Theor Appl Climatol (2011) 104:57–69
DOI 10.1007/s00704-010-0321-8

ORIGINAL PAPER

Spatio-temporal variability of seasonal rainfall in western equatorial Africa

Amin K. Dezfuli

Received: 13 March 2010 / Accepted: 27 July 2010 / Published online: 24 August 2010

Abstract This study reveals homogeneous sub-regions over the poorly studied area of western equatorial Africa (10S–7N and 7E–30E). Monthly totals of 141 stations covering the period 1955–1984 are used. The stations are grouped based on the similarity of their interannual rainfall variability. In addition to annual totals, four different seasons are examined separately for regionalization, an approach that has lacked in previous studies. The four 3-month seasons are defined as follows: January–February–March (JFM), April–May–June (AMJ), July–August–September (JAS), and October–November–December (OND). Two different algorithms are applied and compared: the rotated principal component analysis (RPCA) in conjunction with Ward's method, and the RPCA in conjunction with *k*-means method. The principal components that explain about 65% of total variance are retained and then varimax rotated. The corresponding scores are utilized as input for cluster analysis. Using Ward's method, five sub-regions are recognized for AMJ, JAS and OND and 4 sub-regions for JFM and annual data. The regions are geographically well distributed over the area and consist of roughly the same number of stations. The *F*-test is used to evaluate the homogeneity of each sub-region. The results show that all sub-regions are strongly homogeneous. Assuming the same number of clusters, the *k*-means method provides comparable spatial patterns with those of Ward's method. However, there are some differences, which are more evident in JAS and OND. Like Ward's method, the values of *F*-ratio for the *k*-means algorithm also confirm the homogeneity of all seasons/sub-regions. The interannual variability of rainfall for each season/sub-region is also provided and compared.

A. K. Dezfuli (✉)
Department of Meteorology, Florida State University,
Tallahassee, FL, USA
e-mail: ak06e@fsu.edu

1 Introduction

Equatorial Africa is a crucial region for climate studies, as it occupies the largest equatorial land mass in the world. The factors governing rainfall variability of the western and eastern parts of this region are different. Although eastern equatorial Africa has been well studied, very little is known about the rainfall patterns and atmospheric circulation of the western equatorial Africa (WEA), which is the region of concern in this study.

The WEA experiences the world's most intense thunderstorms and the highest frequency of lightning flashes (Zipser et al. 2006), while it receives considerably less rainfall than other equatorial regions. Orographic effects have a strong impact on the convective regime in the region (Jackson et al. 2009). Interannual variability of rainfall in this region is markedly heterogeneous, in contrast to the rest of Africa where coherent fluctuations encompass areas on the scale of 1,000 km. The link between WEA rainfall and the global oceans is highly seasonally and regionally specific (Balas et al. 2007). That is, the SST/rainfall link varies from season to season and this link differs considerably over relatively small geographical distances within WEA. The previous regionalization studies either focus on a part of WEA, or their approach is more subjective and arbitrary. Therefore, WEA lacks an objective spatio-temporal analysis, which is a fundamental practice in climatology and will enhance our understanding of regional climate processes.

Different techniques have been employed in the literature to define homogeneous sub-regions. Ward's method has been widely accepted as the best-performing hierarchical clustering technique for climate studies (e.g., Gong and Richman 1995; Unal et al. 2003), although the fuzzy hierarchical clustering has also been successfully applied (Dezfuli et al. 2010). In order to avoid the noise in original data, the Ward's method is usually combined with the rotated principal component (RPC) analysis (Busuioc et al. 2001; Jebari et al. 2007; Domroes et al. 1998; Raziei et al. 2008; Kamara and Jackson 1997; Baeriswyl and Rebetez 1997). The RPC analysis has also been used alone to present the dominant modes of rainfall variability (White et al. 1991; Comrie and Glenn 1998; Ogallo 1989).

The *k*-means algorithm, which is perhaps the most popular non-hierarchical clustering technique, serves as an alternative method for climate regionalization. This method has been shown to outperform the hierarchical methods (Gong and Richman 1995) while performing equally well with fuzzy clustering, principal components, and principal components coupled with *k*-means clustering (Wilson et al. 1992). However, Corte-Real et al. (1998) suggested using the principal components coupled with *k*-means method, once the total number of clusters is chosen. Rao and Srinivas (2006) examined the *k*-means and three hierarchical methods (single linkage, complete linkage, and Ward's) for regionalization of watersheds. They proposed a hybrid-clustering algorithm by which Ward's method is used to initialize the *k*-means algorithm. Although, all these approaches have been claimed to be successful in their own cases, none of them has been demonstrated to yield universally acceptable results. Therefore, the problem of choosing a best clustering method for climate regionalization is still unresolved and is more of a subjective decision.

Several studies have attempted to find spatial patterns of rainfall variablity in western equatorial Africa or nearby regions. Nicholson (1980, 1981) defined rainfall zones in West Africa. She based the latitudinal regionalization on mean annual rainfall, while annual variability, seasonal distribution, and coherence of variation served as secondary criteria. Following up, Balas et al. (2007) found the spatial patterns in WEA based only on annual rainfall. Applying the RPC analysis, Ogallo (1989) investigated the spatio-temporal characteristics of seasonal rainfall over eastern equatorial Africa (12S–5N and 28E–42E), which lies to the east of the present study region. He grouped ninety stations into 26 sub-regions, the majority of which have a length of less than 1° in any horizontal direction. Using annual rainfall, Janicot (1992) revealed five sub-regions over West Africa (approximately, 5S–25N and 15W–25E). Three of those are located in the northwestern section of our study region. Most recently, Djomou et al. (2009) analyzed the spatial variability of boreal summer rainfall in West Africa (0–30N and 20W–30E). They used Ward's method and found four sub-regions, one of which overlaps the northern part of our area.

These studies suggest a need for a comprehensive spatio-temporal analysis over the poorly studied area of WEA (10S–7N and 7E–30E). To do that, the stations will be grouped based on their interannual rainfall variability. In addition to annual totals, four different seasons are examined separately for regionalization, an approach that was lacking in previous studies. Also, an objective approach is essential to improve the reliability of spatial patterns. We resolve this issue by using two techniques: the RPC in conjunction with Ward's method and with *k*-means clustering algorithm. That also allows us to compare these two techniques.

The paper is organized as follows. Section 2 describes the data. Section 3 gives a brief overview of RPC analysis, Ward's method, and *k*-means clustering. Section 4 provides the homogeneous sub-regions along with their interannual rainfall variability. The paper will end with a Summary and Conclusions in section 5.

2 Data

This study utilizes rain-gauge data. The data are obtained from Dr. Nicholson's NIC131 African precipitation data set, which includes monthly totals for 141 stations over western equatorial Africa (10S–7N and 7E–30E). The period 1955–1984 is chosen, because it demonstrates the minimum amount of missing data for most of the stations. The data over this period are carefully quality controlled and the stations with more than about 10% of missing data are omitted, particularly if they represent the wet months. This reduces the number of stations to 107. The remaining missing data are interpolated linearly in order to have a continuous time series for each station. The resulting gauge network is geographically well-distributed, except for some southern and eastern parts of the study area (Fig. 1).

3 Methodology

Principal component analysis (PCA) involves a statistical procedure that transforms a number of possibly intercorre-

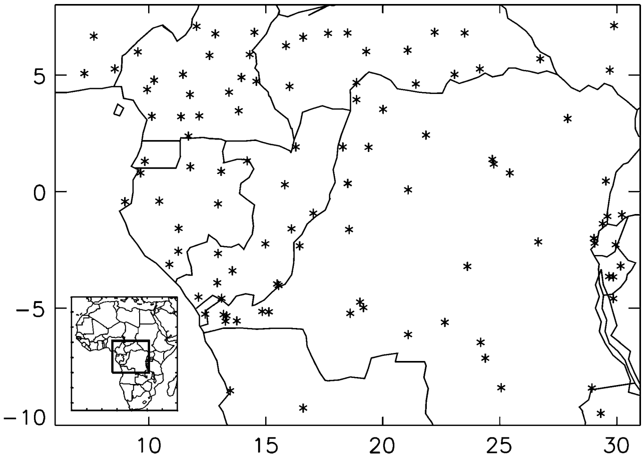

Fig. 1 Spatial network of rain-gauge stations in study area

lated variables into a smaller number of mutually uncorrelated variables (principal component, PCs). The PCs are linear combinations of the original variables. The first PC accounts for as much of the total variance in the original data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Depending on how the data matrix is defined, different modes of decomposition for a PCA can be specified (Richman 1986). The T-mode that involves a fixed parameter is used in this study, with the parameter being seasonal or annual precipitation totals. The data matrix has a dimension of $K\times N$, where N is the number of stations and K is the number of years. Performing PCA requires either correlation or covariance matrix of the data set to be computed. Use of the correlation matrix is preferred, because it ignores the role of absolute values of precipitation and thus it allows stations to be grouped only based on similarity of their interannual variability of rainfall (Comrie and Glenn, 1998). The first M PCs (leading PCs) of the correlation matrix that explain most of the total variance are retained for further analysis. The number of leading PCs to be extracted is determined by inspection of the scree plot of eigenvalues. The leading PCs are usually rotated when physical interpretation rather than data compression is a primary goal (Wilks 2006). That gives a more accurate representation of the dominant spatial modes and thus is suggested for climate regionalization (Muñoz-Díaz and Rodrigo 2004; Rogers and McHugh 2002). The widely accepted method of varimax is used for rotation of PCs. The scores of rotated PCs are then computed. This results in an $M\times N$ matrix, which is considered as the input of cluster analysis. The cluster analysis aims to classify N stations into C groups based on M variables. Two clustering techniques are utilized: Ward's method which is the most commonly used hierarchical method and k-means clustering which is the most popular non-hierarchical method for climate studies.

As an agglomerative hierarchical algorithm, Ward's method begins with N single-member clusters, and successively merges clusters together, until all the stations are in a single cluster after $N-1$ steps. However, unlike the other hierarchical techniques, it does not operate on the distance matrix. This method is designed to minimize the increase of the within-group sum of the squared errors. Like other step-wise hierarchical procedures, it does not guarantee that the results at a given level will be the best minimum variance solution for that particular number of groups, although experience with the method has shown that the solutions are usually satisfactory. One advantage of Ward's method is that it usually does not leave single-member clusters after a reasonable number of stages, and it tends to produce clusters with roughly the same number of entities. A crucial step in hierarchical clustering is to choose the number of clusters, i.e., the level at which the merging process stops. Different algorithms have been suggested to tackle this problem called "cluster validity", which is perhaps "the most difficult and frustrating part of cluster analysis", as noted by Jain and Dubes (1988). However, there is no universally accepted objective solution to this problem and it has remained more of a subjective choice. Each cluster analysis case may have its own criteria for cluster validity. Here, the following criteria are considered for climate regionalization:

- Avoid having too few or too many stations in each cluster, so that each sub-region maintains a reasonable size.
- Guarantee geographically contiguousness of the sub-regions.
- Have a reasonable number of clusters.
- Assure homogeneity of the sub-regions.
- Find consistency between geographically distribution of sub-regions and some known physical features over the region.

There may be some interrelationships among these criteria. For example, the first criterion may somehow guarantee the third one.

A general drawback of hierarchical clustering techniques is that they contain no prevision for reallocation of entities that may have been misclassified at any early stage. In other words, once a station has been assigned to a sub-region it will stay in that sub-region. This problem does not exist in non-hierarchical techniques including k-means. The k-means algorithm begins with a prespecified number of clusters, C. The entities are initially assigned randomly to C clusters. The centroids of clusters are calculated and membership of the entities is updated with being reallocated to the nearest centroid. This process is repeated until some convergence criterion is met, that is, until two consecutive iterations generate the same cluster assignment. Since cluster validity problem is generally less severe in hierarchical approach (Fovell and Fovell 1993), the optimum number of sub-regions obtained by Ward's method is also used in k-means. In fact, this is another advantage of Ward's method, that it does not require prespecification of the number of clusters.

Once the clusters are determined, the homogeneity of each cluster is examined by an F-test. To do that, the variance in time, which determines the year-to-year regional fluctuations, and the mean spatial variance between rainfall anomalies within the region are estimated. The relative importance of these two components is then assessed by an F-test (Kraus 1977; Nicholson 1986). If the sub-regions appear to be inhomogeneous, with some changes, the entire clustering process will be repeated, until a satisfactory level of homogeneity is reached.

4 Results

4.1 Homogeneous sub-regions

The analyses are performed for annual and four seasonal precipitation totals. The four 3-month seasons are defined as follows: winter (January–February–March, JFM), spring (April–May–June, AMJ), summer (July–August–September, JAS), and fall (October–November–December, OND). The decision to use the seasonal division is based on linear correlation between months and on an examination of month-to-month changes in the rainfall distribution and atmosphere circulation features. Several seasonal divisions were tested and this was found to be the most reasonable for the region as a whole. The PCA is applied to the 30× 30-correlation matrix and the PCs with eigenvalues greater than or equal to one are retained. There are 12 leading PCs for each case as shown in Table 1. The percentages of the total variance explained by these PCs are 70%, 66.4%, 63.8%, and 66.3% for JFM, AMJ, JAS, and OND, respectively, and 67.6% for the annual data. The selected PCs are varimax rotated and their corresponding scores are obtained for all cases. The 107×12-matrix of scores will be used as the input for cluster analysis.

The dendrogram obtained by Ward's method for JFM is shown in Fig. 2a. Four sub-regions with roughly the same number of stations are clearly recognized. As a further step of quality control, and in order to obtain spatially contiguous sub-regions, some of the stations are omitted. The number of stations eliminated from each sub-region is presented with the values in parentheses in Table 2. For this season, nine stations are omitted, and this reduces the total number of remaining stations from 107 to 98. For each sub-region, the geographical distribution of the remaining stations is depicted in Fig. 3a. Sub-region 1 stretches down from the very northeast part, and is split up into two bands: the first one meets the western border of the study area at the equator and the second one moves toward the southern edge. Sub-region 2 is split up into two divisions, which are both confined between the equator and the southern border. The eastern division, however, contains a greater number of stations and occupies a larger area. Sub-region 3 lies on the southwest corner and stretches along the coast to south of the equator. The southern half of the region has a sparse network of stations. Sub-region 4 is located to the northwest part, with a dense distribution of stations all over the region. The critical value of F at the 1% probability level is 1.7. That is, for values exceeding this threshold, the probability that temporal variance in regional series can be accounted for by random fluctuations in a few stations, is less than 1%. Values outside of parentheses of Table 2 provide the estimated F-ratio of different sub-regions. The values for all four sub-regions, especially for regions 3 and 4, are substantially greater than 1.7. In other words, the mean rainfall series for each sub-region is indeed representative for that region as a whole.

For the same number of clusters, k-means method is also applied for JFM. The results for the 99 remaining stations are shown in Fig. 4a. The spatial distributions of sub-regions for both methods are very similar, where regions 1, 2, 3, and 4 obtained by k-means occupy fairly the same areas as regions 4, 3, 1, and 2 of Ward's method, respectively. The values estimated for F-ratio are also close to those for the corresponding regions of Ward's method, suggesting a significant spatial homogeneity at 1% probability for all four sub-regions.

Similar analysis is carried out for AMJ. The dendrogram of Ward's cluster analysis for this season is presented in Fig. 2b,

Table 1 Eigenvalues and percentages of explained variance of 12 unrotated PCs for annual and seasonal data

PC	Eigenvalues					Percentage of variance				
	JFM	AMJ	JAS	OND	ANN	JFM	AMJ	JAS	OND	ANN
1	4.45	3.04	2.85	3.26	3.05	14.8	10.1	9.5	10.9	10.2
2	2.85	2.45	1.94	2.83	2.50	9.5	8.2	6.5	9.4	8.3
3	1.93	1.96	1.83	1.81	1.98	6.4	6.5	6.1	6.0	6.6
4	1.76	1.89	1.76	1.79	1.77	5.9	6.3	5.9	6.0	5.9
5	1.66	1.75	1.65	1.66	1.69	5.6	5.8	5.5	5.5	5.7
6	1.55	1.53	1.49	1.59	1.57	5.2	5.1	5.0	5.3	5.2
7	1.42	1.45	1.44	1.39	1.50	4.7	4.8	4.8	4.6	5.0
8	1.15	1.35	1.35	1.27	1.41	3.8	4.5	4.5	4.2	4.7
9	1.10	1.20	1.29	1.16	1.37	3.7	4.0	4.3	3.9	4.6
10	1.07	1.17	1.23	1.12	1.24	3.6	3.9	4.1	3.7	4.1
11	1.05	1.10	1.18	1.02	1.13	3.5	3.7	3.9	3.4	3.8
12	1.00	1.01	1.11	1.01	1.07	3.3	3.4	3.7	3.4	3.6
Total% of variance						70.0	66.4	63.8	66.3	67.6

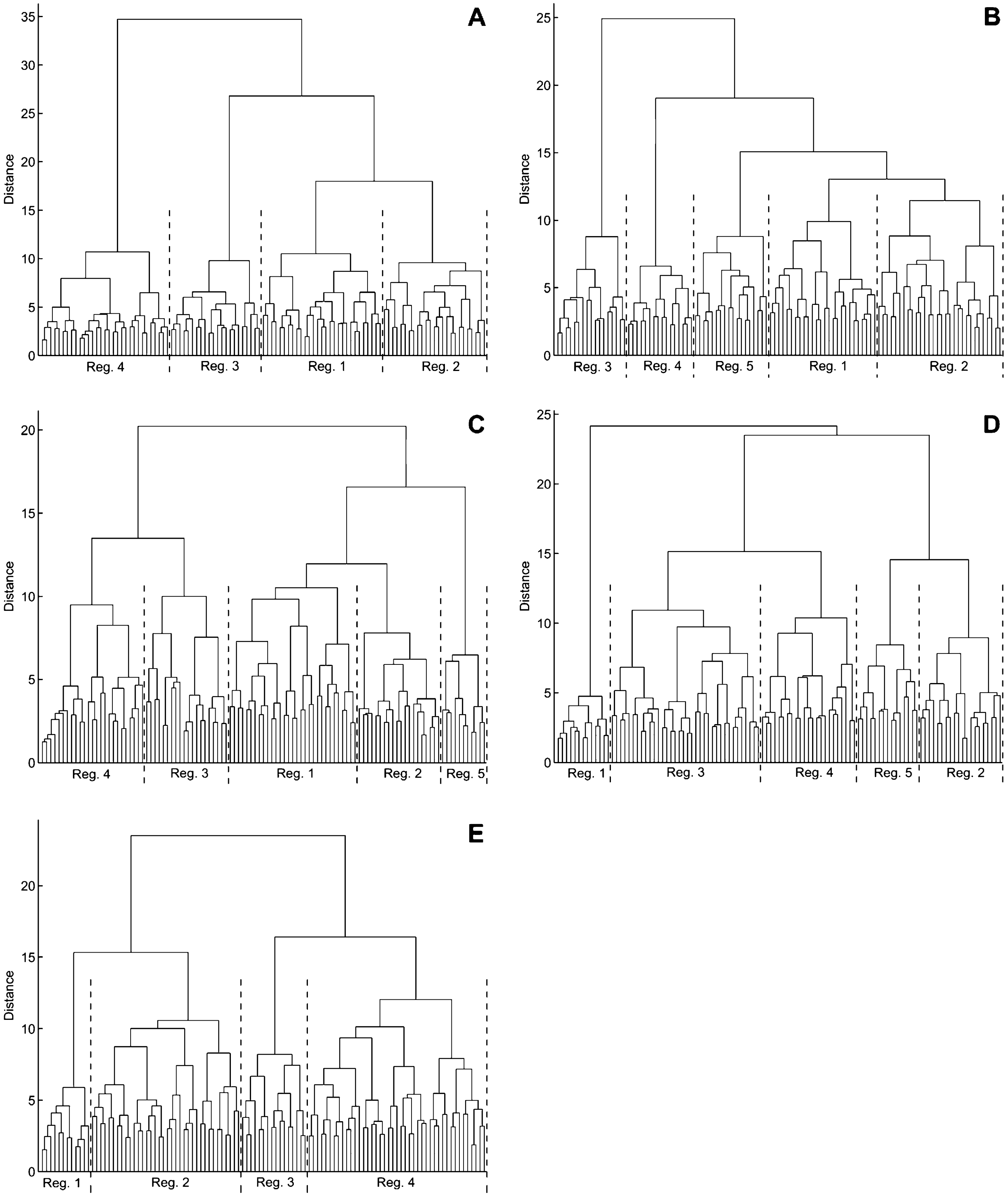

Fig. 2 Dendrogram obtained by Ward's method for **a** JFM, **b** AMJ, **c** JAS, **d** OND, and **e** annual data. Scores of the leading rotated PCs of 107 stations are the inputs of the cluster analysis

in which five sub-regions can be recognized. As given in Table 2, the total number of eliminated stations is ten. The spatial pattern of the 97 remaining stations is shown in Fig. 3b. Sub-region 1 occupies most of the area between the equator and 5N. Sub-region 2 is split up into two divisions: the first one lies to the north of sub-region 1 and the second one displaying a C-shape pattern straddles the equator. Sub-region 3 covers parts of the south, east, and center of the study area, having a dense station network only in the central area. Sub-region 4 develops over a narrow strip along the coast between the equator and 10S. Sub-region 5 is located in the southeastern corner with the same latitude range as sub-region 4. The *F*-ratios for this season as shown in Table 2 are significant at 1% level, confirming homogeneity of all sub-regions, especially regions 4 and 5.

Setting the number of clusters to five, the *k*-means method for the 97 remaining stations provides comparable geographical patterns with the sub-regions of Ward's method (Fig. 4b). Particularly, sub-regions 1, 3, and 4 of *k*-means appear over the same area of sub-regions 4, 3, and 5 of Ward's method, respectively. As expected, the values of *F*-ratio also exceed by far the minimum acceptance level for all sub-regions (Table 2).

The dendrogram of JAS (Fig. 2c) suggests five sub-regions and results in elimination of 17 stations after applying the contiguousness constraint. The spatial distribution of the 90 remaining stations is shown in Fig. 3c. Compared to other seasons, there are more sub-regions that are divided into two divisions, i.e., sub-regions 1, 2, and 3. The reason could be related to the fact that the boreal summer is a dry season over the region (Nicholson 1988) and that makes the influence of local features on precipitation more pronounced. Therefore, it will be more difficult to recognize contiguous homogeneous regions. Sub-region 1 has a northern and a southern part; the second one consists of only a few stations in a large area. Sub-region 2 lies mostly in the eastern half of the region. Sub-regions 3 and 4 stretch along the coast, and region 3 straddles region 4. Sub-region 5 is located in the far eastern part between the equator and about 7S. All sub-regions are homogeneous as shown in Table 2, and region 4 and 5 have the greatest *F*-ratio among the others.

Using the *k*-means algorithm and the same number of sub-regions (i.e., 5), the cluster analysis is repeated for JAS. The spatial pattern of the 87 remaining stations is depicted in Fig. 4c. The corresponding sub-regions in the southern hemisphere for both clustering methods are more similar than those located in the northern hemisphere. Sub-region 1 occupies exactly the same area as sub-region 5 of Ward's method. Region 2 and the southern division of regions 3 and 5 also spread over fairly the same areas as region 4 and southern division of region 3 and 1 of Ward's method, respectively. The values of *F*-ratio seen in Table 2 confirm that the time series of mean rainfall totals can be representative of each sub-region. The homogeneity of region 1 and 2 is more significant.

Figure 2d shows the dendrogram of Ward's cluster analysis for OND. This dendrogram suggests five sub-regions, the spatial pattern of which is shown in Fig. 3d. To maintain the contiguousness condition, 92 statitions are contributed and the rest are omitted. Sub-region 1 is confined between about 0–5S, meeting the Atlantic Ocean at the equator. Sub-region 2 appears over the eastern section between the equator and 10S. Sub-region 3 occupies most of the north and northwest of the study area. Sub-region 4 stretches diagonally from northeast to southwest and sub-region 5 is located along the coast between the equator and 10S. The values of *F*-ratio presented in Table 2 suggest that the mean rainfall totals of all sub-regions can be representative and region 3 and 5 have the greatest homogeneity.

Table 2 Results of *F*-ratio test for Ward's and *k*-means methods

Region	Ward's					*k*-means				
	JFM	AMJ	JAS	OND	ANN	JFM	AMJ	JAS	OND	ANN
1	8.3	5.4	4.5	2.5	6.2	22.7	13.7	9.7	6.9	3.2
	(3)	(3)	(2)	(3)	(0)	(5)	(0)	(2)	(0)	(4)
2	8.2	6.6	6.3	7.0	4.2	16.2	7.0	17.0	15.9	6.7
	(2)	(2)	(5)	(1)	(6)	(0)	(1)	(4)	(1)	(1)
3	15.7	3.9	5.4	13.0	4.5	7.2	4.1	8.4	4.4	7.4
	(0)	(3)	(8)	(4)	(5)	(2)	(5)	(3)	(3)	(2)
4	23.0	13.4	12.9	3.2	8.3	8.8	10.1	6.2	3.3	4.8
	(4)	(0)	(2)	(5)	(3)	(1)	(2)	(6)	(4)	(9)
5		9.5	8.7	19.2			4.8	4.0	14.1	
		(2)	(0)	(2)			(2)	(5)	(2)	
Remaining stations	98	97	90	92	93	99	97	87	97	91

The values in parentheses are the number of omitted stations

Values outside of parentheses are *F*-ratio of the remaining stations

The total number of remaining stations is also presented

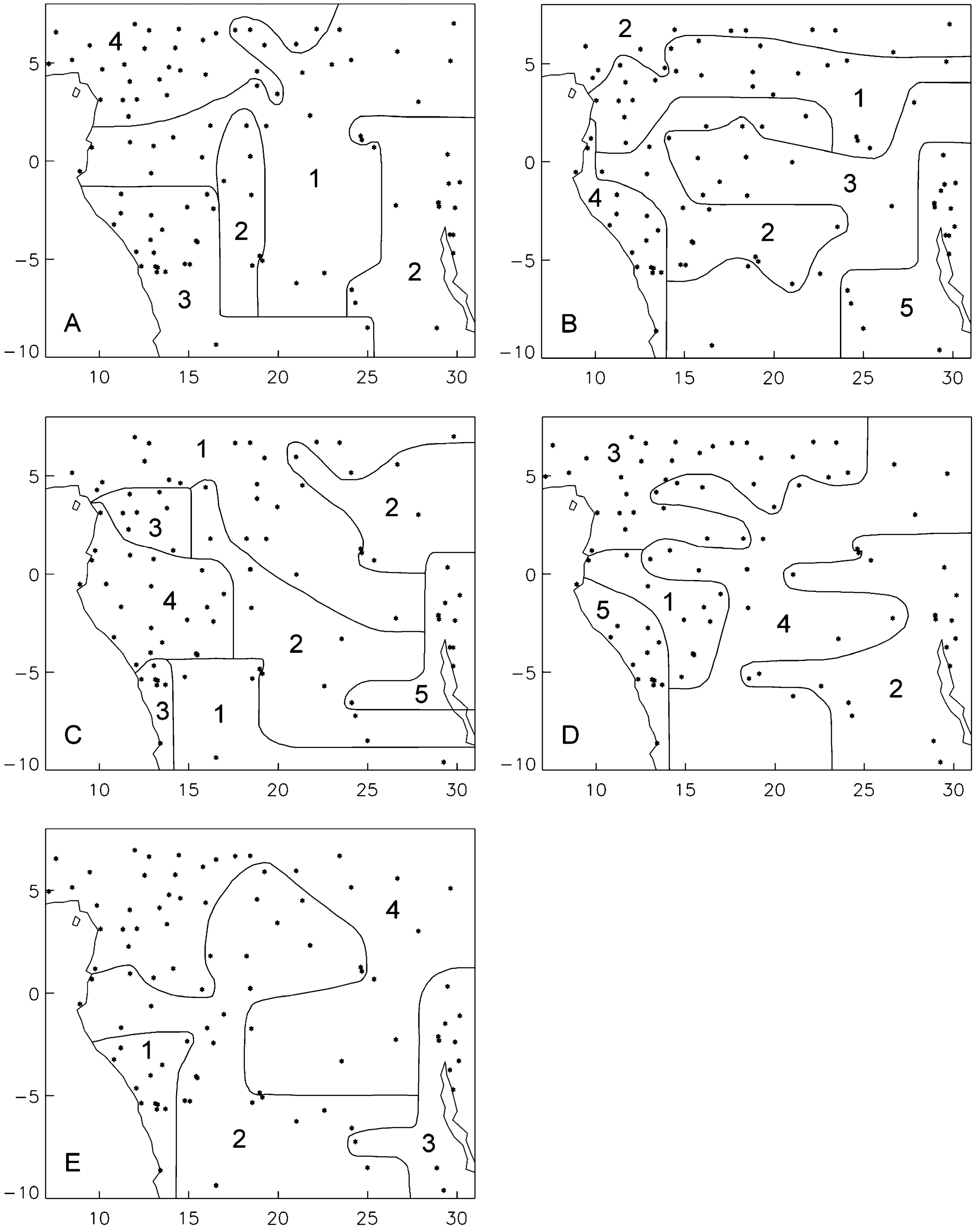

Fig. 3 Spatial distribution of sub-regions for **a** JFM, **b** AMJ, **c** JAS, **d** OND, and **e** annual data, using Ward's method

The spatial distribution of the five sub-regions resulting from *k*-means clustering is depicted in Fig. 4d. The contiguousness is reached with contributing 97 stations. This method provides somewhat different results from Ward's method. The difference is more evident in the area along southwest–northeast diagonal, where merging region 3 and 4 of *k*-means covers approximately the same area as region 1 and 4 of Ward's. However, region 1, 2, and 5 of *k*-means lie over fairly similar zone as region 2, 5, and 3 of Ward's, respectively. The *F*-test confirms the homogeneity of all sub-regions obtained by *k*-means and region 2 and 5 have the largest values of *F*-ratio (Table 2).

The dendrogram of annual analysis is shown in Fig. 2e. This Figure suggests compartmentalizing the area into four

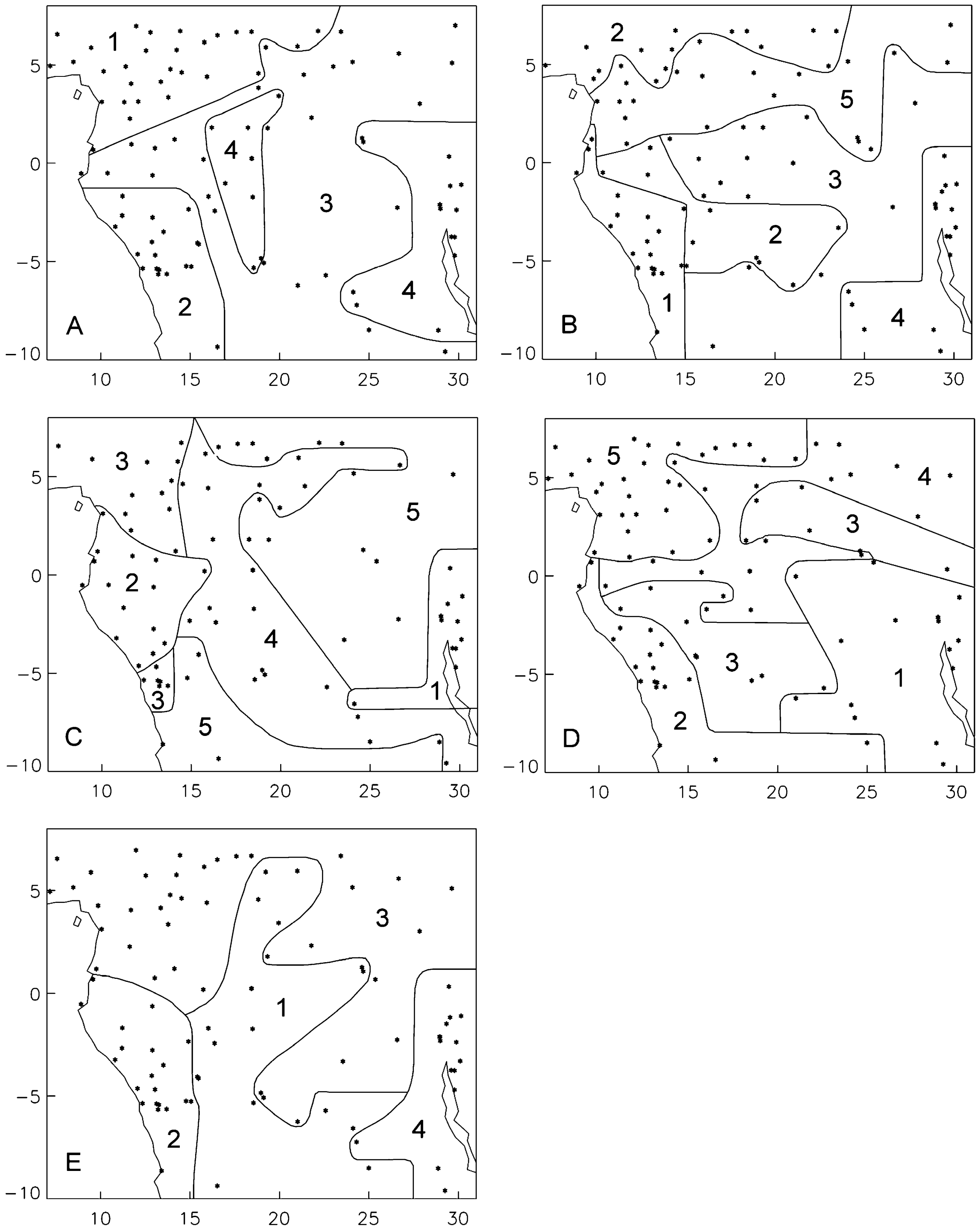

Fig. 4 As in Fig. 3, but for *k*-means method

sub-regions: two with larger areas (sub-region 2 and 4) and two with smaller areas (sub-region 1 and 3). The spatial pattern of the sub-regions for the 93 remaining stations is presented in Fig. 3e. Sub-region 1 lies along the coast in the southern hemisphere. Sub-region 3 covers a strip in the eastern edge of study area between the equator and 10S. The rest of area is occupied by sub-region 2, which lies mostly over south and central areas, and sub-region 4, which lies mostly in the northeast and northwest. The values of *F*-ratio for all sub-regions (Table 2) are greater than the theoretical threshold at the 1% probability level, confirming their homogeneity.

As shown in Fig. 4e, the *k*-means method for the 91 remaining stations gives a spatial distribution similar to the

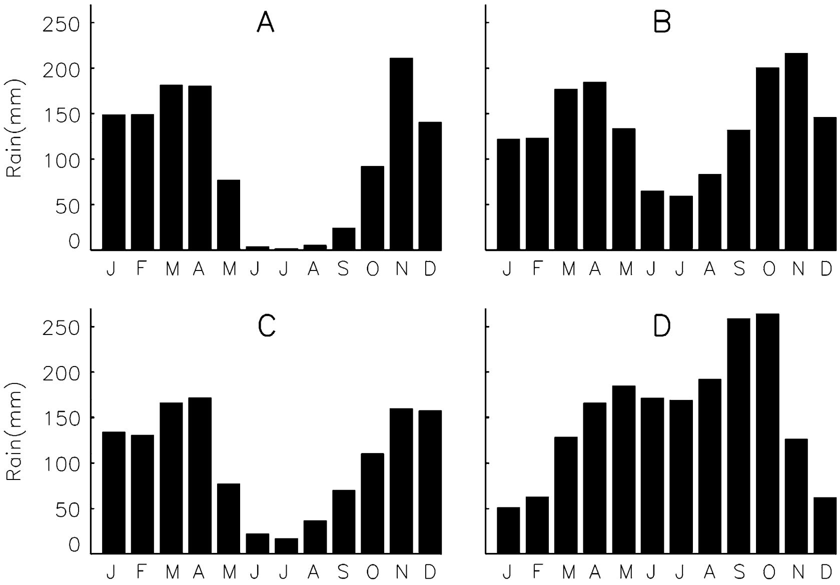

Fig. 5 Annual cycle of rainfall for **a** region 1, **b** region 2, **c** region 3, and **d** region 4, defined by Ward's method

sub-regions of Ward's method. Sub-regions 1, 2, 3, and 4 of *k*-means correspond to sub-regions 2, 1, 4 and 3 of Ward's method, respectively. Table 2 shows the estimated *F*-ratio for the annual data. All values are greater than the threshold, suggesting that all sub-regions are homogeneous.

4.2 Annual cycle

The annual cycle of rainfall for each sub-region, based on the analysis annual data, is shown in Fig. 5 (Ward's method) and Fig. 6 (*k*-means method). For sub-region 1, 2, and 3 of Ward's method (and their corresponding regions of *k*-means, i.e., sub-region 2, 1, and 4, respectively), the minimum amount of rainfall appears during June-July-August and the rainy season is from November through April with the peak occurring in November, March, and April. These three sub-regions lie mostly in the southern hemisphere. Sub-region 4 of Ward's method, however, which lies mostly in the northern hemisphere, presents a different annual cycle. Its dry season is observed during December-January-February, while most of the precipitation falls between March and November. However, September, and October have the largest values. As shown in Fig. 6, sub-region 3 of *k*-means behaves similarly to sub-region 4 of Ward's method, as expected.

4.3 Interannual variability

The time series of regional rainfall for each sub-region is determined. However, in an effort to further evaluate how representative they are, the rainfall time series of all stations

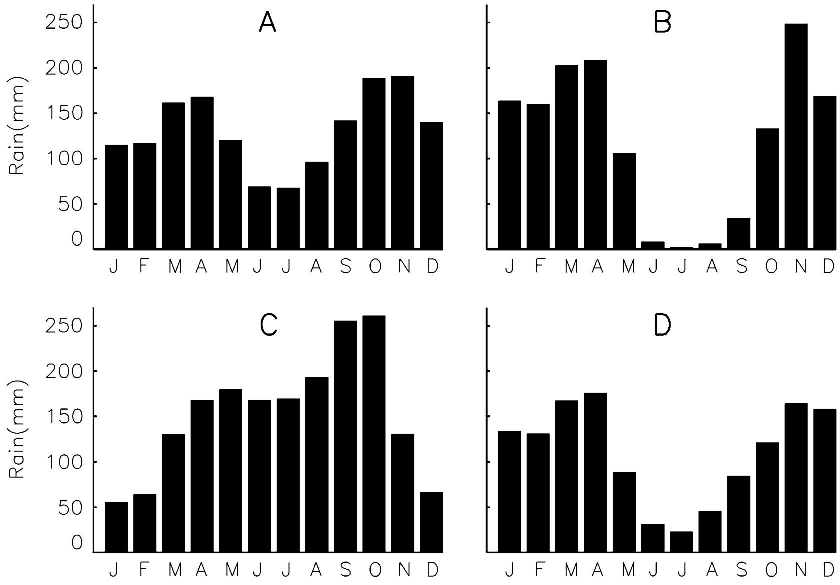

Fig. 6 As in Fig. 5, but for *k*-means method

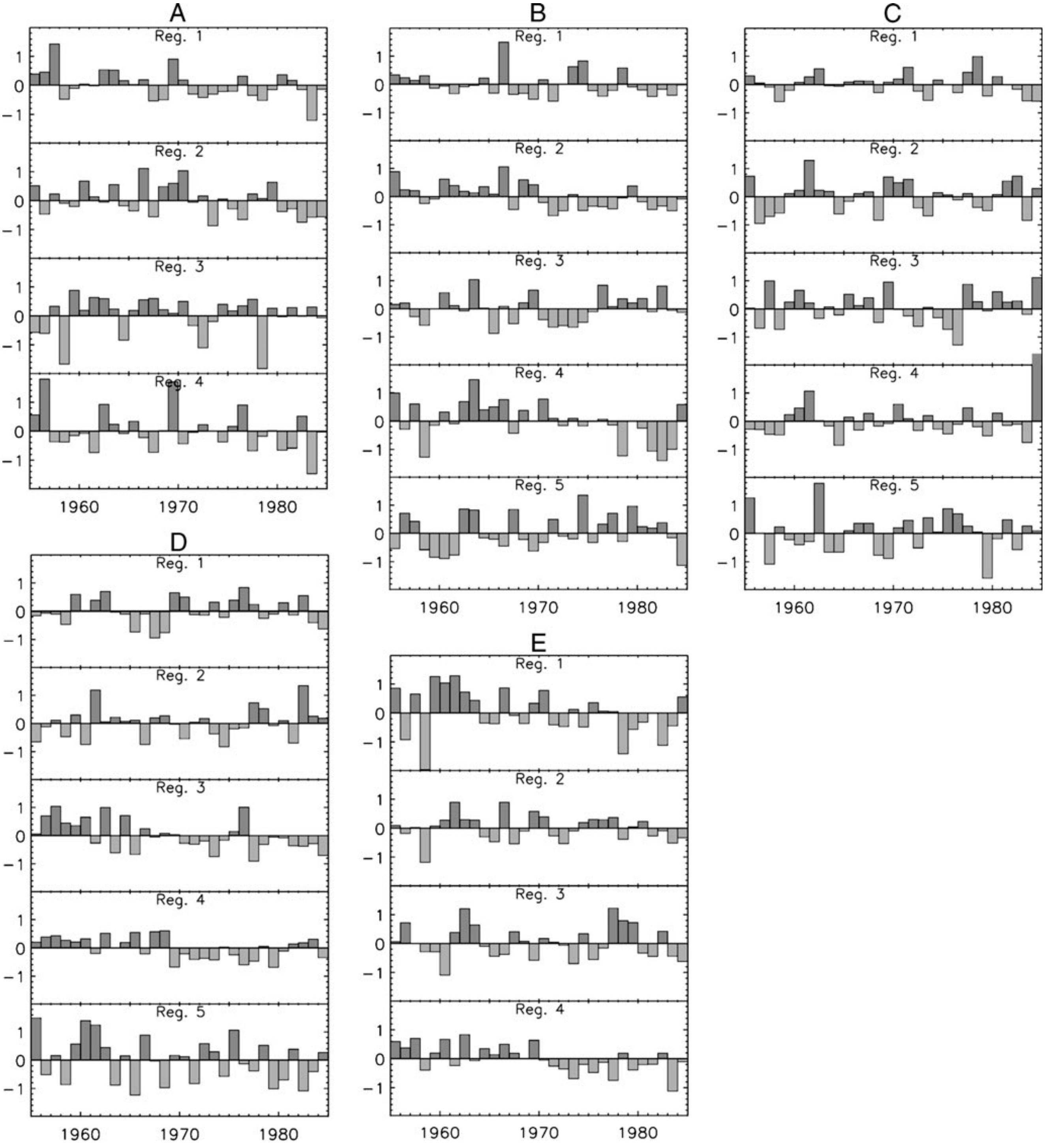

Fig. 7 Rainfall anomalies of different sub-regions using Ward's method for **a** JFM, **b** AMJ, **c** JAS, **d** OND, and **e** annual data

are correlated with that of each sub-region. A few stations are found either to show a very low correlation with all sub-regions or to have a higher correlation with a sub-region other than the one to which they have been assigned. In the first case, which may be due to local influence or poor data quality, the stations are eliminated. In the second case, the stations are reallocated to the correct sub-region. The regional rainfall is then recalculated. However, due to the small number of omitted or reassigned stations, only minor differences appear in time series, comparing with the original data. The results are depicted in Figs. 7 and 8 for Ward's and *k*-means methods, respectively.

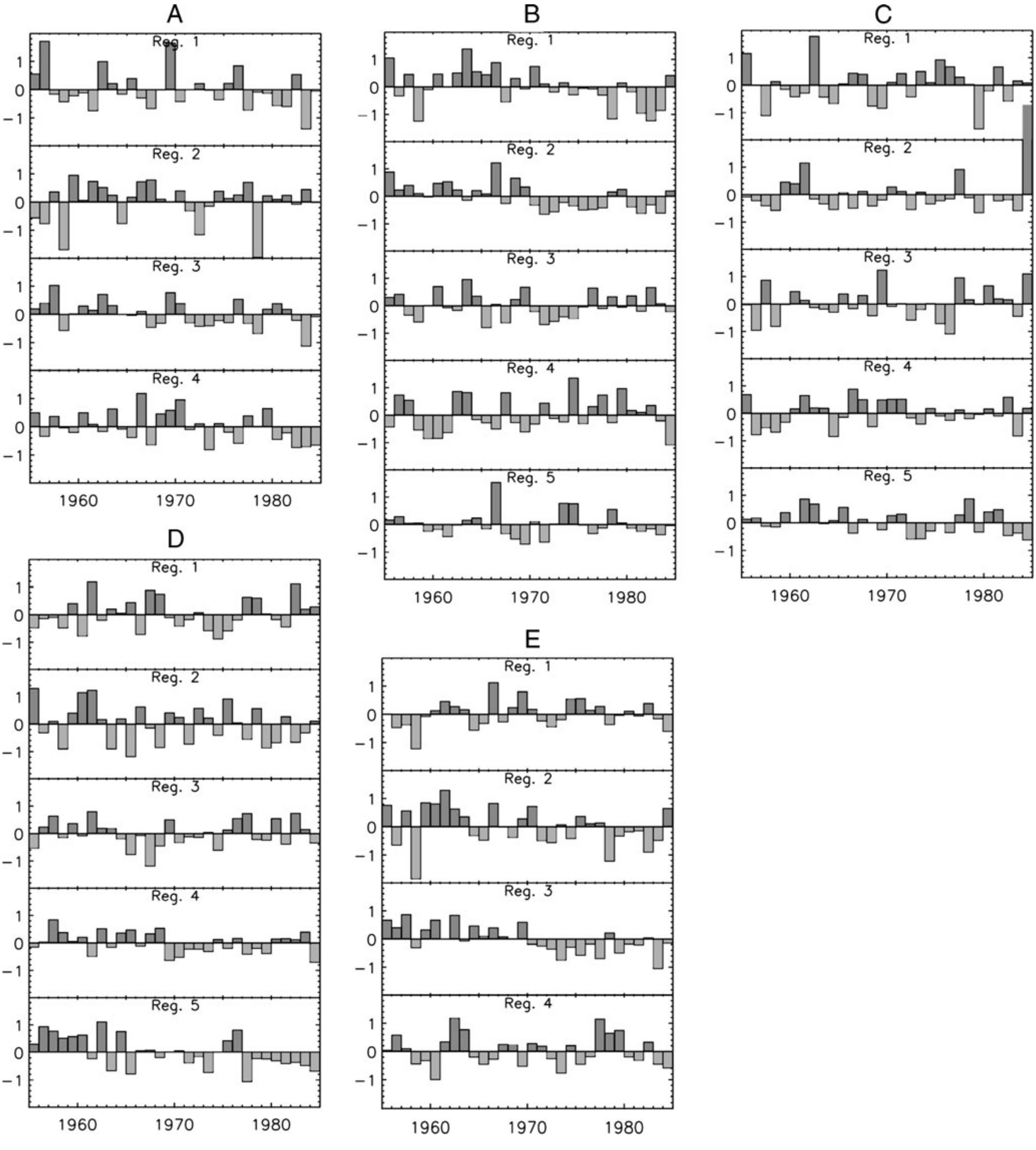

Fig. 8 As in Fig. 7, but for *k*-means method

5 Summary and conclusions

The RPC analysis coupled with two cluster techniques (Ward's and *k*-means) is employed to find homogeneous sub-regions over western equatorial Africa. This robust approach improves our understaning of spatio-temporal variability of WEA rainfall that has not been objectively assessed in previous studies. The total number of clusters obtained by Ward's method is used to initialize the *k*-means method. The optimum number of sub-regions is four for JFM and annual data, and five for AMJ, JAS, and OND. The size and geographical distribution of sub-regions vary from season to season. However, two regions that appeared in most cases are consistent with topography and terrestrial features. The first one develops on a strip along the Atlantic Ocean coastline, starting roughly from the equator and stretching southward. The second one lies on the Mitumba Mountains along the eastern border of study area and within the same latitudes as the first region. The results in southern and middle parts may be involved with more uncertainties due to the sparse geographical distribution of the stations.

Both Ward's and *k*-means methods are found to be appropriate for regionalization and they provide fairly similar results. However, there are minor differences that are more evident for JAS. The sub-regions obtained for this season are also less homogeneous and contiguous. This could be caused by the fact that the boreal summer is a dry season over the region (Figs. 5 and 6) and that makes the influence of local features more pronounced. Also, the noise in rainfall time series can play a bigger role in a dry season.

Previous work by Nicholson (1986) on the entire continent of Africa found the regions delineated for the western equatorial zone were not strongly homogeneous. That was the main motivation of revisiting that region in the current study. Using the same homogeneity test (*F*-test), our results demonstrate significant improvement, where all sub-regions for each season were found to be homogeneous. The improvement owes much to the methodology and intense quality control. Our analysis enables us to compare the regionalization of all seasons. The seasonal variation of the sub-region borders suggests that the climate features driving the interannual rainfall varibability may not be the same for all seasons. This comparative approach was not attempted before since the other studies either used a regionalization based only on annual rainfall (Janicot 1992; Nicholson 1986; Balas et al. 2007) or only one season of rainfall (Djomou et al. 2009).

With the reason that the stations are grouped based on the similarity of their interannual rainfall variability (not other criteria such as annual cycle), the mean rainfall of each sub-region represents the year-to-year rainfall variation. This is of crucial value for exploring the atmospheric circulation factors and teleconnection indices that determine the rainfall variability.

Acknowledgement This work was supported by the National Science Foundation, grants ATM 0004479, 016820, and 0813930, and also by the National Atmospheric and Oceanic Administration grant NA08OAR4310731. I would like to acknowldege Dr. Sharon Nicholson from Florida State University for her valuable comments on the manuscript and for providing me with the rainfall data. I also wish to thank Douglas Klotter for helping me with IDL programming.

References

Baeriswyl P, Rebetez M (1997) Regionalisation of precipitation in Switzerland by means of principal component analysis. Theor Appl Climatol 58:31–41

Balas N, Nicholson SE, Klotter D (2007) The relationship of rainfall variability in West Central Africa to sea-surface temperature fluctuations. Int J Climatol 27:1335–1349

Busuioc A, Chen D, Hellström C (2001) Temporal and spatial variability of precipitation in Sweden and its link with the large-scale atmospheric circulation. Tellus Ser A 53:348–367

Comrie AC, Glenn EC (1998) Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. Clim Res 10:201–215

Corte-Real J, Qian B, Xu H (1998) Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation. Int J Climatol 18:619–635

Dezfuli AK, Karamouz M, Araghinejad S (2010) On the relationship of regional meteorological drought with SOI and NAO over southwest Iran. Theor Appl Climatol 100:57–66

Djomou ZY, Monkam D, Lenouo A (2009) Spatial variability of rainfall regions in West Africa during the 20th century. Atmos Sci Let 10:9–13

Domroes M, Kaviani M, Schaefer D (1998) An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical methods. Theor Appl Climatol 61:151–159

Fovell RG, Fovell MYC (1993) Climate zones of the conterminous United States defined using cluster analysis. J Climate 6:2103–2135

Gong X, Richman MB (1995) On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J Climate 8:897–931

Jackson B, Nicholson SE, Klotter D (2009) Mesoscale convective systems over western equatorial Africa and their relationship to large-scale circulation. Mon Wea Rev 137(4):1272–1294

Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice Hall, Englewood Cliffs

Janicot S (1992) Spatiotemporal variability of West African rainfall. Part I: regionalization and typings. J Climate 5:489–497

Jebari S, Berndtsson R, Uvo C, Bahri A (2007) Regionalizing fine time-scale rainfall affected by topography in semi-arid Tunisia. Hydrol Sci J 52(6):1199–1215

Kamara SI, Jackson IJ (1997) Identification of agro-hydrologic regions in Sierra Leone. Theor Appl Climatol 57:49–63

Kraus EB (1977) Subtropical droughts and cross-equatorial transports. Mon Wea Rev 105:1009–1018

Muñoz-Díaz D, Rodrigo FS (2004) Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. Ann Geophys 22:1435–1448

Nicholson SE (1980) The nature of rainfall fluctuations in subtropical West Africa. Mon Weather Rev 108:473–487

Nicholson SE (1981) Rainfall and atmospheric circulation during drought periods and wetter years in West Africa. Mon Weather Rev 109:2191–2208

Nicholson SE (1986) The spatial coherence of African rainfall anomalies: interhemispheric teleconnections. J Clim Appl Meteorol 25:1365–1381

Nicholson SE (1988) Land surface-atmosphere interaction: physical processes and surface changes and their impact. Prog Phys Geogr 12:36–65

Ogallo LJ (1989) The spatial and temporal patterns of the East African seasonal rainfall derived from principal component analysis. Int J Climatol 9:145–167

Rao AR, Srinivas VV (2006) Regionalization of watersheds by hybrid-cluster analysis. J Hydrol 318:37-56

Raziei T, Bordi I, Pereira LS (2008) A precipitation-based regionalization for Western Iran and regional drought variability. Hydrol Earth Syst Sci 12:1309–1321

Richman MB (1986) Rotation of principal components. J Climatol 6:293–335

Rogers JC, McHugh MJ (2002) On the separability of the North Atlantic oscillation and Artic oscillation. Clim Dyn 19:599–608

Unal Y, Kindap T, Karaca M (2003) Redefining the climate zones of Turkey using cluster analysis. Int J Climatol 23:1045–1055

White D, Richman M, Yarnal B (1991) Climate regionalization and rotation of principal components. Int J Climatol 11:1–25

Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Academic Press, Burlington

Wilson LL, Lettenmaier DP, Skyllingstad E (1992) A hierarchical stochastic model of large scale atmospheric circulation patterns and multiple station daily precipitation'. J Geophys Res 97:2791–2809

Zipser EJ, Cecil DJ, Liu C, Nesbitt SW, Yorty DP (2006) Where are the most intense thunderstorms on earth? Bull Amer Meteor Soc 87:1057–1071