Detection of homogeneous precipitation regions at seasonal and annual time scales, northwest Iran

Mohammad Arab Amiri, Mohammad Saadi Mesgari and Christian Conoscenti

ABSTRACT

Detection of homogeneous climate areas is a challenging issue, which can be affected by different criteria. One of the most prominent factors is choosing the time scale, which can lead to different spatial and temporal patterns. Total precipitation is a key factor in climatological studies, and studying its distribution is of utmost importance. The combination of principal components analysis and cluster analysis is used for homogeneous precipitation areas' detection. Hence, the spatial pattern of total precipitation was investigated in northwestern Iran during the past two decades (1991–2010) on seasonal and annual time scales. The results of clustering on each time scale were validated, and well-defined clusters were investigated and compared with each other. Two homogeneous sub-regions were recognized in spring, the best period for depicting homogeneous precipitation clusters at seasonal resolution. The annual pattern of precipitation delineated three clusters in the study region. Finally, the characteristics of the well-clustered maps reveal the importance of time scale in detection of homogeneous precipitation sub-zones.

Key words | cluster analysis, GIS, principal component analysis, time scale, total precipitation

Mohammad Arab Amiri (corresponding author)
Department of Geographic Information System,
Faculty of Geodesy and Geomatics Eng.,
K. N. Toosi University of Technology,
Tehran, Iran
E-mail: *mohamadamiri89@yahoo.com*

Mohammad Saadi Mesgari
Department of Geographic Information System,
Faculty of Geodesy and Geomatics Eng., and
Center of Excellence in Geospatial Information Technology (CEGIT),
K. N. Toosi University of Technology,
Tehran,
Iran

Christian Conoscenti
Department of Earth and Marine Sciences,
University of Palermo,
Palermo,
Italy

INTRODUCTION

Regional studies on spatial and temporal climate variability are as vital as those of global studies, especially in large countries with different climate regimes (Türkeş *et al.* 2009). Moreover, the changing spatio-temporal patterns of the individual climatic variables are region specific, and vary from one region to another (Qian & Qin 2006) because climate variables vary in time and space, and their spatio-temporal behavior depends on spatial and temporal scales. Some researchers have incorporated spatial and temporal information, and have found that trends for different climate variables differ significantly from region to region (Adamowski *et al.* 2013). Therefore, studying the spatial and temporal variability of climate variables at regional scale is of utmost importance.

Temporal trend analysis and spatial interpolation methods have been widely used in spatio-temporal climate variability studies (Shahid 2009; Santos *et al.* 2010; Martins *et al.* 2012). The majority of the previously performed researches in climate trend analysis focused on long-term trend detection of the main climate variables such as precipitation and temperature (Haylock & Nicholls 2000; Griffiths *et al.* 2003; Qian & Qin 2006; Shahid 2009; Martins *et al.* 2012; Taxak *et al.* 2014; Arab Amiri *et al.* 2016). Regarding spatial variability, a set of possible local and regional factors can contribute to the delineation of homogeneous climatic sub-zones (Adamowski *et al.* 2013). Moreover, homogeneous regions with similar behavior in terms of climatological variables can play an important role in decision-making procedures. This is because identification of homogeneous climate areas at regional scale can be considered an important issue in spatial and temporal analysis of climate time series, and have been proved to be extremely

doi: 10.2166/wcc.2017.088

important in a wide range of fields, such as agriculture, hydrology, energy, ecology, engineering and natural resources management. Hence, spatial variability analysis of climate data is an issue of great importance.

A classification of homogeneous climate sub-regions can be obtained based on the spatial and temporal patterns in climate time series (Vicente-Serrano *et al.* 2015). Principal component analysis (PCA) is a commonly used approach for investigating the spatial variability of climate time series (Türkeş *et al.* 2009; Santos *et al.* 2010; Martins *et al.* 2012; Gocic & Trajkovic 2014; Vicente-Serrano *et al.* 2015; Arab Amiri & Mesgari 2016), and a combination of PCA and cluster analysis (CA) is the most widely used approach for delineation of homogeneous climate sub-zones. Hence, Richman (1986), Dinpashoh *et al.* (2004), Huth (2006), Raziei *et al.* (2008), Vicente-Serrano *et al.* (2015), and other researchers have widely utilized the procedure in detection of homogeneous climate zones.

Another issue in spatio-temporal variability analysis is the time scale or temporal resolution of the data, which can affect the analysis results (Wu *et al.* 2015). The choice of temporal resolution is as important as the choice of spatial scale because the time scale is the key factor when working with periodic data. Furthermore, using different time scales can influence the results of the analysis (Maurya 2013). In other words, the delineated homogeneous sub-zones vary when analyzed at multiple temporal resolutions (Wu *et al.* 2013). Moreover, it is noteworthy to investigate how the choice of time scale may affect the results of the spatial variability analysis at regional scale. Therefore, in this study, the spatial patterns in the precipitation data were investigated on seasonal and annual time resolutions.

Spatial variability analysis of climate parameters in large countries with different climate regimes is of great importance. Iran is a large country with different climate regimes, and the climate of more than 80% of Iran is classified as semi-arid and arid. Thus, the country experiences droughts as well as floods (Madani 2014; Tabari *et al.* 2014). Having considered the great effects of climate changes on the natural physical environment, the significance of climate variability studies over Iran could be highlighted (Tabari *et al.* 2014).

Due to the sparsity of the weather station network, the relatively short period of data availability, different recorded lengths of meteorological data, and missing values in the climatic dataset across Iran, only a few studies about spatial and temporal variability analysis of meteorological variables have been carried out (Dinpashoh *et al.* 2004; Raziei *et al.* 2008; Zoljoodi & Didevarasl 2013). The main objective of this research is to investigate the effects of using different time scales on homogeneous sub-regional climatic areas revealed through spatial regionalization. It is of great importance to study how the choice of temporal scale may affect the results of the spatial variability analysis at regional scale in the study region. For this purpose, a combination of PCA and CA was applied to the precipitation data in northwestern Iran between the years 1991 and 2010. Consequently, time series and geographic information system (GIS) analysis methods can be used in order to evaluate the spatial variability of climatic time series. The paper is structured as follows. The study area and the data used for the analysis are presented in the first part of the next section, and the methodological framework is described in the second part of this section. The main results of the study are presented and discussed in the following section, and the last section provides the conclusions and recommendations for future work.

MATERIALS AND METHODS

Study area and data

Northwest Iran has been selected for this study. Six provinces are located within the study area, including West Azerbaijan, East Azerbaijan, Ardabil, Gilan, Zanjan, and Kurdistan. The study region extends between the latitudes 34° 35′ N and 40° 2′ N, and the longitudes 43° 59′ E and 50° 41′ E, with a total area of approximately 169,723 km^2 including Lake Urmia, and is about 10% of the total land area of Iran. According to the latest census in 2011, this area accounts for around 16% of the national population (Figure 1). The Thiessen polygon map based on geographic coordinates of all synoptic stations is depicted in Figure 1. The polygons can be used to define the area influenced by each station for illustrating the spatio-temporal patterns present in the data set (Wu *et al.* 2015).

Monthly total precipitation records from 24 synoptic stations out of 38 existing stations in the northwestern corner of Iran, for the time interval from January 1991 to

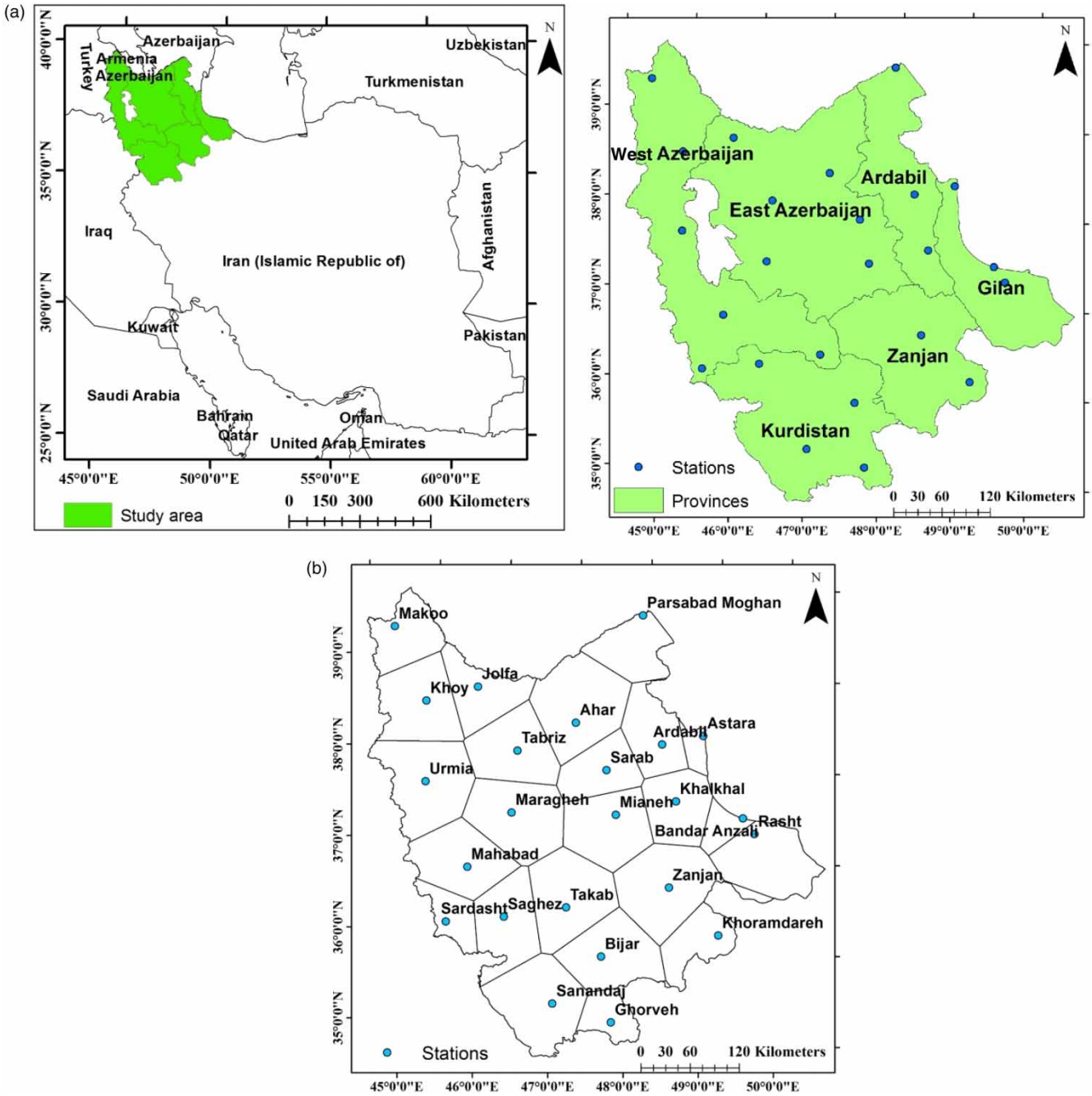

Figure 1 | (a) Location of the study area and (b) the Thiessen polygon map of the study region to define the influence area of each station.

December 2010, recorded by the Islamic Republic of Iran Meteorological Organization (IRIMO), were used in this study. The 24 selected stations have no missing data. The geographic coordinates of the chosen stations, and mean and standard deviation of the time series of the total precipitation amount for each station, are presented in Table 1.

The data used in this study are the monthly total rainfall values recorded by synoptic stations. Since the main objective of the study is evaluating the effects of using different time scales on the results of defining homogeneous areas in terms of precipitation, we aggregated the monthly data in two levels (i.e., seasonal and yearly). It should be noted

Table 1 | Geographical descriptions, mean and standard deviation of annual total precipitation amount of the synoptic stations used in the study

				Annual total precipitation amount	
Station name	**Longitude (dd)**	**Latitude (dd)**	**Altitude (m)**	**Mean (mm)**	**Standard deviation (mm)**
1. Aahar	47.067	38.433	1,390.5	283.445	58.110
2. Ardabil	48.283	38.250	1,332.0	286.720	58.918
3. Astara	48.850	38.367	−21.1	1,357.610	235.200
4. Bandar Anzali	49.450	37.483	−23.6	1,662.000	260.270
5. Bijar	47.617	35.883	1,883.4	335.435	74.170
6. Ghorveh	47.800	35.167	1,906.0	347.265	74.090
7. Jolfa	45.667	38.750	736.2	210.015	67.870
8. Khalkhal	48.517	37.633	1,796.0	367.705	58.340
9. Khoramdareh	49.183	36.183	1,575.0	300.400	82.270
10. khoy	44.967	38.550	1,103.0	255.310	54.470
11. Mahabad	45.717	36.767	1,351.8	411.170	114.050
12. Makoo	44.433	39.333	1,411.3	303.605	70.720
13. Maragheh	46.267	37.400	1,477.7	292.475	91.060
14. Mianeh	47.700	37.450	1,110.0	274.045	67.140
15. Urmia	45.050	37.667	1,328.0	308.600	105.740
16. Parsabad Moghan	47.917	39.650	31.9	267.935	63.650
17. Rasht	49.617	37.317	−8.6	1,306.475	268.160
18. Saghez	46.267	36.250	1,522.8	458.340	142.390
19. Sanandaj	47.000	35.333	1,373.4	397.560	97.980
20. Sarab	47.533	37.933	1,682.0	249.520	42.060
21. Sardasht	45.483	36.150	1,556.8	874.255	199.480
22. Tabriz	46.283	38.083	1,361.0	245.955	57.100
23. Takab	47.100	36.400	1,817.2	318.420	90.410
24. Zanjan	48.483	36.683	1,663.0	290.420	65.100

that for the computation of cumulated seasonal precipitation, we arranged the time series as hydrological seasons: winter (January, February, March), spring (April, May, June), summer (July, August, September) and autumn (October, November, December) (Raziei *et al.* 2008). Therefore, the precipitation data series were used in this study at two different levels, namely seasonal and annual.

An extensive pre-processing procedure was applied to precipitation time series data sets, before using data series in the study. For this purpose, five homogeneity tests, including the standard normal homogeneity test, the Buishand range test, the Pettitt test, the Von Neumann ratio test, and the double-mass curve analysis were applied to the time series of the total precipitation amount at the 5% significance level. Finally, all the used tests recognized all series as sufficiently homogeneous data sets for entering into the analysis.

Methodological framework

The methodology used in this study is summarized in Figure 2.

Principal component analysis

PCA is a statistical multivariate technique that has been widely applied in various applications such as climatology and meteorology for dimensionality reduction. The

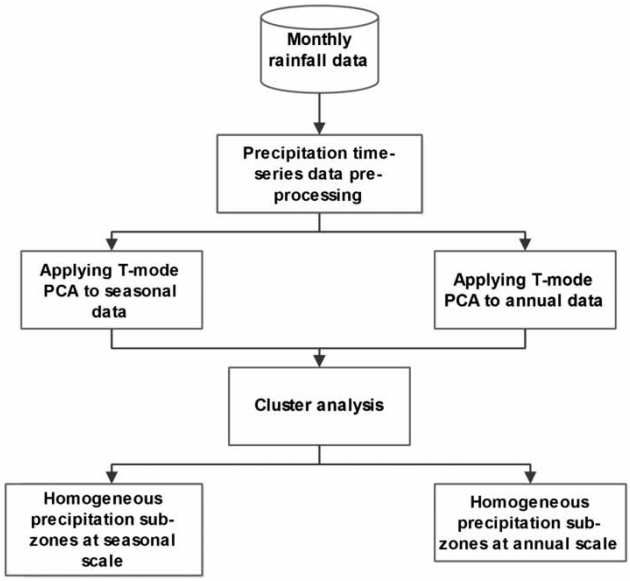

Figure 2 | The methodology used in this study.

method is used for forming new orthogonal uncorrelated variables that are linear combinations of the original highly inter-correlated variables (Vicente-Serrano *et al.* 2015). PCA uses the characteristic equation for computing the eigenvalues and eigenvectors. PCA computations can be done based on either the correlation matrix or the covariance matrix of the observed variables. Moreover, PCA is not being affected by the lack of independence between the original variables, and using normalized data series is not necessary for analysis by such a method (Kalayci & Kahya 2006; Santos *et al.* 2010). Consequently, PCA can be used for capturing information about spatial or temporal co-variability pattern in climate data.

We used the T mode PCA defined by Richman (1986) for matrix configuration. The T-mode PCA can be applied to meteorological data in order to derive general spatial patterns (Vicente-Serrano *et al.* 2015). For this purpose, a data matrix where rows correspond to the stations and columns for the observations is constructed for analyzing precipitation time series. Therefore, the formed matrix is used as input data arrangement for PCA.

Having applied PCA to the precipitation data, the obtained principal components (PCs) were rotated for achieving more stable spatial patterns (Santos *et al.* 2010). The Varimax method is the most widely used technique for producing rotated PCs, and attempts to maximize the inter-correlation between the variables and the components (Ercan *et al.* 2008; Santos *et al.* 2010). Hence, in this study, the Varimax procedure was used for rotating the obtained PCs in order to achieve more spatially localized PCs.

Cluster analysis

The spatial structure of the precipitation data can be more simplified by determining regions which have similar behavior in different time spans. The output of PCA (the rotated PC scores) can be used subsequently in CA. Hence, clustering is applied to the PC scores for homogeneous precipitation area detection.

Choosing the right clustering method for categorizing the stations into groups with similar recorded observations is another issue of great importance. For this purpose, stations can be regarded as objects and the Varimax rotated PC scores can be treated as attributes, and the clustering algorithm can partition objects (stations) into sub-groups based on their similarities (Wu *et al.* 2015). Gong & Richman (1995) concluded that the k-means as a non-hierarchical technique outperformed hierarchical methods such as the Wards's algorithm when working with precipitation data sets (Santos *et al.* 2010). Fovell & Fovell (1993) also noted that the Euclidean distance is the most relevant mathematical criterion or distance measure for classification purposes. From a spatial perspective, k-means starts with the initialization of randomly chosen k stations as cluster centroids; then, the sum of squared errors between each station and its corresponding cluster centroids is computed (Wu *et al.* 2015). In other words, k-means clustering aims to minimize the within-cluster sum of squares by iteratively assigning each station to the closest cluster centroid and re-calculating new cluster centroids. Thus, k-means clustering method was applied to the rotated PC scores in order to classify the stations into groups with minimum within-group variability and maximum between-group variability.

The variance ratio criterion for cluster validation

There are many approaches for cluster validation and determining how many clusters are needed to optimize the definition of homogeneous objects/observations in various applications, such as Ball-Hal, Banfeld-Raftery, C index, and variance ratio

criterion (VRC) (Liu *et al.* 2010). One of the most commonly used methods is the VRC defined by Caliński & Harabasz (1974). The index has some advantages over other indices, such as its ease of calculation and its effectiveness in different clustering approaches. The VRC evaluates the cluster validity based on the average between- and within-cluster variances (Liu *et al.* 2010); and can be used to determine the proper number of clusters in CA. The VRC is defined as follows:

$$VRC_k = \frac{(BCV/(K-1))}{(WCV/(N-K))} \quad (1)$$

where BCV is the overall between-cluster variance (inter-cluster), and WCV is the overall within-cluster variance (intra-cluster) with regard to all clustering variables; K is the number of clusters, and N is the number of observations/objects. The BCV is defined as:

$$BCV = \sum_{k=1}^{K} n_k \|C_k - C\|^2 \quad (2)$$

where C_k is the center of cluster k, C is the overall mean of data, $\|.\|^2$ is the Euclidean distance between the two vectors.

The overall within-cluster sum of squares WCV is defined as:

$$WCV = \sum_{k=1}^{K} \sum_{x \in C_k} ||x - C_k||^2 \quad (3)$$

where x is a data point, and C_k is the kth cluster.

Well-defined clusters are determined by maximizing the value of this index; because, a large between-cluster sum of squares and a small within-cluster sum of squares lead to a large VRC, and the larger the VRC, the better the data clustering. Therefore, the optimal number of clusters is determined by maximizing VRC (Liu *et al.* 2010).

RESULTS AND DISCUSSION

As stated earlier, the data used in this study were the recorded monthly total precipitation at synoptic stations in the studied area. Having applied the T-mode PCA to the input data, the Varimax procedure was used for rotating the obtained PC scores. Before rotating the obtained PC scores, we used the Kaiser's criterion to decide how many PCs should be retained for rotation. Therefore, PCs corresponding to all eigenvalues greater than one were retained and rotated with the Varimax method. The number of PCs retained, the cumulative percentage of total variation, the explained percentage of variances for either un-rotated PCs, and Varimax rotated PCs for each time scale, are shown in Table 2.

The rotated PC scores were used as inputs for CA. For this purpose, the K-means clustering algorithm was used to identify homogeneous precipitation sub-regions in the study area. The optimum number of clusters was determined through a trial and error procedure; i.e., we applied the K-means algorithm six times (one cluster to six clusters) for each time scale, and the optimum number of clusters

Table 2 | Explained variance (%) by the loadings with and without rotation for total precipitation in each time scale

			Principal components							
Time scale	Time period	Explained variance (%)	PC-1	PC-2	PC-3	PC-4	PC-5	PC-6	PC-7	Cumulative percentage of total variation
Seasonal	Winter	Un-rotated	42.57	13.81	10.05	7.15	6.32	4.74		84.64
		Varimax rotated	25.14	12.13	11.37	16.68	12.44	6.88		
	Spring	Un-rotated	54.19	18.05	6.56	4.88				83.66
		Varimax rotated	24.95	17.99	24.31	16.42				
	Summer	Un-rotated	29.45	17.05	14.00	7.21	6.60	5.56	4.50	84.37
		Varimax rotated	17.40	11.21	14.21	11.85	13.55	9.02	7.13	
	Autumn	Un-rotated	62.92	9.14	6.19	4.39				82.65
		Varimax rotated	32.59	8.57	25.23	16.25				
Annual	Year	Un-rotated	50.34	13.24	9.34	5.60				78.51
		Varimax rotated	34.55	12.15	14.75	17.06				

were determined by means of calculating the VRC index. Hence, the number of clusters which maximize the VRC index was considered as the optimum number of clusters for each time scale. Variance decomposition for the optimal classification of homogeneous precipitation regions and the optimum number of clusters, are shown in Table 3.

In the following sections, the properties of each delineated sub-region in each temporal resolution are briefly described, and the effects of using different temporal scales investigated. The best time scale for homogeneous precipitation area detection in the studied area in each time scale is introduced and the results of clustering homogeneous areas in different time scales compared.

Spatial regionalization at seasonal time scale

The seasonal pattern of total precipitation is shown in Figure 3. The studied area was classified into six, two, five, and six homogeneous sub-regions in winter, spring, summer, and autumn, respectively. Tables 2 and 3 also show that the maximum amount of the VRC index belongs to spring with only four PCs retained which explains more than 83% of the total variance. The results of clustering in spring illustrate that the study area is divided into two separate regions, namely, the northern and eastern parts, and the southern part.

The average amounts of the recorded total precipitation and the standard deviations for the delineated clusters for each season are shown in Figure 4. In winter, the average amount of total rainfall varied in a range between 68 mm (the first and the fourth clusters) and 395 mm (the fifth cluster). The second cluster also showed the second highest mean precipitation at about 322 mm. In spring, the first cluster (the north and east of the study region) showed an average amount of total rainfall at about 127 mm. It is also noteworthy that there is a little difference between the average total rainfall values of the second cluster at just above 113 mm in comparison with the first cluster.

In summer, the average values of precipitation were less than 30 mm in four out of five delineated clusters; while the whole average of the recorded precipitation in the second sub-region was about 267 mm, the highest amount of average precipitation in comparison with other clusters. In autumn, the average cumulative rainfall ranged between approximately 76 mm (the fifth sub-region) and 505 mm (the third sub-zone). Autumn also showed the most recorded average precipitation amount in comparison with other seasons.

In general, the investigation of variances within and between clusters indicated that the definition of homogeneous sub-regions in spring are more reasonable compared with other seasons. The other well-defined clusters at seasonal time resolution were defined in autumn with relatively high values of the VRC index. Winter and summer, on the other hand, showed meaningless delineated clusters; and it can be perceived that the delineated clusters are not meaningful in the northwest of Iran in winter and summer. The overall pattern of precipitation showed that there are four distinct parts over the study region, including the eastern part, southern part, northwestern part, and northern part. The results of homogeneous precipitation area delineation can be used in water resources

Table 3 | Variance decomposition for the optimal classification, VRC, and the optimum number of clusters for each time scale

Time scale	Time period	Variance decomposition	Between-classes variance	Within-class variance	Total variance	VRC	Number of clusters
Seasonal	Winter	Absolute (mm^2)	7.69	3.43	11.12	8.06	6
		Percent (%)	69.12	30.88	100.00		
	Spring	Absolute (mm^2)	3.66	3.73	7.39	21.53	2
		Percent (%)	49.46	50.54	100.00		
	Summer	Absolute (mm^2)	8.34	5.39	13.73	7.35	5
		Percent (%)	60.76	39.24	100.00		
	Autumn	Absolute (mm^2)	4.32	1.11	5.43	13.99	6
		Percent (%)	79.54	20.46	100.00		
Annual	Year	Absolute (mm^2)	5.50	3.18	8.68	18.16	3
		Percent (%)	63.37	36.63	100.00		

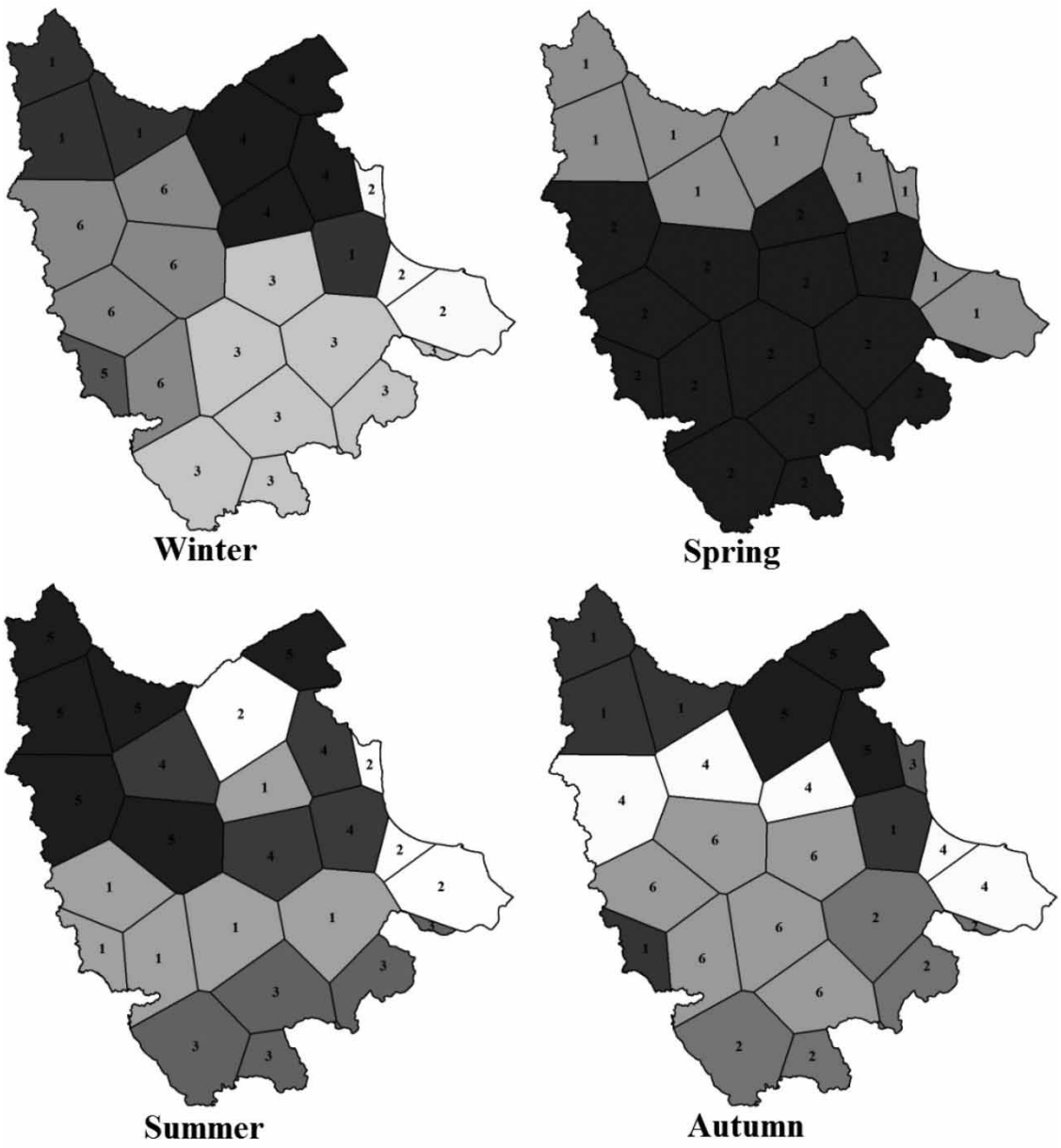

Figure 3 | The delineated clusters at seasonal time scale; the numbers within each polygon and the used grayscales are for visually discriminating the clusters.

management, irrigation control, rain-fed agriculture, and soil erosion prevention (Shahid 2009).

As an example, Figure 5 depicts the spatial distribution of the first four rotated PC scores across the study area during spring. The first PC score, PC-1, has the highest positive score over the southern part of the region, including the regions with lower average amounts of precipitation in spring. The second PC score has high positive values in the northern part and low positive and negative values in the southern part. The third PC score has the highest negative score over the southern part, and low negative and positive values in the northern part. The PC-4 score also has high positive values in the eastern part, including regions with the highest amount of precipitation during spring. Finally, the patterns of all PC scores are identical, delineating the two clusters, as two distinctive sub-regions

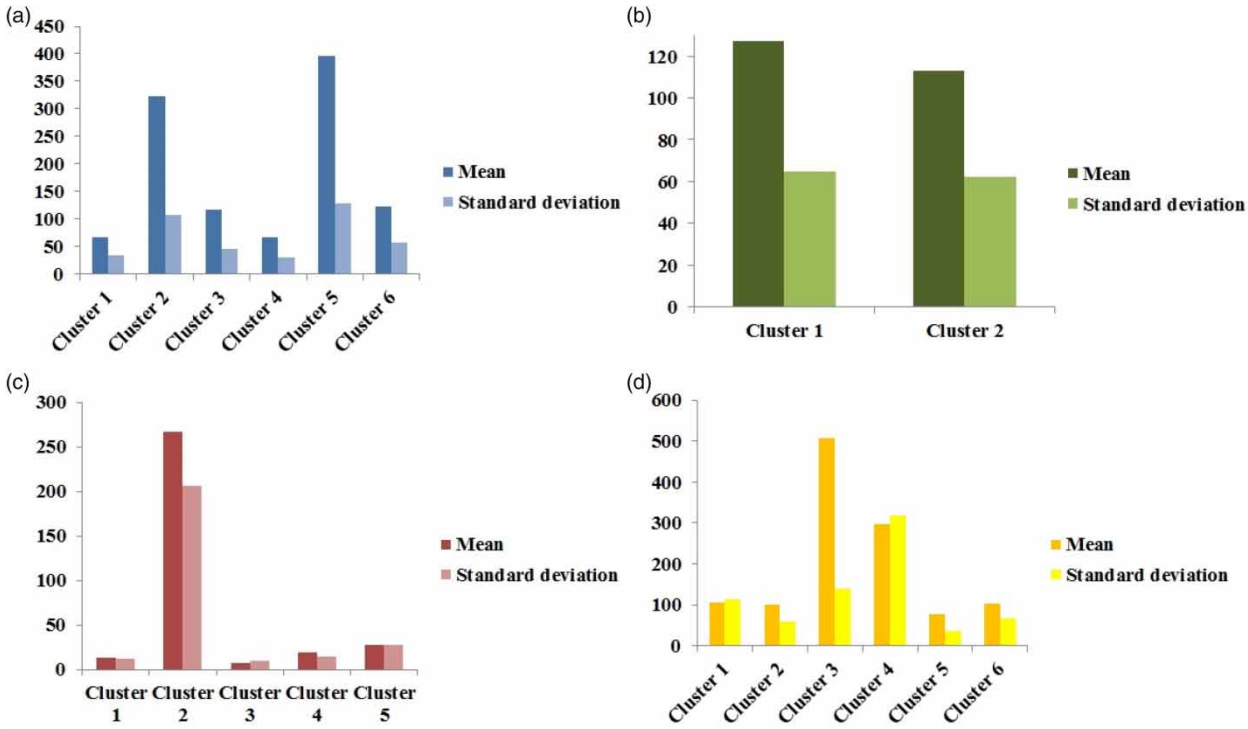

Figure 4 | Mean and standard deviation of the recorded total precipitation amounts for each cluster in: (a) winter, (b) spring, (c) summer, and (d) autumn.

with different spatial variability of total precipitation in spring.

Spatial regionalization at yearly time scale

The precipitation-based regionalization using annual data sets is shown in Figure 6. The study area was classified into three clusters using annual time series. Tables 2 and 3 also showed that the four retained PCs explained more than 78% of the total variation, and the VRC index is relatively high as a result of using annual data sets.

The study area was divided into three homogeneous precipitation sub-regions at yearly resolution. The northern part of the study area and the domain of influence belonging to Khalkhal and Ardebil stations was classified as the first cluster. The eastern region of the study area which is located near the Caspian Sea was categorized as the second cluster. The third cluster consists of the central, the southern, and western parts of the study area, and is the biggest cluster.

Figure 7 illustrates the average amount of annual total precipitation for each station and also for each cluster. The average amount of yearly precipitation in the first cluster varied between approximately 210 mm (Jolfa station) and 367 mm (Khalkhal station). The stations situated in this cluster seem to be more homogeneous than in other clusters; and the average amount of precipitation for the first cluster is about 281 mm. The second cluster with an extremely high amount of precipitation consists of three stations near the Caspian Sea. The average total yearly precipitation in this cluster is about 1,442 mm. The third cluster, where average precipitation varied between 246 mm (Tabriz station) and 874 mm (Sardasht station), had an average amount of yearly precipitation at about 359 mm. Therefore, the three clusters depicted at yearly resolution are distinctive sub-regions with different annual precipitation variability and characteristics.

The spatial distributions of all four rotated PC scores are shown in Figure 8. The first PC score has the highest positive score over the southern and central parts (cluster 3), including regions with moderate amounts of annual precipitation. The PC-2 score, on the other hand, has the highest positive

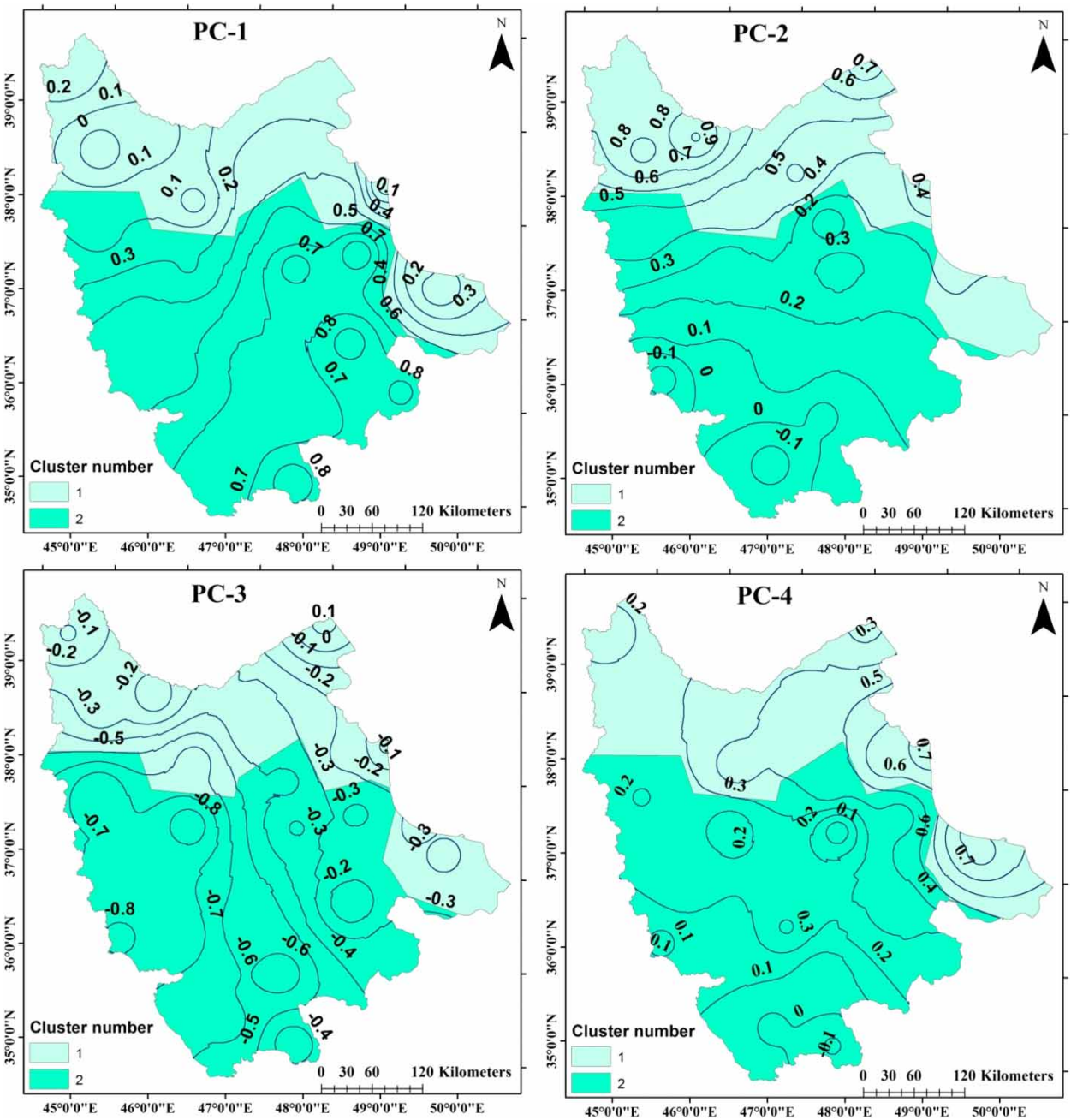

Figure 5 | Varimax rotated scores of total precipitation in spring; note that the background map shows clusters delineated through PCA and CA.

values in the eastern part (cluster 2), which is the region with the highest amount of annual total precipitation. Interestingly, the PC-3 score depicts the highest amount of positive score in the first cluster. Thus, the spatial distribution of the third PC score delineates regions categorized as cluster 1. The PC-4 score has the lowest negative values in northwestern and northeastern parts, including regions with the lowest amount of annual total precipitation. Finally, the spatial distributions of all PC scores delineate the borders of the obtained clusters at yearly resolution.

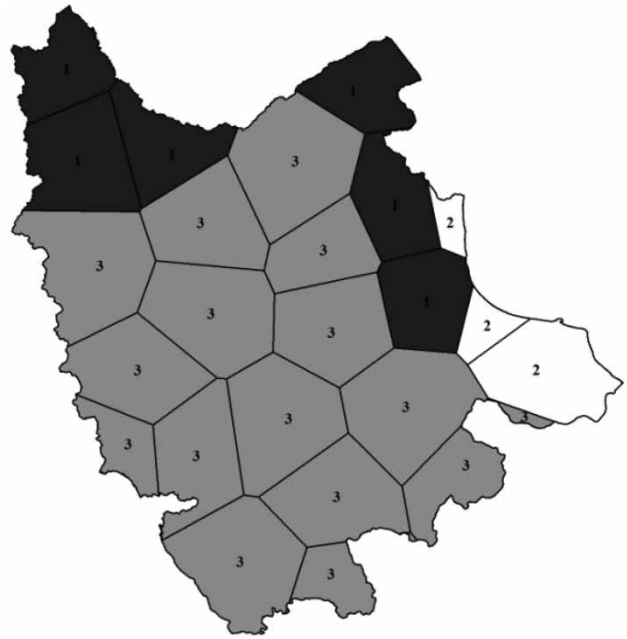

Figure 6 | The delineated clusters at yearly time scale; the numbers within each polygon and the used grayscales are for visually discriminating the clusters.

CONCLUSIONS

The spatial variability of total precipitation was investigated in northwestern Iran using precipitation time series from 24 synoptic stations during the period of 1991–2010. The spatio-temporal pattern of precipitation was analyzed using the T-mode PCA. The PCA was applied to time series of the total precipitation at different time scales, and then the CA was used for clustering the rotated PC scores. The number of clusters delineated by K-means algorithm was selected based on the VRC index in each time scale.

The results reveal the importance of using different time scales in changing the spatial pattern of total precipitation over the study region; i.e., selecting different time scales (seasonal and annual) and also different durations in each time scale (such as spring and winter in the seasonal time resolution) lead to different spatial patterns of total precipitation across the study area. The clustered maps of total precipitation in each time scale were compared using the VRC index.

Having compared the homogeneous sub-regions in each time scale, well-defined regionalized maps could be determined. The results of the analysis at seasonal time scale depict that spring had the most well-defined clusters in comparison with other seasons. The precipitation pattern of spring depicts that there are two separate sub-regions across the region, including the northern and eastern parts, and the southern part, where the latter had a slightly lower average amount of precipitation during spring. Moreover, the spatial characteristics of the annual precipitation delineate three homogeneous sub-regions;

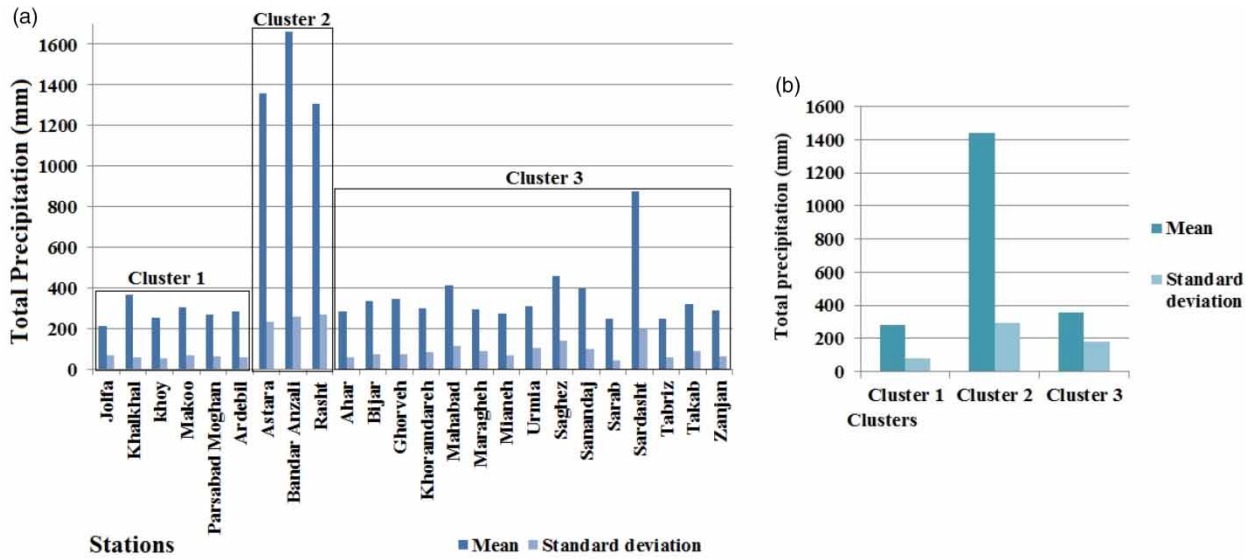

Figure 7 | Mean and standard deviation of annual total precipitation amounts: (a) for each station and (b) for each cluster.

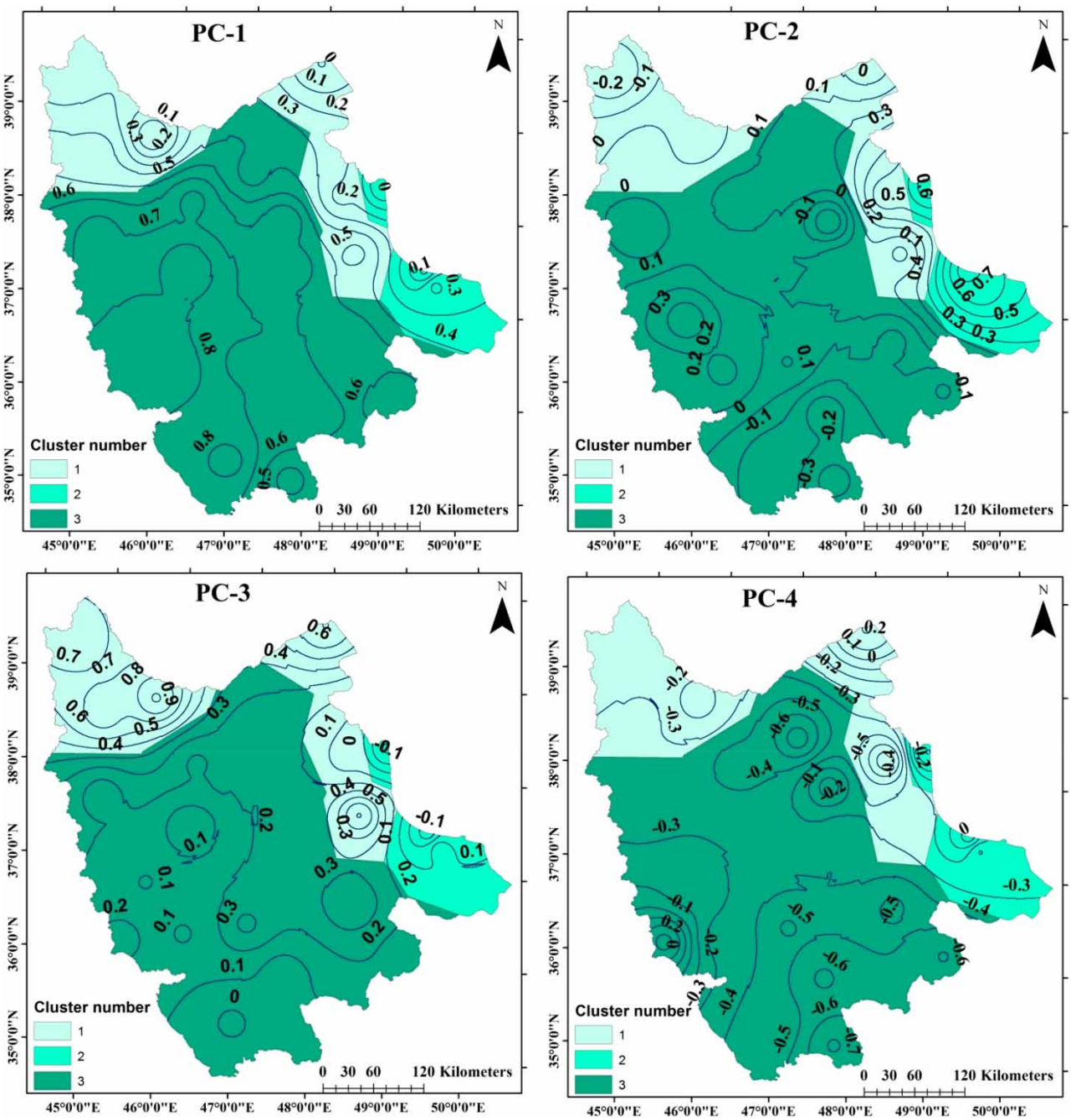

Figure 8 | Varimax rotated PC scores of the annual total precipitation; note that the background map shows clusters delineated through PCA and CA.

and the regions classified in the eastern part had the most annual precipitation in comparison with the other two clusters. The delineated homogeneous sub-zones can be used for agriculture, disaster mitigation, soil degradation and desertification prevention programs, water resources planning and management in the context of regional climatic change.

Here we provide recommendations for future work. The topographic characteristics of the region are an important factor that should be considered in defining the boundaries

of homogeneous precipitation sub-regions to provide more accurate clusters. More research using longer time series is needed to further approve the accuracy of the spatially homogeneous areas detected at each time scale in this study. The use of satellite images and other sources of raw data could also be beneficial in order to decrease the uncertainty arising due to the sparsity of climate stations. Furthermore, other variables can be used in order to give more accurate results regarding precipitation-based regionalization. However, these recommendations were not considered in this study because the main purpose of this study was to only investigate the impacts of time scale on the precipitation-based spatial regionalization.

REFERENCES

Adamowski, J., Adamowski, K. & Prokoph, A. 2013 Quantifying the spatial temporal variability of annual streamflow and meteorological changes in eastern Ontario and Southwestern Quebec using wavelet analysis and GIS. *Journal of Hydrology* **499**, 27–40.

Arab Amiri, M. & Mesgari, M. S. 2016 Spatial variability analysis of precipitation in northwest Iran. *Arabian Journal of Geosciences* **9** (11), 1–10.

Arab Amiri, M., Amerian, Y. & Mesgari, M. S. 2016 Spatial and temporal monthly precipitation forecasting using wavelet transform and neural networks, Qara-Qum catchment, Iran. *Arabian Journal of Geosciences* **9** (5), 1–18.

Caliński, T. & Harabasz, J. 1974 A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods* **3** (1), 1–27.

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. & Mirnia, M. 2004 Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology* **297** (1), 109–123.

Ercan, K., Mehmet, C. & Osman, A. 2008 Hydrologic homogeneous regions using monthly streamflow in Turkey. *Earth Sciences Research Journal* **12** (2), 181–193.

Fovell, R. G. & Fovell, M.-Y. C. 1993 Climate zones of the conterminous United States defined using cluster analysis. *Journal of Climate* **6** (11), 2103–2135.

Gocic, M. & Trajkovic, S. 2014 Spatiotemporal characteristics of drought in Serbia. *Journal of Hydrology* **510**, 110–123.

Gong, X. & Richman, M. B. 1995 On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of Climate* **8** (4), 897–931.

Griffiths, G., Salinger, M. & Leleu, I. 2003 Trends in extreme daily rainfall across the South Pacific and relationship to the South Pacific Convergence Zone. *International Journal of Climatology* **23** (8), 847–869.

Haylock, M. & Nicholls, N. 2000 Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. *International Journal of Climatology* **20** (13), 1533–1541.

Huth, R. 2006 The effect of various methodological options on the detection of leading modes of sea level pressure variability. *Tellus A* **58** (1), 121–130.

Kalayci, S. & Kahya, E. 2006 Assessment of streamflow variability modes in Turkey: 1964–1994. *Journal of Hydrology* **324** (1), 163–177.

Liu, Y., Li, Z., Xiong, H., Gao, X. & Wu, J. 2010 Understanding of internal clustering validation measures. In: *2010 IEEE 10th International Conference on Data Mining (ICDM)*, IEEE, pp. 911–916.

Madani, K. 2014 Water management in Iran: what is causing the looming crisis? *Journal of Environmental Studies and Sciences* **4** (4), 315–328.

Martins, D., Raziei, T., Paulo, A. & Pereira, L. 2012 Spatial and temporal variability of precipitation and drought in Portugal. *Natural Hazards and Earth System Sciences* **12** (5), 1493–1501.

Maurya, R. 2013 *Effects of the Modifiable Temporal Unit Problem on the Trends of Climatic Forcing and NDVI Data Over India.* Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands.

Qian, W. & Qin, A. 2006 Spatial-temporal characteristics of temperature variation in China. *Meteorology and Atmospheric Physics* **93** (1), 1–16.

Raziei, T., Bordi, I. & Pereira, L. 2008 A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrology and Earth System Sciences* **12** (6), 1309–1321.

Richman, M. B. 1986 Rotation of principal components. *Journal of Climatology* **6**, 293–335.

Santos, J. F., Pulido-Calvo, I. & Portela, M. M. 2010 Spatial and temporal variability of droughts in Portugal. *Water Resources Research* **46** (3), 1–13.

Shahid, S. 2009 Spatio-temporal variability of rainfall over Bangladesh during the time period 1969–2003. *Asia-Pacific Journal of Atmospheric Science* **45** (3), 375–389.

Tabari, H., AghaKouchak, A. & Willems, P. 2014 A perturbation approach for assessing trends in precipitation extremes across Iran. *Journal of Hydrology* **519**, 1420–1427.

Taxak, A. K., Murumkar, A. & Arya, D. 2014 Long term spatial and temporal rainfall trends and homogeneity analysis in Wainganga basin, Central India. *Weather and Climate Extremes* **4**, 50–61.

Türkeş, M., Koç, T. & Sariş, F. 2009 Spatiotemporal variability of precipitation total series over Turkey. *International Journal of Climatology* **29** (8), 1056–1074.

Vicente-Serrano, S., Chura, O., López-Moreno, J., Azorin-Molina, C., Sanchez-Lorenzo, A., Aguilar, E., Moran-Tejeda, E., Trujillo, F., Martínez, R. & Nieto, J. 2015 Spatio-temporal

variability of droughts in Bolivia: 1955–2012. *International Journal of Climatology* **35** (10), 3024–3040.

Wu, X., Zurita-Milla, R. & Kraak, M.-J. 2013 Visual discovery of synchronisation in weather data at multiple temporal resolutions. *The Cartographic Journal* **50** (3), 247–256.

Wu, X., Zurita-Milla, R. & Kraak, M.-J. 2015 Co-clustering geo-referenced time series: exploring spatio-temporal patterns in Dutch temperature data. *International Journal of Geographical Information Science* **29** (4), 624–642.

Zoljoodi, M. & Didevarasl, A. 2013 Evaluation of spatial-temporal variability of drought events in Iran using Palmer drought severity index and its principal factors (through 1951–2005). *Atmospheric and Climate Sciences* **3**, 193–207.

First received 18 June 2016; accepted in revised form 3 June 2017. Available online 5 August 2017