International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I12P216
Volume 69 Issue 12, 136-143 , December, 2021

Original Article

Spatio-Temporal Rainfall Variability Analysis, Case Study: KSA

Salma M. Elsherif[1], Alaa El-Zawahry[2], ahmed H. Soliman[3]

[1,2,3]*Irrigation and Hydraulics Department, Faculty of Engineering, Cairo University, Egypt*

[1]salma.m.elsherif.93@cu.edu.eg, [2]alaa.zawahry@gmail.com, [3]a.soliman@cu.edu.eg

Abstract *- Rainfall amount and distribution are varied spatially and temporally all over the world. Moreover, the rainfall variability may significantly vary within the same local region. Identification of rainfall amount and pattern is one of the main challenges facing all hydrologic analysts. Several approaches are available nowadays to deal with the variability of data sets. Some of these approaches can be simply applied, while others are more complicated and maybe not appropriate to be used to handle rainfall variability. So, this paper is devoted to presenting a comprehensive framework of rainfall variability analysis and handling to be followed. The framework is built by combining the K-means approach with some newly developed techniques as part of this research to enhance the results of the current approaches and convert them to be more dynamic. The built framework is tested using rainfall data collected from more than 280 rainfall gauges distributed all over the Kingdom of Saudi Arabia, which has high diversity with no defined pattern, neither spatially nor temporal. The testing results confirmed that the framework is a very powerful tool and gives robust results.*

Keywords — *K-means, KSA Rainfall, Rainfall Variability, Spatial Clustering, Two-Step Clustering.*

I. INTRODUCTION

Rain is one of the main water resources on which many economic development activities, such as agricultural, municipal, industry… etc., depend. Therefore, any change in rainfall amount or distribution has a direct impact on these activities. The more significant the change is, the more influential the impacts are. Changes in rainfall amount and distribution can lead to extreme events from drought to flood or vice versa, passing by all phases between moderate to mild changes. Accordingly, understanding the rainfall trend, distribution, and variability over space and time is pivotal for associated water management and development aspects. In contrast, dealing with the available rainfall data in studies and applications without taking into consideration this variability can lead to uncertain outputs. In several cases, these outputs are vital and cannot be ignored. On the contrary, many applications, depending on the rainfall data, not only fulfilling the purpose but also perform effectively and efficiently, as a reflection of the consideration of the variation on distribution and amount of rainfall rather than simple basic statistics.

Kingdom of Saudi Arabia's (KSA) climate is classified as a composition between arid and semi-arid regions[1]. Based on that classification, the change scheme precipitation has been affected due to the dry climate. Several studies of the rainfall distribution over KSA have shown high variability temporally and spatially [2] – [4]. Some previous studies used a limited number of rainfall stations. Specifically, the rainfall distribution over KSA was studied using the rainfall data of 29 stations distributed over KSA [5]. Meanwhile, another previous research used the data of 28 stations[6]. Both studies stated that spring is the rainiest season and that the southwestern region receives the highest amount of rainfall over all the seasons. On the contrary, other researchers identified the east of the middle region of KSA as the region that receives the highest amount of rainfall, then comes the southwestern region[7]. Nevertheless, this study agreed with the two previous ones regarding the spring having the highest amount of rainfall, which is also confirmed by another research that stated that the highest rainfall occurs during winter and spring seasons from November to April [8].

Additionally, several studies focused on studying specific regions and/or administrative districts of KSA. One of these studies stated that within Riyadh city, which is located in the middle east area of KSA, the rainfall data analysis showed that over time there was no defined change pattern, and within the city's area, rainfall witnessed spatial variability [9]. Whilst, another study analyzed the rainfall data of 37 years (1970 – 2006) over Dhahran city located in the east of KSA [10]. This analysis showed that the rainy seasons are spring and winter, in order, while autumn and summer are dry seasons with almost no rain events. As well, one more study analyzed the rainfall distribution and pattern of the storms over Jeddah city in the west of KSA [11]. The study exhibited that KSA, in general, is facing an increasing trend in both; the frequency and intensity of the rainfall events and that the southwestern region has heavy rainfall events compared to the other regions. Moreover, Jeddah city has the same increasing trend, and its wet season includes the spring and winter seasons. Meanwhile, the summer and autumn represent the dry season. Furthermore, regarding the extreme precipitation events (EPEs) over KSA, it was found that the regions with the highest occurrence of the EPEs are the northeastern, middle, and southwestern, while their frequency increases during the winter and spring seasons[12].

To deal with the variation and by applying the concept of clustering and zoning the rainfall data, reference [13] analyzed the rainfall pattern of the data after clustering according to the cardinal directions only (i.e., North-East, North-West, South-East, and South-West) without taking into consideration the rainfall characteristics. However, there was a single rainfall station located in the South-East region (El-Robaa El-Khali) used to represent the rainfall characteristic in this region. Besides, the author used 30 stations in the study over the whole area of KSA, which made each of the stations' representatives for around 72,000 Km^2 of the area – if divided equally.

Furthermore, reference [14] used 27 stations to define separate zones representing the different rainfall patterns within KSA. The study started with applying the principle components analysis, which failed to assign some of the stations to any of the zones for their low scores. Subsequently, the correlation was the key factor to assign the remaining stations to either of the zones. Yet, four stations came out with relatively low correlation and were grouped in one last zone. One of the strategic points in this study is that the analysis was not relying on the rainfall data only but also the temperature data. On the other hand, the number of stations and their data may be counted as insufficient to take account of the total area of KSA. On the contrary, reference [15] ended up using 269 stations over the Kingdom to investigate the regional distribution for frequency analysis and the determination of the best distribution that fits the maximum daily rainfall data. Moreover, the study used the K-Nearest-Neighbors (KNN) method to divide the total area of KSA into regions by applying it in every trial with different numbers of regions and depending on the visualization to make an initial decision. Afterward, the regions were reshaped to reach homogenous regions and took into consideration the elevations of the stations. Yet, the final regions overlap spatially, do not represent a specific area of the Kingdom, not even an administrative region, and were defined depending on extreme data.

Nonetheless, this paper aims to establish a clear, well-organized framework to be followed to deal with data variability over any area. Starting with the data collection process to have data that cover the study area and well-represent the characteristics and nature of the data variation over its area. Afterward, these data are screened and reviewed to pursue the first step of the clustering. Clustering of the data is performed in two consecutively steps; spatial clustering then temporal analysis within-cluster to test the need to re-clustering/sub-clustering according to the data variation within the same cluster.

II. DATA COLLECTION, SCREENING, AND ANALYSIS

A. Data Collection and Screening

Rainfall data of the stations distributed over KSA were collected with a total number of 336 stations with daily rainfall datasets distributed over KSA, as shown in Fig. 1. Significantly, data screening is an essential step before performing any analysis on the raw data. Starting with the detection of any possible outlier which might be found among datasets due to several reasons such as measurement error or human error transferring the data. Out of the logical sense, some outliers could be recognized manually during the first phase of the data screening (i.e., too large readings). Successively, to recognize the outliers statically among the rainfall daily time series for every station, Interquartile Resistance Rule was used. Firstly, the datasets showed the right skewness. Thus they were transformed to be as much as possible closer to the normal distribution. This rule puts cutoffs using the first quartile Q1 (the 25th percentile) and the third quartile Q3 (the 75th percentile), the range is defined $Q\_1- \mu (Q\_3-Q\_1) \leq x \leq Q\_1+ \mu (Q\_3-Q\_1)$ [16]. In this paper, μ is taken equal to 3 so that any data point lays out of the range is considered as an extreme outlier [17].

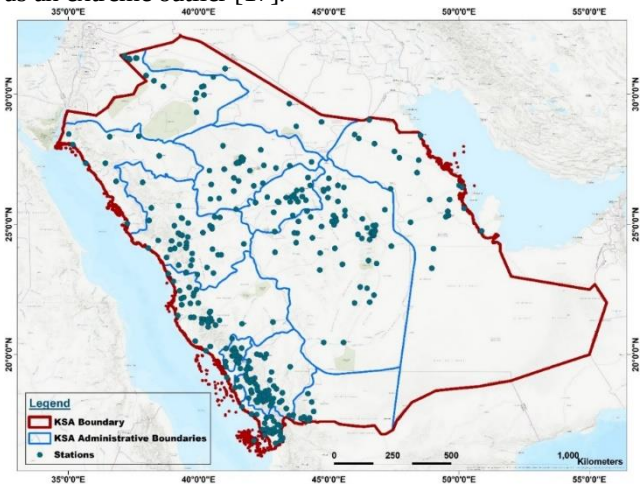


Fig. 1: Distribution of The Collected Daily Rainfall StationsOver KSA

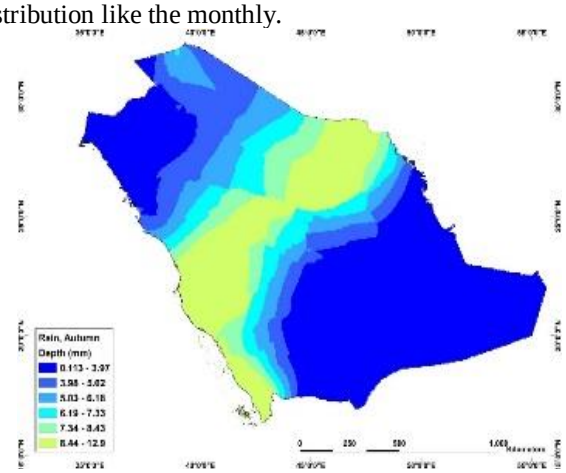
The records availability differs from station to station, moving in-between a years' window from 1960 to 2018 (the maximum covered period is 59 years). However, in order to build up a statically stable ground between stations to carry on estimations and models, stations should have more than 10 years of rainfall data [18], [19]. Yet, some other publications have put a specific threshold for their studies, such as [15]. Eventually, 12% of the 336 stations were eliminated for having less than the lower limit of available data (less than 10 years of records) to end up with 284 stations with records availability, as demonstrated in Table 1.

Table 1: Records Availability Among Screened Stations

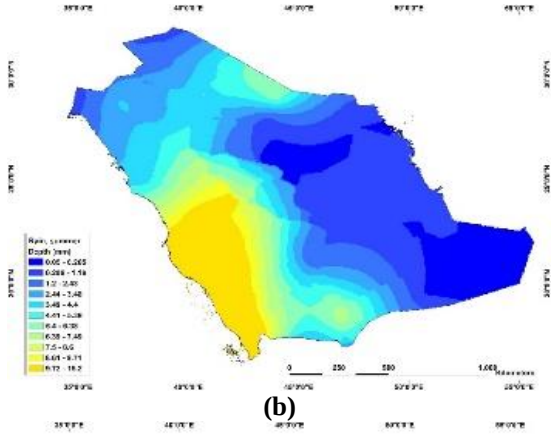
#Years with available daily records	#Stations
>=50	109
40-50	75
30-40	48
20-30	18
10-20	34
Total	284

B. Data Analysis

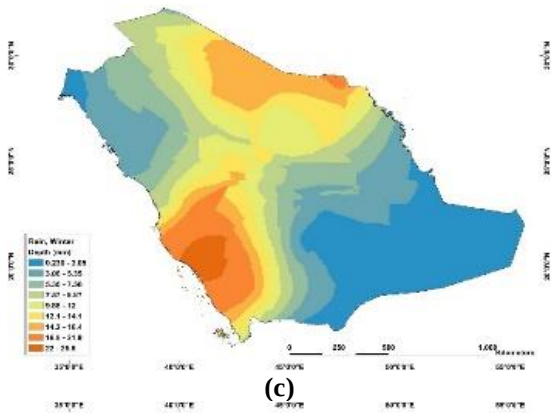
The collected rainfall stations are well distributed over the entire KSA area, with a separation distance between every two stations ranging from 650 m to 1750 km. While, with the nature of KSA of having mountains in the southwestern area and off-shore area along the red sea, the stations' elevation varied from 6 m a.m.s.l. to 2605 m a.m.s.l. On the other hand, KSA (as a dry country) with composite natural of mountains running along the shoreline and deserts; daily rainfall datasets tend to have extreme events, flash floods, together with many of zero values (the average number of rainy days per year is around 13 days). Furthermore, these extreme events and flash floods tend not to last for more than one day. The analysis has shown that the average daily events depth for the stations is about 11 mm, while the maximum daily rainfall depth is 248 mm. The reason behind the right skewness of the daily events histogram is the huge difference between the average depth of the events and the extreme events. On the monthly scale, Fig. 2 illustrates the average daily depth of the events over KSA for (a) autumn, (b) summer, (c) winter, and (d) spring. As shown in Fig. 2, the rainiest season is spring, then come winter, summer, and autumn. Additionally, spatially wise, the southwestern region of KSA has the highest amount of rainfall over the four seasons. Simultaneously, the northeastern region comes second while the gap between its rainfall amount and the amount in the southwestern region increases from autumn reaching the spring. Meanwhile, the northwestern region comes in the third rank, and lastly, the southeastern region presents a very low amount of rainfall due to its nature and containing Robaa El Khali, which is deserted. On the annual scale (Fig. 3), the average annual depth for all stations is 110 mm – in accordance with what [20] stated that the average annual rainfall depth over KSA was found to be around 100 mm and around 101.3 mm over its capital. In comparison, the maximum average annual depth per station is 495 mm. Whilst, the average annual maximum depth for all stations is around 305 mm, and the highest annual depth per station is 1630 mm. Besides, the regions kept the same ranks concerning the annual rainfall distribution like the monthly.



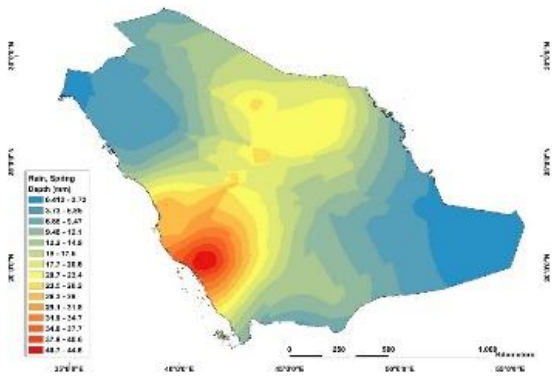
(a)



(b)

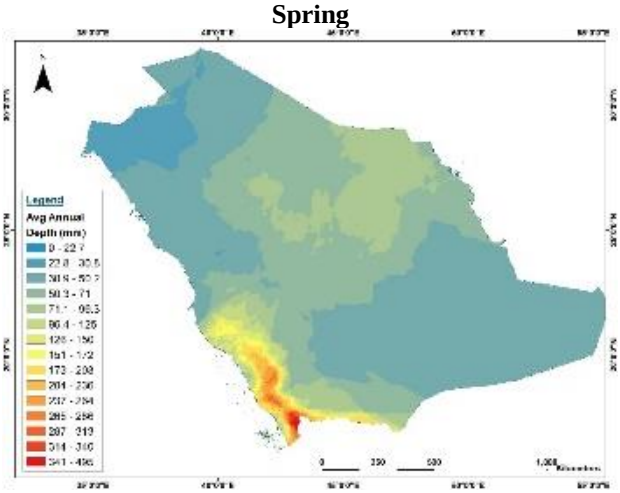


(c)

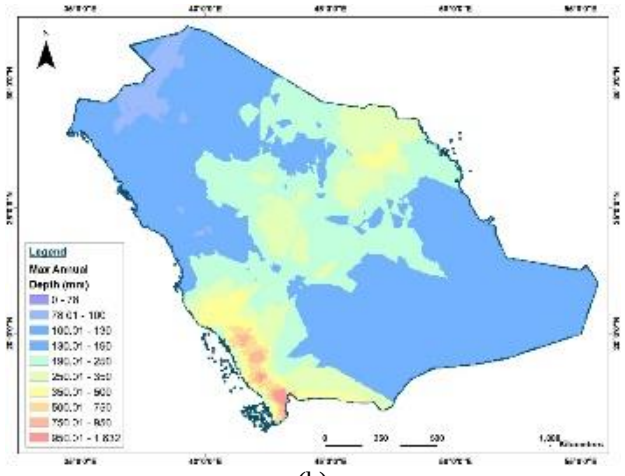


(d)

Fig. 2: Seasonality Effect on Rainfall Distribution Over KSA – (a) Autumn, (b) Summer, (c) Winter, and (d) Spring



(a)



(b)

Fig. 3: (a) Average and (b) Max. Annual Rainfall Distribution Over KSA

III. METHODOLOGY AND APPROACHES

A. Data Spatial Clustering

Several spatial clustering methods were applied in several studies, including the Gaussian Mixture Models (GMM), Agglomerative Hierarchical Clustering, and K-means method.

Gaussian Mixture Models (GMM)clustering method defines the cluster by two parameters (i.e.mean and standard deviation). Subsequently, the closer the point of data to the mean of a particular cluster, the higher the percentage of assigning this point to this cluster. However, this approach allows the clusters to overlap spatially. With this conclusion, this method is not applied in this paper, whose objective is to have defined separated clusters as an output.

The agglomerative Hierarchical Clustering method deals with all the data points as separated clusters. Every successive step combines two clusters depending on the smallest average linkage between them in a hierarchical form. The critical weakness point about this method is that the output depends on the arrangement of the data. Thus, several iterations of data arrangements should be adopted. So; the larger the number of the data points, the more iterations needed to cover all possible solutions. Not only that, some of the output clusters are not spatially logical.

The K-means method is widely used as a clustering method in data mining science. More specifically, it has been applied as the first step of clustering in the studies performed by [21] and [22] on rainfall data in the USA and Mexico, respectively. The method works on minimizing the distance between each element and the cluster's centroid. Additionally, K-means clustering method applications are not limited to rainfall and/or water resources but also, it is extended to cover different aspects as medical applications (e.g., reference [23] used K-means clustering method to study skin diseases). K-means clustering method is selected to be used in this paper because of its reliability and flexibility.

To test the performance of the K-means clustering method, several trials are conducted using the collected rainfall data. The analysis exhibited that the method is sensitive to the initial positioning of the centroids of the clusters, the number of repetitions, and the number of clusters into which the data will be divided. Thus, each of these aspects is studied to determine the optimum approach.

Regarding the initial cluster position and repetition, it was found that there are several methods for the initialization of the centroids of the clusters [24]. The following methods were selected to be applied and tested in this paper:

***a) Random Centroids*:** *centroids to be placed randomly without any constraints except being within the domain of the data [25].*

***b) Random Partitions*:** *data are assigned to clusters randomly then the centroids are calculated to proceed with the rest of the algorithm's steps [26].*

***c) Maxmin*:** *data is sorted descending according to their distance from the global centroid, and according to the number of the clusters, the top points on the list are taken as their initial centroids [27].*

***d) Closest to the boundaries*:** *initial centroids are chosen from the data points according to their position from the boundaries of the data domain. The closest points to the boundaries are taken as initial centroids [28].*

***e) K-means++*:** *the approach is to choose the first centroid randomly from the data points, and the second centroid is the closest point to it. The next step is to weigh the points by their distance to the closest centroid to them, and the next centroid is the point with the highest weight. This final step is repeated till all the initial centroids are determined [29].*

With the purpose of covering all the possible outputs to reach the optimum clustering, several studies performed the method for several iterations [24], [30]. For example, reference [31] studied a number of initialization methods and repeated the approach for 100 iterations which improved the results, but they did not define a pattern for the repetition. In this paper, a repetition pattern is developed to avoid being trapped in local optimum results and to work within the available domain. Moreover, the pattern was developed to depend on the number of iterations (n), which were taken 10, 100, 1000, and 10000 to test the effect of the number of iterations.

Hence the number of clusters of the data (K) plays a vital role in reaching the final outputs of the approach; several methods with different approaches and concepts were adopted in the previous studies [32] and also this paper:

1) Rule of thumb: this simple method relies on practical experience, not a mathematical theory. It is carried out by applying the following equation:$K \sim \sqrt{n/2}$Where n is the number of data elements, however, it can be used to put the threshold according to the number of data, but the spatial distribution of the data and the distance in-between play the main roles to make a decision not only the number of the data.

2) Elbow method: the approach of this method is to calculate the total in-between distance for different numbers of clusters and plot them on a graph. Lastly, choose the number of clusters after which the change in total distance is milder (i.e., an elbow is formed). By this, the number of clusters will be selected, after which increasing this number does not lead to a bigger change.

3) Information Criteria for Selection: these methods are used for selecting among models with different numbers of parameters. They seek to balance the increase in likelihood due to additional parameters by introducing a penalty term for each parameter. This approach is applicable to the K-means clustering method as increasing the number of clusters (K) results in decreasing the total distance, hence, increasing the likelihood. The techniques applied are Akaike's information criterion (AIC), modified Akaike's information criterion (AICm), Bayesian information criterion (BIC), and Final Prediction Error (FPE). All techniques define the best model as the model with the least value.

Based on the aforementioned methods and approaches and in order to facilitate and enhance the application of the K-means clustering method to reach reliable results, an algorithm is built as a part of this paper, as presented in Fig. 4.

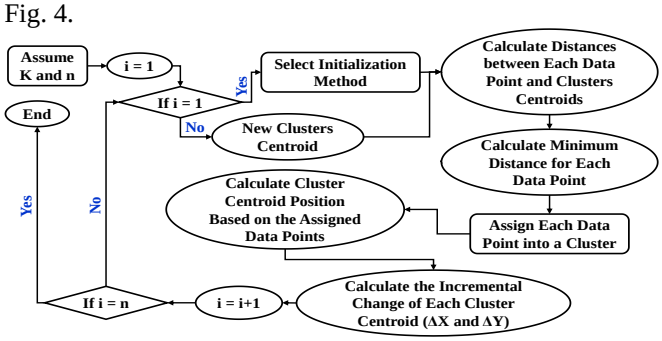


Fig. 4: Developed Algorithm for Enhanced K-means Clustering Method

As can be depicted from Fig. 4, the algorithm starts with the assumption of the number of clusters (K) and the number of iterations (n). At the first iteration, the cluster centroid position is calculated using the selected initialization method. After that, the distance between each rainfall station location and each cluster centroid is calculated. Consequently, each rainfall station is assigned to a specific cluster based on its minimum distance to this cluster centroid. After grouping rainfall stations into clusters, the exact cluster centroid is calculated based on the coordinates of its rainfall stations. In order to assure that the calculated coordinates of each cluster centroid are in the best position, a newly developed technique (developed as a part of this study) is applied depending on the number of iterations used. The newly developed technique depends on calculating the distance between each cluster centroid and the farthest boundary of the studied domain boundaries in both horizontal and vertical directions (L_X and L_Y). Then, the incremental distance (ΔX and ΔY) is calculated as follows:

$$\Delta X = \frac{L_X}{n}, \Delta Y = \frac{L_Y}{n}$$

Where "n" is the number of iterations.

At successive iterations after the first iteration, a new initial position of each cluster centroid is calculated automatically as follow:

$$X_{new} = X_{old} + \Delta X \times i$$
$$Y_{new} = Y_{old} + \Delta Y \times i$$

Where "i" is the iteration number.

At each iteration, the calculated clusters centroid positions should be extracted to be used in other analyses, which will be presented in the following sections.

B. Two-Step Clustering

The previous clustering process depends on the spatial distribution of the stations. However, these clusters are tested for sub-clustering temporally according to the rainfall data and their characteristics using the Two-Step Clustering approach. This method was developed by [33] as a modification of the BIRCH method [34] and has been used since then in several data mining disciplines.

The method is divided into two successively steps; the first step is to construct the cluster feature tree, which deals at the root stage with each of the data points as a cluster feature, then the approach is to end up at the leaf nodes with dense regions with almost equal size. Dealing with these dense regions will make it easier and more efficient for the next step rather than dealing with the whole data elements. The second step is clustering these cluster features hierarchically depending on the distance measure (DM). Lastly, the determination of the number of clusters is carried out by two steps: applying the information criteria method (BIC or AIC) to determine the maximum number of clusters, then, depending on the elbow method of the ratio of distance measured the final number of clusters is selected. The modification Chiu made to the regular BIRCH method is that the distance measure is derived from a probabilistic model; Log-likelihood. The distance measure is calculated as the reduction in the log-likelihood resulting from merging any two cluster features, while it is calculated using the Euclidean distance in the BIRCH method.

This method has several advantages, including dealing with both continuous and multinomial large datasets, noise handling, and automatically determining the number of clusters. Yet, the challenge of using the method is that the order of data may lead to different results. Thus, to overcome this, in this paper, data are re-ordered with every trial for a hundred iterations.

IV. RESULTS AND DISCUSSIONS

After applying the K-means algorithm with a different number of clusters and with the above-mentioned initialization methods, the summation of distances between rainfall station location and its cluster centroid is calculated as summarized in Fig. 5. The results reveal that for two clusters, all methods gave the same cluster configuration due to the clear spatial separation between the two clusters. However, along with increasing the number of clusters, the configuration of the clustering gave different results from one method to another for the sake of increasing the on-border stations.

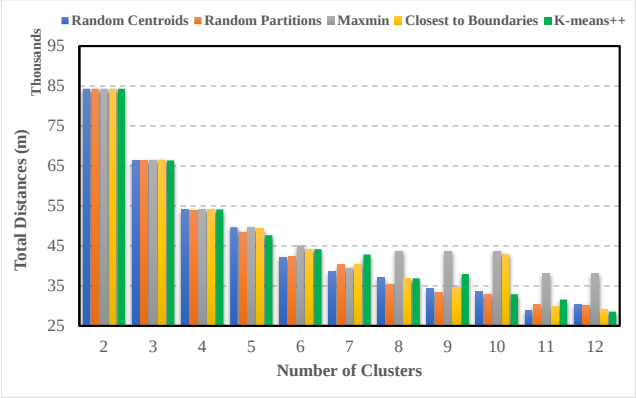


Fig. 5: Variability of Distances Summation corresponding to Each Cluster Initiation Positioning Method at Several Number of Clusters

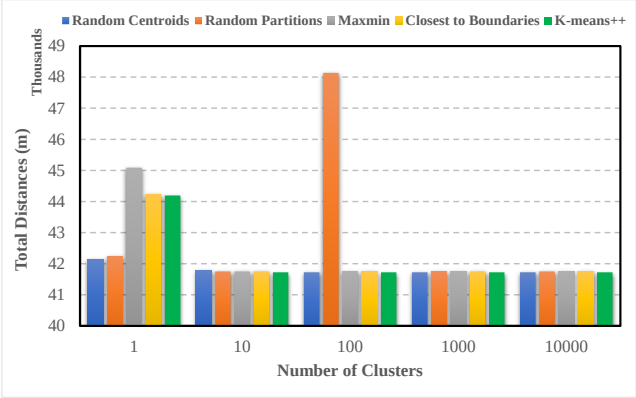


Fig. 6: Variability of Distances Summation corresponding to Each Cluster Initiation Positioning Method at K = 6

As can be depicted from Fig. 5, Random Centroids and Random Partitions methods showed no defined scheme to judge or predict the outputs. Meanwhile, Maxmin and Closest to the Boundaries methods gave clusters with higher total distance, especially with increasing the scattering of the data distribution between the clusters and increasing the number of clusters. Lastly, K-means++ showed better overall performance over Maxmin and Closest to the Boundaries, yet, its approach starts with choosing a random point, and that can lead to different outputs with every trial.

Depending on this conclusion, the repetition pattern was applied with a different number of iterations and different initialization methods for a different number of clusters. Fig. 6 represents the total distance after applying the different initialization methods and repeated with different numbers of iterations for the number of 6 clusters. While Fig. 7 shows the results for the K-means++ method.

In the comparison, Random Centroids and Random Partitions methods gave unpredictable results; for some number of clusters, they reach the minimum distance while for others, they do not, regardless of the number of iterations. This may be explained by the fact that the initialization for the start of the repatriation is completely random. On the other hand, Maxmin and Closest to the boundaries methods perform better while increasing the number of repetitions as that gives them a better chance to be located in the middle of the domain where some clusters exist. However, their performance gets worse by increasing the number of clusters as more and more clusters are drafted from the boundaries. Lastly, K-means++ gives the best performance for not being completely random (the first step only then its approach is to choose by weighting). However, as the selection of the first centroids is random between the data elements, for some clusters, it needed a higher number of iterations than others to reach the minimum total distance. Yet, it reaches the minimum total distance faster by increasing the number of iterations compared with the other methods while increasing the number of clusters.

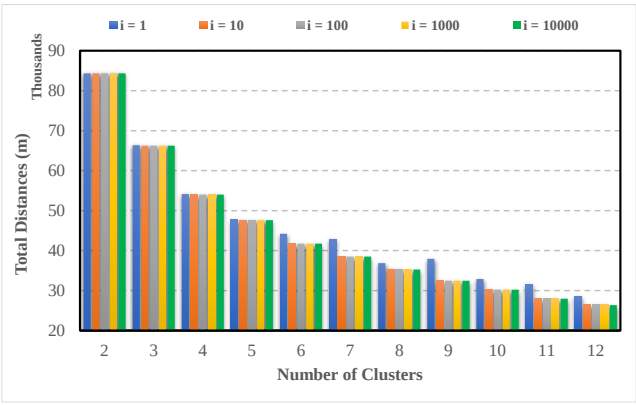


Fig. 7: Variability of Distances Summation corresponding to Several Number of Clusters Using K-means++ Method

The determination of the number of clusters started with putting a threshold of 12 clusters (applying the rule of thumb). Afterward, the elbow method was applied, and results are exhibited in Fig. 7 and Table 2. As concluded, increasing the number of clusters results in a small mild change in the total distance at the end. However, the elbow is observed at K = 4 and 6 as shown in Fig. 8 and with no significant difference to decisive when it comes to choosing between them. Thus, decision-making could not be dependent on this method only.

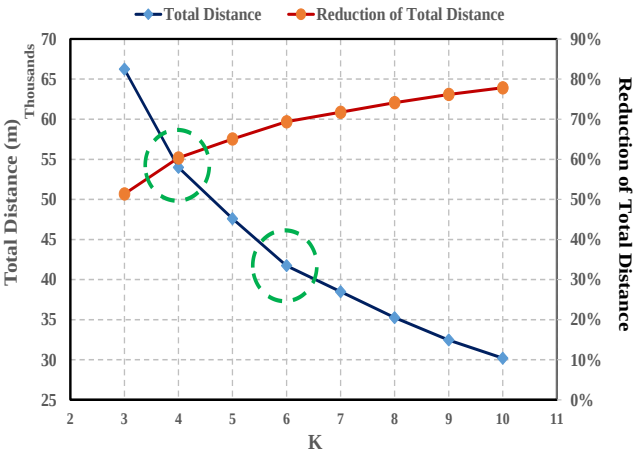


Fig. 8: Results of Elbow Method Application

As highlighted in Table 2, for K= 3, 4, and 5, the values for all techniques are small and relatively close to each other. Thus, the final decision for the number of clusters was made depending on merging the elbow method with the information criteria techniques. Consequently, K = 4 is chosen as the number of clusters with the elbow and least information criteria values. Finally, Fig. 9presents the final selected clusters (K=4) distributed spatially over KSA (excluding El-Robaa El-Khali).

Table 2: Records availability among screened stations

K	AIC	AIC_m	BIC	FPE
1	-1.887	-1.873	-1.874	0.152
2	-2.059	-2.016	-2.033	0.128
3	**-2.966**	**-2.88**	**-2.928**	**0.052**
4	**-2.977**	**-2.834**	**-2.926**	**0.051**
5	**-2.988**	**-2.773**	**-2.924**	**0.05**
6	-2.211	-1.908	-2.134	0.11
7	-0.726	-0.32	-0.636	0.484
8	-1.165	-0.641	-1.062	0.312
9	-1.635	-0.978	-1.519	0.195
10	-1.745	-0.939	-1.616	0.175
11	-1.475	-0.505	-1.334	0.229
12	0.068	1.219	0.222	1.07

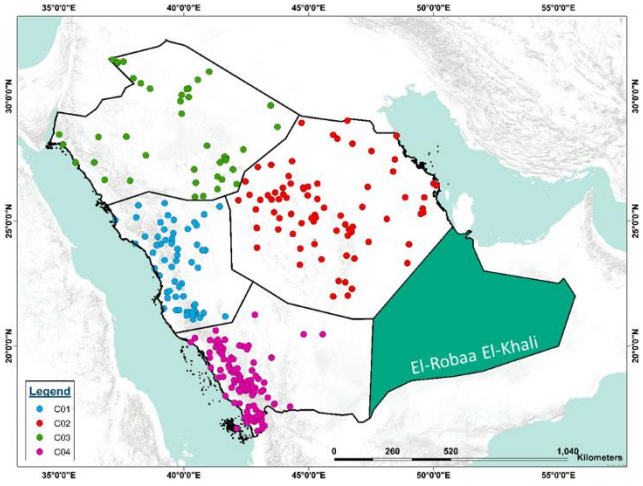


Fig. 9: Final clusters (K=4) distributed over KSA

Afterward, the Two-Step clustering method was applied to the four clusters. Several iterations were undertaken, and, incidentally, the outputs were not different for this study. Applying BIC and AIC to determine the maximum number of clusters resulted in a maximum number of clusters of one – the clusters do not need to be sub-clustered/re-clustered.

After reaching a defined, separated four clusters distributed over KSA's area and to demonstrate the rainfall variability over them, for each cluster, the average, maximum daily rainfall record, and the average number of rainy days are presented in Table 3. Additionally, Table 4 lists the average and maximum annual rainfall records.

Table 3: Daily data analysis for clusters

Cluster	Average Daily depth (mm)	Max. Daily depth (mm)	Average rainy days per year
1	11.200	190.4	9
2	8.332	202	11
3	7.861	140	8
4	12.974	248.6	18

Table 4: Annual data analysis for clusters

Cluster	Average Annual depth (mm)	Max. Annual depth (mm)
1	68.44	638.5
2	67.96	720.5
3	47.24	286
4	173.08	1781.98

V. CONCLUSION

In conclusion, the Kingdom of Saudi Arabia's rainfall distribution shows variability spatially and temporally. The analysis of this variability came along with the conclusions conducted in the literature. It can be summarized that; the highest amounts of rainfall occur over the southwestern and northeastern regions. Nevertheless, on the temporal scale, daily rainfall data distribution shows right skewness with many zeros and flash floods occurring with different patterns and durations. While the seasonality affects the rainfall pattern and amount, the amount of precipitation decreases from spring to winter to summer to autumn.

Consequently, with the witnessed variability, clustering of the rainfall data according to their spatial and temporal diversities is essential. Yet, the way the data are located spatially, density, and spread along the study area affect the complexity of the procedure of the clustering process to reach the optimum output. It is proven that increasing either the number of clusters or the density of the data elements around/on the clusters' borders increases the probability of having different outputs. Besides, more elements around/on the borders arises the question about the need to divide these data elements into more than one cluster. With all these concerns taken into consideration and by taking on a sensitivity analysis, the K-means method is recommended for spatial clustering with K-means++ initialization method and optimized repetition pattern to avoid being trapped in a local minimum and to work within the total available domain. Yet, when using the K-means clustering method to determine the number of clusters, the rule of thumb is not reliable, and the elbow method and information criteria methods should be combined to make a definitive decision.

REFERENCES

[1] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, Present and future Köppen-Geiger climate classification maps at 1-km resolution, Sci. data, 5 (2018) 180-214.

[2] M. S. ALYamani and Z. Sen, Regional variations of monthly rainfall amounts in the Kingdom of Saudi Arabia, Earth Sci., 6 (1) (1993).

[3] P. Wan, POINT RAINFALL CHARACTERISTICS OF SAUDI

ARABIA, Proc. Inst. Civ. Eng., 61 (1) 91976) 179–187.
[4] M. El-Nesr, A. Abdulrahman, and M. Abu-Zreig, Analysis of evapotranspiration variability and trends in the Arabian Peninsula, Am. J. Environ. Sci., 6 (6) (2010) 535–547.
[5] S. Hag-elsafi and M. El-Tayib, Spatial and statistical analysis of rainfall in the Kingdom of Saudi Arabia from 1979 to 2008, Weather, 71 (10) (2016) 262–266.
[6] A. Mashat and H. Abdel Basset, Analysis of rainfall over Saudi Arabia, J. King Abdulaziz Univ. Metrol. Environ. Arid L. Agric. Sci., 142 (592) (2011) 1–40
[7] A. O. Alamodi, A. S. Mashat, and H. M. Abdel Basset, On the relation between atmospheric pressure systems and rainfall prediction over the Kingdom of Saudi Arabia, Project, (2008).
[8] B., Bashir, andH., Fouli, Studying the spatial distribution of maximum monthly rainfall in selected regions of Saudi Arabia using geographic information systems, Arabian Journal of Geosciences, 8(11) (2015) 9929–9943.
[9] M. A., Al-Saleh, Variability and frequency of daily rainfall in Riyadh, Saudi Arabia. The Geographical Bulletin, 39(1) (1997) 48–57.
[10] S., Rehman, Temperature and rainfall variation over Dhahran, Saudi Arabia,(1970–2006), International Journal of Climatology: A Journal of the Royal Meteorological Society, 30(3) (2010) 445–449.
[11] M. M. A., Abdullah,A. M., Youssef, F., Nashar, and E. A., AlFadail, Statistical Analysis of Rainfall Patterns in Jeddah City, KSA: Future Impacts, In Rainfall. IntechOpen, (2019).
[12] R. M., Atif, M., Almazroui, S., Saeed, M. A., Abid, M. N., Islam, and M., Ismail, Extreme precipitation events over Saudi Arabia during the wet season and their associated teleconnections, Atmospheric Research, 231(February 2019), 104655. https://doi.org/10.1016/j.atmosres.2019.104655
[13] M., Almazroui, Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009, Atmospheric Research, 99(3–4) (2011) 400–414.
[14] M. Almazroui, R. Dambul, M. N. Islam, and P. D. Jones, Principal components-based regionalization of the Saudi Arabian climate, Int. J. Climatol., 35 (9) (2015) 2555–2573.https://doi.org/10.1002/joc.4139
[15] W. M., Abdeen, A. G., Awadallah, and N. A. Hassan, Investigating regional distribution for maximum daily rainfall in arid regions: a case study in Saudi Arabia, Arabian Journal of Geosciences, 13(13) (2020). https://doi.org/10.1007/s12517-020-05413-8.
[16] D. C., Hoaglin, andB., Iglewicz, Fine-tuning some resistant rules for outlier labeling, Journal of the American Statistical Association, 82(400) (1987) 1147–1149.
[17] D. C., Montgomery, G. C., Runger, andN. F. Hubele, Engineering statistics. John Wiley & Sons, (2011).
[18] J. K., Eischeid, P. A., Pasteris, H. F., Diaz, M. S., Plantico, andN. J., Lott, Creating a serially complete, national daily time series of temperature and precipitation for the western United States, Journal of Applied Meteorology, 39(9) (2000) 1580–1591.
[19] M. F., Villazón, andP., Willems, Filling gaps and daily dis accumulation of precipitation data for the rainfall-runoff model, in Proc. 4th Int. Sci. Conf. BALWOI, (2010) 25–29.
[20] A., Masood, J., Bahrawi, andA., Elfeki, Modeling annual rainfall time series in Saudi Arabia using first-order autoregressive AR(1) model, Arabian Journal of Geosciences, 12(6) (2019). https://doi.org/10.1007/s12517-019-4330-3
[21] J. L., Morales, F. A., Horta-Rangel, I., Segovia-Domínguez, A. R., Morua, andJ. H., Hernández, Analysis of a new spatial interpolation weighting method to estimate missing data applied to rainfall records,Atmósfera, 32(3) (2019) 237–259.
[22] R. S. V., Teegavarapu, Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbor classification and cluster-based interpolation methods, Hydrological Sciences Journal, 59(11) (2014) 2009–2026.
[23] S., Gopalakrishnan, Dr. B., Ebenezer Abishek, Dr. A., Vijayalakshm, and Dr. V., Rajendran, Analysis, And Diagnosis Using Deep-LearningAlgorithm On Erythemato-Squamous Disease, International Journal of Engineering Trends and Technology, 69(3) (2021)52–57.
[24] M. E., Celebi, H. A., Kingravi, andP. A. Vela, A comparative study of efficient initialization methods for the k-means clustering algorithm, Expert Systems with Applications, 40(1) (2013) 200–210.
[25] E. W., Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications, Biometrics, 21(1965) 768–769.
[26] S., Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory, 28(2) (1982) 129–137.
[27] T. F., Gonzalez, Clustering to minimize the maximum intercluster distance, Theoretical Computer Science, 38(1985) 293–306.
[28] S., Sieranoja, andP. Fränti, Random projection for k-means clustering, International Conference on Artificial Intelligence and Soft Computing, Springer (2018) 680–689.
[29] D., Arthur, andS., Vassilvitskii, k-means++: the advantages of careful seeding,. SODA'07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, (2007) 1027–1035.
[30] C. M., Poteraş, M. C., Mihăescu, andM., Mocanu, An optimized version of the K-Means clustering algorithm, Federated Conference on Computer Science and Information Systems, IEEE, (2014) 695–699.
[31] P., Fränti, andS. Sieranoja, How much can k-means be improved by using better initialization and repeats?, Pattern Recognition, 93(2019) 95–112.
[32] T. M., Kodinariya, andP. R., Makwana, Review on determining the number of Cluster in K-Means Clustering, International Journal of Advanced Research in Computer Science and Management Studies, 1(6) (2013) 2321–7782.
[33] T., Chiu, D., Fang, J., Chen, Y., Wang, andC., Jeris, A robust and scalable clustering algorithm for mixed type attributes in the large database environment. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2001) 263–268.
[34] T., Zhang, R., Ramakrishnan, and M. Livny, BIRCH: An efficient Data Clustering Method for Very Large Databases, Proceedings of the 1996 ACM SIGMODInternationalConferenceonManagementof Data,(1997) 103–114.