Journal of Advanced Informatics in Water, Soil, and Structure

journal homepage: wss.torbath.ac.ir

Research Article

Hydrodynamic Analysis and Clustering of Piezometers of Baghmalek Plain Aquifer Using the K-Means Algorithm

Pooria Sheikhy*[1], Hamid Reza Majedi[1], Hasan Daneshian[1], Kazem Hemmadi[1]

[1]*Khuzestan Water and Power Authority*

ARTICLE INFO

Received date: 13 Dec 2025
Accept date: 08 Mar 2026
Published date: 11 Mar 2026

Keywords:
Groundwater, Baghmalek aquifer, Hydrodynamic clustering, K-means algorithm, Orange software

Abstract

Groundwater is essential for drinking, agriculture, and industry, especially in arid regions like Iran. The Baghmalek plain aquifer, covering 49 km² in northeastern Khuzestan, faces serious challenges from over-extraction, climate change, and falling water levels. This study applies the K-Means algorithm to analyze hydrodynamic behavior and cluster piezometers in this aquifer based on water table fluctuations. Water level data from five active piezometers were collected over 144 months (October 2011–September 2023), normalized, and processed in Orange software. The K-Means analysis grouped the piezometers into three distinct clusters. Silhouette index values (0.5–0.685) confirmed good clustering quality, indicating strong hydrodynamic similarity within each group. Long-term hydrographs and groundwater flow direction maps validated these results. These findings support redesigning the groundwater monitoring network and optimizing new piezometer placements. Additionally, identifying zones with similar hydrodynamic behavior can guide the strategic siting of production wells and targeted abstraction planning, promoting more sustainable management of the Baghmalek aquifer.

Homepage: wss.torbath.ac.ir

*Corresponding Author:
Sheiky, Pooria
Email: pooriasheikhy@yahoo.com
ORCID: 0009-0007-7509-2869

https://doi.org/10.22048/wss.2026.588846.1032

How to cite this article:
Sheikhy, P., Majedi, H. R., Daneshian, H., & Hemmadi, K. (2026). Hydrodynamic Analysis and Clustering of Piezometers of Baghmalek Plain Aquifer Using the K-Means Algorithm. *Journal of Advanced Informatics in Water, Soil, and Structure*, *1*(2), 332-338.

1- Introduction

Water is recognized as the most critical resource for human survival on Earth. Surface and groundwater resources are essential factors in meeting human needs, agricultural and industrial uses, ecosystem sustainability, and societal development. Iran, with an average annual precipitation of 250 millimeters, is geographically located in the arid and semi-arid belt (Alizadeh, 2010). Groundwater resources, as one of the most vital components of the hydrological cycle, play a central role in meeting these demands. In semi-arid and arid regions such as Iran, where more than 60 percent of consumed water is supplied from groundwater sources, sustainable management of these resources is critical (Madani, 2014).

Khuzestan Province is among the provinces where most areas exhibit an arid and semi-arid climate. Frequent droughts and population growth in this province have made the need for water resources increasingly apparent, leading to shortages in water supplies, particularly in groundwater resources (Majidi et al., 2024). Given the aforementioned issues, challenges such as overexploitation, pollution, climate change, and water level fluctuations threaten the stability of aquifers. These problems not only lead to declining groundwater levels but also affect local ecosystems and economic development (Gleeson et al., 2012).

Observation wells and piezometers serve as key monitoring tools that enable continuous measurement of groundwater levels and subsequently facilitate the understanding of flow patterns within the aquifer (Freeze and Cherry, 1979). Identification of similar patterns in the hydrodynamic behavior of piezometers can contribute to a better understanding of the hydrogeological structure of the aquifer, the identification of regions with similar hydraulic characteristics, and the optimization of monitoring networks (Kumar et al., 2018). In this context, data mining techniques, and particularly clustering methods, provide powerful tools for the analysis of hydrogeological data.

The K-Means algorithm is one of the most widely used clustering methods that, due to its simplicity, high speed, and interpretable results, has been employed in numerous hydrogeological studies (MacQueen, 1967). This algorithm divides data into K clusters based on similarity, enabling the identification of homogeneous groups of piezometers. Several studies have demonstrated the effectiveness of this method in analyzing groundwater data. For example, Ramos-Leal et al. (2007) utilized the K-Means algorithm to classify groundwater wells based on water quality. Additionally, Moeck et al. (2020) applied this method to identify spatial patterns in groundwater level data.

A review of the literature shows that the application of clustering methods, such as K-means, has become increasingly common in groundwater studies for classifying monitoring wells and delineating hydrodynamic zones. However, the successful application of these methods depends largely on the interpretability of the resulting clusters and their consistency with the hydrogeological setting of the study area. Accordingly, the present study was conducted with the aim of analyzing and clustering the piezometers of the Baghmalek Plain aquifer using the K-means algorithm, in order to gain a better understanding of aquifer behavior and to identify the hydrodynamic patterns of the active piezometers, considering the hydrogeological and hydrodynamic conditions as well as statistical indices. This integrated approach enables the clustering results to be interpreted not only as statistical partitions, but also as physically meaningful hydrodynamic zones, which can provide more practical insight for groundwater management.

2- Case study

The Baghmalek plain, as one of the characteristic alluvial plains in Khuzestan Province, is located in the northeastern part of the province, covering an approximate area of 49 square kilometers. Geographically, the plain extends between longitudes 49° 49' 17" to 49° 56' 11" E and latitudes 31° 30' 38" to 31° 38' 19" N, situated approximately 140 kilometers from the provincial capital, Ahvaz.

In this region, two important and predominantly seasonal rivers, named Gelal and Padrazan, flow. The Gelal River enters the plain from the north and, after joining the Padrazan River in the southern section, merges with the Abbas River. In terms of geographical boundaries, the Baghmalek plain is bordered by the Izeh County to the north and the Ramhormoz County to the south and southwest. The climate of the Baghmalek region is predominantly warm and semi-arid, with an annual average precipitation that is relatively higher compared to other areas of Khuzestan Province.

From a geological perspective, the Baghmalek region is situated in the folded Zagros sedimentary-structural zone, which is recognized as one of the most active tectonic sectors of the Zagros. Stratigraphically, this region is primarily influenced by the Bakhtiari and Gachsaran formations, which surround the Baghmalek plain and play a key role in determining the hydrogeological characteristics of the area. Examination of data from deep production and exploration wells indicates that the main groundwater resources are concentrated in relatively coarse-grained alluvial deposits adjacent to the Bakhtiari formation. These alluvial deposits act as the principal groundwater reservoirs in the region (Fig. 1). The dominant aquifer type in the Baghmalek plain is unconfined; however, in the southern portions as well as some western and southwestern areas, evidence of confined aquifer conditions has been observed.

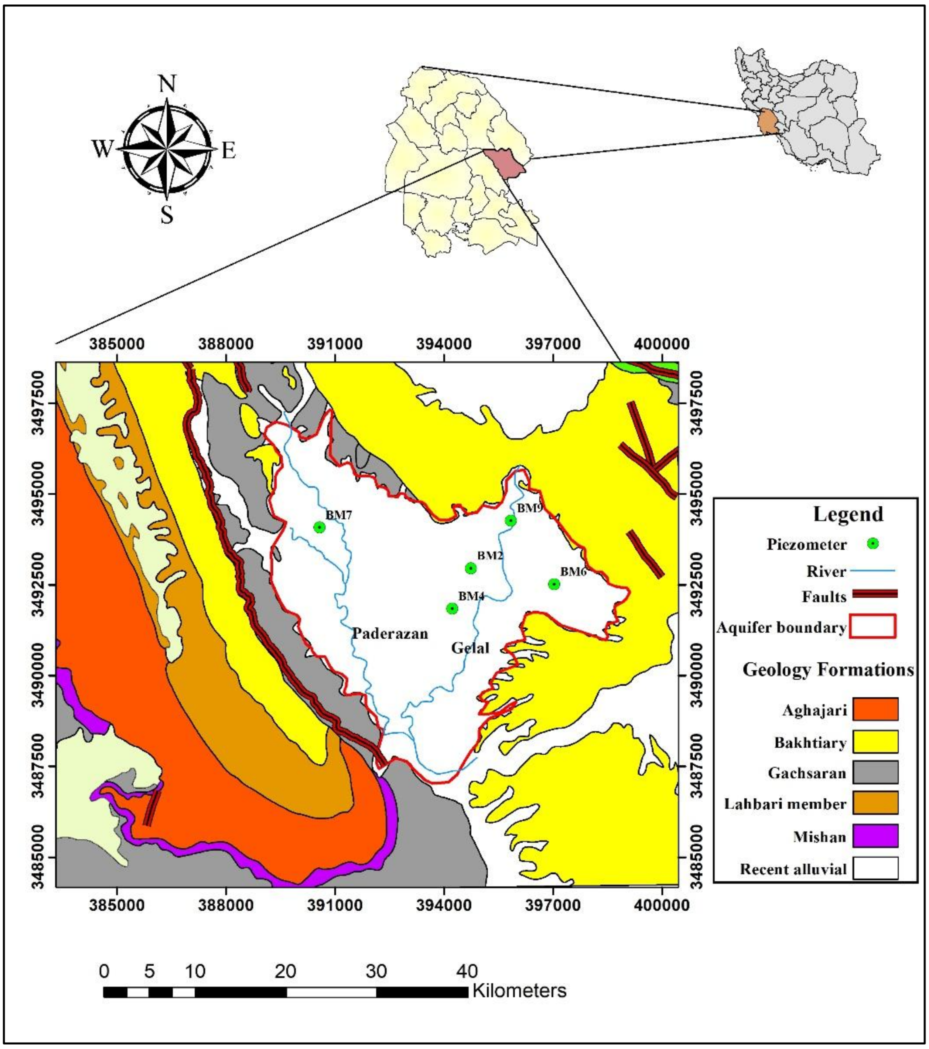

Fig. 1 Geological map and boundary of the Baghmalek plain aquifer

3- K-Means Clustering of Baghmalek Plain Piezometers

Clustering is one of the most important data analysis methods in the field of unsupervised machine learning, whose objective is to group a set of data points based on their similarity or proximity to one another. In this method, data are assigned to different clusters such that members of each cluster exhibit maximum similarity to one another and minimum similarity with members of other clusters. Clustering serves as a powerful tool for identifying hidden patterns, classifying complex data, and reducing data dimensionality (Xu and Wunsch, 2005).

One of the most widely used clustering methods is the K-Means algorithm, which is employed in numerous scientific studies due to its simplicity and high efficiency. The objective of this algorithm is to partition data into K distinct clusters by maximizing intra-cluster similarity and minimizing inter-cluster similarity. By assigning each data point to the nearest cluster center and updating the cluster centers, the algorithm seeks to minimize the sum of squared distances from data points to their corresponding cluster centers. The objective function of the algorithm is expressed as:

$$J = \sum \sum \|pi - ck\|^2 \quad \text{Eq} \quad (1)$$

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

In order to identify similar behavioral patterns among the active piezometers installed in the Baghmalek plain aquifer, the clustering method and K-Means algorithm were employed. To this end, monthly water table level data of the piezometers over 144 months (from October 2011 to September 2023), comprising 12 water years, were collected and, after normalization, analyzed using the K-Means algorithm in the Orange software environment (Fig. 2). As previously mentioned, the objective of this process was to separate the piezometers based on the trend and fluctuations of the water table and to identify groups with similar hydrodynamic behavior.

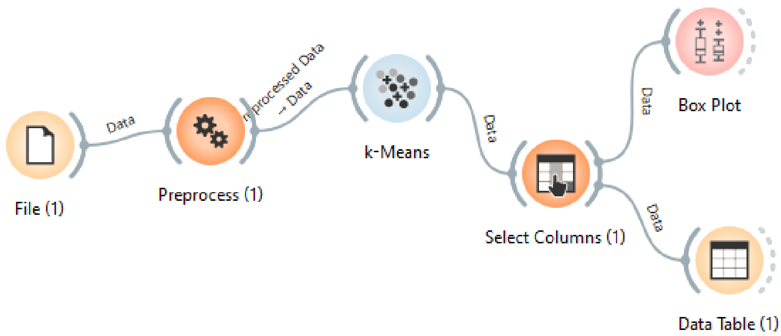

Fig. 2 Workflow of the clustering procedure and K-Means algorithm in the Orange software environment

4- Results and Discussion

The clustering and segregation of the Baghmalek plain piezometers resulted in the classification of the target piezometers into 3 distinct clusters (Table 1). Based on Table 1, the piezometers of the Baghmalek aquifer were divided into three clusters: C1, C2, and C3, where piezometers Bm-6 and Bm-9 were placed in cluster C1, piezometers Bm-2 and Bm-7 in cluster C2, and piezometer Bm-4 in cluster C3.

The highest silhouette index value corresponds to piezometer Bm-9 in cluster C1 with a value of 0.685, indicating very good agreement of this piezometer with its cluster and appropriate distinction from other clusters. Additionally, piezometers Bm-6 and Bm-2, with silhouette values of 0.654 and 0.621 respectively, demonstrate satisfactory clustering quality. This indicates that these samples have been correctly assigned to their clusters and exhibit high similarity with other members of their respective clusters.

In contrast, the lowest silhouette value corresponds to piezometer Bm-4 in cluster C3 with a value of 0.5. Although this value is still positive and acceptable, it indicates that, compared to other samples, this piezometer is somewhat on the boundary between clusters and its distinction from other clusters is less pronounced. Nevertheless, the absence of negative silhouette values among the samples indicates the lack of misclassified samples and confirms the overall validity of the results.

One of the methods for examining the fluctuation and hydrodynamics of piezometers is the construction of detailed hydrographs for each piezometer over different time intervals. Following the clustering of the Baghmalek plain piezometers using the K-Means algorithm and the analysis of the silhouette index for each cluster, detailed long-term hydrographs were constructed for each piezometer in order to investigate their hydrodynamic behavior. Figures 3 through 7 show the detailed long-term hydrographs of piezometers Bm-6, Bm-9, Bm-2, Bm-7, and Bm-4, respectively. Based on the constructed detailed hydrographs, piezometers that were placed in similar clusters by the K-Means algorithm exhibit identical water table fluctuation patterns, and the decline and rise behavior of each piezometer closely resembles that of the corresponding piezometer within each cluster. This finding provides confirmation of the hydrodynamic clustering of the piezometers based on the K-Means algorithm.

Groundwater flow direction is one of the most important hydrogeological parameters in examining aquifer behavior. Naturally, any pattern or hydrodynamic behavior exhibited by any of the piezometers in the aquifer has a significant impact on the groundwater flow direction. Any change in the water level of piezometers can indicate a change in the direction or intensity of groundwater flow. In other words, the hydrodynamic analysis of piezometers not only reveals the dynamic behavior of the aquifer but also enables accurate mapping and interpretation of the groundwater flow direction. Therefore, piezometers that exhibit different hydrodynamic behavior cannot display identical and similar flow directions.

Table 1 Clustering of observation wells in the Baghmalek aquifer based on the K-means algorithm

Piezometer	Cluster	Silhouette Index
Bm-2	C2	0.62
Bm-4	C3	0.50
Bm-6	C1	0.65
Bm-7	C2	0.54
Bm-9	C1	0.68

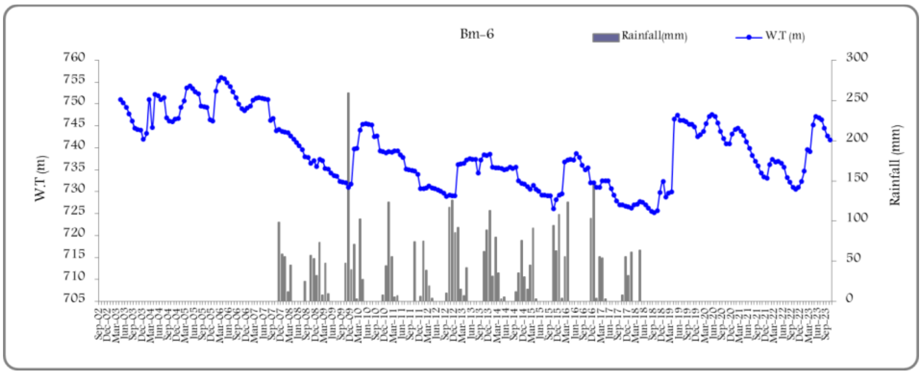

Fig. 3 Detailed hydrograph of piezometer Bm-6

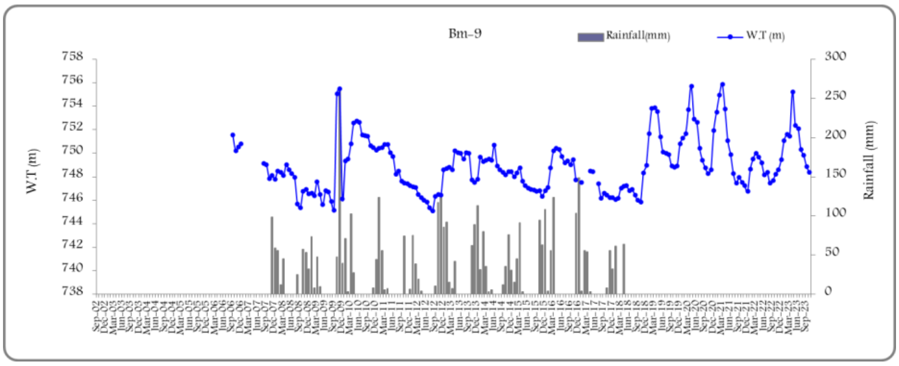

Fig. 4 Detailed hydrograph of piezometer Bm-9

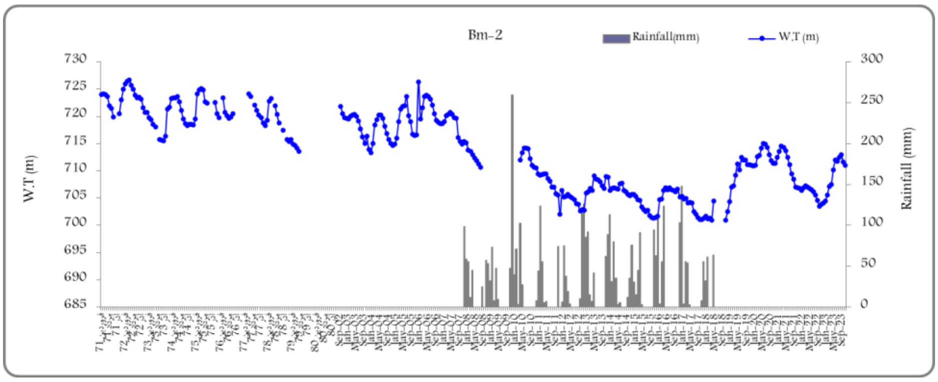

Fig. 5 Detailed hydrograph of piezometer Bm-2

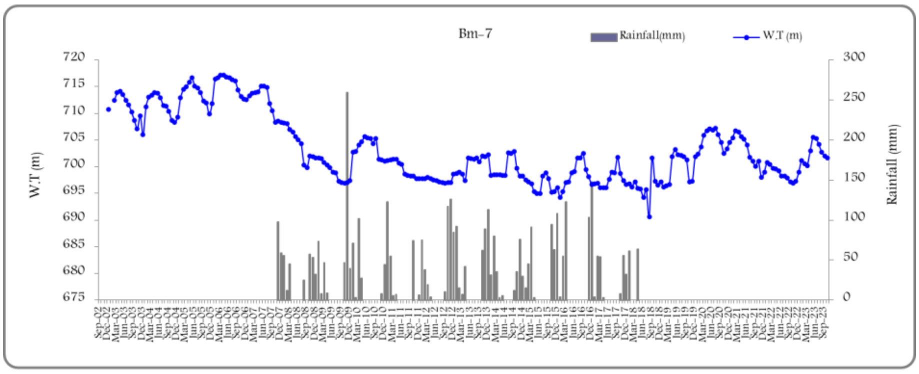

Fig. 6 Detailed hydrograph of piezometer Bm-7

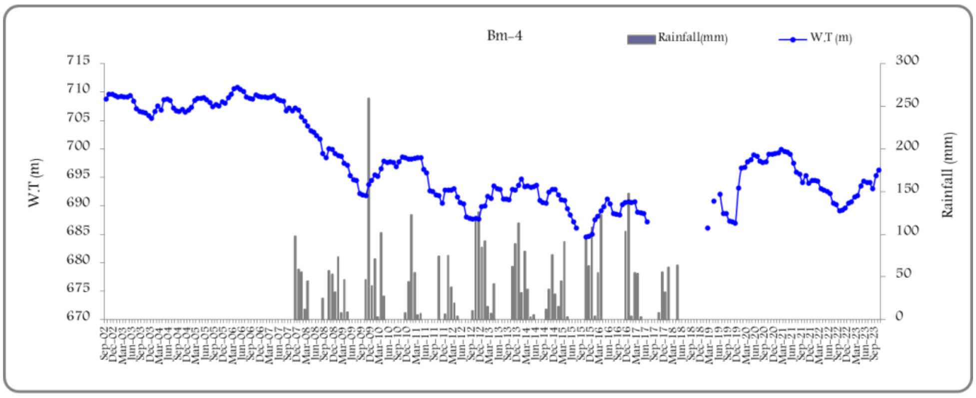

Fig. 7 Detailed hydrograph of piezometer Bm-4

Accordingly, in order to verify the clustering performed by the K-Means algorithm, the groundwater flow direction map of the Baghmalek aquifer was also constructed (Fig. 8). Based on the flow direction map of the Baghmalek aquifer, piezometers that have been placed in the same cluster exhibit the same flow direction. This uniform pattern in flow direction also serves to confirm the validity of the performed clustering.

To independently validate the clustering results and evaluate the hydrodynamic behavior of the Baghmalek aquifer, linear regression analysis was performed on the groundwater-level time series of the piezometers over the monitoring period. The regression model was defined as $y_t=a+bt$, where b represents the monthly slope of water-level change and the p-value indicates the significance level of the hypothesis test for the existence of a trend. The results of this analysis are presented in Table2.

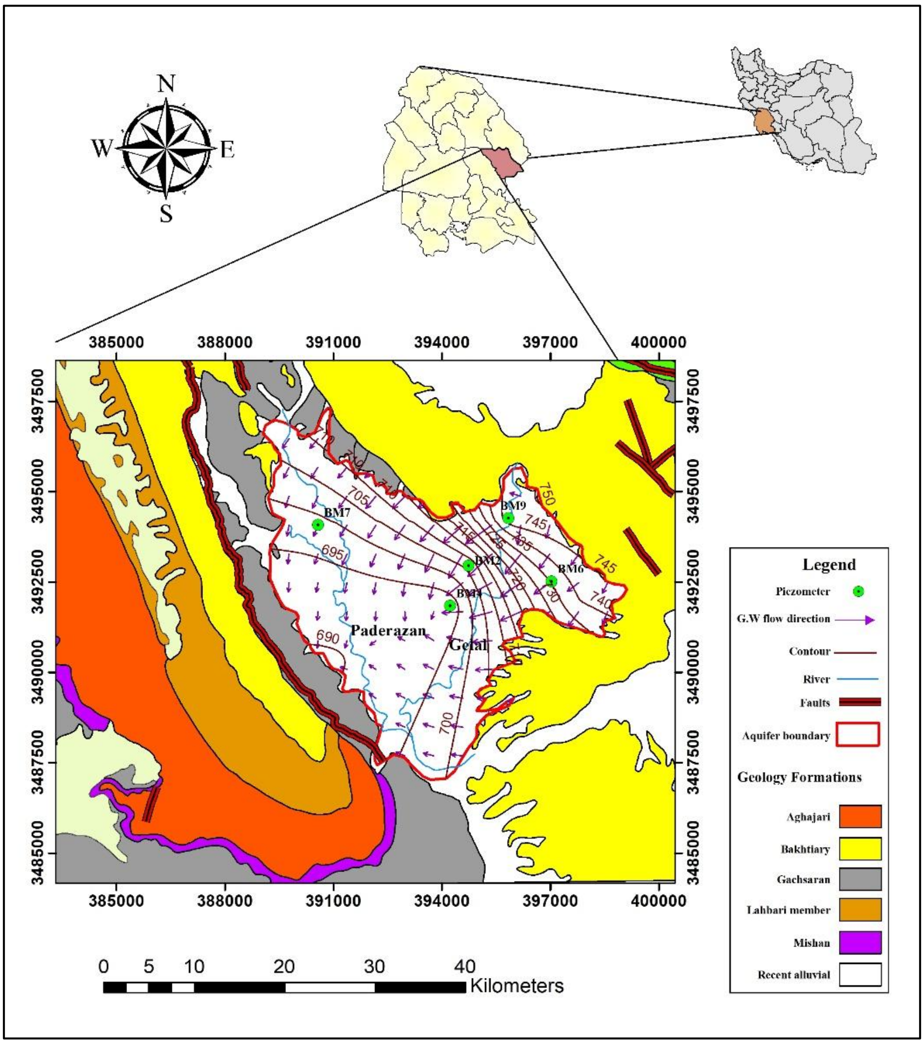

Fig. 8. Groundwater flow direction map of the Baghmalek plain aquifer

Table 2. Linear regression parameters and trend analysis of the Baghmalek aquifer piezometers

Piezometer	N (months)	Monthly slope(b)	Intercept (a)	R^2	p-value
Bm-2	144	0.0857	703.61	0.291	2.4×10^{-8}
Bm-4	144	0.0835	687.63	0.286	3.4×10^{-8}
Bm-6	144	0.1074	731.4	0.173	3.4×10^{-5}
Bm-7	144	0.0613	697.14	0.206	4.8×10^{-6}
Bm-9	144	0.0308	747.98	0.116	8.3×10^{-4}

As shown in Table 2, all piezometers exhibit a mild upward trend that is statistically highly significant ($p<0.001$). Piezometers Bm-2 and Bm-4, which belong to the same cluster, show very similar trend slope values (0.0857 and 0.0835 m/month, respectively) and comparable coefficients of determination. In contrast, piezometer Bm-9 has the lowest upward slope rate (0.0308 m/month), indicating its location in a hydrodynamically distinct zone. This strong correspondence between the linear trend parameters and the extracted clusters independently confirms the validity of the clustering results.

5- Conclusions

The present research was conducted with the aim of analyzing and identifying similar hydrodynamic patterns among the active piezometers in the Baghmalek plain aquifer using the K-Means clustering algorithm. Using monthly water table level data over 12 water years (October 2011 to September 2023) and after data normalization, three clusters (C1, C2, and C3) were identified as the optimal clusters for the piezometric wells. The following results were obtained from the clustering of the Baghmalek plain piezometers:

1. The silhouette index analysis revealed that all clusters fall within the acceptable range (0.5 to 0.68). The values of this index indicate that the cluster boundaries are appropriate and that the hydrodynamic separation of the piezometric wells in the studied aquifer has been performed correctly.
2. To confirm the hydrodynamic clustering of the piezometers by the K-Means algorithm, detailed long-term hydrographs were constructed for each piezometer, revealing that piezometers within the same cluster exhibit similar water table fluctuations.
3. By constructing the groundwater flow direction map, it was found that piezometers within the same cluster also share the same flow direction. This finding demonstrates a good agreement between the clustering results and the hydrogeological conditions of the region.

Overall, the clustering results indicate that piezometers placed within the same cluster exhibit similar behavior in terms of water table variations over time, and the data have been effectively divided into three distinct clusters.

4. The linear regression analysis revealed that groundwater levels in all piezometers exhibit a mild upward trend (ranging from 0.03 to 0.11 m/month) that is statistically highly significant ($p<0.001$). The physical correspondence of the clusters with the groundwater flow direction, along with their statistical consistency with the regression trend-slope parameters, demonstrates the effectiveness of this framework in validating the spatiotemporal patterns of groundwater systems under data-scarce conditions.

Therefore, the K-Means algorithm has been able to identify an appropriate structure in the data. This finding can be utilized in the process of upgrading the groundwater monitoring network and designing locations for new piezometer installations. Furthermore, the hydrodynamic clustering from this research leads to the segregation of the aquifer into hydrodynamically uniform zones, which enables more accurate siting for the drilling of new production wells and the determination of optimal abstraction areas.

Acknowledgments

The authors would like to acknowledge the Khouzestan Water and Power Authority and the Department of Comprehensive Water Resources Studies of Khouzestan for their cooperation in providing the data used in this paper.

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. Data collection, analysis, and manuscript preparation were performed collaboratively. All authors read and approved the final manuscript.

References

Alizadeh, A. (2010). *Principles of Applied Hydrology* (28th ed.). Astan Quds Razavi Publications. (In Persian)

Freeze, R.A., & Cherry, J.A. (1979). *Groundwater*. Prentice-Hall, Englewood Cliffs, NJ, 176, 161-177.

Freeze, T., Wada, Y., Bierkens, M.F., & Van Beek, L.P. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature, 488*(7410), 197-200.

Kumar, S., Venkatesh, A.S., Singh, R., Udayabhanu, G., & Saha, D. (2018). Geochemical signatures and isotopic systematics constraining dynamics of fluoride contamination in groundwater across Jamui district, Indo-Gangetic alluvial plains, India. *Chemosphere, 205*, 493-505.

MacQueen, J. (1967). Multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297).

Madani, K. (2014). Water management in Iran: what is causing the looming crisis? *Journal of Environmental Studies and Sciences, 4*(4), 315-328.

Majidi, H., Sheikhi Baghlari, P., & Daneshian, H. (2024). Predicting groundwater level changes in the Baghmalek plain aquifer using the MODFLOW model in response to climate change and groundwater abstraction. *First Conference on Hydrology and Water Resources, Ahvaz*. (In Persian)

Moeck, C., Grech-Cumbo, N., Podgorski, J., Bretzler, A., Gurdak, J.J., Berg, M., & Schirmer, M. (2020). A global-scale dataset of direct natural groundwater recharge rates: A review of variables, processes and relationships. *Science of the Total Environment, 717*, 137042.

Ramos-Leal, J.A., Martínez-Ruiz, V.J., Rangel-Mendez, J.R., & Alfaro De La Torre, M.C. (2007). Hydrogeological and mixing process of waters in aquifers in arid regions: a case study in San Luis Potosi Valley, Mexico. *Environmental Geology, 53*(2), 325-337.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks, 16*(3), 645-678.