Digest 2012, December 2012, 1609-1622

Classification of Annual Precipitations and Identification of Homogeneous Regions using K-Means Method†

Mahmut FIRAT*
Fatih DİKBAŞ**
Abdullah Cem KOÇ***
Mahmud GÜNGÖR****

ABSTRACT

Reliable and correct estimation of hydrological and meteorological processes is one of the major problems in regions with insufficient hydrologic information and data. The classification of the hydrological variables and determination of homogeneous regions are the most important steps of regional studies. The purpose of this study is to classify the annual total precipitation series and to identify the homogeneous regions by K-Means method. The K-means method, which is the simplest and most commonly used clustering method, divides a data set into clusters by minimizing the sum of the Euclidean distance between each feature vector and its closest cluster centre. The annual precipitation records and longitude, latitude and altitude values obtained of 188 stations operated by the National Meteorology Works (DMI) in Turkey were considered for clustering analysis. The number of clusters was determined as 7. Moreover, the regional homogeneity test was applied for testing the homogeneity of regions identified by clustering analysis.

Keywords: *Clustering, k-means method, annual precipitation, homogeneity test.*

1. INTRODUCTION

Accurate modeling of hydrological and meteorological processes is very important in water resources planning and management, safety design and management of dams, spillways and other water structures. To achieve this aim enough amount of data should be needed. May be the most important but the hardest step for the regional estimation studies is determination of the homogenous regions. Usually geographically close stations accepted to be in the same region. But this kind of classified regions cannot be said as hydrologicaly homogenous. Transfer of information between two regions is impossible so for more trustworthy regional study, hydrologicaly similar regions should be determined. Hierarchical and non-hierarchical classification techniques are used to determine and classifying the homogenous regions. Classification techniques especially used to determine homogenous regions in regional flood frequency analysis [1, 2, 3, 4, 5, 6, 7] and classification of precipitation variable and determining homogenous precipitation regions

* İnönü University, Malatya, Turkey - mahmut.firat@inonu.edu.tr
** Pamukkale University, Denizli, Turkey - f_dikbas@pau.edu.tr
*** Pamukkale University, Denizli, Turkey - a_c_koc@pau.edu.tr
**** Pamukkale University, Denizli, Turkey - mgungor@pau.edu.tr
† Published in Teknik Dergi Vol. 23, No. 3 July 2012, pp: 6037-6050

[8, 9]. Demirel [10] aimed classifying Turkish watersheds and forming similar regions by hierarchical classifying method. In the study which was performed by Turan [11] Turkish river productivities classified by Ward method. Kahya et al. [12] aimed classification of Turkish rivers geographically by K-Means method. Demirel et al. [13] used K-Means method in the classification of basic drought components in Turkish watersheds. Isik and Singh [14] classified the Turkish watersheds according to the mean monthly discharges by using the Ward and K-Means methods. Classification of annual precipitations and determining homogenous regions in Turkey with nonhierarchical (K-Means) classification method is the aim of this study. To this aim annual precipitation of 188 meteorological stations all around Turkey were used. Furthermore L-Moments based homogeneity test was applied to check homogeneity.

2. NON-HIERARCHICAL CLUSTERING METHOD (K-MEANS)

Clustering analysis is a multivariable statistical technique which is used categorizing and separating data into homogenous subsets. K-Means method is one of the well known clustering techniques and it has a nonhierarchical structure [7, 15]. K-Means method can divide and classify C number clusters from D number variables belonging the data set X and N number properties vector [6, 16]. In this method process starts by determination of the C number cluster centers and each member of the data set is allocated to the nearest cluster by means of the similarity criterion. After the appointment cluster centers are recalculated and some members can be assigned to another cluster. This process is repeated until the changing stopped [16]. If N is the number of properties vector and d is the number of variables the X data set for a problem can be defined as $X = \{x_k | k = 1,2,.....N\}$.

In this data set kth property vector can be written as $x_k = [x_{k1,} x_{k2} .... x_{kd}]$, $x_k \in R^d$ [17].

In K-Means method minimization of aim function is intended which is given in Eq. (1) when the data set separates into subclusters. Minimizing this equation requires minimizing the distance between each property vector and the nearest cluster center thus ensuring similar structured data collecting in the same cluster. In the literature Euclidean distance which is given in Eq. (2) is used as a distance measure [16].

$$J(S:X) = \sum_{i=1}^{C} \sum_{k=1}^{N} d_{ik}^2 (x_k, s_i) \tag{1}$$

d_{ik}^2 which appears in the Eq. (1) is worked out as follows.

$$d_{ik}^2 = \left\| x_k^{(i)} - s_i \right\|^2 \tag{2}$$

$$s_i = \frac{1}{N} \sum_{k=1}^{N} x_k^{(i)} \tag{3}$$

Initiative of the Euclidean distance matrix is given below.

$$\begin{bmatrix} d(1,1) & d(1,2) & d(1,3) & \ldots & d(1,C) \\ d(2,1) & d(2,2) & & & \\ d(3,1) & & & & \\ & & & & \\ d(N,1) & d(N,2) & d(N,3) & \ldots & d(N,C) \end{bmatrix} \tag{4}$$

Where, C is the number of clusters, S is the matrix containing the cluster centers $(S = \{s_1, s_2, ...s_C\})$, s_i is the center of i^{th} cluster, $x_k^{(i)}$ is the k^{th} property vector of the i^{th} cluster, d_{ik}^2 is the distance between the k^{th} property vector of the i^{th} cluster and center of the i^{th} cluster. Perhaps the most important advantage of the K-Means methods is the simplicity of its application. On the other hand performance of the method is dependent to the cluster number and cluster centers selected initially is the foremost disadvantage. Procedure steps for the K-Means method; (i) Random determination of the number of clusters and their centers, (ii) Calculation of the distance from the cluster center for each variable, (iii) Appointment of the each variable to the nearest cluster, (iv) Minimization of the aim function, (v) Recalculation of the cluster centers and forming new clusters, (vi) Iteration repeated until no change in the cluster memberships.

3. DISCORDANCY AND REGIONAL HOMOGENEITY TESTS

Discordancy measure given in the Eq. (5) is used to find the discordant stations in the regions which are determined after the clustering process [18, 19].

$$D_i = \frac{1}{3} N_i (u_i - \overline{u})^T A^{-1} (u_i - \overline{u}) \tag{5}$$

$$u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}] \tag{6}$$

$$\overline{u} = \frac{1}{N} \sum_{i=1}^{N} u_i \tag{7}$$

$$A = \sum_{i=1}^{N} (u_i - \overline{u})(u_i - \overline{u})^T \tag{8}$$

Where, D_i is the measure of discordancy, N is the number of property vectors, u_i is the vector containing L-moment ratios for the station i, $\overline{u}$ is the regional average, T is the transpose of a vector or a matrix, A is the covariance matrix, $t^{(i)}$ is the *L-Cv* moment ratio of station *i*, $t_3^{(i)}$ is the *L-Cs* moment ratio of station *i*, $t_4^{(i)}$ is the *L-Ck* moment ratio of station *i*. The station *i* can be neglected if the D_i measure is high which is calculated by using the Eq. (5). For the regions that station number is greater than 15 in case the D_i value for a station were greater than 3 that station is best be eliminated from the region [18, 19]. Homogeneity test (H test) which is based on the L-Moment ratios (*L-Cv variability*, *L-Cs skewness* and *L-Ck kurtosis*) is used for testing the homogeneity of the regions after the cluster analysis. Groups which are obtained by the cluster analysis are evaluated by means of this test. Variations in the L-Moments for a group of stations in a region are compared. This variation can be defined as the ratio of weighted standard deviation of regional L-moment ratios and data length of each station. $L-Cv$ ratio is advised to be used of its effect on the variance forecast is greater than $L-Cs$ and $L-Ck$ ratio variations [18, 19]. L-moment ratios calculated for N gauging stations $(t^{(i)}, t_3^{(i)}, t_4^{(i)})$ and regional mean L-moment ratios ($t^R = \sum_{i=1}^{N} n_i t^{(i)} / \sum_{i=1}^{N} n_i$, $t_3{}^R = \sum_{i=1}^{N} n_i t_3{}^{(i)} / \sum_{i=1}^{N} n_i$, $t_4{}^R = \sum_{i=1}^{N} n_i t_4{}^{(i)} / \sum_{i=1}^{N} n_i$) can be written for a region. Weighted Standard deviation (V_1) based on the $L-Cv$ ratios can be determined by the Eq. (9) [18, 20, 21].

$$V_1 = \left[\frac{\sum_{i=1}^{N} n_i (L-Cv^{(i)} - \overline{L-Cv})^2}{\sum_{i=1}^{N} n_i} \right] \tag{9}$$

Where, $L-Cv^{(i)}$ is the L-moment ratio for the station *i*, $\overline{L-Cv}$ is the regional mean L-moment ratio, n_i is the data length of the station *i*, V_1 is the weighted Standard deviation based on the $L-Cv$ ratios. In this study four parameter Kappa distribution is used and adjusted to the regional mean L-moment ratios to evaluate the homogeneity measure instead of two and three parameter distributions. Kappa distribution represents many distributions used in the frequency analysis of hydrologic phenomena. Parameters of the Kappa distribution are location parameter, scale parameter and two shape parameters Advantage of the Kappa distribution is less restrictive according to the other two or three parameter distributions. Distribution function $F(x)$ and probability density function $f(x)$ of the Kappa distribution are given in the Eqs. (10) and (11) respectively [18].

$$F(x) = \left\{ 1 - \alpha \left[1 - \frac{\omega}{\delta}(x - \xi) \right]^{1/\omega} \right\}^{1/\alpha} \qquad \omega \neq 0 \quad \alpha \neq 0 \tag{10}$$

$$f(x)=\frac{1}{\delta}\left[1-\frac{\omega}{\delta}(x-\xi)^{1/k}\right]^{1/\omega-1}\left\{1-\alpha\left[1-\frac{\omega}{\delta}(x-\xi)\right]^{1/\omega}\right\}^{1/\alpha-1} \quad (11)$$

$$\begin{array}{ll} \xi+\delta/\omega \leq x<\infty & \omega \leq 0 \;\; ve \;\; \alpha<0 \\ \xi+\delta(1-\alpha^{-\omega})/\omega \leq x \leq \xi+\delta/\omega & \omega>0 \;\; ve \;\; \alpha>0 \\ -\infty<x \leq \xi+\delta/\omega & \omega \leq 0 \;\; ve \;\; \alpha>0 \\ \xi+\delta(1-\alpha^{-\omega})/\omega \leq x<\infty & \omega>0 \;\; ve \;\; \alpha<0 \end{array} \quad (12)$$

Where, $f(x)$ is the probability density function, ξ is the location parameter of Kappa distribution, δ is the scale parameter of the Kappa distribution, ω and α are shape parameters of the Kappa distribution. Kappa distribution was used to produce 500 homogenous data regions by means of L-moment ratios and simulation was carried out. Variability was calculated for every produced region and Standard deviation and mean of these variations were calculated. H measure was obtained by using Eq. (13) to test homogeneity.

$$H_k=\frac{V_k-\mu_{Vk}}{\sigma_{Vk}} \qquad k=1,2,3 \quad (13)$$

Where, H_k is the regional homogeneity measure (H_1, H_2, H_3), H_1 is the regional homogeneity measure for L-variability, H_2 is the regional homogeneity measure for L-skewness, H_3 is the regional homogeneity for L-kurtosis, σ_{Vk} is the standard deviation of the values obtained by the simulation, μ_{Vk} is the mean of the values obtained by the simulation, V_k is the weighted Standard deviation calculated by using the regional statistics and data (V_1, V_2 and V_3). Results of the calculated H values are evaluated; (i) If $H<1$ then cluster is "*accepted as homogenous*", (ii) If $1 \leq H < 2$ then, cluster "*may be homogenous*", (iii) If $H \geq 2$ then cluster is "*accepted as definitely heterogeneous*" [18, 19].

4. STUDY SITE AND DATA

In this study, total annual precipitation data taken from the stations throughout Turkey operated by The General Directorate of State Meteorological Affairs (DMI) were used for cluster analysis. As selection of the stations and the data to be used is largely influential on the results of cluster analysis, the missing value status, observation period and reliability of data were investigated for all observed precipitation stations. It is proposed in the papers in literature that the stations to be used in regional estimation studies and analysis should have a statistically significant number of data [18]. By considering this, 188 precipitation observation stations were selected having data between 01.10.1967 and 30.09.1998 and the locations of these stations are presented in Figure 1. It is stated in literature that variables with different scales influence the clustering results and that the data should be normalized with appropriate transformation functions [20, 21]. Also in this study, the data were normalized by using the transformation functions given in Equations (14), (15) and (16) before being used in clustering analysis.

$$P_{yi} = (P_i - P_{\min})/(P_{mak} - P_{\min}) \tag{14}$$

$$Y_{yi} = (Y_i - Y_{\min})/(Y_{mak} - Y_{\min}) \tag{15}$$

$$Z_{yi} = Z_i / Z_{mak} \tag{16}$$

Where, P_i is the precipitation in station *i*, P_{yi} is the normalized precipitation in station I, P_{mak} is the maximum precipitation, P_{min} is the minimum precipitation, Y_i is the latitude or longitude of station I, Y_{yi} is the normalized latitude or longitude of station I, Y_{mak} is the maximum latitude or longitude, Y_{min} is the minimum latitude or longitude, Z_i is the normalized elevation of station i and Z_{mak} is the maximum elevation.

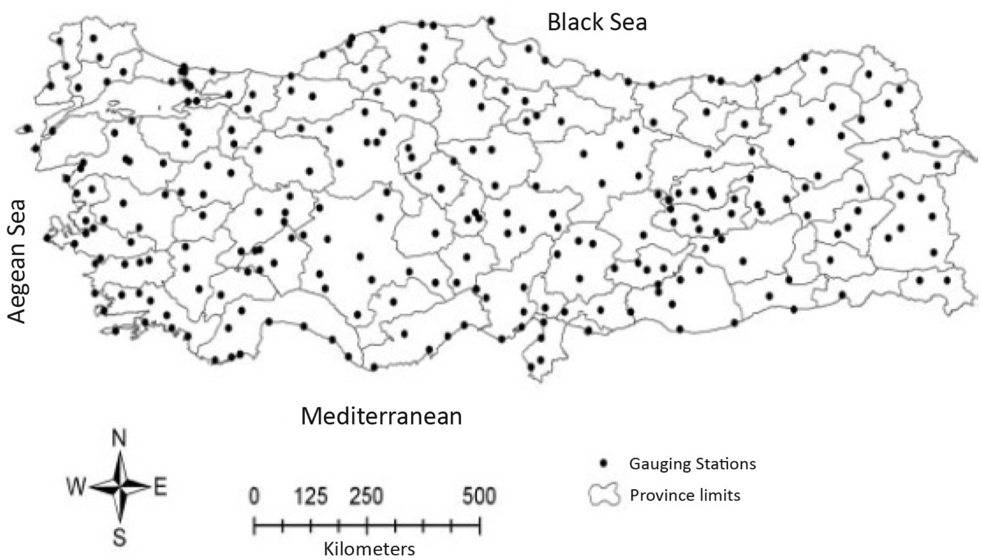

Figure 1. Rain gauge stations used in the clustering analysis

5. RESULTS AND CONCLUSION

The data used in the clustering of total annual precipitations by using K-Means method consists of 188 property vectors and 4 variables (a 188x4 matrix). As a result of the clustering made by using this data set, the number of clusters was found to be 7. Matlab R.14 software was used in the clustering of the precipitation series and determination of homogeneous regions. Figure 2 shows the locations of the stations determined by using K-Means method.

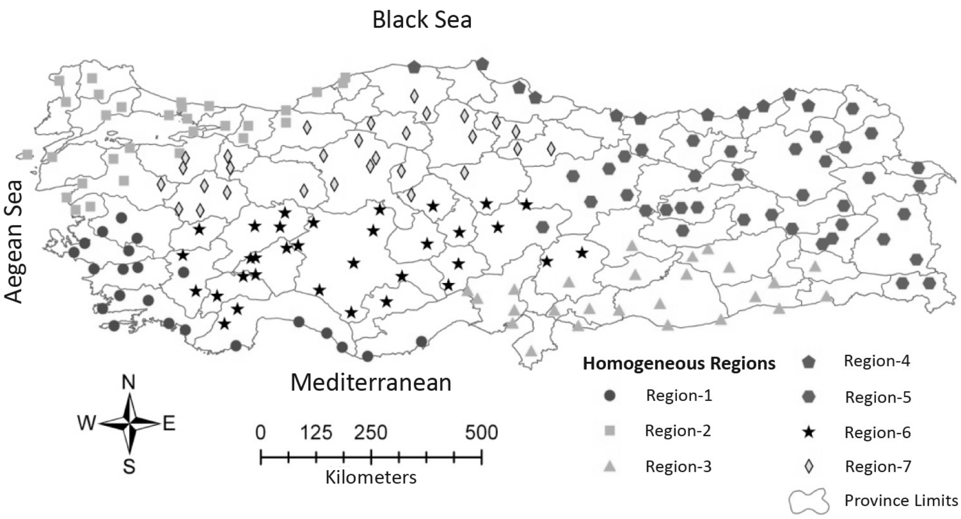

Figure 2. Separation of stations in the regions which are determined by K-Means method

The results in Figure 2 shows that the Region 1 covers the mid and west parts of the Mediterranean Region, Region 2 covers the Marmara Region and the west part of Black Sea Region, Region 3 covers the Southeast Anatolian Region, Region 4 covers the Mid and East Black Sea Regions, Region 5 covers East Anatolian Region and the inner parts of East Black Sea Region, Region 6 covers the southern part of the Central Anatolian Region and the Region 7 covers the northern part of Central Anatolian Region and the inner parts of Western Black Sea Region. The regional average $P_{ort}^{R}(mm)$ and maximum value $P_{m}^{R}(mm)$, regional average of standard deviation $\sigma_{ort}^{R}(mm)$ and its maximum value $\sigma_{m}^{R}(mm)$ and the mean of the coefficient of variation ${C_v}_{ort}^{R}$ and its maximum value ${C_v}_{m}^{R}$ for the annual precipitations of the stations in the determined regions are calculated and presented in Table 1. To be able to use the regions determined by using cluster analysis in future studies, the discordancy test for regional homogeneity should be applied. For testing the homogeneity of the 7 regions and for applying the discordancy test, H (H_1, H_2 and H_3) and D values proposed by Hosking and Wallis [18, 19] are calculated and presented in Table 2.

Table 1. The Statistics Calculated for the Clusters Determined by Using K-Means Method

Region	Number of Stations	$P_{ort}^{R}\,(mm)$	$P_{m}^{R}\,(mm)$	$\sigma_{ort}^{R}\,(mm)$	$\sigma_{m}^{R}\,(mm)$	$C_{v\,ort}^{\;R}$	$C_{v\,m}^{\;R}$
1	22	745.25	1167.99	254.80	169.53	0.23	0.29
2	30	729.37	1211.48	216.95	145.78	0.20	0.30
3	25	645.04	1099.73	267.50	166.76	0.26	0.34
4	11	1204.04	2203.28	286.16	158.83	0.14	0.21
5	38	562.19	1287.51	311.19	117.71	0.20	0.29
6	33	452.94	745.00	151.37	90.29	0.20	0.31
7	29	517.92	1490.18	274.93	93.05	0.18	0.23

Table 2. Regional Homogeneity Test for the Clusters Determined by Using K-Means Method

Region	Number of Stations	Number of Station-Year	H Test			Discordancy Station Number (D)
			H_1	H_2	H_3	
1	22	682	-1.532	-2.330	-1.872	17232 (3.40)
2	30	930	-0.462	-1.056	-0.884	17110 (3.08)
3	25	775	-0.600	-1.672	-1.588	-
4	11	341	1.854	-0.781	-0.726	-
5	38	1178	1.173	0.667	0.155	17088 (3.65) 17668 (3.43)
6	33	1023	-0.092	-0.799	-0.912	17926 (4.69)
7	29	899	-3.300	-3.556	-4.024	17748 (4.05)

In the evaluation of discordancy test results calculated for the stations, it is proposed that the stations with a D value of 3 and higher could be left out of assessment in the future regional studies as they could not pass the discordancy test when the number of stations in the cluster is higher than 15 [18, 19]. According to the results given in the table, the D values for the stations 17088, 17110, 17232, 17668, 17748 and 17926 are higher than the limit. The comparison of the regional homogeneity tests calculated for the regions show that, only the H_1 value calculated for cluster 7 is higher than 2 and that the H_1 values calculated for the remaining clusters are lower than 2. The H_1 values calculated for the clusters 2, 3, and 6 are lower than 1 and they are evaluated as “acceptably homogeneous”.

As the H_1 values for the clusters 1, 2 and 5 are between 1 and 2, these clusters are evaluated as "possibly homogeneous". But, as the H_1 value calculated for Cluster 7 is higher than 2, it is seen that this cluster is not sufficiently homogeneous. According to these results, homogeneity was not obtained for only one cluster in the classification of total annual precipitation and determination of hydrologically homogeneous regions and it can be stated that the K-means method is generally successful in the classification of total annual precipitation. In this study, solutions with Ward's method which is a hierarchical clustering method were also made for evaluating the results of K-Means method and making comparisons. In the solutions, the number of clusters was determined to be 7 and regional homogeneity tests were applied for each cluster. Table 3 shows the results of regional homogeneity and discordancy tests for the determined regions.

Table 3. Regional Homogeneity Test for the Clusters Determined by Using Ward's Method

Region	Number of Stations	Number of Station-Year	*H* Test			Discordancy Station Number (D)
			H_1	H_2	H_3	
1	14	434	2.370	-0.825	-0.962	-
2	42	1302	1.042	0.393	-0.029	17088 (3.58) 17172 (3.09) 17668 (3.53)
3	26	806	-0.303	-0.925	-0.703	-
4	19	589	-3.044	-3.015	-3.312	-
5	41	1271	-0.636	-1.704	-1.775	17824 (3.13) 17926 (5.48)
6	27	837	-1.517	-2.168	-1.795	-
7	19	589	-1.503	-2.173	-2.115	-

The results of the Ward's method indicate that the H_1 values calculated for the regions 1 and 4 are higher than the limit value and these regions were evaluated as "definitely heterogeneous". When we look at the H_1 values of the other regions, we evaluate that the regions 3 and 5 are "acceptably homogeneous" and the regions 2 and 6 are "possibly homogeneous". According to the results given in Table 2, only one region was seen to be "definitely heterogeneous", 3 regions "acceptably homogeneous" and 3 regions "possibly homogeneous". These results show that, while 1 region was found to be "heterogeneous" in the K-means method, 2 regions were "heterogeneous" in the Ward's method.

6. RESULT

In this study, K-Means method was used for clustering the total annual precipitation throughout Turkey and for determining the homogeneous regions. For this, the data of 188 precipitation observation stations with sufficient observation periods were used and the number of clusters was found to be 7 as a result of cluster analysis. Regional homogeneity test was applied based on L-moments method and the discordancy test for the determined regions. It was seen from the discordancy test results that the D values calculated for 6 stations were higher than the limit value. According to the results of the regional homogeneity tests, as only the H_1 value calculated for region 7 was higher than 2, it was evaluated to be “definitely heterogeneous” and it was seen that this region does not have sufficient homogeneity. The H_1 values for the remaining clusters were lower than 2 and while the clusters 2, 3 and 6 were “acceptably homogeneous”, the clusters 1, 2 and 5 were “possibly homogeneous”. In this study, the precipitation series were also classified by using Ward’s method and regional homogeneity test was applied for the determined regions. According to the results of the regional homogeneity tests applied for the regions determined by using both methods, only 1 region was found to be “heterogeneous” in the K-Means method and 2 regions were found “heterogeneous” in the Ward’s method.

According to these evaluations, it can be stated that the results obtained by using K-Means method in the clustering of total annual precipitations are at an acceptable level.

Acknowledgements

This study was supported by TÜBİTAK within the project 107Y318. The writers would like to thank TÜBİTAK for the support, and to The General Directorate of State Meteorological Affairs (DMI) for their contribution in obtaining information and to the Editor and the referees for their valuable ideas and suggestions.

Symbols

A	: Covariance matrix
C	: Number of clusters
d	: Number of variables
d_{ik}^2	: Distance between the k^{th} property of i^{th} cluster vector and center of i^{th} cluster
D_i	: Discordancy criterion
$F(x)$	: Total distribution function
$f(x)$	: Probability density function
H_k	: Regional homogeneity criteria (H_1, H_2, H_3)
H_1	: Regional homogeneity criteria for L-Cv

H_2	: Regional homogeneity criteria for L-Cs
H_3	: Regional homogeneity criteria for L-Ck
$L-Cv$	: L-Moment ratio (L-variability)
$L-Cs$	: L-Moment ratio (L-skewness)
$L-Ck$	: L-Moment ratio (L-Kurtosis)
$L-Cv^{(i)}$	: L-moment ratio for ith station (L-variability),
$\overline{L-Cv}$	: Regional mean L-moment ratio
N	: Number of property vectors
n	: Length of data for the station
P_i	: Total precipitation of i^{th} station
P_{yi}	: Normalized precipitation of i^{th} station
P_{mak}	: Maximum precipitation
$P_{\min}$	: Minimum precipitation
P_{ort}^R	: Regional average of annual precipitations for the stations which are in the regions determined by the clustering analysis
P_m^R	: Maximum value of annual precipitations for the stations which are in the regions determined by the clustering analysis
S	: Matrix which contains the cluster centers
s_i	: Center of the i^{th} cluster
T	: Transpose of a vector or a matrix
$t^{(i)}$	: *L-Cv* L-moment ratio of the station *i*
$t_3^{(i)}$	: *L-Cs* L-moment ratio of the station *i*
$t_4^{(i)}$	: *L-Ck* L-moment ratio of the station *i*
t^R	: Regional mean of the L-Cv L-moment ratio
t_3^R	: Regional mean of the L-Cs L-moment ratio
t_4^R	: Regional mean of the L-Ck L-moment ratio
u_i	: The vector containing the i^{th} station's L-moment ratios
$\overline{u}$	: Regional average of the L-moment ratios

V_k	: Weighted standard deviation calculated from the regional statistics and data
V_1	: Weighed standard deviation based on the L-Cv ratios
V_2	: Weighed standard deviation based on the L-Cs ratios
V_3	: Weighed standard deviation based on the L-Ck ratios
Y_i	: Latitude or longitude of the i^{th} station
Y_{yi}	: Normalized latitude or longitude of the i^{th} station
Y_{mak}	: Maximum latitude or longitude
$Y_{\min}$	: Minimum latitude or longitude
Z_i	: Elevation of the i^{th} station
Z_{yi}	: Normalized elevation of the i^{th} station
Z_{mak}	: Maximum elevation
X	: Data set
x_k	: k^{th} property vector of the data set
$x_k^{(i)}$	: k^{th} property vector of the i^{th} set
σ_{Vk}	: Standard deviation of the values obtained by simulation
σ_{ort}^R	: Regional average of the standard deviation of annual precipitations stations in the regions which are determined by the cluster analysis
σ_m^R	: Maximum value of the standard deviations of annual precipitations stations in the regions which are determined by the cluster analysis
C_{vort}^R	: Regional average of the coefficient of deviation of annual precipitations stations in the regions which are determined by the cluster analysis
C_{vm}^R	: Maximum value of the coefficient of deviations of annual precipitations stations in the regions which are determined by the cluster analysis
μ_{Vk}	: Average of the values obtained by simulation
ω ve α	: Shape parameters for the Kappa distribution
ξ	: Location parameter for the Kappa distribution
δ	: Scale parameter for the Kappa distribution

References

[1] Mosley M. P., Delimitation of New Zealand Hydrologic Regions. Journal of Hydrology, 49, 173–192, 1981.

[2] Acreman, M.C., Sinclair, C. D., Classification of Drainage Basins According to Their Physical Characteristics, An Application for Flood Frequency Analysis in Scotland. Journal of Hydrology, 84(3-4), 365-380, 1986.

[3] Burn, D.H., Cluster Analysis as Applied to Regional Flood Frequency. Journal of Water Resources Planning and Management, 115, 567–582. 1989.

[4] Burn, D. H., Catchment Similarity for Regional Flood Frequency Analysis using Seasonality Measures. Journal of Hydrology, 202, 212–230, 1997.

[5] Lecce, S.A., Spatial Variations in the Timing of Annual Floods in the Southeastern United States. Journal of Hydrology, 235, 151–169, 2000.

[6] Burn, D.H., Goel, N.K., The Formation of Groups for Regional Flood Frequency Analysis. Hydrological Sciences Journal, 45(1), 97–112, 2000.

[7] Burn, D. H., Zrinji, Z., and Kowalchulk, M., Regionalization of Catchments for Regional Flood Frequency Analysis. Journal of Hydrologic Engineering, 2(2), 76–82, 1997.

[8] Guttman, N.B., The use of L-Moments in the Determination of Regional Precipitation Climates. Journal of Climate 6, 2309–2325, 1993.

[9] Soltani, S., Modarres, R., Classification of Spatio-Temporal Pattern of Rainfall in Iran Using A Hierarchical and Divisive Cluster Analysis. Journal of Spatial Hydrology, 6(2), 1-12, 2006.

[10] Demirel, M.C. 2004. Cluster Analysis of Streamflow Data over Turkey. Master of Science Thesis. İstanbul Technical University, 119p.

[11] Turan, A., Türkiye Akarsu Verimlerinin Küme Analizi ile Sınıflandırılması. Sakarya University, School of Natural and Applied Sciences, Master Thesis, 155p, 2005.

[12] Kahya, E., Demirel, M.C., Piechota, T.C., Spatial Grouping of Annual Streamflow Patterns in Turkey. Hydrology Days, 169-176, 2007.

[13] Demirel, M.C., Mariano, A.J., Kahya, E., Performing K-Means Analysis to Drought Principal Components of Turkish Rivers. Hydrology days, 145-151, 2007.

[14] Isik, S., Singh, V.P., Hydrologic Regionalization of Watersheds in Turkey. Journal of Hydrologic Engineering. 13(9), 824-834, 2009.

[15] Lin, G-F., Chen, L-H., Identification of Homogeneous Regions for Regional Frequency Analysis using the Self-Organizing Map. Journal of Hydrology, 324, 1-9, 2006.

[16] Rao, A., Srivinas, V.V., Regionalization of Watersheds by Fuzzy Cluster Analysis. Journal of Hydrology, 318, 57-79, 2006.

[17] Hall, M.J., Minns, A.W., The Classification of Hydrologically Homogeneous Regions. Hydrological Sciences Journal, 44(5), 693-704, 1999.

[18] Hosking, J.R.M., Wallis, J.R., Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, Cambridge.1997.

[19] Hosking, J.R.M., Wallis, J.R., Some Statistics Useful in Regional Frequency Analysis. Water Resources Research, 29 (2), 271–281, 1993

[20] Cannarozzo, M., Noto, L.V., Viola, F., La Loggia, G., Annual Runoff Regional Frequency Analysis in Sicily. Physics and Chemistry of the Earth, 34, 679–687, 2009.

[21] Lim, Y.H., Voeller, D.L., Regional Flood Estimations in Red River using L-Moment-Based Index-Flood and Bulletin 17B Procedures. Journal of Hydrologic Engineering, 14(9), 1002-1016, 2009.