DOI:10.28974/idojaras.2025.3.5

IDŐJÁRÁS

Quarterly Journal of the HungaroMet Hungarian Meteorological Service
Vol. 129, No. 3, July – September, 2025, pp. 339–355

K-means clustering of precipitation in the Black Sea Region, Türkiye

Aslı Ulke Keskin[1], **Gurkan Kır**[1], and **Utku Zeybekoglu** [2,*]

[1]*Faculty of Engineering, Ondokuz May University*
55139, Atakum, Samsun, Türkiye

[2]*Boyabat Vocational School of Higher Education*
Sinop University
57200 Boyabat, Sinop, Türkiye

**Corresponding Author e-mail: utkuz@sinop.edu.tr*

(Manuscript received in final form May 31, 2024)

Abstract— In recent years, there has been a significant uptick in the frequency of disasters stemming from the impacts of global climate change. In response, both nationally and internationally, various studies are being conducted to mitigate these effects. Classifying regions affected by climate change into similar classes based on climate parameters is crucial for applying consistent methodologies in studies conducted within these regions. This approach will help determine the most appropriate strategies for mitigating the effects of climate change in these regions. The study utilized observational records of annual precipitation from 31 stations in the Black Sea Region, sourced from the Turkish State Meteorological Service, covering the data spans the period between 1982 and 2020. Cluster analysis was conducted using the k-means algorithm. The optimal cluster among those formed was determined through the silhouette index analysis. The study suggests that the optimal number of clusters is 2.

Key-words: clustering, precipitation, silhouette analysis, k-means

1. Introduction

Climate, characterized by extreme values of meteorological parameters such as precipitation, temperature, and wind, represents the collective state of the atmosphere for a specific location over a defined time interval (*Demircan et al.,* 2017). Until the mid-1950s, long-term averages of these parameters were generally assumed to remain unchanged. As we progress into the 20th century, rapid industrial developments have resulted in the unplanned consumption of natural resources, increased environmental pollution in tandem with population growth, and the release of substantial amounts of greenhouse gases into the atmosphere. Consequently, greenhouse gases, with their heat-retaining capacity, have begun to induce changes in climate parameters over time. These alterations in climate parameters are commonly referred to as global climate change (*Turkes*, 2010; *Ozkoca*, 2015). On a global scale, climate change manifests its local effects through various disasters such as floods, droughts, and storms. Global climate change, the focus of numerous articles in recent years, is intensifying its impact day by day, negatively affecting human life in economic and social spheres. Consequently, studies aimed at understanding climate change and implementing measures to address it are becoming increasingly crucial. Most studies conducted in the context of climate change reveal that there are changes in the intensity and distribution of precipitation throughout the year, rather than significant changes in total annual precipitation values. In this context, the change in standard duration precipitation for Izmir meteorological station on a station basis was studied by *Karahan* (2012), on a regional scale for the Southeastern Anatolia Region by *Karahan et al.* (2008), for the Aegean Region by *Karahan* (2011, 2019), and for the Eastern Black Sea Region meteorological stations by *Karahan et al.* (2015). The mentioned studies used different methods, and it was shown that there were changes in precipitation intensities in the first and second half of the measurement period. Similarly, *Zeybekoglu* and *Karahan* (2018) presented increasing and decreasing trends in standard duration rainfall intensities for all meteorological stations operated by the Turkish State Meteorological Service (TSMS) at the national scale. Additionally, *Karahan* (2022) conducted a detailed analysis of the increasing and decreasing trends in rainfall intensities, as well as the dates of occurrence of deterioration in rainfall intensity.

The classification of regions based on similar climate parameters is believed to contribute to various studies such as combating climate change, protecting water resources, and land use planning. *Erinc* (1949) classified precipitation and temperature data from 53 meteorological stations in Türkiye into four different climate zones using the Thornthwaite method. This study marks the first comprehensive and detailed classification of Türkiye's geography with sufficient data. *Turkes* (1996) classified precipitation data of Türkiye using the Normalization Procedure method proposed by Kraus in 1977. In the study, seven different regions were identified during the period 1930–1993. *Kulkarni* and *Kripalani* (1998) identified similarity classes of Indian rainfall data using the

Fuzzy c-means method. They divided 306 meteorological observation stations into four different clusters using rainfall data for the period 1871–1984. *Unal et al.* (2003) determined the similarity classes of temperature and precipitation data covering the period between 1951–1998 in Türkiye using five different clustering methods. The study concluded that the Ward's method was the most effective among the preferred methods. *Soltani* and *Modarres* (2006) categorized precipitation data from 28 meteorological stations in Iran into similar classes using hierarchical and non-hierarchical clustering methods. The study identified eight different classes, employing Ward's method and the k-means algorithm. *Sonmez* and *Komuscu* (2008) utilized the k-means algorithm in their study to identify precipitation regions in Türkiye. They analyzed monthly total precipitation series obtained from 148 meteorological stations covering the period 1977-2006, identifying six different precipitation regions. *Sahin* (2009) utilized monthly average temperature, monthly relative humidity, and monthly total precipitation data from 150 meteorological stations to determine similar climate classes in Türkiye. They employed the Ward's, Kohonen artificial neural network, and fuzzy artificial neural network methods to identify seven different regions. *Dikbas et al.* (2012) identified six different precipitation regions in Türkiye using the fuzzy c-means method with records from 188 stations spanning 1967–1998. *Sahin* and *Cigizoglu* (2012) determined sub-climatic and sub-precipitation regime classes in Türkiye using the Ward's and fuzzy artificial neural network methods. They analyzed precipitation, temperature, and humidity data from 232 meteorological stations for the period 1974–2002, identifying seven precipitation regime zones and seven climate zones. *Fırat et al.* (2012) identified 7 different regions with similar characteristics using the k-means method for the similarity classes of annual total precipitation measured at 188 precipitation observation stations in Türkiye, covering the period 1967–1998. *Iyigün et al.* (2013) conducted a cluster analysis study using precipitation, temperature, and relative humidity data, employing the Ward's method. The data were obtained from 244 meteorological stations in Türkiye, covering the period from 1970 to 2010. As a result of the study, *Iyigün et al*. (2013) identified 14 different clusters. *Rau et al.* (2017) divided the rainfall data of the Peruvian Pacific slope and coast into regions with similar characteristics. They utilized the regional vector method and the k-means algorithm, identifying nine different rainfall regions. *Zeybekoglu* and *Ulke Keskin* (2020) conducted a clustering analysis of precipitation intensity series using the fuzzy c-means algorithm, incorporating latitude, longitude, and altitude values of observation stations. They found that 95 meteorological observation stations in Türkiye formed five different clusters. Additionally, the authors clustered various hydrometeorological parameters of the same study area using clustering algorithms and silhouette index analysis (*Kir at al.*, 2023a, b, c). According to temperature observations, clusters with similar characteristics were determined using k-means and FCM algorithms. According to silhouette index analysis, the optimal number of clusters was determined as 5 for k-means and

4 for FCM; however, the 5-cluster approach suggested by k-means was deemed the most ideal distribution (*Kir et al.,* 2023a). Similar stations based on wind speed characteristics were determined using both K-Means and FCM. According to silhouette index analysis, the optimal number of clusters was determined as 5 and 4 for k-means and FCM, respectively; however, the 5-cluster approach determined by k-means was suggested as the most ideal distribution (*Kir et. al.,* 2023b). In the authors' research (*Kir et al.,* 2023c), clusters with similar characteristics were formed using FCM based on precipitation records from stations in the region. According to the silhouette index analysis of the clusters they created, they concluded that the most appropriate approach is a 4-cluster distribution (*Kir et al.,* 2023c).

In the literature, numerous studies have been conducted in Türkiye and abroad on the determination of climate classes. When examining these studies, it is evident that precipitation and temperature are predominantly emphasized as climate parameters. Additionally, the evaluation of results obtained by using fuzzy c-means and k-means methods together with silhouette analysis is not very common in climate studies (*Kir,* 2021). Accordingly, the aim of this study is to cluster meteorological observation stations in the Black Sea Region with similar characteristics using the k-means algorithm based on precipitation records. The most appropriate number of clusters was determined through Silhouette index analysis for different cluster numbers with the k-means algorithm.

2. Materials and methods

2.1. Materials

In this study, annual total precipitation records for the period 1982–2020 (39 years) from 31 observation stations operated by Turkish State Meteorological Service in the Black Sea Region were utilized. It was ensured that the data had a record length of at least 30 years for statistical adequacy (*Kite,* 1991).

The observation stations used in the study are located in 17 different provinces in the Black Sea Region. Eleven of the stations are located in the Western Black Sea (Düzce, Akçakoca, Bolu, Zonguldak, Bartın, Amasra, Kastamonu, İnebolu, Bozkurt, Tosya, Sinop), 10 in the Central Black Sea (Samsun, Bafra, Çorum, Osmancık, Amasya, Merzifon, Tokat, Zile, Ordu, Ünye), and the remaining 10 in the Eastern Black Sea region (Giresun, Şebinkarahisar, Trabzon, Akçaabat, Gümüşhane, Bayburt, Rize, Pazar, Artvin, Hopa).

In the Black Sea Region, where the precipitation regime is oriented in the north-south direction, precipitation is observed in all seasons. The precipitation regime in areas close to each other can exhibit significant differences. While the average annual precipitation in Rize, one of the provinces with the highest precipitation in the region, is 2284 mm, the average annual precipitation in Trabzon, which is right next to it, is 846 mm. The precipitation rate in the Black

Sea Region is high in the east (e.g., Rize: 2284 mm; Hopa: 2329 mm), but it decreases towards the Central Black Sea with the decrease in elevation (e.g., Samsun: 716 mm; Amasya: 465 mm; Çorum: 450 mm). In the Western Black Sea, the precipitation rate increases again with elevation (e.g., Zonguldak: 1227 mm; Bartın: 1051 mm). Additionally, when the data used in the study are evaluated, the average annual precipitation of the Black Sea Region is 901 mm. The geographical distribution of the stations is shown in *Fig. 1*, while the geographical location of the stations and basic statistical information on meteorological observations are provided in *Tables 1* and *2*.

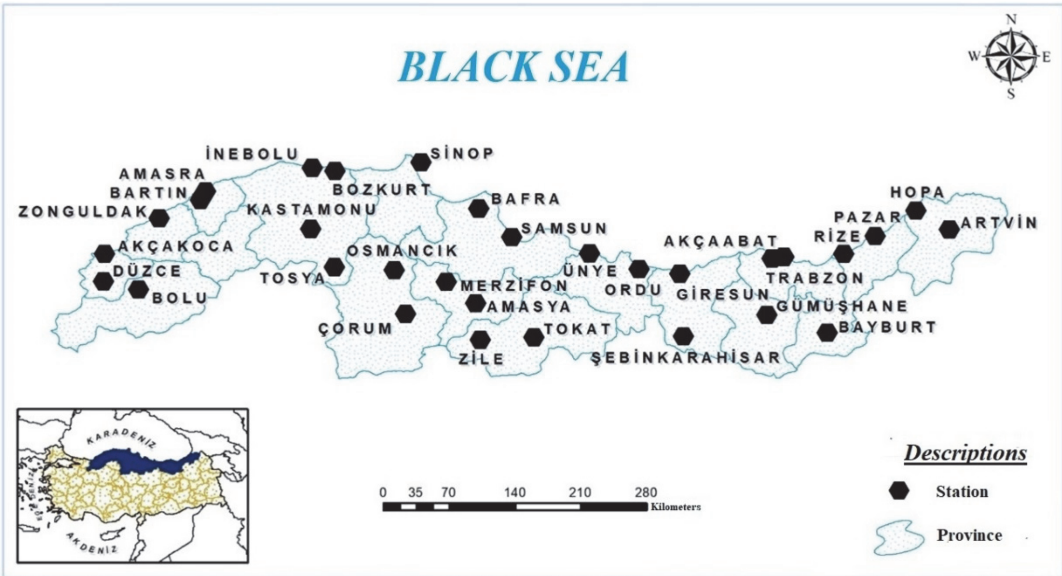

Fig. 1. Spatial distribution of the stations.

Table 1. Geographical details of the stations

Station	Latitude (N)	Longitude (E)	Elevation (m)
Düzce	40°50'37.3"	31°08'55.7"	146
Akçakoca	41°05'22.2"	31°08'14.6"	10
Bolu	40°43'58.4"	31°36'07.9"	743
Zonguldak	41°26'57.3"	31°46'40.5"	135
Bartın	41°37'29.3"	32°21'24.8"	33
Amasra	41°45'09.4"	32°22'57.7"	73
Kastamonu	41°22'15.6"	33°46'32.2"	800
İnebolu	41°58'44.0"	33°45'49.0"	64
Bozkurt	41°57'34.9"	34°00'13.3"	167
Tosya	41°00'47.5"	34°02'12.1"	870
Çorum	40°32'46.0"	34°56'10.3"	776
Osmancık	40°58'43.3"	34°48'04.0"	419
Sinop	42°01'47.6"	35°09'16.2"	32
Amasya	40°40'00.5"	35°50'07.1"	409
Merzifon	40°52'45.5"	35°27'30.6"	754
Samsun	41°20'39.0"	36°15'23.0"	4
Bafra	41°33'05.4"	35°55'28.9"	103
Tokat	40°19'52.3"	36°33'27.7"	611
Zile	40°17'45.6"	35°53'25.8"	719
Ordu	40°59'01.7"	37°53'08.9"	5
Ünye	41°08'34.8"	37°17'34.8"	16
Giresun	40°55'21.7"	38°23'16.1"	38
Şebinkarahisar	40°17'13.9"	38°25'09.5"	1364
Gümüşhane	40°27'35.3"	39°27'55.1"	1216
Trabzon	40°59'54.6"	39°45'53.6"	25
Akçaabat	41°01'57.0"	39°33'41.4"	3
Bayburt	40°15'16.9"	40°13'14.5"	1584
Rize	41°02'24.0"	40°30'04.7"	3
Pazar	41°10'39.7"	40°53'57.5"	78
Artvin	41°10'30.7"	41°49'07.3"	613
Hopa	41°24'23.4"	41°25'58.8"	33

Table 2. Basic statistics of rainfall records (mm)

Station	Mean	Std. Dev.	Min.	Max.	Var.	Skew.
Düzce	818.43	124.54	527.0	1084.9	0.15	0.05
Akçakoca	1127.62	175.42	742.6	1460.7	0.16	-0.06
Bolu	562.52	87.20	382.5	754.5	0.16	0.05
Zonguldak	1226.74	187.03	818.8	1740.1	0.15	0.76
Bartın	1051.11	161.86	753.1	1350.3	0.15	0.14
Amasra	981.62	180.33	660.6	1412.6	0.18	0.53
Kastamonu	521.17	119.95	338.2	870.5	0.23	0.93
İnebolu	1053.78	136.43	728.0	1330.0	0.13	-0.26
Bozkurt	1185.50	238.58	498.2	1595.7	0.20	-1.00
Tosya	476.32	100.40	250.8	735.5	0.21	0.34
Çorum	450.19	88.94	242.9	633.8	0.20	0.06
Osmancık	423.41	117.67	234.6	794.4	0.28	0.82
Sinop	718.52	133.87	333.3	1008.1	0.19	-0.30
Amasya	465.32	88.68	293.4	682.0	0.19	0.58
Merzifon	444.35	93.29	225.1	703.3	0.21	0.50

Table 2. continued

Station	Mean	Std. Dev.	Min.	Max.	Var.	Skew.
Samsun	716.47	93.16	562.8	999.1	0.13	0.86
Bafra	763.16	162.47	424.0	1141.4	0.21	0.37
Tokat	444.26	72.10	313.3	593.0	0.16	0.09
Zile	444.82	90.04	237.4	639.0	0.20	0.25
Ordu	1058.36	128.71	787.2	1433.8	0.12	0.64
Ünye	1185.51	160.85	906.6	1532.8	0.14	0.44
Giresun	1308.07	170.71	970.7	1743.4	0.13	1.09
Şebinkarahisar	568.64	91.91	345.8	741.9	0.16	-0.12
Gümüşhane	472.08	84.11	311.0	651.0	0.18	0.34
Trabzon	846.55	111.99	594.4	1044.6	0.13	-0.39
Akçaabat	721.37	111.71	494.0	1017.4	0.15	0.32
Bayburt	464.34	75.35	318.2	614.6	0.16	-0.03
Rize	2284.35	273.76	1694.0	3097.1	0.12	0.73
Pazar	2105.38	360.58	1326.8	2905.0	0.17	0.34
Artvin	721.41	132.51	425.1	1005.9	0.18	-0.12
Hopa	2329.73	372.30	1685.3	3379.5	0.16	1.07

When analyzing *Table 2*, the station with the lowest average annual total precipitation is Osmancık with 423.41 mm, while the station with the highest average annual total precipitation is Hopa with 2329.73 mm.

2.2. K-means algorithm

The k-means clustering algorithm (*Xin et al.,* 2011) is one of the simplest unsupervised and hard clustering algorithms. It is used to classify a given dataset into various clusters (*Vani et al.,* 2019).

Process steps of the algorithm:

- Step 1: Random centers are selected.
- Step 2: The distance between the centroids and the data points is calculated.
- Step 3: Data points are assigned to clusters based on the minimum Euclidean distance measure:

$$J_{KM}(X;V)=\sum_{i=1}^{c}\sum_{j=1}^{n} D_{ij}^{\ 2}, \tag{1}$$

where V is a given centroid, c is the number of clusters, n is the number of iteractions, and D_{ij} is the Euclidean distance between each data points and centroid

- Step 4: New centroids are calculated:

$$V_i = \sum_{i=1}^{n_i} \frac{D_{ij}}{n_i}; 1 \leq i \leq c. \quad (2)$$

- Step 5: Check whether the new centroids are equal to the old centroids.
- Step 6: If the new centroids and the old centroids are equal, the algorithm terminates; otherwise, it goes back to step 2.

Input: V is the number of centroids (centers); x and y are the distance center values between the centroid and the data points; D_{ij} is the Euclidean distance between each data point and the centroids; c is the number of clusters, and n is the number of iterations.
Output: Number of clusters.

Advantages of the algorithm (*Vani et al.,* 2019):

- It is easy to understand and simple to implement.

Disadvantages of the algorithm(*Vani et al.,* 2019):

- Not effective for overlapping clusters.
- Ineffective for clustering heterogeneous data.
- Provides a local optimum of the squared error function.
- Randomly choosing the cluster centers may not yield optimal results.

2.3. Silhouette index analysis

In this method developed by *Rousseeuw* (1987), the suitability of each element in the dataset to the cluster to which it is assigned is defined by the silhouette index value obtained between [-1, +1]. A positive silhouette index value indicates that the element is assigned to the correct cluster, while a negative value indicates that the element is assigned to the wrong cluster. The silhouette index value indicates the degree of membership to the cluster to which the element is assigned. For example, a silhouette index value of +1 means that the element is definitely assigned to the correct cluster, while -1 means that the element is definitely assigned to the wrong cluster. The silhouette index value is calculated by the following formula (*Sonmez* and *Komuscu*, 2008; *Gunay Atbas,* 2008):

$$S(i) = \frac{min\{b(i,m)-a(i)\}}{max\{a(i), min(b(i,m))\}}, \quad (3)$$

where $a(i)$ refers to the average distance between point i and all other points in the same cluster, $b(i, m)$ is the average distance between point i and all points in cluster m.

3. Results

In this study, the k-means algorithm was utilized to identify clusters with similar characteristics using annual average temperature observations from 31 stations in the Black Sea Region, covering the period between 1982 and 2020. Analyses were performed using MATLAB R2016a. The maximum number of clusters was chosen as 5, which is less than the square root of the number of stations (*Karahan*, 2011, 2019; *Pal* and *Bezdek* 1995; *Zhang et al.,* 2008). Before cluster analysis of the observation records, these data were standardized using in the following formula (*Unal et al.,* 2003):

$$z = \frac{x_i - \bar{x}}{s}, \tag{4}$$

where x_i represents the standardized observation at rank i, $\bar{x}$ is the mean of the dataset. s is the standard deviation of the dataset, and z is defined as the standardized value (*Unal et al.,* 2003).

In the classification conducted for each cluster number from 2 to 5, using the k-means algorithm with the maximum number of clusters set at 5, the resulting clusters for the Black Sea Region wind speed series are illustrated in *Figs 2–5*, while statistical summary information of the clusters is provided in *Tables 3–6*.

The clusters formed when the number of clusters is selected as 2 are shown in *Fig. 2*. Upon analysis of the results, cluster A consists of 28 stations located in the Western, Central, and Eastern Black Sea regions. Cluster B consists of 3 stations located only in the Eastern Black Sea coastal region. *Table 3* presents the maximum, minimum, mean, and standard deviation values of annual total precipitation for the identified clusters.

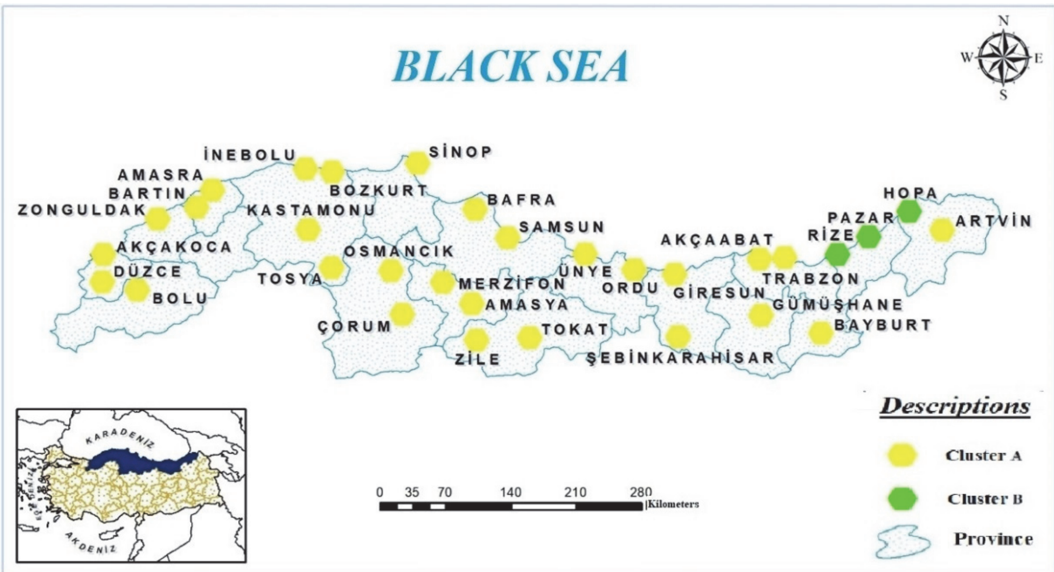

Fig. 2. Spatial distribution of stations for 2 clusters.

Table 3. Statistical summary of precipitations for 2 clusters distribution (mm)

Cluster	Min.	Max.	Mean	Std. Dev.
A	423.41	1308.07	757.92	293.15
B	2105.38	2329.73	2239.82	118.62

The clusters obtained when the number of clusters is selected as 3 are shown in *Fig. 3*. Upon analysis of the results, it is observed that clusters A and B are separated as two sub-clusters of cluster A in the previous distribution. Additionally, it is noted that the Rize, Pazar, and Hopa stations, which maintained their integrity in the previous distribution, are assigned to cluster C. Thus, cluster A consists of 18 stations located in the Western, Central, and Eastern Black Sea regions. Cluster B consists of 10 stations located in the Western, Central, and Eastern Black Sea coastal areas. Cluster C consists of 3 stations located only in the Eastern Black Sea coastal area. *Table 4* presents the maximum, minimum, mean, and standard deviation values of annual total precipitation for the clusters.

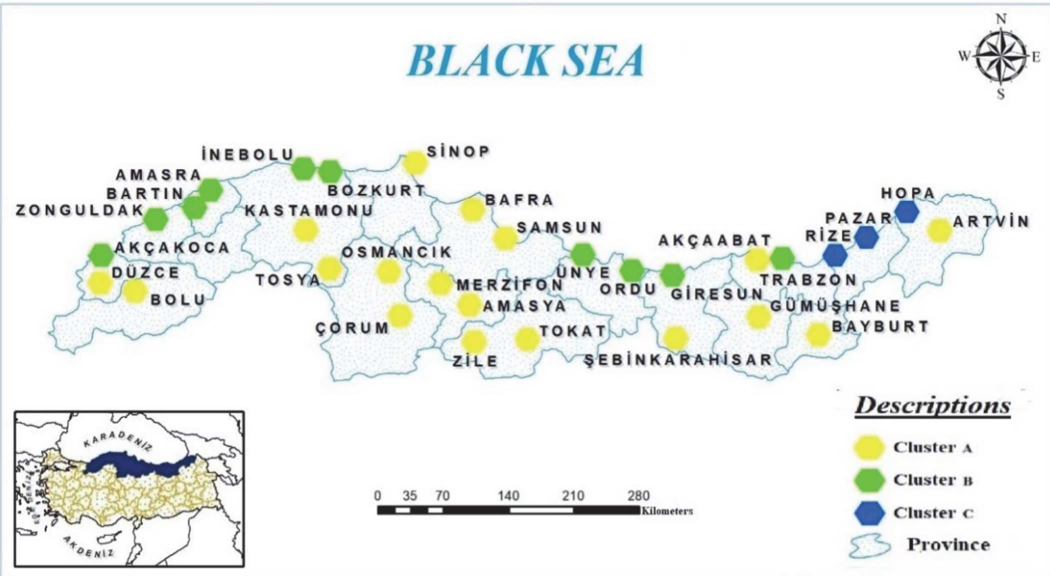

Fig. 3. Spatial distribution of stations for 3 clusters.

Table 4. Statistical summary of precipitations for 3 clusters distribution (mm)

Cluster	Min.	Max.	Mean	Std. Dev.
A	423.41	818.43	566.49	135.95
B	846.55	1308.07	1102.49	133.00
C	2105.38	2329.73	2239.82	118.62

The clusters formed when the number of clusters is selected as 4 are shown in *Fig. 4*. Upon analysis of the results, it is observed that clusters C and D are separated as two sub-clusters of cluster C in the previous distribution. Additionally, it is noted that the Trabzon station, which was in cluster B in the previous distribution, is assigned to cluster A. Thus, Cluster A consists of 19 stations located in the Western, Central, and Eastern Black Sea regions. Cluster B consists of 9 stations located in the Western, Central, and Eastern Black Sea coastal areas. Clusters C and D consist of 1 and 2 stations, respectively, located only in the Eastern Black Sea coastal zone. Cluster C consists of only the Pazar station. Cluster D consists of the Rize and Hopa stations. *Table 5* presents the maximum, minimum, mean, and standard deviation values of annual total precipitation for the identified clusters.

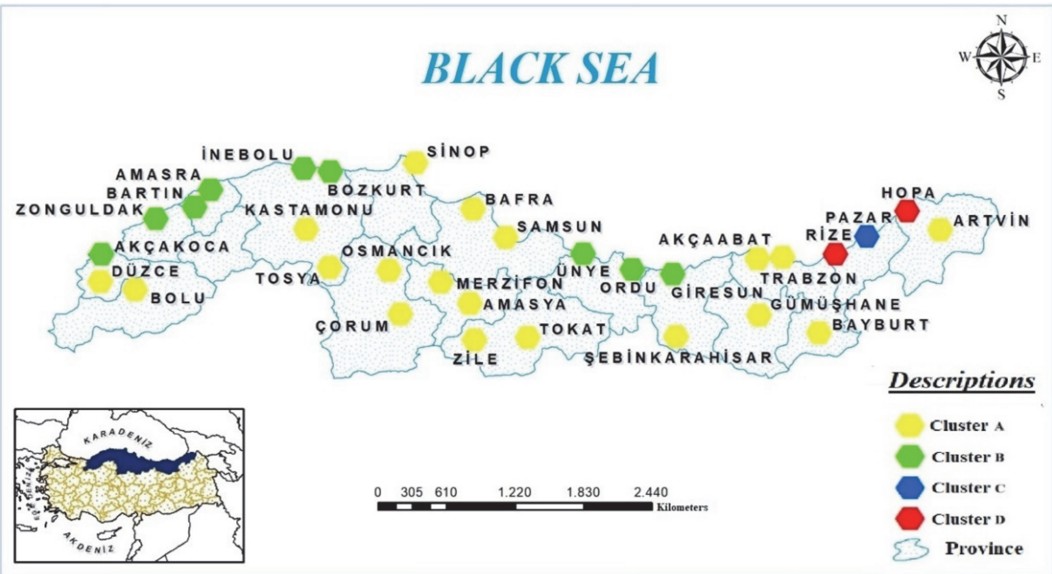

Fig. 4. Spatial distribution of stations for 4 clusters.

Table 5. Statistical summary of precipitations for 4 clusters distribution (mm)

Cluster	Min.	Max.	Mean	Std. Dev.
A	423.41	846.55	581.23	146.91
B	981.62	1308.07	1130.92	103.93
C	2105.38	2105.38	2105.38	-
D	2284.35	2329.73	2307.04	32.09

The clusters formed for cluster number 5 are shown in *Fig. 5*. Upon analysis of the results, it is observed that clusters A, C, and E are separated as three sub-clusters of cluster A in the previous distribution. Additionally, Rize, Pazar, and Hopa stations maintain integrity again and form cluster D. Thus, cluster A consists of 9 stations located in the inland areas of the Western, Central, and Eastern Black Sea regions. Cluster B consists of 9 stations located in the coastal areas of the Western, Central, and Eastern Black Sea regions. Cluster C consists of 7 stations located in the Western, Central, and Eastern Black Sea regions. Cluster D consists of 3 stations located in the Eastern Black Sea coastal area. Cluster E consists of 3 stations located in the inland areas of the Western and Eastern Black Sea regions. The maximum, minimum, mean, and standard deviation values of annual total precipitation for the identified clusters are presented in *Table 6*.

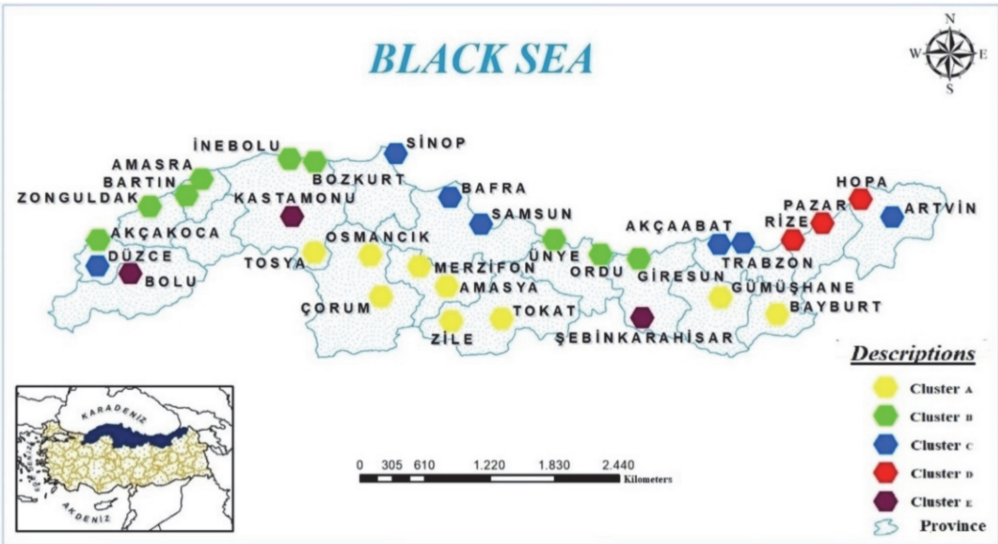

Fig. 5. Spatial distribution of stations for 5 clusters.

Table 6. Statistical summary of precipitations for 5 clusters distribution (mm)

Cluster	Min.	Max.	Mean	Std. Dev.
A	423.41	476.32	453.90	16.89
B	981.62	1308.07	1130.92	103.93
C	716.47	846.55	757.99	53.98
D	2105.38	2329.73	2239.82	118.62
E	521.17	568.64	550.78	25.82

Clusters were identified using the k-means algorithm for each cluster number starting from 2 up to 5, which was determined as the maximum number of clusters. Silhouette index analysis was used to assess the accuracy of the clusters and to determine the optimum number of clusters. First, silhouette index values of the results obtained for each cluster number were calculated. Then, the average silhouette index values and the number of negative silhouette index values of each cluster were determined. The optimum number of clusters was determined based on the lowest negative silhouette index value in silhouette analysis (*Sonmez* and *Komuscu*, 2008).

Silhouette index values of the stations in the clusters determined from 2 clusters to 5 clusters with k-means are presented in *Fig. 6* and *Table 7*, while the average index value and the number of negative silhouette index values of the clusters are presented in *Table 7*.

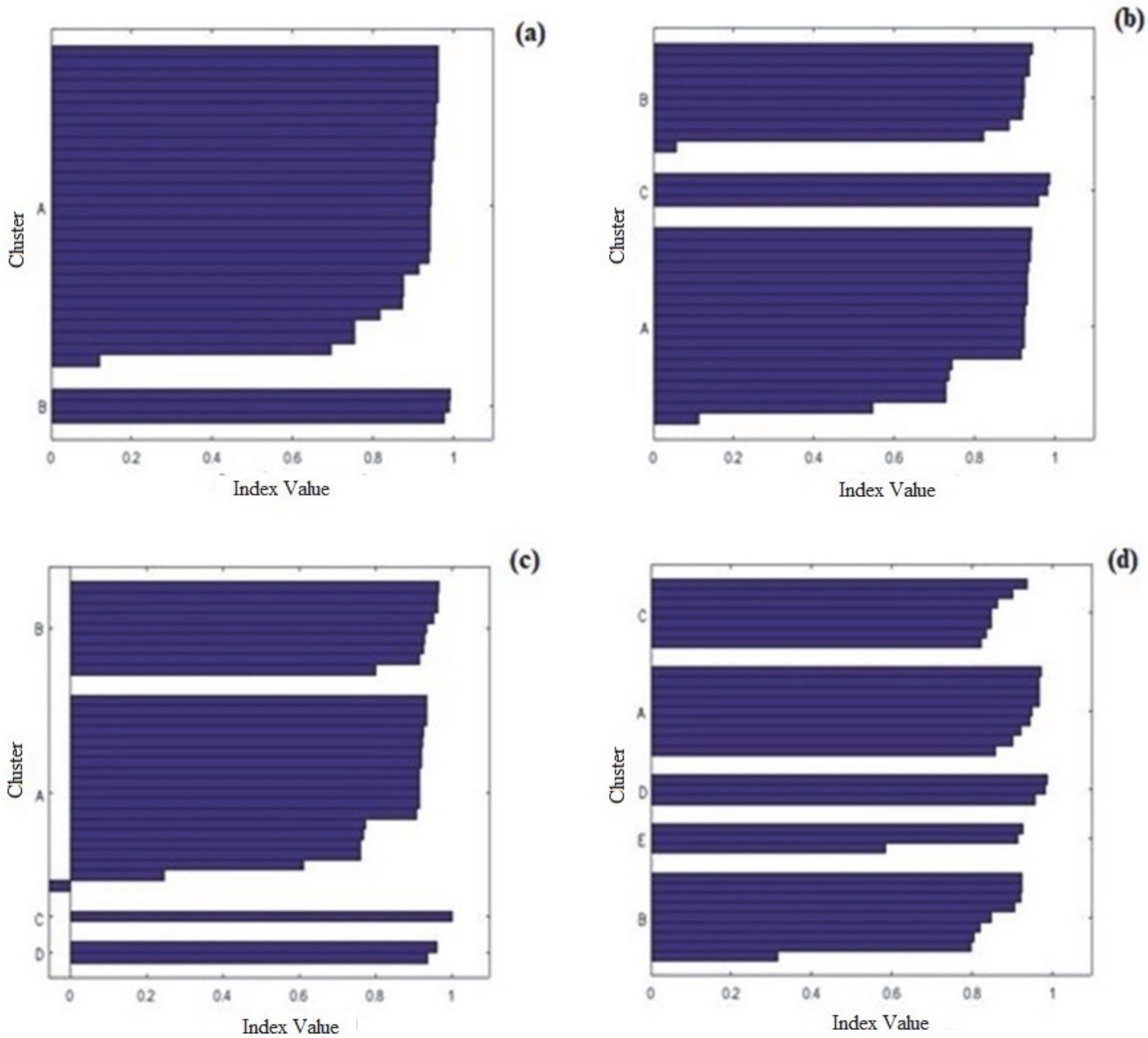

Fig. 6. Silhouette index values of the stations in the clusters.

Table 7. Silhouette index values of the stations in the clusters

Station	No of Clusters			
	2	**3**	**4**	**5**
Düzce	0.956[A]	0.113[A]	0.245[A]	0.900[C]
Akçakoca	0.817[A]	0.945[B]	0.966[B]	0.922[B]
Bolu	0.956[A]	0.940[A]	0.934[A]	0.928[E]
Zonguldak	0.697[A]	0.923[B]	0.952[B]	0.905[B]
Bartın	0.877[A]	0.918[B]	0.926[B]	0.797[B]
Amasra	0.913[A]	0.821[B]	0.802[B]	0.317[E]
Kastamonu	0.951[A]	0.942[A]	0.933[A]	0.585[A]
İnebolu	0.875[A]	0.920[A]	0.928[B]	0.806[B]
Bozkurt	0.754[A]	0.937[A]	0.963[B]	0.924[B]
Tosya	0.946[A]	0.934[A]	0.924[A]	0.858[A]
Çorum	0.943[A]	0.926[A]	0.916[A]	0.972[A]
Osmancık	0.939[A]	0.916[A]	0.906[A]	0.920[A]
Sinop	0.962[A]	0.737[A]	0.769[A]	0.834[C]
Amasya	0.945[A]	0.931[A]	0.921[A]	0.944[A]
Merzifon	0.942[A]	0.924[A]	0.914[A]	0.967[A]
Samsun	0.962[A]	0.744[A]	0.775[A]	0.823[A]
Bafra	0.961[A]	0.546[A]	0.610[A]	0.935[C]
Tokat	0.942[A]	0.924[A]	0.914[A]	0.967[C]
Zile	0.942[A]	0.924[A]	0.914[A]	0.968[A]
Ordu	0.872[A]	0.923[B]	0.933[B]	0.820[B]
Ünye	0.754[A]	0.937[B]	0.963[B]	0.924[B]
Giresun	0.108[A]	0.886[B]	0.916[B]	0.849[B]
Şebinkarahisar	0.956[A]	0.939[A]	0.933[A]	0.915[E]
Gümüşhane	0.946[A]	0.932[A]	0.923[A]	0.901[A]
Trabzon	0.952[A]	0.057[B]	-0.057[A]	0.863[C]
Akçaabat	0.962[A]	0.728[A]	0.761[A]	0.848[C]
Bayburt	0.945[A]	0.930[A]	0.921[A]	0.948[A]
Rize	0.993[B]	0.988[C]	0.936[D]	0.987[D]
Pazar	0.978[B]	0.960[C]	1.000[C]	0.957[D]
Artvin	0.962[A]	0.728[A]	0.761[A]	0.849[C]
Hopa	0.990[B]	0.983[C]	0.959[D]	0.982[D]
Average silhouette index value	**0.893**	0.837	0.844	0.875
Number of negative silhouette index value	-	-	1	-

According to the results of the silhouette index analysis method presented in *Fig. 6* and *Table 7*, the average silhouette index values of the stations were calculated as 0.893, 0.837, 0.844, and 0.875, respectively, when the number of clusters was selected as 2, 3, 4, and 5. If the number of clusters is 3, the negative silhouette index value belongs to Trabzon with -0.057. Among the clusters formed using the precipitation values of the stations in the Black Sea Region, the most appropriate number of clusters is proposed as a 2-cluster distribution, where the average silhouette index value is maximum at 0.893 and there are no negative silhouette index values. Although not supported by the results of the silhouette index analysis, choosing 3 clusters may offer a good alternative to the 2-cluster approach in terms of geographical integrity and precipitation values.

4. Discussion and conclusion

This study identified clusters with similar annual precipitation characteristics among stations in the Black Sea Region using the k-means algorithm. Clustering analysis was performed for 4 different numbers of clusters ranging from 2 to 5, and the optimal number of clusters was determined using the Silhouette index analysis method. As a result of the analysis, according to the k-means and silhouette index analysis methods, the stations in the Black Sea Region were determined to form 2 clusters with similar precipitation characteristics. A 4-cluster approach with stations having negative silhouette index values is not recommended. Based on silhouette index analysis among the clusters determined by FCM with the same dataset, the authors propose a 4-cluster approach. Alternatively, according to the results of the silhouette index analysis, the 2-cluster, 5-cluster, and 3-cluster approaches can be preferred, respectively (*Kir et al.,* 2023c). In *Kir et al.* (2023c), where clusters are formed with FCM, as an alternative to the 2-cluster approach, the 5-cluster and 3-cluster approaches can be preferred based on the silhouette index analysis. The main difference between the results of this study and *Kir et al.* (2023c) using the same data is thought to be due to the algorithms used in clustering analysis.

When comparing the results of this study with the main studies in the literature covering the Black Sea Region (*Turkes*, 1996; *Unal et al.,* 2003; *Iyigun et al.,* 2013; *Zeybekoglu* and *Ulke Keskin,* 2020; *Ozturk et al*., 2017), it is thought that the main reasons for the identification of different clusters are due to the following factors:

- the methods used in cluster analysis,
- hydrometeorological parameters used and observation periods, and
- regional geographical features of the study area such as mountainous terrain, ruggedness, and the parallelism of the mountains to the coast, as well as the sea effect.

As a follow-up to this study:

- It is recommended to incorporate monthly and seasonal precipitation regimes in clustering analyses using precipitation values and to compare the results with the clusters determined by annual precipitation values.
- In addition to precipitation observations, it is proposed to determine climate classes with various combinations not included in the literature by incorporating hydrometeorological parameters such as temperature, wind speed, current, humidity, evaporation, and geographical location information.
- It is recommended to conduct clustering studies that include hierarchical methods such as Ward's method as well as non-hierarchical methods such as fuzzy c-Means.

- It is recommended that cluster analysis should also be conducted for other regions in the geography of Türkiye.

Data availability: The meteorological data used in this manuscript were obtained from the Turkish State Meteorological Service (TSMS) for the master's thesis titled "Evaluation of the meteorological data of the Black Sea Region using clustering analysis methods" written by Gurkan Kir under the supervision of Asli Ulke Keskin.
Author contributions: AUK: Data curation, investigation, methodology, supervision, and writing (original draft, review and editing). GK: Data curation, investigation, methodology, formal analysis, visualization, and writing (original draft, review, and editing). UZ: Investigation, methodology, formal analysis, validation, visualization, supervision, and writing (original draft, review, and editing).
Competing interests: The authors declare that none of the authors has any competing interests.
Acknowledgments: The authors thank the Turkish State Meteorological Service for providing the meteorological data. The authors also thank the reviewers for their constructive criticisms which have considerably improved this manuscript.

References

Demircan, M., Arabacı, H., Coşkun, M., Türkoğlu, N., and *Çiçek, İ.*, 2017: İklim değişikliği ve halk takvimi: Maksimum sıcaklık desenleri ve değişimi. IV. Türkiye İklim Değişikliği Kongresi, İstanbul Medipol Üniversitesi, İstanbul. (In Turkish)

Dikbas, F., Firat, M., Koc, A.C., and *Gungor, M*., 2012: Classication of precipitation series using fuzzy cluster method. Int. J. Climatol. 32(10), 1596–1603. https://doi.org/10.1002/joc.2350

Erinc, S., 1949: The climates of Turkey according to Thornthwaite's classifications*. Ann. Assoc. Am. Geograp. 39*, 26–46. https://doi.org/10.2307/2561098

Fırat, M., Dikbaş, F., Koç, A.C., and *Güngör, M*., 2012: Classification of annual precipitations and identification of homogeneous regions using k-means Method. *Tech. J. 23*(113), 6037–6050.

Günay Atbaş, A.C., 2008: A study on determining the number of clusters in cluster analysis. MSc Thesis, Ankara University, Ankara, Turkey.

Iyigün, C., Türkeş, M., Batmaz, İ., Yozgatlıgil, C., Purutçuoğlu, V., Kartal Koç, E., and *Öztürk, M.Z*., 2013: Clustering current climate regions of Turkey by using a multivariate statistical method. *Theor Appl Climatol. 114*, 95–106. https://doi.org/10.1007/s00704-012-0823-7

Karahan, H., 2011: Determination of Rainfall Intensity-Duration-Frequency Relationships Using Differential Evolution Algorithm. 108Y299, Research Project, The Scientific and Technological Research Council of Turkey (TUBITAK), Final Report, Ankara, Türkiye.

Karahan H., 2019: Determination of Homogeneous Sub-Regions by Using Intensity-Duration-Frequency Relationships and Cluster Analysis: An Application for the Aegean Region. Pamukkale University J. *Eng. Sci. 25*(8), 998–1013. https://doi.org/10.5505/pajes.2019.09365

Karahan, H., 2012: Determining rainfall-intensity-duration-frequency relationship using Particle Swarm Optimization. *KSCE J. Civil Engin. 16*, 667–675. https://doi.org/10.1007/s12205-012-1076-9

Karahan, H., 2022: Kentsel taşkınların kentleşme ve altyapı bağlamında değerlendirilmesi. TMMOB Afet Sempozyumu, Atmosfer ve İklim Kaynaklı Afetler Oturumu, Çağrılı Konuşmacı, 20-22 Nisan 2022. (In Turkish)

Karahan, H., Ayvaz, M.T., and *Gurarslan, G*., 2008: Determination of intensity-duration-frequency relationship by genetic algorithm: Case study of GAP. Teknik Dergi/*Technical Journal of Turkish Chamber of Civil Engineers 19*(92), 4393–4407.

Karahan, H., Bahar, E., and *Zeybekoglu, U*., 2015: Standart süreli maksimum yağış şiddetleri için trend analizi: Doğu Karadeniz Bölgesi için bir uygulama. 7. Kentsel Altyapı Sempozyumu, Bildiriler Kitabı, 227-238, Kasım 2015, Trabzon.

Kir, G., 2021: Evaluation of the meteorological data of the Black Sea Regıon using clustering analysis methods. MSc Thesis, Ondokuz May University, Samsun, Turkey.

Kir, G., Ulke Keskin, A., and *Zeybekoglu, U.*, 2023a: Clustering of Black Sea Region Stations Using K-Means, Fuzzy CMeans and Silhouette Index Analysis Methods. *Resilience 7*(2), 325–337. https://doi.org/10.32569/resilience.1336940

Kir, G., Ulke Keskin, A., and *Zeybekoglu, U.*, 2023b: Clustering of Black Sea Region Meteorology Observation Stations Using Wind Speed Values. *Ordu University Journal of Science and Technology 13*(1), 44–58. https://doi.org/10.54370/ordubtd.1275677

Kir, G., Ulke Keskin, A., and *Zeybekoglu, U.*, 2023c: Clustering of Precipitation in the Black Sea Region with by Fuzzy C-Means and Silhouette Index Analysis. *Black Sea Journal of Engineering and Science 6*(3), 210–218. https://doi.org/10.34248/bsengineering.1296734

Kite, G., 1991: Looking for Evidence of Climatic Change in Hydrometeorological Time Series. Western Snow Conference. Washington, Alaska.

Kulkarni, A. and *Kripalani, R.,* 1998: Rainfall patterns over India: Classification with fuzzy c-means method. *Theor. Appl. Climatol. 59*, 137–146. https://doi.org/10.1007/s007040050019

Ozkoca, T., 2015: Trend analysis of hydrometeorological parameters at middle blacksea region coast band. MSc Thesis, Ondokuz May University, Samsun, Turkey.

Ozturk, M.Z., Cetinkaya, G., and *Aydın, S.*, 2017: Climate Types of Turkey According to Köppen-Geiger Climate Classification. *J. Geography 35*, 17–27. https://doi.org/10.26650/JGEOG295515

Pal, N.R. and *Bezdek, J.C.*, 1995: On Cluster validity for the fuzzy c-means model. *IEEE Transactions On Fuzzy Systems 3*, 370–379. https://doi.org/10.1109/91.413225

Rau, P., Bourrel, L., Labat, D., Melo, P., Dewitte, B, Frappart, F., Lavado, W., and *Felipe, O.*, 2017: Regionalization of rainfall over the Peruvian Pacific slope and coast. Int. J. Climatol. 37, 143–158. https://doi.org/10.1002/joc.4693

Rousseuw, P.J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math. 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Soltani, S. and *Modarres, R.*, 2006: Classification of spatio temporal pattern of rainfall in Iran using a hierarchical and divisive cluster analysis. *J. Spat. Hydrol. 6*(2), 1–12.

Sonmez, I. and *Komuscu, A.*, 2008: Redefinition rainfall regions using k-means clustering methodology and changes of sub period. *Iklim Değişikliği ve Çevre 1*, 38–49.

Sahin, S., 2009: Applying artificial neural networks on determining climate zones and comparison with the Ward?s method. PhD Thesis, Istanbul Technical University, Istanbul, Turkey.

Sahin, S. and *Cigizoğlu, H.K.*, 2012: The sub-climate regions and the sub-precipitation regime regions in Turkey. *J. Hydrol. 450–451*, 180–189. https://doi.org/10.1016/j.jhydrol.2012.04.062

Turkes, M., 1996: Spatial and temporal analysis of annual rainfall variations in Turkey. *Int. J. Climatol. 16*(9), 1057–1076. https://doi.org/10.1002/(SICI)1097-0088(199609)16:9%3C1057::AID-JOC75%3E3.0.CO;2-D

Turkes, M., 2010: Küresel iklim değişikliği: Başlıca Nedenleri, gözlenen ve öngörülen değişiklikler ve etkileri. Uluslararası Katılımlı 1. Meteoroloji Sempozyumu, Devlet Meteoroloji İşleri Genel Müdürlüğü, Ankara. (In Turkish)

Unal, Y., Kındap, T., and *Karaca. M.,* 2003: Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol. 23,* 1045–1055. http://dx.doi.org/10.1002/joc.910

Vani, H.Y., Anusuya, M.A., and *Chayadevi, M.L.*, 2019: Fuzzy Clusterıng Algorıthms-Comparatıve Studıes For Noısy Speech Sıgnals. Ictact *J. Soft Comput. 9*(3), 1920–1926. https://doi.org/10.21917/ijsc.2019.0267

Xin, G.L., Min, F.Y., and *Wen, T.H.*, 2011: Speech Recognition based on K-Means Clustering and Neural Network Ensembles. Proceedings of 7th International Conference on Natural Computation.

Zeybekoglu, U. and *Karahan, H.*, 2018: Investigation of Rainfall Intensity Series of Standart Duration with Trend Analysis Methods. Pamukkale *Univ. J. Engin. Sci. 24*(6), 974–1004. https://doi.org/10.5505/pajes.2017.54265

Zeybekoglu, U. and *Ulke Keskin, A.*, 2020: Defining rainfall intensity clusters in Turkey by using the fuzzy c-means algorithm. *Geofizika 37*(2), 181–195. https://doi.org/10.15233/gfz.2020.37.8

Zhang, Y., Wang, W., Zhang, X., and *Li, Y.,* 2008: A cluster validity index for fuzzy clustering. Information Sciences 178, 1205–1218. https://doi.org/10.1016/j.ins.2007.10.004