Journal of Advanced Informatics in Water, Soil, and Structure 2025 2 (2) 312-321

Journal of Advanced Informatics in Water, Soil, and Structure

journal homepage: wss.torbath.ac.ir

Research Article

Assessment of Groundwater Quality Patterns in the Southern Ramhormoz Aquifer Using K-Means Clustering and GIS

Saeedeh Firouzi[*1], *Pooria Sheikhy Beglari*[1], *Seyed Majid Mousav*[1], *Ali Shahbazi*[1]
1- Khuzestan Water and Power Authority, Khuzestan, Iran.

ARTICLE INFO

Received date: 13 Sep 2025
Accept date: 24 Nov 2025
Published date: 29 Nov 2025

Keywords:
GIS, Groundwater, K-Means clustering, Nitrate, Qualitative classification, Salinity

Abstract

Monitoring the quality of groundwater resources is essential for the sustainable use of these resources. K-Means is one of the most important and widely used algorithms in clustering. In the present study, qualitative classification of wells in the southern part of Ramhormoz Plain aquifer was performed using the K-Means method. Accordingly, water samples were collected from 12 well sampling locations in spring 2022 and, after chemical analysis, were clustered in the Python environment. Qualitative analysis of the data from the sampled wells indicated 3 dominant clusters in this plain. Based on this, the zoning map of Ramhormoz plain was drawn, which showed 3 zones with relatively good, moderate, and undesirable quality. The existence of marked differences in nitrate and salinity levels among clusters also indicates that the factors affecting water quality in the study area have been a combination of natural geological elements and human pressures on groundwater resources. Also, drawing contour maps of total dissolved solids (TDS), electrical conductivity (EC), and nitrate demonstrates the influence of these parameters on K-Means clustering.

Homepage: wss.torbath.ac.ir

*Corresponding Author:
Saeedeh Firouzi
Email: s.firoozi67@yahoo.com
ORCID: 0000- 0003-2646-3575

https://doi.org/10.22048/wss.2026.582946.1028

How to cite this article:
Firoozi,S , Sheikhy Beglari,P , Mousavi,M and Shahbazi,A . (2025). Assessment of Groundwater Quality Patterns in the Southern Ramhormoz Aquifer Using K-Means Clustering and GIS. *Journal of Advanced Informatics in Water, Soil, and Structure*, *1*(2), 312-320. doi: 10.22048/wss.2026.582946.1028

1- Introduction

Water is considered one of the most important assets and vital resources for humans and other living organisms. In recent years, salinization and declining groundwater quality have become a serious environmental problem worldwide. Undoubtedly, monitoring groundwater quality, considering its changes over time (which may occur due to human activities and industrial development), is essential for the sustainable use of water resources (Dinpajouh, 2022). Since there are no fixed and uniform standards for analyzing water quality worldwide, multivariate methods such as cluster analysis and factor analysis are employed. These methods are capable of classifying observation stations with similar water quality based on qualitative parameters into different groups (Kolahdouzan et al., 2011). For example, the clustering of wells in the Almonte plain in Spain is a representative case of this approach. Naranjo-Fernandez et al. (2020), based on water level data from 160 piezometric stations, showed the water level decline in this plain during the statistical period of 1975-2016 in six different clusters.

Firouzi et al. (2023) investigated the hydrochemical and statistical analysis of groundwater resources in the Zeydoun plain using hierarchical cluster analysis (HCA). The existing groundwater wells in this plain were divided into two categories: the first category showed sulfate water type and the second category showed chloride water type. Sheikhi Beglari et al. (2024) examined the qualitative status of groundwater resources in the Qaleh Tal plain aquifer, and the results showed that groundwater quality in this region varied from relatively poor to good. High concentrations of nitrate, electrical conductivity, and total hardness were the main factors reducing quality in relatively poor resources. Firouzi et al. (2025) evaluated the hydrochemistry of groundwater resources in the Ramhormoz plain for drinking and agricultural purposes, and found that groundwater resources are classified as very saline and unsuitable for agricultural use, and qualitative zoning for drinking showed that groundwater resources in the eastern part of the plain are in the good range while in the outlet area (western and southwestern regions) they fall in the non-consumable range.

Javadi et al. (2017) used the K-Means clustering method as a tool for modifying the DRASTIC index assessment in the vulnerability of the Qazvin aquifer with an area of 4,942 km². This method was applied to a large aquifer in northern Iran, and its results were statistically compared with vulnerability estimates obtained from the traditional DRASTIC method. The results showed that using the clustering method, a vulnerability map was created that does not rely on fixed weights and rankings, thus providing a more objective representation of the physical characteristics of the system. This resulted in Pearson correlation coefficients for nitrate, chloride, and TDS layers being 15, 22, and 5 percent higher, respectively, than the traditional DRASTIC method.

Not only does the clustering method have the ability to classify flow in groundwater, but it has also been applied in surface water resources. Liu et al. (2022) used the clustering method to improve the accuracy of short-term prediction models. In this study, a short-term flow prediction framework was proposed by combining the LSTM method with a self-organizing map (SOM), which clusters hydrometeorological factors through the SOM network to classify flow into multiple states with different clustering patterns. Also, in another study, Azhar et al. (2015) classified the water quality of 9 stations in the Muda River basin using the clustering method, and the results showed that the measured variables of river water quality status have a high correlation with water classification by the Environmental Protection Agency.

The aim of this study is to develop a data-driven classification and qualitative zoning of groundwater resources in the southern part of the Ramhormoz Plain by integrating K-Means clustering with GIS-based spatial analysis of EC, TDS, and nitrate. This approach is intended to identify groups of wells with similar hydrochemical behavior, delineate groundwater quality zones, and support interpretation of the relative influence of geogenic and anthropogenic factors on salinity and nitrate contamination.

2- Material and Methods

2-1- K-Means Clustering Method

Clustering is one of the branches of unsupervised learning and is an automatic process during which samples are divided into groups whose members are similar to each other, and these groups are called clusters. The goal of clustering is to find similar clusters of objects among the input samples. Clustering algorithms must possess the following characteristics (Steinley and Brusco, 2007): (1) scalability, (2) ability to work with different attributes, (3) finding clusters with arbitrary shapes, (4) requiring minimal domain knowledge for input parameter specification, (5) insensitivity to input data order, (6) handling of different and high dimensions, and (7) adaptability and ease of use.

The most important and widely used algorithm in clustering is the K-Means algorithm. K-Means is a centroid-based clustering algorithm in which the center of each cluster (centroid) is calculated as the arithmetic mean of the data points belonging to that cluster. These centroids are not necessarily actual data points but represent the average position of the observations within each cluster. In fact, the K-Means of a cluster is the most central element of that cluster. The objective of this method is to reduce sensitivity to large values in the dataset. In this algorithm, each cluster is represented by one of the data points closest to the center; therefore, this method is a vector quantization approach originally derived from signal processing and is popular for cluster analysis in data mining. K-Means clustering aims to partition n observations into k clusters, where each observation belongs to the cluster with the nearest mean, which serves as the prototype of the cluster. In this algorithm, considering the number of data categories (k), the following least squares function is minimized and the data are optimally separated and categorized (Aliesmaeili and Rajabi, 2016):

$$J = \sum_{K} \sum_{p_i} (p_i - c_k)^2 \quad (1)$$

where pi or (yi, xi) is a data sample from the dataset and ck or (cky, ckx) refers to the center of cluster k.

The data used in this research were collected from 12 sampling locations in the spring of 2022. The samples were sent to the laboratory of the Khuzestan Water and Power Authority for qualitative analysis, and the chemical analysis results of the Ramhormoz plain samples are presented in Table 1. In this study, using the K-Means algorithm in Python software and drawing TDS, EC, and nitrate contour maps in GIS software, the qualitative zoning of the Ramhormoz plain was performed.

Table 1 Results of chemical analysis of groundwater samples from Ramhormoz plain

No.	Well	X	Y	EC (µS/cm)	TDS (mg/L)	Nitrate (mg/L)
1	R1	362671	3459942	4030	2655.02	26.498
2	R7	369947	3459931	3860	2542.049	26.269
3	R11	361616	3464822	3990	2628.088	27.478
4	R12	361908	3453381	2650	1818.194	5.914
5	R14	364844	3455349	2190	1488.729	2.4
6	R16	350661	3460515	6830	4500.959	1.8
7	R21	349000	3470690	2660	1753.354	1.551
8	R22	354545	3467005	3930	2590.192	16.371
9	R25	364953	3464153	3830	2523.736	16.456
10	R5	366108	3458040	2930	2015.68	10.021
11	R13	360636	3451499	6240	4117.224	1.53
12	R2	364987	3459979	4060	2671.052	41.645

2-2- Study Area

The study area is located in the southern part of the Ramhormoz plain, between longitudes 49°46′ to 49°9′ E and latitudes 31°41′ to 31°6′ N, covering an area of 405 km² in the eastern part of Khuzestan province. The boundaries of Masjed Soleyman and Haftgel are in the northern part, Jayizan and Omidiyeh in the east, and Dalun-Meydavoud and Baghmalek in the western part of this plain. Figure 1 shows the geological map and the location of the study area and sampling sites.

3- Results and Discussion

Qualitative analysis of groundwater well data using the K-Means clustering method, as shown in Figure 2, led to the separation of the investigated resources into three distinct clusters, each having different and significant hydrochemical characteristics, as shown in Tables 2 and 3. Statistical and content analysis of the qualitative characteristics of wells in each cluster indicates that the K-Means algorithm has successfully revealed the hidden structure in the qualitative data and created groups with relatively homogeneous characteristics within clusters and heterogeneous characteristics compared to other clusters. Detailed examination of EC, TDS, and nitrate concentration parameters in each cluster, as key indicators of groundwater quality, provides a comprehensive view of the qualitative status of wells and the factors affecting them.

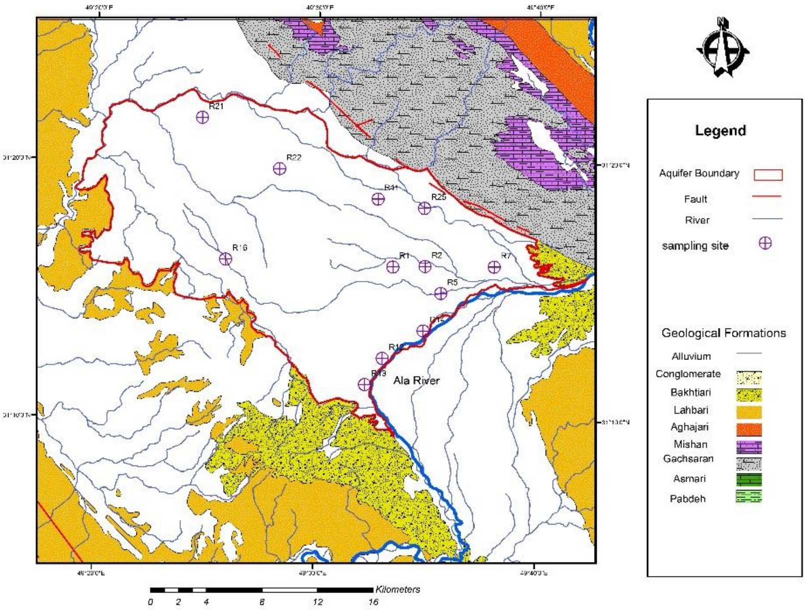

Fig. 1 Geological map and location of sampling sites in Ramhormoz plain

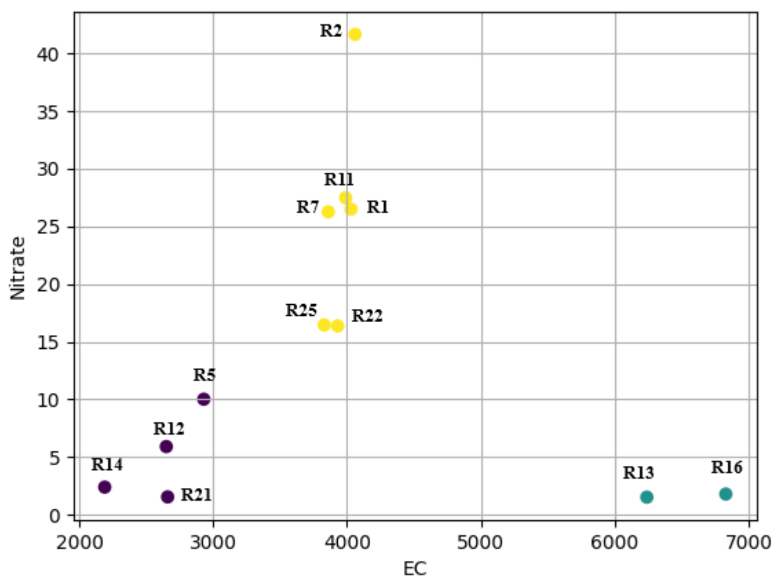

Fig. 2 Cluster plot of groundwater samples from Ramhormoz plain – Spring 2022

According to Table 2, the first cluster (Cluster 0) includes wells with the lowest nitrate values among all data. Nitrate in this group is generally below 10 mg/L and even below 2 mg/L in some cases.

Notable wells in this cluster include R14, R21, and R5. Simultaneously with the low nitrate concentration, electrical conductivity and total dissolved solids values in this group are in the relatively low to moderate range. EC in the wells of this cluster is generally between 2,200 and 2,900 μS/cm, and TDS is around 1,500 to 2,000 mg/L. These characteristics indicate relatively favorable qualitative conditions, free from human pollution in these wells. The origin of such quality can be sought in various factors, including the spatial distance of these wells from surface pollution sources such as agricultural lands or residential areas, the presence of impermeable surface layers, and natural recharge of the aquifer through precipitation or upstream aquifers with high quality. Particularly, low nitrate in wells such as R21 (1.55 mg/L) and R14 (2.4 mg/L) indicates that these areas have not been directly or indirectly affected by human activities, or such effects have been very minimal. Overall, the wells in Cluster 0 are in an acceptable condition from the perspective of public health, drinking safety, and even agriculture.

In the second cluster (Cluster 1), a completely different situation is observed. Although nitrate values in this cluster are very low and even lower than those observed in Cluster 0 (with values around 1.5 to 1.8 mg/L), the electrical conductivity and TDS values simultaneously are at the highest possible level in the entire dataset. For example, well R16 has an EC of 6,830 μS/cm and TDS of approximately 4,500 mg/L, and well R13 also has similar values. Such values clearly indicate very high salinity in the waters of this group. The origin of high salinity in this area, unlike the previous cluster, is likely geological rather than anthropogenic; because the simultaneous occurrence of high salinity with very low nitrate indicates the absence of surface pollution sources such as sewage or chemical fertilizers. Evaporitic formations such as marl, gypsum, salt, or saline shale layers may be present along the recharge path of these wells, causing the dissolution of soluble ions in the water and consequently a significant increase in salinity. Although nitrate, as one of the most important anthropogenic pollutants, is absent in this cluster, water quality from the salinity perspective is considered very poor, and such water is unsuitable for human consumption. From an agricultural perspective, using water with TDS above 3,000 mg/L is only possible for salt-tolerant plants and in many cases causes severe reduction in crop yields. This situation demonstrates that low water quality does not necessarily always mean pollution, but can result from geological origin and natural hydrogeochemical structures of the region.

The third cluster (Cluster 2) contains the largest number of wells and is in a more critical qualitative condition compared to the other two clusters. In this cluster, nitrate concentration in most wells is between 16 and 27 mg/L, and in a specific case (well R2), the nitrate value has reached more than 41 mg/L, which is very close to the permissible limit set by the World Health Organization (50 mg/L) for drinking water. Electrical conductivity and TDS values in this group are also in a relatively high range, with averages of approximately 3,900 μS/cm and 2,600 mg/L, respectively. This simultaneous increase in nitrate with increasing salinity may be due to the concurrent presence of anthropogenic pollution sources in the region as well as favorable geological characteristics for pollution transfer to the aquifer depth. There is a high probability that wells in this cluster are located near agricultural lands with high application of nitrogenous fertilizers, or that urban or rural sewage systems have directly or indirectly entered groundwater resources. Such a situation is considered a warning for public health, as increasing nitrate in groundwater resources, in addition to adverse effects on water quality, also has serious health consequences, including the occurrence of blue baby syndrome (methemoglobinemia). This cluster represents areas whose aquifers have experienced the greatest pressure from human activities and require continuous monitoring, corrective measures, and exploitation restrictions.

Table 2 K-Means clustering results for wells in Ramhormoz plain

Cluster	Mean EC	Mean TDS	Mean Nitrate	No. of Wells
0	Lower	Lower	Low to moderate	4 wells
1	Very high	Very high	Very low	2 wells
2	High	High	High to very high	6 wells

Table 3 Descriptive statistics of K-Means clustering for wells in Ramhormoz plain

Parameter	**Cluster 0**	**Cluster 1**	**Cluster 2**
Nitrate (mg/L)	1.5 to 10	Very low – below 2	High (16 to 41)
Mean EC-TDS	Low	Very high	High
Pollution level	Low	Low	High
Water quality	Relatively good	Moderate – high salinity from geology and non-agricultural activity	Unsuitable – nitrate pollution from agricultural activity
Well names	R5, R12, R14, R21	R13, R16	R1, R2, R7, R11, R22, R25

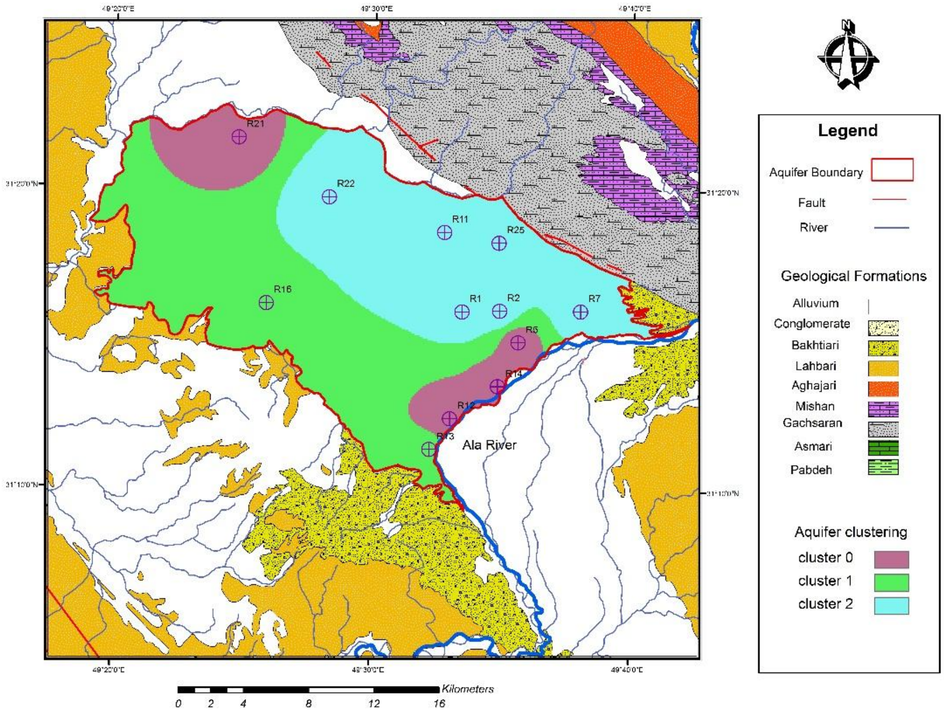

Fig. 3 Qualitative zoning map of Ramhormoz plain using K-Means clustering with GIS

The qualitative analysis of groundwater well data in the Ramhormoz plain using the GIS zoning method confirms the K-Means clustering results. All zoning maps were produced using the Kriging interpolation method in ArcGIS. Kriging was selected because it accounts for the spatial autocorrelation among groundwater quality data and provides statistically based estimates by modeling the spatial structure of the data through a semi-variogram. Figure 3 shows the spatial zoning of groundwater samples based on K-means clustering, which is generally consistent with the spatial distribution patterns of TDS, EC, and nitrate shown in Figures 4–6. Figures 4 and 5 show the TDS and EC contour maps, respectively, indicating that the highest values of these two parameters are mainly observed in the southern part of the plain. According to Figure 6, nitrate concentrations in the southern part of the plain are relatively low; therefore, this area is classified as Cluster 1 based on the clustering results. In the northern part of the plain, the spatial pattern is not uniform. A limited portion of the northern area shows lower TDS and EC values and belongs to Cluster 0, whereas the remaining northern areas are assigned to Cluster 2 and are characterized by relatively higher nitrate concentrations, along with higher EC and TDS values compared with the Cluster 0 portion of the northern plain. This clarification indicates that the apparent inconsistency results from the spatial variability within the northern sector rather than a contradiction between the clustering results and the contour maps.

A detailed comparison of the clusters with international standards and hydrogeological contexts reveals distinct degradation pathways. Cluster 1 represents a geogenic salinity facies; despite having the lowest nitrate levels (1.5–1.8 mg/L), it exhibits the highest TDS (>4500 mg/L) and EC values, exceeding the thresholds for most agricultural and domestic uses. This indicates that high salinity in this zone is not driven by human activities but likely by the dissolution of evaporitic formations (e.g., marl, gypsum, or salt layers). Conversely, Cluster 2 highlights the impact of anthropogenic pressures. In this group, nitrate concentrations reach up to 41 mg/L—approaching the WHO maximum permissible limit of 50 mg/L—coupled with high EC (~3900 µS/cm). This simultaneous elevation of nitrate and salinity suggests a combination of agricultural runoff and possibly domestic sewage infiltration. Cluster 0, with the lowest nitrate (<10 mg/L) and relatively low EC, remains the most preserved zone, likely due to its distance from pollution sources or natural protection by impermeable layers. These findings align with regional studies which distinguish between natural mineralization and human-induced contamination in Iranian aquifers, further validating the K-Means algorithm's ability to objectively classify complex groundwater data.

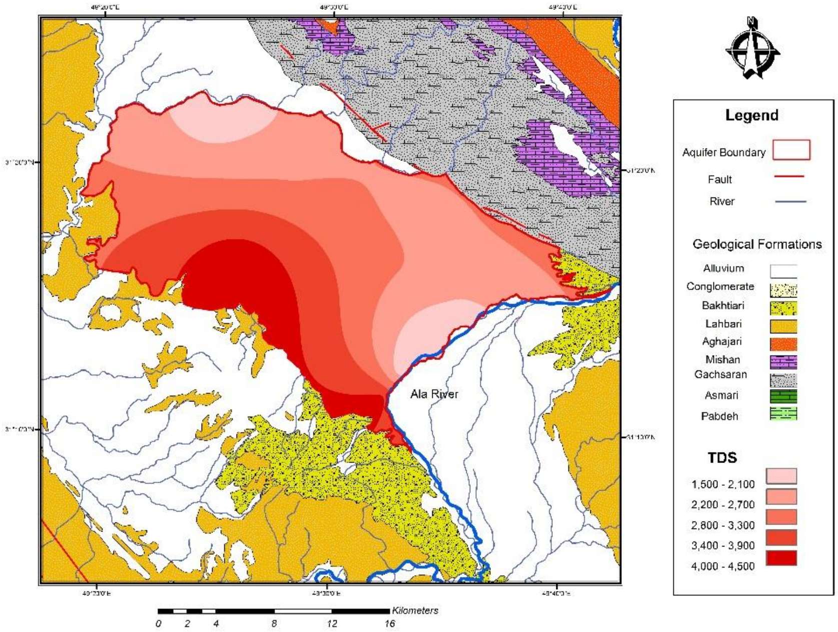

Fig. 4 Total dissolved solids (TDS) contour map of Ramhormoz plain

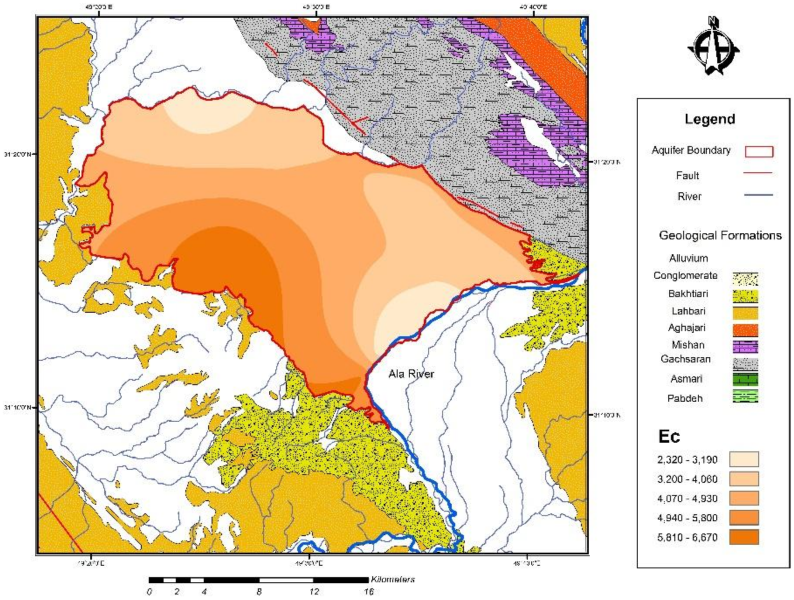

Fig. 5 Electrical conductivity (EC) contour map of Ramhormoz plain

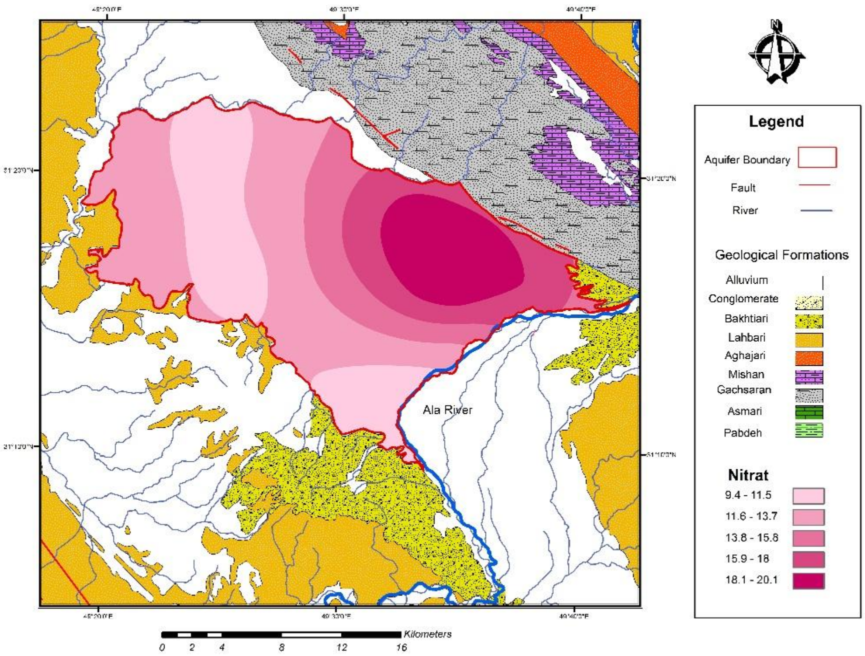

Fig. 6 Nitrate contour map of Ramhormoz plain

4- Conclusions

Based on the clustering results, it can be concluded that the K-Means algorithm, despite its simplicity and high speed, has effectively revealed the hidden qualitative structure in groundwater data and separated areas with high, moderate, and low quality from each other. Detailed analysis of the statistical characteristics of each cluster provides valuable information for water resource management and lays the groundwork for optimal decision-making regarding exploitation, protection, or restoration of damaged aquifers. The existence of clear differences in nitrate and salinity levels among clusters also confirms that the factors affecting water quality in the study area are a combination of natural geological elements and human pressures, and it is necessary for these two aspects to be jointly considered in future management studies. Particularly, distinguishing areas with naturally high salinity but no signs of human pollution from areas with simultaneously high salinity and nitrate plays a key role in future water resource protection policies.

Based on the findings of this study, the following recommendations are proposed for future research and management actions:

Establishing a more complete network of qualitative sampling wells with better distribution across the plain and regular, precise monitoring of chemical changes in the groundwater aquifer for salinity and nitrate assessment.

Creating a statistical database of nitrate and total dissolved solids from the entire plain over a long-term period to investigate qualitative changes in the Ramhormoz plain.

Preparing land use maps of the plain and identifying the contribution of nitrate pollution from agriculture and animal husbandry in the Ramhormoz plain.

Acknowledgments

The authors would like to thank the Khouzestan Water and Power Authority and the Deputy for Comprehensive Water Resources Studies of Khouzestan for their cooperation in providing the data for this article.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors Contribution Statement

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed collaboratively. All authors read and approved the final manuscript.

References

Aliesmaeili, S., & Rajabi Behjat, A. (2016). A novel method for data stream clustering using an algorithm in intrusion detection systems. *First International Conference on Innovation in Science and Technology*. Barcelona, Spain.

Azhar, S. C., Aris, A. Z., Yussof, M. K., Ramli, M. F., & Juahir, H. (2015). Classification of river water quality using multivariate analysis. *Procedia Environmental Sciences*, 30, 79–84.

Dinpajouh, Y. (2022). Clustering of observation wells in the Khoy plain aquifer in terms of water quality using the K-Means method. *Hydrogeology*, 7(1), 25–41.

Firouzi, S., Kalantari, N., & Mohammadi, H. (2023). Integration of hydrochemical and statistical methods for qualitative evaluation of groundwater resources in Zeydoun plain, Khuzestan province. *Aquifer and Qanat*, 4(2), 187–204.

Firouzi, S., Mousavi, M., & Karimi, A. R. (2025). Hydrochemical evaluation of groundwater resources in Ramhormoz plain for determining agricultural and drinking water quality. *First National Conference on Hydrogeology of Iran*. Tabriz.

Javadi, S., Hashemy, S. M., Mohammadi, K., Howard, K. W. F., & Neshat, A. (2017). Classification of aquifer vulnerability using K-means cluster analysis. *Journal of Hydrology*, 549, 27–37.

Kolahdouzan, A., Dinpajouh, Y., Abbaspour, D., & Ghorbanian, M. (2011). Investigation of groundwater quality change trends in Miandoab plain using the Mann-Kendall method. *Water and Soil Science*, 2(25), 221–235.

Liu, Y., Ji, C., Wang, Y., Zhang, Y., Hou, X., & Xie, Y. (2022). Quantifying streamflow predictive uncertainty for the optimization of short-term cascade hydropower stations operations. *Journal of Hydrology*, 605, Article 127376.

Naranjo-Fernandez, N., Guardiola-Albert, C., Aguilera, H., Serrano-Hidalgo, C., & Montero-Gonzalez, E. (2020). Clustering groundwater level time series of the exploited Almonte-Marismas aquifer in southwest Spain. *Water*, 12(4), 1–20.

Sheikhi Beglari, P., Majedi, H., & Daneshian, H. (2024). Zoning and qualitative assessment of groundwater resources in Qaleh Tal plain aquifer using the IRWQI index. *First Conference on Hydrology and Water Resources*. Ahvaz.

Steinley, D., & Brusco, M. J. (2007). Initializing k-means batch clustering: A critical evaluation of several techniques. *Journal of Classification*, 24(1), 99–121.