273**[b]	**0.6573**[b]
	Maximum	**−0.6499**[b]	**0.6260**[b]	**−0.9638**[a]	−0.0413	0.0413	**0.9459**[a]	**−0.6040**[b]	**0.7843**[a]
Winter	Longitude	**0.9173**[a]	−0.0794	**0.9774**[a]	−0.0224	−0.0026	**−0.9290**[a]	**−0.9454**[a]	0.0368
	Latitude	−0.0145	**0.7175**[b]	**0.8881**[a]	−0.0939	**0.8635**[a]	0.4149	0.0570	**0.6358**[b]
	Altitude	**0.8187**[a]	0.1597	**0.9411**[a]	−0.2703	**−0.5731**[b]	**−0.8006** [a]	**−0.7092**[b]	**−0.5239**[b]
	Average	**0.9243**[a]	0.1896	**0.9374**[a]	0.2895	**0.8794**[a]	−0.3949	**−0.9244**[a]	0.0649
	SD	**0.5459**[b]	**−0.7788**[a]	0.4667	**0.8194**[a]	**0.7015**[b]	−0.4454	−0.4481	**−0.7397**[b]
	CV	−0.3919	**−0.8545**[a]	**−0.7929**[a]	0.4915	**−0.6843**[b]	0.2039	**0.8093**[a]	**−0.5666**[b]
	Maximum	0.1420	**−0.9328**[a]	0.1100	**0.8692**[a]	**0.7091**[b]	−0.0286	0.2442	**−0.7611**[a]
Pre-monsoon	Longitude	**−0.7201**[b]	−0.4886	**0.7722**[a]	−0.1991	**0.5536**[b]	−0.1590	**−0.8105**[a]	−0.2309
	Latitude	−0.0650	0.0041	**0.8222**[a]	−0.2707	**−0.6040**[b]	**−0.5628**[b]	**0.9201**[a]	−0.1182
	Altitude	**−0.7208**[b]	−0.2853	**0.5827**[b]	−0.1866	**0.7616**[a]	0.2920	**−0.9294**[a]	−0.0046
	Average	**−0.9167**[a]	0.0789	**0.9377**[a]	−0.0173	0.2005	**−0.8768**[a]	0.3142	**−0.9133**[a]
	SD	**−0.7805**[a]	**0.5821**[b]	**0.7525**[a]	**0.6002**[b]	**0.8179**[a]	−0.3793	−0.2355	**−0.8866**[a]
	CV	**0.6275**[b]	**0.6159**[b]	**−0.8269**[a]	0.3099	**0.6009**[b]	0.4808	**−0.6251**[b]	0.4728
	Maximum	**−0.5142**[b]	**0.8111**[b]	0.3368	**0.9084**[a]	**0.6562**[b]	−0.4224	**−0.6465**[b]	−0.4503
Annual	Longitude	**0.8977**[a]	0.0933	**−0.8526**[a]	0.4239	−0.0512	**−0.9165** [a]	**0.9690**[a]	−0.0333
	Latitude	**0.5356**[b]	−0.1160	−0.3140	**0.9093**[a]	**−0.6345**[b]	0.1320	**0.6319**[b]	**0.5449**[b]
	Altitude	0.4389	0.2788	**−0.8948**[a]	0.2181	**0.5438**[b]	−0.2769	**0.8493**[a]	−0.2091
	Average	**0.8331**[a]	0.4967	**−0.8526**[a]	−0.4913	−0.2717	**−0.8868**[a]	**0.9648**[a]	−0.1619
	SD	−0.2310	**0.9389**[a]	−0.4392	**−0.8793**[a]	**0.8901**[a]	−0.3158	0.1589	**−0.9434**[a]
	CV	**−0.8712**[a]	0.4210	**0.8004**[a]	**−0.5114**[b]	**0.9212**[a]	0.1844	**−0.7563**[a]	**−0.5982**[b]
	Minimum	**0.8775**[a]	0.2870	**−0.9403**[a]	−0.0206	**−0.7276**[b]	**−0.5187**[b]	**0.8605**[a]	−0.2914
	Maximum	−0.2387	**0.9470**[a]	**−0.6255**[b]	**−0.6826**[b]	**0.6202**[b]	**−0.5155**[b]	0.0350	**−0.8563**[a]

Notes: Entries shown in bold indicate significant principal components. CV: co-efficient of variation; SD: standard deviation.
[a] Strongly significant.
[b] Moderately significant.

longitude and the mean is opposed by latitude, which emphasizes the fact that mean rainfall increases when moving northwards. Moreover, this study justifies the fact that rainfall is largely influenced by longitude, latitude, the mean, SD and CV. Similarly, in an earlier study longitude and latitude have shown a positive correlation with the third harmonics of the seasonal and annual rainfall of Iran (Sabziparvar *et al*., 2015).

5 | CONCLUSIONS

This study examined the spatial patterns of monthly, seasonal and annual rainfall in the northwestern arid lands of India by delineating four clusters through hierarchical cluster analysis (HCA) and investigating the relative influence of eight geographical and statistical parameters on delineated clusters. The means and medians of rainfall were the lowest for cluster I, located in the western portion, whereas cluster IV, with the highest mean rainfall, was situated towards the eastern portion. Skewness and kurtosis values indicated the presence of normality in rainfall clusters for all periods, except for June and the post-monsoon. The co-efficient of variation (CV) suggested minimum temporal variability (35–42%) for annual rainfall. Box–whisker plots confirmed normality in the monsoon and annual rainfall. The mean monthly rainfall showed strong to

moderate relationships with longitude, latitude, the standard deviation (SD), maximum rainfall and CV. Similarly, the mean seasonal and annual rainfall was correlated with longitude, latitude, altitude, the SD, and maximum and minimum rainfall. The multicollinearity of the variables was evidenced from the results of the Gleason–Staelin redundancy measure and Bartlett's test of sphericity, which suggested the appropriateness of the input variables at performing principal component analysis (PCA). The PCA resulted in two to three significant principal components (PCs) explaining the major variance for monthly (75.36–88.44% for June, 69.41–83.53% for July, 74.60–91.83% for August and 73.70–86.58% for September), seasonal (63.62–90.39% for monsoon, 81.82–93.30% for post-monsoon, 75.74–88.60% for winter and 64.31–77.77% for pre-monsoon) and annual (71.48–90.73%) rainfall. On unit circle plots of the first two PCs, PC 1 was termed the "mean rainfall component". In monthly and seasonal rainfall, close linkages for maximum rainfall and the SD were observed. Also, longitude and means were opposed by the CV. The outcome of this study emphasizes the fact that rainfall increases at higher longitudes and latitudes, where rainfall variability decreases. The findings are useful for policy-makers to formulate adequate guidelines for planning and management of unusual rainwater quantities in a water-deficient study area.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous reviewers for their constructive and useful comments, which help to improve the earlier version of this manuscript.

ORCID

Deepesh Machiwal https://orcid.org/0000-0002-5659-9783

REFERENCES

Amissah-Arthur, A. and Jagtap, S.S. (1999) Geographic variation in growing season rainfall during three decades in Nigeria using principal component and cluster analyses. *Theoretical and Applied Climatology*, 63, 107–116.

Bartlett, M. (1950) Tests of significance in factor analysis. *British Journal of Psychology*, 3(2), 77–85.

Darand, M. and Daneshvar, M.R.M. (2014) Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environmental Processes*, 1, 517–532.

Davis, J.C. (2002) *Statistics and Data Analysis in Geology*. Singapore: John Wiley & Sons.

Dillon, R. and Goldstein, M. (1984) *Multivariate Analyses: Methods and Applications*. New York: Wiley.

Everitt, B.S. and Dunn, G. (1991) *Applied Multivariate Analysis*. London: Edward Arnold.

Frich, P., Alexander, L.V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A. M.G. and Peterson, T. (2002) Observed coherent changes in climatic extremes during the second half of twentieth century. *Climate Research*, 19, 193–212.

Gadgil, S. and Iyengar, R.N. (1980) Cluster analysis of rainfall stations of the Indian peninsula. *Quarterly Journal of the Royal Meteorological Society*, 106, 873–886.

Gleason, T.C. and Staelin, R. (1975) A proposal for handling missing data. *Psychometrika*, 40(2), 229–252.

Golian, S., Saghafian, B., Sheshangosht, S. and Ghalkhani, H. (2010) Comparison of classification and clustering methods in spatial rainfall pattern recognition at Northern Iran. *Theoretical and Applied Climatology*, 102, 319–329.

Gupta, A., Kamble, T. and Machiwal, D. (2017) Comparison of ordinary and Bayesian kriging techniques in depicting rainfall variability in arid and semi-arid regions of north-west India. *Environmental Earth Sciences*, 76, 512. https://doi.org/10.1007/s12665-017-6814-3.

Guttman, N.B. (1993) The use of L-moments in the determination of regional precipitation climates. *Journal of Climate*, 6, 2309–2325.

Haines, H.A. and Olley, J.M. (2017) The implications of regional variations in rainfall for reconstructing rainfall patterns using tree rings. *Hydrological Processes*, 31, 2951–2960.

Harman, H.H. (1960) *Modern Factor Analysis*. Chicago: University of Chicago Press.

Jebari, S., Berndtsson, R., Uvo, C. and Bahri, A. (2009) Regionalizing fine time-scale rainfall affected by topography in semi-arid Tunisia. *Hydrological Sciences Journal*, 52(6), 1199–1215.

Kaiser, H.F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.

Kar, A., Garg, B.K., Singh, M.P. and Kathju, S. (Eds.). (2009) *Trends in Arid Zone Research in India*. Jodhpur: Central Arid Zone Research Institute.

Lin, G.-F., Chang, M.J. and Wu, J.T. (2017) A hybrid statistical downscaling method based on the classification of rainfall patterns. *Water Resources Management*, 31(1), 377–401.

Liu, C.W., Lin, K.H. and Kuo, Y.M. (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *The Science of the Total Environment*, 313, 77–89.

Machiwal, D., Dayal, D. and Kumar, S. (2017) Long-term rainfall trends and change points in hot and cold arid regions of India. *Hydrological Sciences Journal*, 62(7), 1050–1066.

Mares, M.A. (Ed.). (1999) *Encyclopedia of Deserts*. Norman, OK: University of Oklahoma Press.

Medina-Cobo, M.T., García-Marín, A.P., Estévez, J., Jiménez-Hornero, F.J. and Ayuso-Muñoz, J.L. (2017) Obtaining homogeneous regions by determining the generalized fractal dimensions of validated daily rainfall data sets. *Water Resources Management*, 31(7), 2333–2348.

Modarres, R. and Sarhadi, A. (2011) Statistically-based regionalization of rainfall climates of Iran. *Global and Planetary Change*, 75, 67–75.

Moharana, P.C., Santra, P., Singh, D.V., Kumar, S., Goyal, R.K., Machiwal, D. and Yadav, O.P. (2016) ICAR-Central Arid Zone Research Institute, Jodhpur: erosion processes and desertification in the Thar desert of India. *Proceedings of the Indian National Science Academy*, 82(3), 1117–1140.

Nnaji, C.C., Mama, C.N. and Ukpabi, O. (2016) Hierarchical analysis of rainfall variability across Nigeria. *Theoretical and Applied Climatology*, 123(1–2), 171–184.

Otto, M. (1998) Multivariate methods. In: Kellner, R. , Mermet, J.M. , Otto, M. and Widmer, H.M. (Eds.) *Analytical Chemistry*. Weinheim, Germany: Wiley-VCH.

Sabziparvar, A.A., Movahedi, S., Asakereh, H., Maryanaji, Z. and Masoodian, S. A. (2015) Geographical factors affecting variability of precipitation regime in Iran. *Theoretical and Applied Climatology*, 120(1–2), 367–376.

Şen, Z. and Habib, Z. (2001) Monthly spatial rainfall correlation functions and interpretations for Turkey. *Hydrological Sciences Journal*, 46(4), 525–535.

StatSoft. (2004) *STATISTICA (data analysis software system), version 6*. Available at: www.statsoft.com.

Suhaila, J. and Jemain, A.A. (2009) A comparison of the rainfall patterns between stations on the East and the West coasts of Peninsular Malaysia using the smoothing model of rainfall amounts. *Meteorological Applications*, 16(3), 391–401.

USEPA. (1998). *Guidance for Data Quality Assessment: Practical Methods for Data Analysis. Quality Assurance Division*, EPA QA/G-9, version QA97. Washington: United States Environmental Protection Agency (USEPA), pp. 2.3-3–2.3-5.

Ward, J.H., Jr. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244.

How to cite this article: Machiwal D, Kumar S, Meena HM, Santra P, Singh RK, Singh DV. Clustering of rainfall stations and distinguishing influential factors using PCA and HCA techniques over the western dry region of India. *Meteorol Appl*. 2019;26: 300–311. https://doi.org/10.1002/met.1763