Studying Patterns of Rainfall and Topographical Clustering for Kingdom of Bahrain: An Application of Big Data

UNEB GAZDER
Department of Civil Engineering
University of Bahrain
Sakhir, 32038
BAHRAIN

Abstract: - Rainfall is an important aspect for urban planners, especially for professionals related to highway design. This study was aimed at studying the patterns of rainfall in Kingdom of Bahrain using cluster analysis with topographical data. The rainfall data was collected from the website meteorological directorate of ministry of transportation and telecommunication of Bahrain, which consisted of approximately 14000 days, thus referred to as big data. K-means clustering was used to identify the patterns. Moreover, Geographic Information System (GIS) was used for topographical clustering of elevation map of Bahrain. This was done to identify the areas which should receive more attention with regards to drainage due to the topography of Bahrain. The results of this study showed that the time periods which receive comparatively heavier rainfall are around January while those with significantly lower rainfall are during May and October. The topographical analysis showed that the northern parts of the country are supposed to be paid more attention to avoid flooding and, consequently, damage to infrastructure. Moreover, results of the study could be used by the urban and transportation planners in Bahrain for design of efficient drainage resulting in economic benefits to the society.

Key-Words: - rainfall, cluster analysis, big data analysis, GIS, topography

Received: August 11, 2023. Revised: December 14, 2023. Accepted: February 22, 2024. Published: April 3, 2024.

1 Introduction

Rainfall patterns and their changes have been commonly associated with the climatic changes occurring on a global level [1]. Studying rainfall patterns is a major concern for sustainable urban development [2]. Rainfall and urban development are considered interlinked to each other as one can affect the other [3,4]. The issue is more pressing for the highways as accumulation of water on the highways can result in structural and functional damage [5]. Such damages result in higher construction and maintenance costs, economic losses to travelers and, in some cases, loss of life [6-9].

There are two important aspects of studying rainfall patterns, namely, spatial and temporal. Spatial patterns refer to the study of areas which may be exposed to higher risk of flooding accumulation due to higher rainfall and/or lower elevations [10]. Temporal patterns refer to the identification of the season/time of the year during which heavier rainfalls are expected [11]. These factors would eventually impact the drainage design and emergency management activities.

Bahrain is considered one of the Arid countries with low rainfall intensity and quantities [12]. Due to this reason, effects of rainfall are often overlooked by urban and transport planners. However, consequent damages have been caused to the infrastructure, even with low rainfall, should not be ignored [13]. Despite this, there are few studies available on the analysis of rainfall patterns, especially those combining the temporal and topographical patterns.

The analysis technique used in this study is clustering analysis. Clustering aims to find the best partitioning of the data based upon some criteria of optimality. The criteria can be in the form of an objective function or the number of clusters to be achieved [14,15]. It has been commonly applied in the government and private sector for the identification of potential risks and opportunities [16]. The applications of clustering algorithms have been reported in the fields of statistics, computer science and machine learning for solving classical as well as new problems, such as traveling salesman and bioinformatics [17].

The clustering algorithms work on distance or similarity metric for development of clusters [18]. Some of the common clustering algorithms include k-means, hierarchical, self-organizing maps, expectation maximization, Density-based spatial clustering of applications with noise (DBSCAN), non-spherical clusters by using minimum spanning tree over KD-tree-based micro-clusters (MCMST), Fuzzy C-mean, and VIASCKDE Index. The selection of clustering algorithm depends on several factors such as popularity/familiarity, flexibility, applicability to the available dataset, and handling of multidimensional problems [19]. K-means and hierarchical clustering have been used more often than other available algorithms [20].

Due to its importance and multi-disciplinary impacts, this study is aimed at studying the spatial and temporal patterns of rainfall in Bahrain. Cluster analysis and Geographic Information System (GIS) was used for this purpose. The objectives of this study include classifying different periods of year according to the rainfall data and determining areas in Bahrain which are exposed to higher risk of flooding and accumulation. The results of this study are expected to enable better understanding of the rainfall patterns, possibly resulting in changes in infrastructure planning and management practices. Moreover, the readiness of emergency management services can also be improved with the identification of critical times and locations, subjected to higher rainfall and flooding.

2. Study Area and Dataset

Bahrain is an Archipelago of 33 islands, located at 16 Km from Saudi Arabia and 21 Km from Qatar. Its land area is estimated to be approximately 700 Km^2 [21]. Similar to other neighboring countries, Bahrain has also traditionally relied on petroleum for its trade. However, with the depletion of its reservoirs, it is currently facing scarcity of resources which makes the application of sustainability concepts crucial for long-term development of the nation [22].

Weather of Bahrain is usually hot and humid with temperatures going close to 50^0C during summer. The average temperatures are approximately close to 30^0C throughout the year [23]. Rainfall in Bahrain is less frequent and heavy rainfalls are seldom expected [24]. Due to this fact, the aspect of highway drainage is overlooked at the time of design and construction which then results in higher investments at later stages [25]. Hence, it is important that existing rainfall patterns are studied, and future planning is done accordingly to save economic losses to the country, especially in the current situation of resource scarcity.

Big data is referred to as a massive dataset with hidden, potentially complex trends and patterns [26]. It was crucial for this study to utilize a large amount of data for identifying rain patterns over the course of several years. The dataset for this study was taken from the website of ministry of transportation and telecommunications of Bahrain. The ministry has a meteorological directorate which is responsible for monitoring weather conditions in Bahrain. The data recorded by the sensors installed at the geographic location of Bahrain International Airport which is in Muharraq. It covered the period from January 1944 to December 2014. An initial analysis of the data showed that rainfall has not increased/decreased substantially throughout the above-mentioned period. Hence, it can be expected that the analysis done on the available data is valid for the present year as well.

The average rainfall was observed to be 0.2mm per day and 6.5mm per month. Maximum rainfall was 67.9mm which was observed in March 1995. The data had a standard deviation of 2mm. Table 2 shows more detailed month-wise statistics of the available data. It can be clearly observed that the variation in the monthly data seems quite significant for some of the months. This is further highlighted by the box plots shown in Fig. 1. The vast difference in rainfall data of different months, in terms of their magnitude and variation, justifies the need of the current study.

Table 1. Statistics of Available Data

	Mean	**Minimum**	**Maximum**	**Std.Dev.**
January	18.84	0.00	135.90	24.15
February	14.05	0.00	106.80	22.60
March	14.25	0.00	139.20	22.74
April	6.95	0.00	69.90	12.79
May	1.03	0.00	11.90	2.37
June	0.00	0.00	0.00	0.00
July	0.00	0.00	0.00	0.00
August	0.01	0.00	1.00	0.12
September	0.00	0.00	0.00	0.00
October	0.32	0.00	8.90	1.32
November	7.91	0.00	101.60	20.11
December	14.81	0.00	119.60	24.35

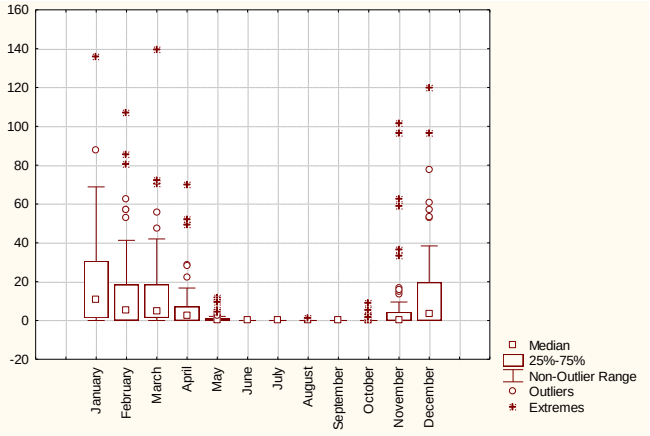

Fig. 1. Comparison of month-wise data

3. Analysis Methods

The clusters can be used to conveniently identify the patterns in the available dataset which has been the reason for their utilization in different fields including identification of faults. This was the prime reason for selection of this technique in the present study [27]. As an initial step, Analysis of Variance (ANOVA) was performed to check if there is a significant difference in the rainfall patterns of different months. The results of this test are shown in Table 2, which confirms that there is a significant difference in the rainfall patterns of different months at 5% level of confidence. This paves the way for further investigation done with clustering analysis in this study.

Table 2. ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	38700	11	3518	15	0.00	1.8
Within Groups	183449	792	231			
Total	222149	803				

Two types of cluster analysis were applied, to identify the season (months) with higher rainfall, namely, hierarchical tree and K-means clustering. Hierarchical clustering works by finding out the Euclidean distance (Equation (1)) between the samples of data and clustering the samples which have the least distance between them [28].

$$E_d(p, q) = \sum(p_i - q_i)^2 \qquad (1)$$

Where E_d is the Euclidean distance between datasets p and q.

On the other hand, K-means clustering is done by specifying the number of clusters to be made from the data and clustering algorithm will develop those clusters on the K-means distance between the samples. The clusters are developed based upon the maximization of distances between means (centroids) of clusters [29]. The former can be considered under unsupervised while the latter is part of supervised clustering approach.

As stated earlier, Bahrain is a small island, hence, the amount of rainfall does not change between parts of country. Therefore, it seemed logical to identify the areas which have higher potential of flooding based on topography. Geographic Information System (GIS) was used to find the parts of the country which have significantly lower elevation and, consequently, more likely to face water accumulation. GIS is a platform which provides integration between the spatial data of an area and its non-spatial characteristics. GIS is a popular platform to analyze spatial data based on non-spatial attributes [30]. Its advantages include integration of data from different sources and the ability to generate quarries and filters for the feature of interest. Due to these applications, GIS has been used for topographical analysis [31], which is the same task for this it is used in this study. An open-source software, namely QGIS, was used for this purpose. The maps and elevations for Bahrain were taken from Open Street Maps and United Stated Geological Survey (USGS) website.

4. Results

Hierarchical Tree clustering was first applied to find out the possible clusters based on rainfall data. The results are shown in Fig. 2. This figure shows that the season from May to October has a similar rainfall pattern while other months are significantly different than these months as well as each other.

Based on this finding, it was decided to proceed with seven clusters for K-means clustering. The results of K-means clustering are shown in Table 3. These results also conform with the Hierarchical Tree clusters. Therefore, it can be said that May-October is a season of lower rainfall with an average of 0.23mm while other months are part of the higher rainfall season. Among which, January had the highest rainfall with an average of 18.84mm. Table 3 shows more detailed statistics for each cluster.

For topographical clustering, street map of Bahrain was acquired from opensteetmap.org. Then, a layer was created using the elevations of roads and streets for Bahrain. The source of elevation data is already mentioned in the previous section. Clustering analysis involved classification of points on the highways and streets according to their

elevations. QGIS was used for performing topographical analysis.

Fig. 3 shows the topographical clustering of points in Bahrain which was done by using GIS. The colors are graduated from red (for lowest elevations) to white (for highest elevations). It can be observed that lower elevations are clustered along the coastal areas. Hence, flooding could be easily avoided in these areas with efficient drainage systems at lower cost since the outlet (sea) is closer to these areas. There is a higher clustering of low elevation points in the northern part as compared to the southern part, hence, this area requires more attention.

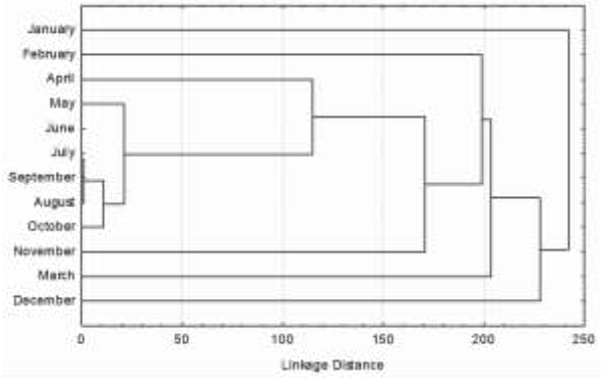

Fig. 2. Hierarchical Tree Clustering

Table 3. K-Means Clustering

	Jan	Feb	Mar	Apr	Dec	May-Oct	Nov
Jan	0.00	1158	1175	874	1320	915	1112
Feb	34	0.00	862.	591	1176	697	1000
Mar	34	29	0.00	619	905	708	868
Apr	29	24	24	0.00	776	205	582
Dec	36	34	30	27	0.00	798	1035
May-Oct	30	26	26	14	28	0.00	453
Nov	33	31	29	24	32	21	0.00

Table 4. Statistics of Clustered Data

Cluster	Mean	Minimum	Maximum	Std.Dev.
Jan	18.84	0.00	135.90	24.15
Feb	14.05	0.00	106.80	22.60
Mar	14.25	0.00	139.20	22.74
Apr	6.95	0.00	69.90	12.79
Dec	14.81	0.00	119.60	24.35
May-Oct	0.23	0.00	11.90	1.16
Nov	7.91	0.00	101.60	20.11

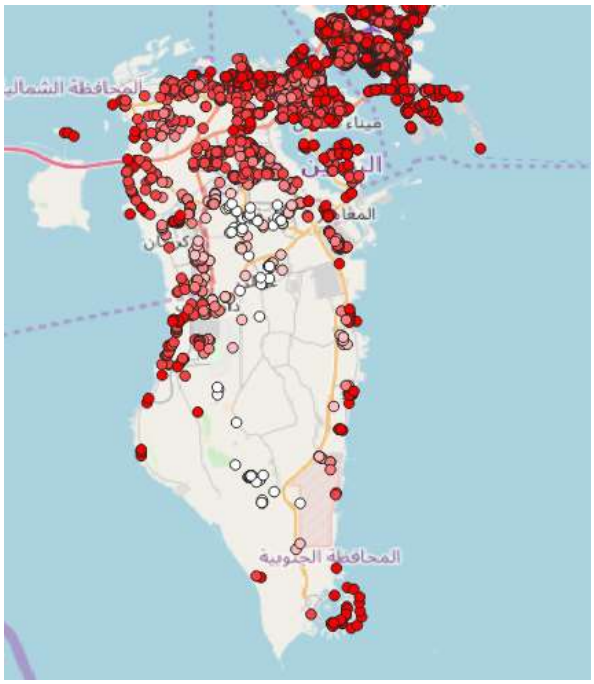

Fig. 3. Spatial Clustering

It should be noted that Hasanean and Almazroui [32], in a previous study, have recorded higher rainfall quantities for Saudi Arabia, especially in the adjoining areas of Bahrain. However, they have also emphasized the fact that the rainfall patterns are changing in the region. The identification of precipitation patterns has a global impact, since these arid regions (including Bahrain and Saudi Arabia) comprise ¼ of the land area and are characterized by strong intermittent hydrologic regime [33].

Another distinct feature of this study is the resemblance of the rainfall patterns in Bahrain with European countries with higher precipitation during winter season. This pattern has been linked to the climate change [34] which seems to be evident in arid countries such as Bahrain.

5. Conclusions

This study aimed at identifying the seasons (months) and areas of Bahrain which have higher potential of urban flooding. Meteorological data for a 60-year period was used for the analysis which is referred to as the big data. Clustering techniques, namely, Hierarchical Tree and K-means, were applied, on the big data of rainfall time series, identify the season of higher rainfall while GIS was used to identify areas which are at a higher risk of flooding due to topography of the country.

Clustering analysis showed that the following clusters can be made based on rainfall data; January, February, March, April, May-October, November

and December. May to October is a period with significantly lower rainfall. GIS analysis showed that coastal areas are at a higher risk of flooding, especially in the northern part of the country. The findings of this study are expected to be beneficial for planning authorities in Bahrain. They can be used for better preparedness against the risks of infrastructure damage due to urban flooding. One of the possible applications is the necessary cleaning of drainage systems in the area in the time and zones identified by the clusters. Another possible application of the results of this study is the assignment of resources (such as pumps) in the vicinity of the identified zones. The results of this study can also be linked with the climatic change patterns in the region and used to study the hydrologic regimes.

It is recommended for future studies to combine the spatial clustering data with temporal clusters for a more detailed analysis. GIS platform can be further used to study run-off patterns and designing of efficient drainage systems. Other possible directions of research could be to employ other machine learning and computational intelligence techniques for prediction of rainfall data.

References

[1] J. L. Martel, F. P. Brissette, P. Lucas-Picher, M. Troin, and R. Arsenault, Climate Change and rainfall intensity–duration–frequency curves: Overview of Science and Guidelines for Adaptation, *Journal of Hydrologic Engineering*, 26(10), 2021 03121001. M. H. Dore, Climate Change and Changes in Global Precipitation Patterns: What do We Know?, *Environment International*, 31(8), 2005, pp. 1167-1181.

[2] I. Abatcha, A. Mustapha, and A. Barkindo, Comprehensive Analysis of Rainfall Variability in Urban Maiduguri, Nigeria: Implications for Climate Resilience and Sustainable Development, *International Journal of Environment and Climate Change*, 14(3), 2024, 149-159.

[3] D. Mu, P. Luo, J. Lyu, M. Zhou, A. Huo, W. Duan, ... and X. Zhao, Impact of Temporal Rainfall Patterns on Flash Floods in Hue City, Vietnam, *Journal of Flood Risk Management*, 14(1), 2021, e12668.

[4] M. Hemmati, B. R. Ellingwood, and H. N. Mahmoud, The Role of Urban Growth in Resilience of Communities under Flood Risk, *Earth's Future*, 8(3), 2020, e2019EF001382.

[5] D. Lu, S. L. Tighe, and W. C. Xie, Impact of Flood Hazards on Pavement Performance, *International Journal of Pavement Engineering*, 21(6), 2020, 746-752.

[6] H. C. Z. Qing, Impact of Flooding on Highway [J], *Meteorological Monthly*, 9, 2000.

[7] R. K. Dahal, S. Hasegawa, T. Masuda, and M. Yamanaka, Roadside Slope Failures in Nepal during Torrential Rainfall and Their Mitigation, *Disaster Mitigation of Debris Flows, Slope Failures and Landslides*, 2006, pp. 503-514.

[8] P. A. Pisano, L. C. Goodwin, and M. A. Rossetti, US Highway Crashes in Adverse Road Weather Conditions, In *24th Conference on International Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, New Orleans, LA, January, 2008.

[9] A. M. Youssef, B. Pradhan, and N. H. Maerz, Debris Flow Impact Assessment caused by 14 April 2012 Rainfall along the Al-Hada Highway, Kingdom of Saudi Arabia using High-Resolution Satellite Imagery, *Arabian Journal of Geosciences*, 7(7), 2014, pp. 2591-2601.

[10] S. Ghosh, D., Das, S. C. Kao, and A. R. Ganguly, Lack of Uniform Trends but Increasing Spatial Variability in Observed Indian Rainfall Extremes, Nature Climate Change, 2(2), 2012, 86.

[11] J. Panthi, P. Dahal, M. Shrestha, S. Aryal, N. Krakauer, S. Pradhanang, ... and R. Karki, Spatial and Temporal Variability of Rainfall in The Gandaki River Basin of Nepal Himalaya, *Climate*, 3(1), 2015, pp. 210-226.

[12] N. A. Elagib, and A. S. A. Abdu, Development of Temperatures in The Kingdom of Bahrain from 1947 to 2005, *Theoretical and Applied Climatology*, 101, 2010, pp. 269-279.

[13] P. K. Naik, M. Mojica, F. Ahmed, and S. Al-Mannai, Storm Water Injection in Bahrain: Pilot Studies, *Arabian Journal of Geosciences*, 10(20), 2017, 452.

[14] S. Landau, and I. C. Ster, Cluster Analysis: Overview, Á Á, 11(x12), 2010, x1p.

[15] B. S. Duran, and P. L. Odell, *Cluster Analysis: A Survey* (Vol. 100), Springer Science & Business Media, 2013.

[16] T. J. Roelandt, and P. den Hertog, Cluster Analysis and Cluster-based Policy Making: The State of The Art, *Boosting Innovation: The Cluster Approach*, 1999, pp. 413-427.

[17] R. Xu, and D. Wunsch, Survey of Clustering Algorithms, *IEEE Transactions on Neural Networks*, 16(3), 2005, pp. 645-678.

[18] A. Nagpal, A. Jatain, and D. Gaur, Review Based on Data Clustering Algorithms, In *2013 IEEE Conference on Information & Communication Technologies*, April 2013, pp. 298-303, IEEE.
[19] O. A. Abbas, Comparisons between Data Clustering Algorithms, *International Arab Journal of Information Technology (IAJIT)*, 5(3), 2008.
[20] P. Govender, and V. Sivakumar, Application of K-Means and Hierarchical Clustering Techniques for Analysis of Air Pollution: A Review (1980–2019), *Atmospheric Pollution Research*, 11(1), 2020, pp. 40-56.
[21] S. Mabon, The Battle for Bahrain: Iranian-Saudi Rivalry, *Middle East Policy*, 19(2), 2012, 84.
[22] F. H. Lawson, *Bahrain: The Modernization of Autocracy*, Routledge, 2019.
[23] R. Hasan, S. M. Suliman, and Y. A. Malki, An Investigation into The Delays in Road Projects in Bahrain, *International Journal of Research in Engineering and Science*, 2(2), 2014, pp. 38-47.
[24] H. Radhi, F. Fikry, and S. Sharples, Impacts of Urbanisation on The Thermal Behaviour of New Built Up Environments: A Scoping Study of The Urban Heat Island in Bahrain, *Landscape and Urban Planning*, 113, 2013, pp. 47-61.
[25] J. Saundalkar, First Phase of Bahrain's Sheikh Zayed Highway to go Ahead; Western Bainoona Group involved, *ME Construction News*, available at https://meconstructionnews.com/35443/first-phase-of-bahrains-sheikh-zayed-highway-to-go-ahead-western-bainoona-group-involved, 24 June 2019, accessed on 29th August 2019.
[26] Y. V. Burlachenko, and B. A. Snopok, Methods of Cluster Analysis in Sensor Engineering: Advantages and Faults, Semiconductor Physics Quantum Electronics & Optoelectronics, 2010.
[27] S. Sagiroglu, and D. Sinanc, Big Data: A Review, In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42-47), May, 2013. IEEE.
[28] R. S. Madhulatha, An Overview on Clustering Methods, *IOSR Journal of Engineering*, Vol. 2(4), April 2012, pp: 719-725.
[29] A. K. Jain, Data Clustering: 50 Years beyond K-Means, *Pattern Recognition Letters*, 31(8), 2010, pp. 651-666.
[30] E. Esa, and M. Assen, A GIS based Land Suitability Analysis for Sustainable Agricultural Planning in Gelda Catchment, Northwest Highlands of Ethiopia, *Journal of Geography and Regional Planning*, 10(5), 2017, pp. 77-91.
[31] S. Fotheringham, and P. Rogerson, *Spatial analysis and GIS*, CRC Press, 2014.
[32] H. Hasanean, and M. Almazroui, Rainfall: Features and Variations over Saudi Arabia, A Review, *Climate*, 3(3), 2015, pp. 578-626.
[33] R. B. Ouarda, C., Charron, K. N. Kumar, P. R. Marpu, H. Ghedira, A. Molini, and I. Khayal, Evolution of The Rainfall Regime in The United Arab Emirates, *Journal of Hydrology*, 514, 2014, pp. 258-270.
[34] O. Planchon, H. Quénol, N. Dupont, and S. Corgne, Application of The Hess-Brezowsky Classification to The Identification of Weather Patterns causing Heavy Winter Rainfall in Brittany (France), *Natural Hazards and Earth System Sciences*, 9(4), 2009, pp. 1161-1173.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Uneb Gazder is the sole author of this paper who carried out all tasks related to the study shown in this paper.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The author has no conflicts of interest to declare that are relevant to the content of this article.