



Integrated clustering analysis for delineating seawater intrusion and heavy metals in Arabian Gulf Coastal groundwater of Saudi Arabia

Mohammed Benaafi^a, S.I. Abba^{a,*}, Bassam Tawabini^b, Ismail Abdulazeez^a, Billel Salhi^a, Jamilu Usman^a, Isam H. Aljundi^{a,c}

^a Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

^b College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

^c Department of Chemical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:

Seawater intrusion
Clustering analysis
Groundwater
Arabian gulf
Sustainable groundwater management

ABSTRACT

The intrusion of seawater (SWI) into coastal aquifers is a major concern worldwide, affecting the quantity and quality of groundwater resources. The region of Saudi Arabia that lies along the eastern coast has been affected by SWI, making it crucial to accurately identify and monitor the affected areas. This investigation aimed to map the degree of seawater intrusion in a complex aquifer system in the study area using an integrated clustering analysis approach. The study collected 41 groundwater samples from wells penetrating multi-layered aquifers, and the samples were analyzed for physicochemical properties and major ions. Clustering analysis methods, including Hierarchical Clustering Analysis (double-clustering) (HCA-DC), K-mean (KMC), and fuzzy k-mean clustering (FKM), were employed to evaluate the spatial distribution and association of the groundwater properties. The results revealed that the analyzed GW samples were divided into four clusters with varying degrees of SWI. Clusters A, B, C, and D contained GW samples with very low (f_{sea} of 1.9%), high (f_{sea} of 14.9%), intermediate (f_{sea} of 7.9%), and low (f_{sea} of 5.2%) degrees of SWI, respectively. FKM clustering exhibited superior performance with a silhouette score of 0.83. Additionally, the study found a direct correlation between the degree of SWI and increased concentrations of boron, strontium, and iron, demonstrating SWI's impact on heavy metal levels. Notably, the boron concentration in cluster B, which endured high SWI, exceeded WHO guidelines. The study demonstrates the value of clustering analysis for accurately monitoring SWI and associated heavy metals. The findings can guide policies to mitigate SWI impacts and benefit groundwater-dependent communities. Further research can help develop effective strategies to mitigate SWI effects on groundwater quality and availability.

1. Introduction

Freshwater is an invaluable resource for the planet, and groundwater (GW) is a crucial source of this precious commodity. Groundwater plays a vital role in meeting the rising demand for water, particularly in dry areas with little rainfall, high evaporation,

* Corresponding author.

E-mail address: sani.abba@kfupm.edu.sa (S.I. Abba).

<https://doi.org/10.1016/j.heliyon.2023.e19784>

Received 21 May 2023; Received in revised form 30 August 2023; Accepted 31 August 2023

Available online 1 September 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and minimal surface water [1–3]. The rapid growth of population, urbanization, and industrialization, particularly in coastal areas, has resulted in a shortage of water resources that are no longer suitable for human use. The over-extraction of GW beyond its natural recharge rate results in hydrodynamic and hydrochemical imbalances, which subsequently causes seawater intrusion into aquifer system in coastal areas and elevates the GW salinity. This phenomenon poses a significant threat to the sustainability and availability of freshwater resources in these regions, with implications for both human populations and the environment [4–6].

The salinization of GW represents a significant hazard to the availability and quality of GW resources in coastal areas. It can reduce the availability and quality of GW suitable for drinking and irrigation purposes, negatively impacting the sustainability and resilience of the affected regions. GW resources in numerous coastal cities across the globe are under threat of salinization due to factors such as global warming, SWI, and seawater outflow [7–12]. The increased water demand in coastal regions due to humankind's development has led to severe consequences. The unsustainable over-extraction of GW resources can create hydrochemical and hydrodynamic imbalances, resulting in SWI of coastal regions, leading to the salinization of coastal freshwater resources. This can make it unsuitable for drinking and irrigation, underscoring the need for effective management practices [5]. Global warming further exacerbates the situation which intensifies SWI and exacerbates the risk of GW salinization in coastal areas worldwide [13–15].

SWI is reported as the major risk to GW in more than 501 coastal cities around the world [8,16–18]. Mapping the SWI and GW salinization degree is essential for sustainable GW resource management. There are various techniques documented in literature for detecting SWI in coastal aquifers, including hydrogeology, hydrochemistry, isotopes, and geophysics [19–21]. Researchers have employed multivariate statistical analysis of physical and chemical GW data to evaluate the effect of SWI on coastal aquifers [20,22,23]. Principal component analysis (PCA) and factor analysis are the most used multivariate statistical techniques in SWI studies [19,24,25]. Telahigue et al. [25] utilized PCA as a multivariate statistical technique to identify the main hydrochemical processes affecting the GW chemistry and quality in the southeastern region of Tunisia. Sae-Ju et al. [19] implemented a PCA with hydrochemical facies evolution (HFE) diagram to understand the freshening and intrusion processes in a coastal aquifer in Thailand. Hajji et al. [24] used a PCA model with a hydrochemical facies evolution (HFE) model to determine the most influential factors on GW quality in coastal aquifers in Tunisia. Some scientists have used hierarchical cluster analysis (HCA) in combination with principal component analysis (PCA) to determine the geochemical processes that govern the geochemistry and water quality of GW in coastal regions (e.g., Refs. [10,26]).

This study aims to use HCA-DC, KMC, and FKM clustering methods to determine the extent of SWI in a multi-aquifers system of the Arabian Gulf Coast of Saudi Arabia. The study considers the spatial variability of physicochemical properties and major ions to identify hotspot zones of SWI and support effective management strategies. To our knowledge, it is the first time to combine various clustering techniques to delineate and map the extent of SWI in coastal GW systems. The results of this research can aid decision-makers in effectively managing GW resources and implementing strategies to manage SWI sustainably.

2. Study area

The area under investigation is situated on the west coast of the Arabian Gulf, as shown in Fig. 1, falls within the Qatif province of

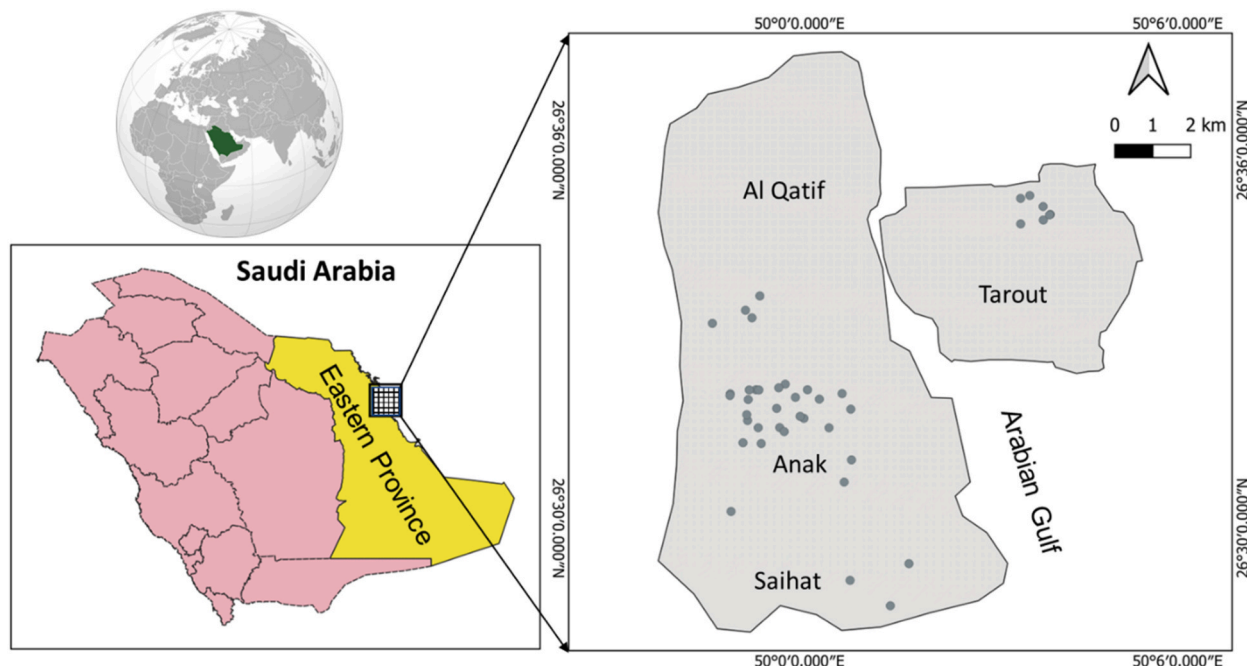

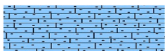


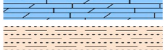




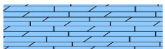
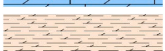
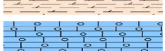
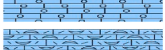
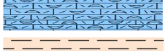



Fig. 1. Map depicts the geographical location of the study area and the selected GW samples.

eastern Saudi Arabia, and encompasses Tarout island and the Qatif coastal oasis. Qatif governorate, which borders Dammam governorate to the south and Al-Jubail governorate to the north, is a developed coastal oasis. The area being studied is a flat coastal plain that gradually increases in elevation towards the west, reaching a height of approximately 12 m. The region under investigation encompasses residential and agricultural zones, with recent urban expansion on farmland. The primary water source for irrigation in Qatif and Tarout Island is the multi-layered aquifer systems, which consist of shallow, middle, and deep aquifers [27,28]. In the studied region, the intermediate and deep aquifers are predominantly utilized for GW extraction due to the high salinity levels of groundwater in the shallow aquifer [29]. Within the study area, numerous wells reach the middle and deep aquifers, which are utilized for agricultural water supply. The aquifers can be categorized into three types: shallow, intermediate, and deep. The maximum depths of these aquifers are 30 m, 80 m, and 130 m below the ground surface, respectively. The water level varies across these aquifers, ranging from less than 1 m in the shallow aquifer to over 5 m in the deep aquifer.

The climate of eastern Saudi Arabia is characterized by hot summers and mild winters. The region experiences a desert climate, with very low rainfall throughout the year, mainly during winter [30]. The average temperature in the region ranges between 20 and 45 °C, with the highest temperatures typically recorded during the summer months. The region is also influenced by the Shamal winds, which are hot and dry winds blowing from the northwest direction. The climate of eastern Saudi Arabia poses a significant challenge for water management, as the region faces increasing demand for freshwater resources and limited surface water sources. The study area's vulnerability to SWI highlights the importance of sustainable water management practices to mitigate the impact of climate on GW resources.

Table 1
Lithological and hydrogeological units in Saudi Arabia's eastern coastal region (after [32,34]).

Formation	Member	Lithology	Description	Thickness (m)	Hydrogeologic Unit
Quaternary Deposits			Eolian Sands and Sabkha	0–10	
Hofuf			Sandy Limestone	0–95	Neogene Aquifer
Dam			Massive Limestone	0–90	
			Dolomitic Limestone		
Hadruk			Sand and Shale	0–90	
			Shale		
			Sandstone		
			Marl		Marl Aquitard
Dammam	Alat		Dolomitic Limestone	0–80	Alat Aquifer
			Dolomitic Shale	0–30	Alat Aquitard
	Khobar		Limestone	0–70	Khobar Aquifer
	Alevolina		Fossiliferous Limestone	0–20	Limestone Aquitard
	Saila Midra		Shale with limestone	0–10	Shale Aquitard
Rus			Marl and chalky limestone	60–100	Shale Aquitard
Umm Er Radhuma (UER)			Dolomitic Limestone	200–400	UER Aquifer

3. Geological and hydrological characteristics

The geology of the study area is illustrated in Table 1. The area’s geological composition consists of seven formations with varying ages and lithology. The oldest one is Umm Er Radhuma Formation, which spans a thickness of 400 m and primarily consists of dolostone, limestone, and dolomitic limestone [31]. The Dammam Formation is approximately 300 m thick and mainly composed of limestone, chalky, and dolomitic limestone. It is one of the most significant aquifers in the region, and its water is typically of good quality and low salinity [32,33]. The Dammam Formation is located above the Rus Formation in a geological sequence. The Dammam Formation is positioned above the Rus Formation in the geological sequence. It comprises two shale members; Midra, Saila, and three limestone members; Alveolina, Khobar, and Alat. These members are arranged in ascending order from the bottom to the top. Midra and Saila consist of shale and are 10 m thick on average [34]. The Alveolina member, on the other hand, is composed of microcrystalline limestone and ranges in thickness from meters to 12 m [34]. The Dammam Formation’s Khobar Member is mainly made up of porous limestone and has a thickness of approximately 40 m [33,35]. The upper section of Dammam Formation which consists of porous dolomitic limestone formed the Alat member of the formation with thickness ranging from 15 m in the exposed area to approximately 70 m underground [34].

The Hadruk Formation is a layer of sandstone, shale and marl that is overlain by the Dammam Formation. It can range in thickness from a few meters to 90 m [32]. The Dam Formation, which sits atop the Hadruk Formation, is consist of limestone rock varied in facies from microcrystalline to massive limestone with voids The thickness of Dam Formation ranges from 40 m in the exposed section to 90 m in the sedimentary basin in the subsurface [34]. The Dam Formation is overlain by the clastic and sandy limestone layers of Hofuf Formation. Quaternary deposits, such as Sabkha, aeolian sand, and marine terraces, can be found along the Arabian Gulf shore and typically range in thickness from 3 to 10 m [34].

The research site has a complex karstified aquifer system with three main aquifers - Neogene, Alat, and Al-Khobar, separated by two aquitards, as shown in Table 1. The Neogene aquifer developed in the clastic layers of the Hadruk Formation is shallow and unconfined in the studied area. It is estimated to have a storativity of 0.01 and a transmissivity range of 25.5–144 m²/h [27,36]. The middle aquifer in the study area is The Alat Aquifer and it formed in the dolomitic limestone section of the Dammam Formation. The aquifer depth varies greatly, ranging from a few meters at higher elevations to over 100 m in Ras Tanura, with an average depth of 25 m. From regional prospect, the aquifer transmissivity ranges from 1.16 m²/h to 310 m²/h, and storativity range from 0.00013 to 0.00002. However, in the study area, it has an approximate transmissivity of 12.96 m²/h [36].

The Khobar Aquifer is formed within the lower section of the Dammam Formation, comprised of karstified and fissured dolomitic limestone. Its top boundaries are the Alat member shale aquitard and other Dammam Formation shale units. It is the main source of drinking and agricultural water in Hassa and Qatif areas. The hydraulic characteristics of the aquifer range from 0.02 to 324 m²/h for

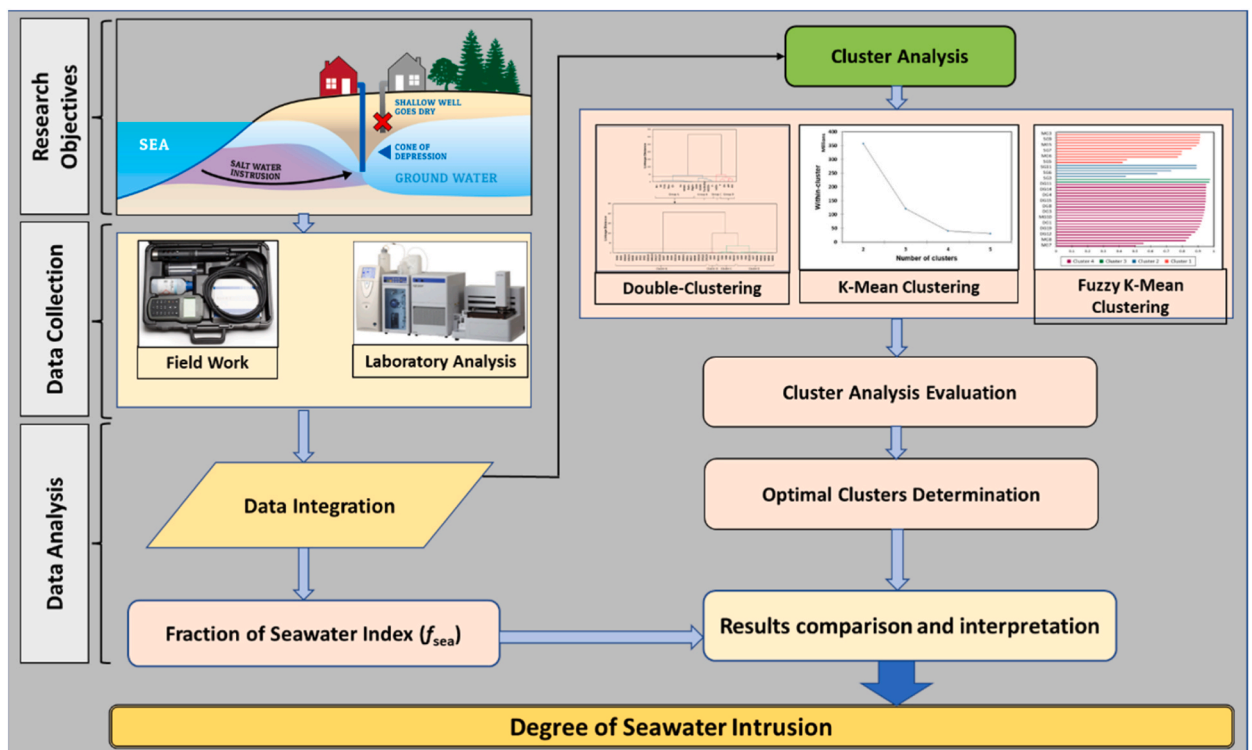


Fig. 2. A flowchart outlining the proposed methodology.

transmissivity and 0.001 to 0.0001 for storativity in the eastern province of Saudi Arabia. The transmissivity of The Khobar Aquifer is about $312 \text{ m}^2/\text{h}$ in the study area [36].

The distribution of the potentiometric level of the aquifers in the study area has been studied by Benaafi et al. [37]. They found that the water level within the shallow aquifer display decline towards the east, northeast, and southeast in the direction of the Arabian Gulf. Similarly, the water level in the deep aquifer also declines towards the Arabian Gulf, but with a noticeable upconing of GW levels observed in the central region of the study area. This rise in the water level can be attributed to the reduction of GW pumping from the deep aquifer in the middle part of the study area. This decrease in GW extraction is due to substituting groundwater with treated wastewater for irrigation purposes.

4. Materials and methods

The proposed methodology for this study follows a systematic and comprehensive approach to assess the degree of seawater intrusion in the research area (Fig. 2). The process begins with identifying research objectives and the design of the study, followed by data collection through fieldwork and laboratory analysis. The gathered data then undergoes a thorough analysis and interpretation phase, integrating various sources of information to ensure a holistic understanding of the issue. A key aspect of the methodology involves implementing cluster analysis, which includes double-clustering, K-Mean clustering, and Fuzzy K-Mean clustering techniques. These clustering methods serve to group data points based on their similarities and provide valuable insights into the spatial patterns of seawater intrusion. Following the clustering process, the results are evaluated to determine the optimal number of clusters that best represent the data.

A Fraction of Seawater Index (f_{sea}) is employed to further refine the analysis, allowing for a more accurate and quantitative determination of the extent of seawater intrusion. Finally, the study compares and interprets the results of the various clustering methods and the f_{sea} index to provide a comprehensive understanding of the degree of seawater intrusion in the area. This methodology, with its multi-faceted approach, ensures a robust and reliable assessment of seawater intrusion, allowing for the development of informed strategies to mitigate its impacts. The cluster analysis results were compared with the conventional index of SWI (fraction of seawater index (f_{sea})) to have an effective statistical technique for SWI mapping.

4.1. Water sampling

In March 2021, a total of 41 wells that served as sources of water for both human consumption and irrigation in Qatif City and Tarout Island, eastern province of Saudi Arabia, were sampled, as depicted in Fig. 1. The GW samples were obtained from wells tapped in the shallow and deep aquifer in the study area. The shallow wells ranged in depth from 5 m to 10 m. And the depth of the production wells ranges from 70-m to 130-m and produce water from the deep aquifers [27,37]. Water samples for GW were obtained from the wells once the electrical conductivity (EC) levels had stabilized after pumping—the sampling process adhered to USEPA's guidelines [38].

On-site measurements of water physical and chemical parameters (temperature, pH, EC, ORP, and DO) were performed using a portable meter equipped with GPS (Hanna HI9829). This was followed by collecting GW samples in high-density bottles and filtered on site using $0.45 \mu\text{m}$ membranes. The filtered samples were subsequently utilized to create two distinct 250 ml sampling kits to analyze anions and major cations, and trace elements. The GW samples were acidified with ultrapure HNO_3 (e.g., concentrated, enough to achieve a pH of less than 2) for the analysis of main cations and trace elements. Unfiltered samples (100 ml of unfiltered samples were used to study the water isotopes (^{18}O and ^2H)). For further analysis, all the samples were stored at 4°C .

4.2. Laboratory analysis

At the Environmental Laboratories of King Fahd University of Petroleum and Minerals, high-performance ion chromatography (IC) was utilized to analyze major ions (Na^+ , K^+ , Ca^{2+} , Mg^{2+} , Cl^- , F^- , Br^- , SO_4^{2-} , and NO_3^-). The laboratory measurement of bicarbonate carried out via acid titration and TDS through gravimetric analysis (TDS). Quality control was ensured by analyzing blind, duplicate, and blank GW samples to validate the accuracy and calibration of the equipment. Cross-checking was performed to verify the ionic charge balance of all GW samples (equation (1)) [39].

$$\text{charge balance}(\%) = \frac{\sum \text{meq Cations} - \sum \text{meq Anions}}{\sum \text{meq Cations} + \sum \text{meq Anions}} \quad (1)$$

The charge balance for the GW samples ranged from 0.3% to 5%, and an average value of 3.5%. Notably, all the GW samples exhibited a charge balance that was less than the accepted criterion of 10%. Stable isotopes (^{18}O and ^2H) were sampled and tested following Clark's protocol [40]. The primary analysis of ^{18}O and ^2H was conducted at the Stable Isotope Laboratory of King Fahd University and Minerals, Saudi Arabia. The isotope analysis results were expressed in per mill (‰) compared to Vienna Standard Mean Ocean Water (V-SMOW), with a precision of $\pm 0.2\%$.

Heavy and trace metals in tested groundwater (B, Al, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Cd, Ba, and Pb) were measured using ICP-MS (US EPA 6020 method). The water samples were filtered using 0.45 m filter paper to remove the suspended particles and acidified by nitric acid to decrease the chemical variance. The quality of elemental analysis was monitored via reference standard materials and duplicates to ensure the reliability of the results. Analytical results were accurate to within 5% RSD.

5. Statistical analysis

5.1. Hierarchical cluster analysis (HCA)

GW studies have been widely used Hierarchical Cluster Analysis (HCA) to identify the hierarchical grouping level of water samples based on differences or similarities in their characteristics [41–43]. This statistical approach appears unsuitable for showing the group of parameters (e.g., physicochemical properties and ions) responsible for grouping water samples into classes with various degrees of seawater intrusion. To achieve the study's objective of delineating the degree of salinization and SWI, we applied a double-clustering approach to identify clusters of water samples that belong to one group of variables (the group that reflects salinity). In the first stage, we applied HCA to all datasets of the physicochemical properties of the tested GW. Accordingly, cluster analysis generates four groups of variables. The variables most relevant to salinity (EC, TDS, Na, Ca, Mg, K, Cl, Br, and SO_4) were used as input variables for grouping the observations (GW samples) into clusters with different degrees of SWI. The HCA-DC approach has been evaluated by Mora et al. (2021) for mapping the salinization and trace metal contamination in north-central Mexico. They found that the HCA-DC is a reliable statistical tool for mapping the contaminant hotspot zones and the process that controlled them.

5.2. K-mean clustering method

K-means clustering algorithm is a largely applied technique for clustering huge data. It is sophisticated non-supervised, and non-hierarchical algorithm [44]. In this approach, the data is separated into subgroups and sorted into clusters based on their centroids, resulting in a centroidal clustering technique. The goal of clustering is to reduce the MSSD of each point to its closest cluster mean while ensuring that members of each group share similar characteristics and attributes. According to Han et al. [45] the K-means clustering algorithm consists of four main steps: 1) Cluster number identification, 2) Finding the center point of each cluster, 3) measuring the distance between each data point and its closest centroid, and 4) assigning each data point to the nearest centroid. The algorithm repeats these steps until the position of the cluster centroids no longer changes. The cluster centroids are computed using equation (2) [46].

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

In which μ_k represents the new centroid, N_k denotes the number of instances, and x_q refers to the q-th data in the k-th cluster.

5.3. Fuzzy K-mean cluster analysis

The fuzzy K-means algorithm is an unsupervised, non-hierarchical, and soft clustering technique that divides the data into k clusters based on fuzzy logic conceptualization [47]. It was initially proposed by Dunn [47]. and improved and generalized by Bezdek [48]. The FKM clustering algorithm is implemented by determining the cluster center and then splitting the dataset into clusters by applying equation (3) to minimize the objective function [48].

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ki} \quad (3)$$

where J , c , n , u_{ik} , m , stands for minimized objective function, number of clusters, number of data observations, membership degree of sample k within cluster i , exponential weight of the degree of fuzziness, and overlapping degree between clusters. d_{ki} represents the Euclidean distance between sample x_k and cluster center v_i .

5.4. Silhouette coefficient

Numerous studies within the literature have indisputably confirmed the significance of the silhouette coefficient as a crucial metric to verify the effectiveness of the clustering techniques and determining the optimum clusters [49–51]. The coefficient range is between -1 and 1 , where a high positive value indicates a well-performing clustering analysis. In contrast, negative values imply that the GW samples are not belong to that cluster—the formula used to calculate the silhouette coefficient illustrated in equation (4) [49].

$$S(i) = \frac{b(i) - a(i)}{\text{Max}(a(i), b(i))} \quad (4)$$

The formula for the silhouette coefficient involves the use of $b(i)$. This refers to the average difference between a specific data point (i) and all the other data points in the same cluster. On the other hand, $a(i)$ refers to the lowest average dissimilarity between instance (i) and any other cluster.

5.5. Seawater intrusion index

The fraction of the seawater index (f_{sea}) is a widely used for quantifying the degree of SWI in GW systems. The method proposed by

Table 2

Descriptive statistics of physicochemical variables, major ions, and isotopes.

Statistic	Temperature (°C)	pH	DO mg/L	Eh mV	Turbidity NTU	EC µS/cm	TDS mg/L	Na ⁺ mg/L	Ca ²⁺ mg/L	Mg ²⁺ mg/L	K ⁺ mg/L	Cl ⁻ mg/L	Br ⁻ mg/L	SO ₄ ²⁻ mg/L	HCO ₃ ⁻ mg/L	F ⁻ mg/L	NO ₃ ⁻ mg/L	δ ² H [‰]	δ ¹⁸ O [‰]
Minimum	24.0	6.5	0.4	-205.0	0.5	3072.5	1955.0	237.3	129.7	51.3	9.9	401.8	1.9	294.2	87.4	0.0	5.5	-31.1	-4.3
Maximum	37.4	8.0	5.7	278.0	33.7	20431.8	15560.0	1936.0	704.1	371.7	97.0	3519.8	19.0	1740.2	712.6	3.6	22.5	-20.6	-2.6
Mean	30.9	7.2	1.4	113.4	5.0	6638.1	4536.0	572.6	261.1	110.8	25.2	1064.1	5.8	583.4	242.0	1.2	11.9	-25.7	-3.5

Appelo & Postma. [39], is one of the most commonly implemented approaches for determining the f_{sea} value. In this approach, the concentration of chloride - a conservative ion that does not undergo significant chemical reactions or biological processes - is utilized as the basis for calculating the f_{sea} . The chloride concentration is preferred over other ions due to its conservative behavior, which enables it to be transported through the subsurface without undergoing significant changes in concentration or composition. The f_{sea} value is calculated using equation (5) [39].

$$f_{sea} = \frac{C_{Cl,sam} - C_{Cl,f}}{C_{Cl,sw} - C_{Cl,f}} \tag{5}$$

Where $C_{Cl,sam}$, $C_{Cl,sw}$, $C_{Cl,f}$ represents the chloride concentration in the samples from groundwater, seawater, and freshwater, respectively. The f_{sea} value ranges from 0 to 1, with higher values indicating a greater degree of SWI or salinization in the GW system. A value of 0 indicates no SWI, while a 1 indicates that the GW sample is entirely composed of seawater. The f_{sea} value can be used to assess the overall quality of the GW and identify potential SWI sources, such as over-pumping, coastal erosion, or sea-level rise.

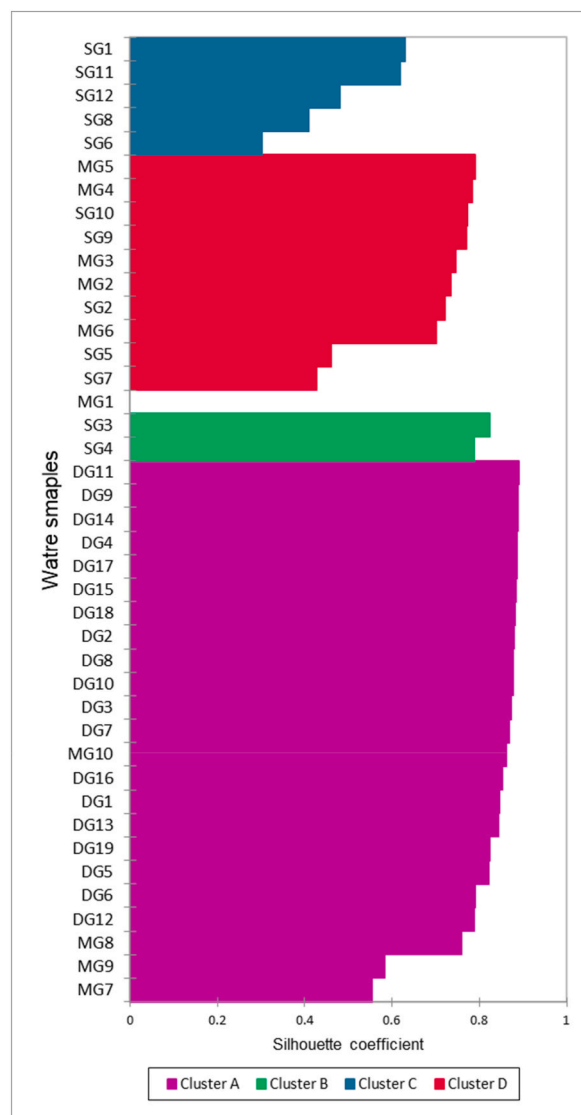


Fig. 3. Silhouette plot of HCA-DC, KMC, and FKM

6. Results and discussion

6.1. Groundwater hydrochemical characterization

Table 2 presents the descriptive statistics (minimum, maximum, and mean) of the physicochemical properties, major ions, and isotopes of the tested GW samples from the eastern coastal region of Saudi Arabia. The temperature of the GW samples ranged from 24 °C to 37.4 °C, which reflects the different depths of the sampled GW. The tested GW samples has pH values ranging from 6.5 to 8.0, indicating neutral water. The DO values of the tested GW sample range from 1.01 to 5.7 mg/L. The electrical conductivity (EC) and total dissolved solids (TDS) concentrations exhibited wide ranges, with minimum values of 3072 $\mu\text{S}/\text{cm}$ and 1955 mg/L and maximum values of 20431 $\mu\text{S}/\text{cm}$ and 15560 mg/L, respectively. Notably, the salinity values of the GW samples are above the accepted limit (1000 mg/L) for drinking purpose (WHO, 2014), indicating its unsuitability for human kind useage.

The groundwater has values of Eh range from -205 mV to 278 mV and turbidity range from 0.5 NTU to 33.7 NTU, respectively. The dominant anions in the GW samples were chloride and sulfate, with values in a range from 401.8 mg/L to 1064.1 mg/L, 294.2 mg/L to 1740.2 mg/L, and average values of 3519.8 mg/L and 583.4 mg/L, respectively. The predominant cations in the study area are Na and Ca, with concentration ranges from 237.3 mg/L to 1936 mg/L, 129.7 mg/L to 704.1 mg/L, and mean values of 572.6 mg/L and 261.1 mg/L, respectively.

6.2. Cluster analysis results

This study employed three clustering techniques - HCA-DC, KMC, and FKM - to group the analyzed GW samples. The clustering results revealed four distinct clusters (A, B, C, and D), as presented in Fig. 3 and Table 3. The silhouette coefficients for each groundwater sample within each cluster are depicted in Fig. 3. The results demonstrate that the highest values are associated with clusters A and B, indicating a strong cohesion within these groups and clear separation from other clusters. Cluster A contained the highest percentage of GW samples (56%), with 46.3% originating from the deep aquifer and 9.7% from the middle aquifer. The deep GW samples in cluster A are predominantly from wells within 2 km–5 km from the Arabian Gulf shore. Clusters B and C were composed of 17.1% of the total samples. They consist of GW samples obtained from the shallow aquifer, and wells found at distances of 2 km–3 km and 3 km–4 km from shoreline, respectively. Cluster D included samples from shallow and middle aquifers, totaling 14.7% of the total samples. The shallow aquifer samples accounted for 12.2%, while the middle aquifer samples accounted for 2.5%. The shallow aquifer GW samples in cluster D were taken from wells that are located within a distance of 4.7 km from the shore, while the middle aquifer samples were mainly from wells in Tarout Island and within a short distance of 1 km–2 km from the sea, as depicted in Table 3.

6.2.1. Double-Clustering approach

The importance of HCA clustering in understanding the hydrochemical processes of GW systems and monitoring GW quality in coastal regions has been studied by several researchers (e.g., Refs. [52,53]). In GW studies, scholars have implemented the HCA clustering technique with a combination of principal component analysis (PCA) [52,53], fuzzy c-mean clustering (FCM) [54], and water quality index (WQI) Das and Panda [55]. In this study, hierarchical cluster analysis (HCA) was applied in two stages to partition the hydrogeochemical data into different groups and subgroups (clusters) exhibiting different levels of SWI intensity. In the first stage, HCA was employed to the thorough dataset to partition the data into a group of variables, and the results are shown in the dendrogram in Fig. 4. Four groups of variables were defined with a linkage distance of approximately 35. Group A contained nine variables, including TDS, EC, Br^- , Na^+ , Cl^- , Ca^{2+} , Mg^{2+} , K^+ , and SO_4^{2-} . Group B contains turbidity, bicarbonate, fluoride, and isotopes. Group C consists of only nitrate ions. Group D consisted of temperature, Eh, dissolved oxygen, and pH. Since group A contained variables that were mostly clustered with TDS and EC, we applied cluster analysis (second HCA) for group A for the purpose of clustering water samples into clusters with different degrees of SWI. Therefore, the HCA-DC was applied to group A of variables to divide the water samples into clusters with varying degrees of SWI and salinization. Fig. 4 and Table 3 show that the water samples were grouped into four clusters (A, B, C, and D). Cluster A comprises 19 water samples from the deep aquifer and four samples from the middle aquifer. Cluster B contained two water samples from the shallow aquifer. Cluster C contained five water samples from a shallow aquifer. Cluster D comprises five water samples from the shallow aquifer and six from the middle aquifer.

6.2.2. K-mean clustering

The effectiveness of K-means clustering in partitioning GW data into groups that facilitate interpretation and pattern recognition

Table 3
GW samples percentage in each cluster.

Cluster	GW samples %	Aquifer Type	Distance from sea (km)
A	46.3	Deep	2–5
	9.7	Middle	3–4
B	4.9	Shallow	2–3
C	12.2	Shallow	3–4
D	12.2	Shallow	3.5–4.7
	14.7	Middle	1–2

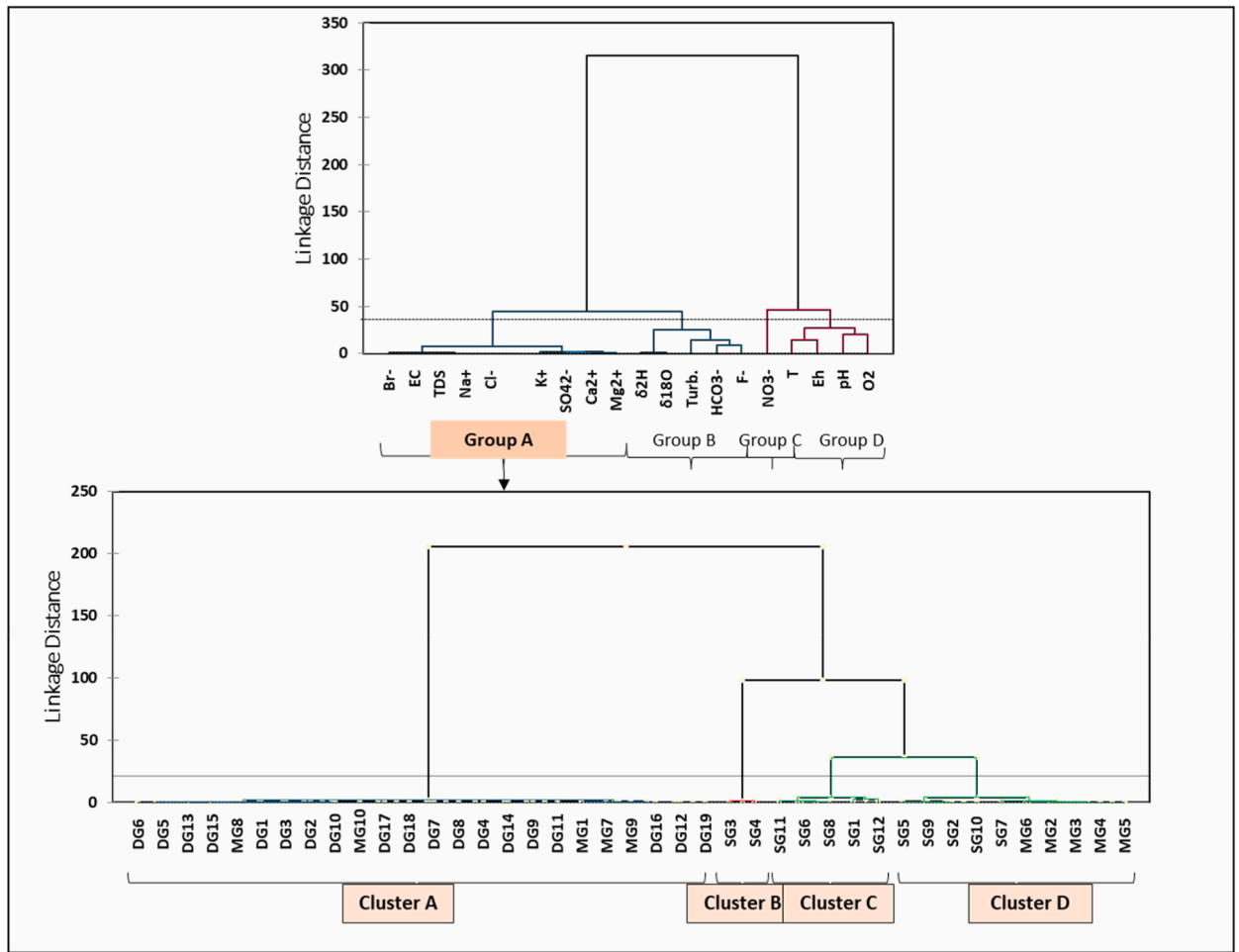


Fig. 4. HCA dendrogram for variables groups and water samples clusters of group A.

has been described in the literature [56]. Scholars have applied K-mean clustering in combination with other techniques, such as fuzzy c-mean, to evaluate GW quality and define regions with suitable drinking water [57]. The K-means clustering technique was also implemented with a self-organizing map for pattern recognition of nitrogen pollution in lake water [58]. The present study employed the K-means clustering to partition the GW hydrochemical data into distinct clusters to identify the spatial pattern of GW salinization and SWI. Fig. 5 displays the results of utilizing the silhouette coefficient to calculate the optimal cluster numbers. After analyzing the data, it was found that the highest silhouette coefficient value, which was 0.73, was associated with four clusters of observations,

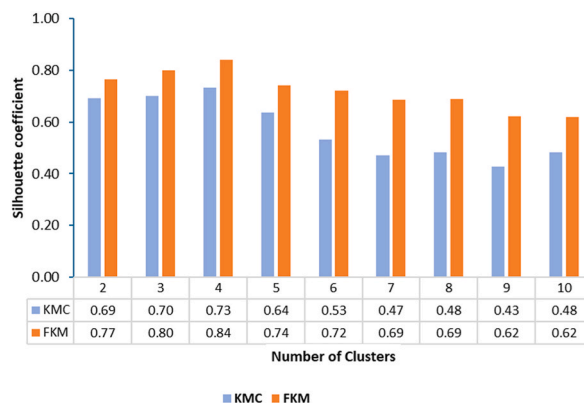


Fig. 5. Silhouette scores of the KMC and FKM.

namely clusters A, B, C, and D. Notably, the GW samples in these four clusters were consistent with the results obtained from the HCA-DC analysis (see Table 3).

6.2.3. Fuzzy K-mean clustering

Fuzzy K-means clustering has been implemented by scholars in GW studies to identify the quality of water and spatial patterns of hydrochemical data [56,57]. Eskandari et al. [56], applied FCM clustering in combination with HCA and KMC to delineate the hydrochemical evolution of a karst aquifer in northwestern Iran. In the current investigation, the FKM algorithm is utilized to partition the GW data into distinct clusters. The clustering analysis incorporated twenty variables, enabling a comprehensive evaluation of the physicochemical properties of the GW samples. As demonstrated in Fig. 5, the maximum silhouette score of 0.84 was obtained for four clusters, indicating that this was the optimal clustering solution for the GW samples. These four clusters were designated as A, B, C, and D.

Notably, the clustering results obtained from the fuzzy K-means algorithm were consistent with those generated by the HCA-DC and KMC methods, as depicted in Table 3. The similarity in clustering results across the different methods indicates the robustness and reliability of the clustering analyses and underscores the consistency of the underlying patterns in the GW data. The information derived from these clustering analyses can inform GW management strategies and practices, such as identifying areas with potentially high risks of SWI and implementing measures to mitigate such risks. Overall, applying the fuzzy K-means clustering algorithm to the GW data proved to be an effective tool for identifying distinct clusters and evaluating the physicochemical properties of the samples.

6.3. Evaluation of cluster analysis

The performance of the applied clustering techniques was evaluated via the silhouette scores, as illustrated in Fig. 6. The Silhouette scores of HCA-DC range from 0.39 for cluster C to 0.80 in cluster A, with an average value of 0.63. KMC techniques show silhouette scores range from 0.49 in cluster C to 0.83 in cluster A, with an average value of 0.69. FKM clustering method shows Silhouette scores within a range from 0.72 in cluster C to 0.94 in cluster A and an average value of 0.83. According to the Silhouette score average values for each type of clustering technique applied in this study, it was found that the FKM clustering performance was higher than the others.

6.4. Spatial analysis

Spatial maps were constructed for the study area’s shallow, middle, and deep aquifers, as illustrated in Fig. 7. Fig. 7A represents the distribution of shallow groundwater (GW) samples across three clusters: B, C, and D. The GW samples in cluster B were distributed close to the Arabian Gulf. They displayed high degree of SWI, revealing a more significant influence of seawater on GW chemistry and salinity. However, samples in clusters C and D, distributed further inland, display a decreasing impact of SWI on GW chemistry. Fig. 7 B depicts the distribution of GW samples of the middle aquifer that belong to clusters A and D. Samples in cluster A, located far from the Arabian Gulf, display no influence yet of SWI on GW quality. However, samples of cluster D occurred in Tarout Island and were near the Arabian Gulf and displayed a slight contamination of GW of the middle aquifer via SWI. Fig. 7C depicts the GW samples from deep

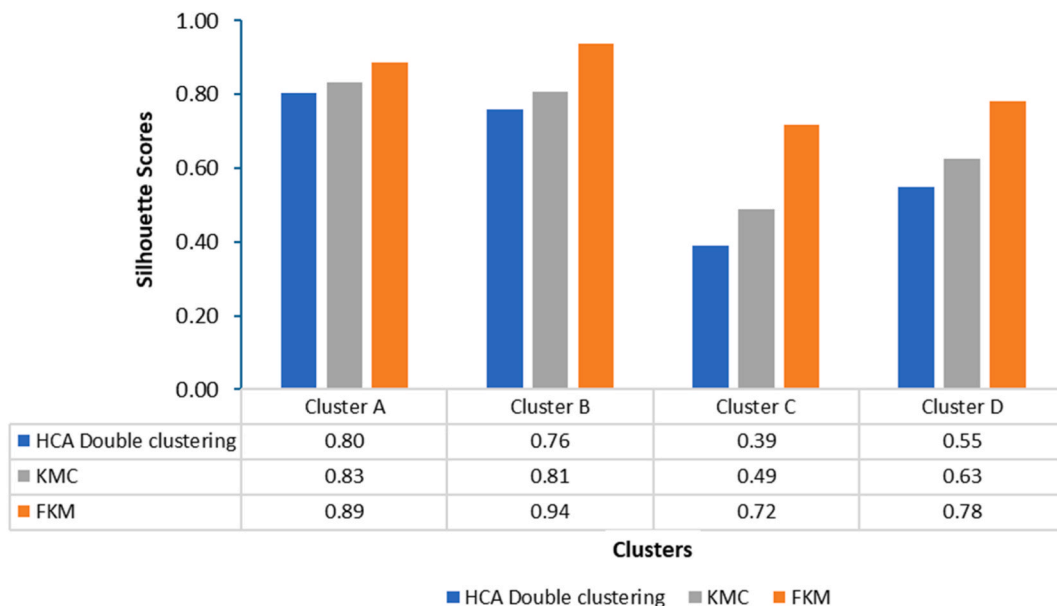


Fig. 6. Silhouette score for each cluster of FKM and K-mean, double-clustering.

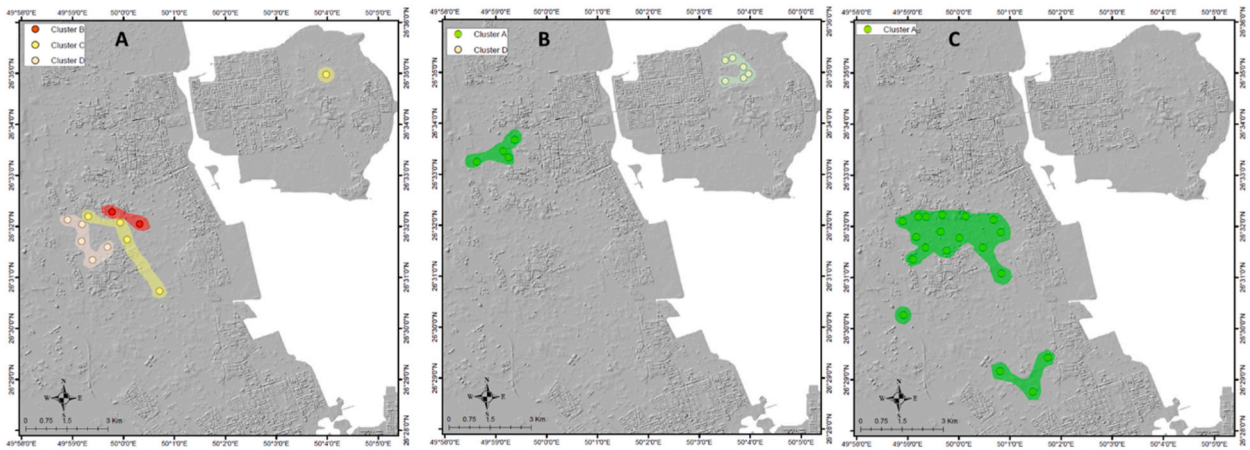


Fig. 7. Maps presenting the spatial distribution of clusters A, B, C, and D within different aquifer depths. A) Shallow aquifer, B) Middle aquifer, and C) Deep aquifer.

aquifers belonging to cluster A. Samples were distributed at different distances from the Arabian Gulf; however, they displayed no risk of SWI on the GW quality. Generally, the spatial maps of the three studied aquifers revealed the distribution of GW samples from each cluster. They showed that shallow GW was being significantly affected by SWI to varying degrees based on its distance from the Arabian Gulf. Nonetheless, SWI slightly affects middle aquifer GW in wells close to the coast and has no impact on wells further inland. The spatial assessment of tested GW confirmed that SWI has not yet affected the deep aquifer.

The utilization of integrated clustering analysis in this study has yielded notable enhancements in mapping seawater intrusion, particularly within the complex, multi-aquifer system of the investigated region. This approach represents a novel application in the field of seawater intrusion mapping, which traditionally has relied on other methodologies. For instance Ref. [59], employed the Inverse Distance Weighted (IDW) method in conjunction with a Geographic Information System (GIS), to delineate areas of seawater intrusion in Tunisia. Similarly [60], leveraged both DRASTIC and GALDIT standard parametric models, as well as their Analytic Hierarchy Process (AHP) counterparts (AHP-DRASTIC and AHP-GALDIT), in conjunction with GIS, to map seawater intrusion within the Mahdia-Ksour Essef aquifer in Tunisia. These traditional techniques necessitate comprehensive data sets encompassing soil characteristics, aquifer matrix properties, aquifer properties, and water quality parameters to adequately assess aquifer system vulnerability. In contrast, the current study’s clustering analysis technique, utilizing foundational data on groundwater physicochemical properties, allowed for effective mapping of seawater intrusion within the studied coastal area.

6.5. Degree of seawater intrusion

The fraction of seawater index (f_{sea}) is a widely used indicator for assessing the degree of SWI in coastal aquifers. The f_{sea} is defined as the fraction of seawater present in the GW sample and is calculated based on the concentration of chloride - a conservative ion that does not undergo significant changes in concentration or composition during subsurface transport [39,61]. In this study, to assess the accuracy of the clustering methods used to map the degree of seawater intrusion in the studied area, we calculated the f_{sea} values of the GW samples within each cluster. Table 4 exhibits the results of the f_{sea} calculation of GW samples, along with statistical values of f_{sea} for each cluster. The table also includes information on the aquifer type, GW depth, and distance from the shoreline.

Cluster A consists of groundwater samples from deep aquifers and wells located 2–5 km from the Arabian Gulf shore, with a value of f_{sea} range from 1.2% to 2.8% and an average of 1.9%. This implies that the deep aquifer in the studied region is yet to be impacted by

Table 4
SWI index (f_{sea}) values for each cluster.

Cluster	Statistics	Fraction of seawater index (f_{sea})	Aquifer Type	Depth (m)	Distance from sea(km)
Cluster A	Min	1.2	Deep aquifer	100–150	2.0–5.0
	Max	2.8			
	Mean	1.9			
Cluster B	Min	14.4	Middle aquifer	70–90	3.0–4.0
	Max	15.4			
	Mean	14.9			
Cluster C	Min	7.2	Shallow Aquifer	1.0–30.0	3.0–4.0
	Max	10.1			
	Mean	7.9			
Cluster D	Min	3.9	Shallow aquifer	1.0–30.0	3.0–4.7
	Max	6.2			
	Mean	5.2			

Table 5
Statistics values of Heavy mineral concentration (mg/L).

Cluster	Statistic	B	Al	Co	V	Mn	Ni	Cu	Zn	As	Fe	Cr	Sr	Pb	Hg	Mo	Cd	Ba
Cluster A	Minimum	0.449	0.000	0.001	0.024	0.000	0.009	0.007	0.000	0.003	0.329	0.007	4.123	0.000	0.000	0.007	0.000	0.021
	Maximum	0.768	0.008	0.002	0.074	0.013	0.029	0.081	0.075	0.034	2.904	0.045	7.022	0.001	0.001	0.017	0.000	0.035
	Mean	0.553	0.004	0.001	0.048	0.003	0.020	0.025	0.022	0.018	1.236	0.027	5.545	0.000	0.000	0.011	0.000	0.029
Cluster B	Minimum	2.377	0.002	0.004	0.148	0.050	0.138	0.147	0.141	0.046	1.787	0.096	19.889	0.000	0.000	0.009	0.000	0.029
	Maximum	2.532	0.002	0.007	0.148	0.089	0.158	0.226	0.186	0.070	2.287	0.101	27.911	0.000	0.000	0.018	0.000	0.038
	Mean	2.454	0.002	0.006	0.148	0.070	0.148	0.187	0.163	0.058	2.037	0.099	23.900	0.000	0.000	0.013	0.000	0.033
Cluster C	Minimum	1.318	0.000	0.003	0.074	0.002	0.039	0.043	0.001	0.044	1.142	0.002	15.010	0.000	0.000	0.021	0.000	0.015
	Maximum	1.698	0.005	0.005	0.126	0.099	0.133	0.146	0.337	0.066	7.263	0.061	22.648	0.000	0.001	0.046	0.000	0.047
	Mean	1.460	0.003	0.004	0.090	0.038	0.088	0.104	0.150	0.059	2.605	0.043	19.260	0.000	0.000	0.031	0.000	0.029
Cluster D	Minimum	0.733	0.000	0.001	0.025	0.000	0.012	0.038	0.000	0.006	0.911	0.003	5.965	0.000	0.000	0.002	0.000	0.022
	Maximum	1.157	0.007	0.004	0.140	0.053	0.090	0.140	0.236	0.044	3.831	0.046	13.765	0.000	0.002	0.043	0.000	0.044
	Mean	0.997	0.002	0.002	0.056	0.012	0.041	0.076	0.062	0.019	2.393	0.022	9.784	0.000	0.001	0.017	0.000	0.034

SWI. Cluster B includes groundwater samples from shallow aquifers 2 km–3 km from the Arabian Gulf. These samples have an f_{sea} range of 14.4%–15.4% and an average value of 14.5%, indicating a significant amount of seawater intrusion in the shallow aquifer. Cluster C consists of groundwater samples extracted from shallow aquifers situated 3–4 km away from the Arabian Gulf. These samples have an f_{sea} range of 7.2%–10.1%, with an average value of 7.9%. This indicates an intermediate level of SWI. Cluster D comprises groundwater samples from shallow and middle aquifers situated from 3 km to 4.7 km and 1 km–2 km from the sea, respectively. The f_{sea} values for cluster D vary from 3.9% to 6.2%, with an average of 5.2%. These values indicate that SWI has a minimal impact on the quality of the groundwater samples.

The clustering results imply that shallow GW in the studied region is characterized by varying degrees of SWI, with a high degree of intrusion observed in wells near the Arabian Gulf. As the distance from the seawater source increases, the degree of SWI in the shallow aquifer tends to decrease. In addition, the clustering analysis showed that shallow groundwater can be divided into three categories depending on their level of SWI: high, intermediate, and low. These findings have significant implications for managing and preserving coastal aquifers, as they highlight the potential risks associated with SWI and the need for effective mitigation measures to safeguard GW resources.

The analysis of SWI in the middle aquifer revealed a low degree of intrusion, indicating a slight impact on the quality of the GW samples. However, sustained pumping from the middle aquifer may exacerbate SWI and severely impact GW quality. In contrast, the deep aquifers remain largely unaffected by SWI, but increased pressure on these aquifers may accelerate SWI. Integrating the results of the SWI analysis with those of the clustering analysis, it was observed that GW samples from the deep aquifer, regardless of their distance from the shore, exhibit similar levels of SWI to GW samples from wells tapping the middle aquifer and located more than 3 km inland. These samples were grouped in cluster A and represented regions with no impact of SWI.

Conversely, GW samples from the middle aquifer close to the coastline exhibited similar levels of SWI to GW samples from the shallow aquifer situated within a distance greater than 3 km from the shoreline (Table 4). These samples were grouped in cluster D and displayed low SWI. In addition, the clustering analysis revealed two distinct clusters, B and C, comprising GW samples from the shallow aquifer. GW samples in cluster B exhibited a high degree of SWI and were near the coastline, within a distance range of 2–3 km. In contrast, GW samples in cluster C were sourced from wells situated further from the coastline, at distances ranging from 3 to 4 km. The results highlight the importance of considering both the clustering and SWI analyses in assessing the quality and sustainability of GW resources in coastal aquifers. Effective management strategies are needed to mitigate the risks associated with SWI and preserve the quality of GW resources. Specifically, measures aimed at reducing pumping pressure in the middle aquifer and limiting GW extraction from the shallow aquifer in areas with high SWI are recommended to minimize the impacts of SWI on GW quality.

6.6. Heavy metal concentration in SWI zones

Table 5 displays the statistics of heavy metal concentrations in the tested groundwater, with strontium (Sr), iron (Fe), and boron being the dominant heavy metals present in significant concentrations. The average concentrations of other metals, including Aluminum (Al), Cobalt (Co), Vanadium (V), Manganese (Mn), Nickel (Ni), Copper (Cu), Zinc (Zn), Arsenic (As), Chromium (Cr), Lead (Pb), Mercury (Hg), Molybdenum (Mo), Cadmium (Cd), and Barium (Ba) were found to be less than 0.2 mg/L. To compare the four clusters of groundwater samples, we observed that clusters B and C had higher concentrations of strontium than clusters D and A, with average values of 23.9 mg/L, 19.2 mg/L, 9.7 mg/L, and 5.5 mg/L, respectively (Fig. 8). Similarly, the concentration of boron metal was also higher in clusters B and C than in clusters D and A, with average values of 2.53 mg/L, 1.69 mg/L, 1.15 mg/L, and 0.76 mg/L, respectively (Fig. 8). The significant increase in strontium and boron concentration with increasing seawater intrusion into the aquifer system indicates the impact of seawater intrusion on groundwater pollution with heavy metals. As the concentration of boron in

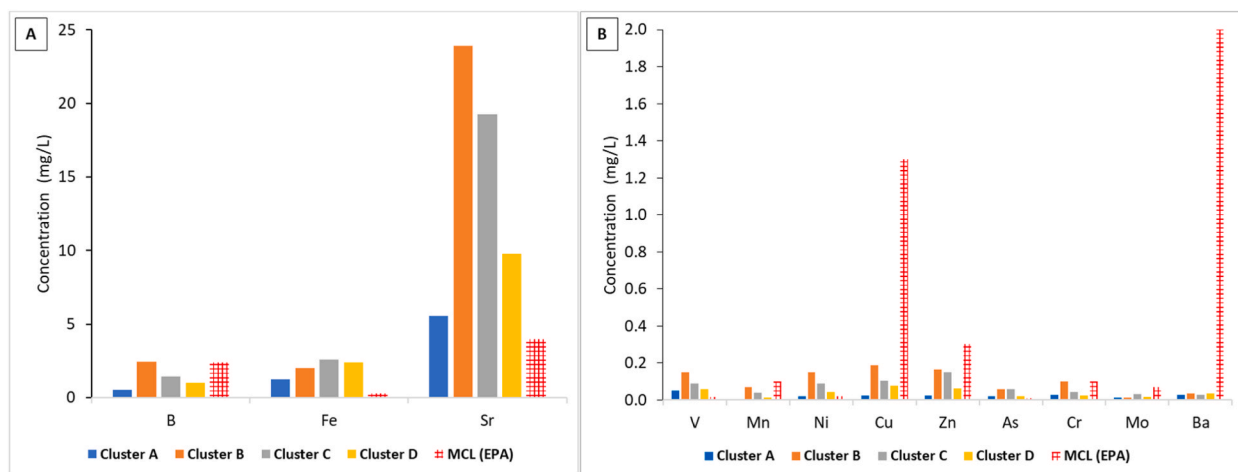


Fig. 8. Average heavy metal concentrations in groundwater samples across four clusters. A) Major metals (B, Fe, and Sr). B) Metals with average values below 0.2 mg/L (MCLs) = maximum contaminant levels.

seawater is high, reaching 4.5 mg/L in the Arabian Gulf [62], seawater intrusion into coastal freshwater aquifers noticeably pollutes the fresh groundwater with boron metal. Similarly, strontium can also be another heavy metal pollutant that intrudes into the fresh groundwater in the coastal aquifer. Conversely, iron metals display average values of 1.2 mg/L, 2.0 mg/L, 2.6 mg/L, and 2.4 mg/L of GW samples in cluster A, B, C, and D, respectively (Table 5 and Fig. 8). These values of iron metal display an increasing trend with increased seawater intrusion, except in cluster B, indicating another hydrochemical process controlling iron concentration in the aquifer system in the study area (e.g., water-rocks interaction).

The current study demonstrates that groundwater samples grouped in clusters with a high rate of seawater intrusion (clusters B and C) are associated with a higher average concentration of strontium and boron metals compared to groundwater samples in clusters with a low rate of seawater intrusion (clusters D and A) as shown in Fig. 8. According to the U.S. Environmental Protection Agency (EPA) guidelines, it is appeared that the concentration of Sr, Fe, V, Ni, and As in groundwater samples in all clusters exceeding the maximum contaminant level as shown in Fig. 8 [63]. However, the concentration of boron metal exceeded the maximum contaminant level (2.4 mg/L) set by the World health organization [64] in groundwater samples of cluster B. The concentration of the other metal, including Mn, Cu, Zn, Cr, Mo, and Ba, in groundwater samples from all clusters display values less than the maximum contaminant level for drinking water.

Generally, groundwater samples in clusters with a high rate of seawater intrusion display a higher concentration of strontium and boron metals, indicating the regions and aquifer's instability for drinking purposes. Therefore, using an integrated clustering approach to delineate seawater intrusion hot spot zones in the current study helped define the heavy metal polluted regions due to seawater intrusion. In the literature, several studies reported elevated concentrations of boron in coastal regions. For instance, Rahman et al. [65] found that the boron concentration in the coastal region of Bangladesh reached 4.1 mg/L. The highest concentration was found in the shallow wells, mostly affected by seawater intrusion. Mora et al. [66] reported increasing of boron concentration during the seawater intrusion process in an aquifer system in Baja California, Mexico. They concluded that the seawater intrusion process mainly controls boron elevated concentration with no effect from the anion exchange process. Wang et al. [67] investigated the heavy mineral distribution in coastal aquifers in Jiangsu Province, China. They found that boron is the major heavy metal with concentration ranges from 0.01 mg/L to 2.25 mg/L. Their results showed that the study area is affected by seawater intrusion with a significant correlation between boron metal and seawater intrusion.

7. Conclusion and Recommendations

The present study employed an integrated clustering approach to map and delineate the degree of SWI in the Arabian Gulf Coast of Saudi Arabia. Specifically, we applied Hierarchical Clustering Analysis (double-clustering) (HCA-DC), K-mean (KMC), and fuzzy k-mean clustering (FKM) algorithms to the physical, hydrochemical, and major ion data of the tested groundwater. This approach enabled a comprehensive evaluation of the groundwater characteristics and provided a robust clustering solution for identifying distinct groups of groundwater samples with similar characteristics. Four clusters (A, B, C, and D) of GW samples were developed via HCA-DC, KMC, and FKM clustering algorithms. The performance of cluster analysis was evaluated using silhouette score criteria. The main conclusions of the current study are shown as follows.

- The integrated clustering techniques (HCA-DC, KMC, and FKM) successfully identified four distinct clusters of groundwater samples with varying degrees of seawater intrusion.
- The HCA-DC, KMC, and FKM clustering techniques effectively delineated the study area's hotspot zones of salinization and seawater intrusion (SWI). Among the methods, FKM displayed the best performance.
- Cluster A, consisting mainly of deep aquifer samples located 2–5 km from shore, showed minimal signs of intrusion with averaged f_{sea} values of 1.9% and low concentrations of boron (0.76 mg/L) and strontium (5.5 mg/L).
- Clusters B and C, composed of shallow aquifer samples within 4 km from shore, exhibited the highest seawater intrusion with averaged f_{sea} values of 14.5% (Cluster B) and 7.9% (Cluster C). These clusters also had the highest boron (2.53 mg/L for Cluster B and 1.69 mg/L for Cluster C) and strontium concentrations (23.9 mg/L for Cluster B and 19.2 mg/L for Cluster C), indicating pollution from seawater.
- Cluster D, consisting of shallow and middle aquifer samples located 3.5–4.7 km from shore, showed low but detectable intrusion with an averaged f_{sea} of 5.2% and intermediate boron (1.15 mg/L) and strontium (9.7 mg/L) levels.
- The study demonstrated that employing integrated clustering algorithms can markedly improve the mapping of seawater intrusion (SWI) extent in coastal areas.
- Groundwater with high seawater intrusion, enriched with strontium and boron, may pose health risks and is unfit for drinking.
- The study's findings can guide remediation plans and monitoring to halt additional seawater intrusion and associated health risks.

The current study's findings demonstrate that mapping the degree SWI in coastal regions can be significantly enhanced using a combination of clustering algorithms. The study suggests that the clustering approaches can help identify areas that require remediation and monitoring to prevent further contamination by SWI. The outcome of this study can serve as a guide for future water management initiatives, particularly in complex areas vulnerable to salinity pollution. Further studies are required to integrate clustering techniques, such as self-organized maps, as alternative solutions for SWI pattern recognition in coastal regions.

Author contribution statement

Mohammed Benaafi: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. S.I. Abba: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper. Bassam Tawabini, Ismail Abdulazeez, Billel Salhi: Performed the experiments; Analyzed and interpreted the data; Wrote the paper. Jamilu Usman, Isam H. Aljundi: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding

This research was funded by the Deanship of Research Oversight and Coordination (DROC) at King Fahd University of Petroleum & Minerals (KFUPM) under the Interdisciplinary Research Center for Membranes and Water Security [Grant Number: INMW2308]. Article Processing Charges (APC) was provided by the Deanship of Research Oversight and Coordination (DROC).

Author contribution statement

Mohammed Benaafi: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. S.I. Abba: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper. Bassam Tawabini, Ismail Abdulazeez, Billel Salhi: Performed the experiments; Analyzed and interpreted the data; Wrote the paper. Jamilu Usman, Isam H. Aljundi: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no conflict of interest.

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.

Acknowledgments

Authors would like to acknowledge all support provided by the Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum and Minerals.

References

- [1] V. Gholami, K.W. Chau, F. Fadaee, J. Torkaman, A. Ghaffari, Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers, *J. Hydrol. (Amst.)* 529 (2015) 1060–1069.
- [2] Z. Li, et al., Groundwater quality and associated hydrogeochemical processes in Northwest Namibia, *J. Geochem. Explor.* 186 (2018) 202–214.
- [3] M. Benaafi, S.I. Abba, I.H. Aljundi, Effects of seawater intrusion on the groundwater quality of multi-layered aquifers in eastern Saudi Arabia, *Molecules* 28 (7) (2023) 3173.
- [4] L. Tulipano, D.M. Fidelibus, A. Panagopoulos, *Groundwater Management of Coastal Karstic Aquifers*, 2005.
- [5] A.D. Werner, et al., Seawater intrusion processes, investigation and management: recent advances and future challenges, *Adv. Water Resour.* 51 (2013) 3–26.
- [6] M. Mastrocicco, N. Colombani, The issue of groundwater salinization in coastal areas of the mediterranean region: a review, *Water (Basel)* 13 (1) (2021) 90.
- [7] A. Batayneh, et al., Hydrochemical facies and ionic ratios of the coastal groundwater aquifer of saudi gulf of aqaba: implication for seawater intrusion, *J. Coast Res.* 30 (1) (2014) 75–87, <https://doi.org/10.2112/JCOASTRES-D-13-00021.1>.
- [8] T. Cao, D. Han, X. Song, Past, present, and future of global seawater intrusion research: a bibliometric analysis, *J. Hydrol. (Amst.)* 603 (2021), 126844, <https://doi.org/10.1016/j.jhydrol.2021.126844>.
- [9] G. De Filippis, L. Foglia, M. Giudici, S. Mehl, S. Margiotta, S.L. Negri, Seawater intrusion in karstic, coastal aquifers: current challenges and future scenarios in the Taranto area (southern Italy), *Sci. Total Environ.* 573 (2016) 1340–1351, <https://doi.org/10.1016/j.scitotenv.2016.07.005>.
- [10] P. Maurya, R. Kumari, S. Mukherjee, Hydrochemistry in integration with stable isotopes ($\delta^{18}O$ and δD) to assess seawater intrusion in coastal aquifers of Kachchh district, Gujarat, India, *J. Geochem. Explor.* 196 (2019) 42–56.
- [11] M.F.A. Al Naem, et al., A study on the impact of anthropogenic and geogenic factors on groundwater salinization and seawater intrusion in Gaza coastal aquifer, Palestine: an integrated multi-techniques approach, *J. Afr. Earth Sci.* 156 (2019) 75–93.
- [12] X. Xu, et al., Characteristics of coastal aquifer contamination by seawater intrusion and anthropogenic activities in the coastal areas of the Bohai Sea, eastern China, *J. Asian Earth Sci.* 217 (2021), 104830.
- [13] Z. Ding, et al., Seawater intrusion impacts on groundwater and soil quality in the northern part of the Nile Delta, Egypt, *Environ. Earth Sci.* 79 (13) (2020), <https://doi.org/10.1007/s12665-020-09069-1>.
- [14] J.A. Torres-Martínez, A. Mora, J. Mahlknecht, D. Kaown, D. Barceló, Determining nitrate and sulfate pollution sources and transformations in a coastal aquifer impacted by seawater intrusion—a multi-isotopic approach combined with self-organizing maps and a Bayesian mixing model, *J. Hazard Mater.* 417 (2021), 126103.
- [15] S.H. Ko, K. Ishida, Z.M. Oo, H. Sakai, Impacts of seawater intrusion on groundwater quality in Htantabin township of the deltaic region of southern Myanmar, *Groundw. Sustain. Dev.* 14 (2021), <https://doi.org/10.1016/j.gsd.2021.100645>.

- [16] H. Abd-Elhamid, A. Javadi, I. Abdelaty, M. Sherif, Simulation of seawater intrusion in the Nile Delta aquifer under the conditions of climate change, *Nord. Hydrol* 47 (6) (2016) 1198–1210.
- [17] D. Han, M.J. Currell, Delineating multiple salinization processes in a coastal plain aquifer, northern China: hydrochemical and isotopic evidence, *Hydrol. Earth Syst. Sci.* 22 (6) (2018) 3473–3491.
- [18] D.T. Vu, T. Yamada, H. Ishidaira, Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam, *Water Sci. Technol.* 77 (6) (2018) 1632–1639.
- [19] J. Sae-Ju, S. Chotpanarat, T. Thitimakorn, Hydrochemical, geophysical and multivariate statistical investigation of the seawater intrusion in the coastal aquifer at Phetchaburi Province, Thailand, *J. Asian Earth Sci.* 191 (2020), 104165.
- [20] M. Balasubramanian, et al., Isotopic signatures, hydrochemical and multivariate statistical analysis of seawater intrusion in the coastal aquifers of Chennai and Tiruvallur District, Tamil Nadu, India, *Mar. Pollut. Bull.* 174 (2022), 113232.
- [21] Z.E.S. Salem, O.M. Osman, Use of major ions to evaluate the hydrogeochemistry of groundwater influenced by reclamation and seawater intrusion, West Nile Delta, Egypt, *Environ. Sci. Pollut. Control Ser.* 24 (4) (2017) 3675–3704, <https://doi.org/10.1007/s11356-016-8056-4>.
- [22] H. Arslan, Application of multivariate statistical techniques in the assessment of groundwater quality in seawater intrusion area in Bafra Plain, Turkey, *Environ. Monit. Assess.* 185 (3) (2013) 2439–2452.
- [23] P.J.S. Kumar, P. Jegathambal, B. Babu, A. Kokkat, E.J. James, A hydrogeochemical appraisal and multivariate statistical analysis of seawater intrusion in point calimere wetland, lower Cauvery region, India, *Groundw Sustain Dev* 11 (2020), 100392.
- [24] S. Hajji, N. Allouche, S. Bourri, A.M. Aljuaid, W. Hachicha, Assessment of seawater intrusion in coastal aquifers using multivariate statistical analyses and hydrochemical facies evolution-based model, *Int J Environ Res Public Health* 19 (1) (2022) 155.
- [25] F. Telahigue, B. Agoubi, F. Souid, A. Kharroubi, Assessment of seawater intrusion in an arid coastal aquifer, south-eastern Tunisia, using multivariate statistical analysis and chloride mass balance, *Phys. Chem. Earth, Parts A/B/C* 106 (2018) 37–46.
- [26] H. Ferchichi, M.F. Ben Hamouda, B. Farhat, A. Ben Mammou, Assessment of groundwater salinity using GIS and multivariate statistics in a coastal Mediterranean aquifer, *Int. J. Environ. Sci. Technol.* 15 (11) (2018) 2473–2492.
- [27] W.A. Abderrahman, M. Rasheeduddin, Management of groundwater resources in a coastal belt aquifer system of Saudi Arabia, *Water Int.* 26 (1) (2001) 40–50.
- [28] M. Rasheeduddin, H. Yazicigil, R.I. Al-Layla, Numerical modeling of a multi-aquifer system in eastern Saudi Arabia, *J. Hydrol. (Amst.)* 107 (1–4) (1989) 193–222.
- [29] C.H. V Ebert, Water Resources and Land Use in the Qatif Oasis of Saudi Arabia, *Geogr Rev*, 1965, pp. 496–509.
- [30] A. Al-Shaibani, Economic potential of brines of Sabkha jayb uwayyid, eastern Saudi Arabia, *Arabian J. Geosci.* 6 (7) (2013) 2607–2618.
- [31] M. Ziegler, Late permian to Holocene Paleofacies Evolution of the arabian Plate and its hydrocarbon occurrences 6 (3) (2001) 445–504.
- [32] L. F. Ramirez Powers, C.D. Redmond, E.L.J. Elberg, Geology of the Arabian Peninsula Sedimentary Geology of Saudi Arabia U.S. Geological Survey Professional Paper, 560-D, 1966, p. 154.
- [33] J.W. Tleel, Surface geology of Dammam dome, eastern province, Saudi Arabia, *Am Assoc Pet Geol Bull* 57 (3) (1973) 558–576.
- [34] R. Weijermars, Surface geology, lithostratigraphy and Tertiary growth of the Dammam Dome, Saudi Arabia: a new field guide, *GeoArabia* 4 (2) (1999) 199–226.
- [35] M.H. Al-Tamimi, Stratigraphical and Microfacies Analysis of the Early Paleogene Succession in the Dammam Dome, Eastern Saudi Arabia, 1985. MS. thesis.
- [36] H.S. Edgell, Aquifers of Saudi Arabia and their geological framework, *Arab J Sci Eng* 22 (1 C) (1997) 3–31.
- [37] M. Benaafi, et al., Integrated hydrogeological, hydrochemical, and isotopic assessment of seawater intrusion into coastal aquifers in Al-Qatif area, eastern Saudi Arabia, *Molecules* 27 (20) (2022) 6841.
- [38] USEPA, Groundwater Sampling Guidelines, Environment Protection Authority, 2000, p. 36, no. April.
- [39] C.A.J. Appelo, D. Postma, Geochemistry, Groundwater and Pollution, CRC press, 2004.
- [40] I. Clark, Groundwater Geochemistry and Isotopes, CRC press, 2015.
- [41] J.C. Egbueri, Heavy metals pollution source identification and probabilistic health risk assessment of shallow groundwater in Onitsha, Nigeria, *Anal. Lett.* 53 (10) (2020) 1620–1638.
- [42] R.A.A. Gyimah, C. Gyamfi, G.K. Anornu, A.Y. Karikari, F.W. Tsyawo, Multivariate statistical analysis of water quality of the Densu River, Ghana, *Int. J. River Basin Manag.* 19 (2) (2021) 189–199.
- [43] A. Mora, J.A. Torres-Martínez, C. Moreau, G. Bertrand, J. Mählknecht, Mapping salinization and trace element abundance (including as and other metalloids) in the groundwater of north-central Mexico using a double-clustering approach, *Water Res.* 205 (2021), 117709.
- [44] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a k-means clustering algorithm, *J R Stat Soc Ser C Appl Stat* 28 (1) (1979) 100–108.
- [45] J. Han, J. Pei, H. Tong, Data Mining: Concepts and Techniques, Morgan kaufmann, 2022.
- [46] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, *ACM Comput. Surv.* 31 (3) (1999) 264–323.
- [47] J.C. Dunn, A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Clusters, 1973.
- [48] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer Science & Business Media, 2013.
- [49] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* 20 (1987) 53–65.
- [50] P. Ardarsa, O. Surinta, Water quality assessment in the lam Pa thao Dam, chaiyaphum, Thailand with K-means clustering algorithm, in: 2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C), IEEE, 2021, pp. 35–39.
- [51] D. Arsene, A. Predescu, B. Pahonțu, C.G. Chiru, E.-S. Apostol, C.-O. Truică, Advanced strategies for monitoring water consumption patterns in households based on IoT and machine learning, *Water (Basel)* 14 (14) (2022) 2187.
- [52] S. Mountadar, A. Younsi, A. Hayani, M. Siniti, S. Tahiri, Groundwater salinization process in the coastal aquifer sidi abed-ouled ghanem (province of el Jadida, Morocco), *J. Afr. Earth Sci.* 147 (2018) 169–177.
- [53] M. Monjerezi, R.D. Vogt, P. Aagaard, J.D.K. Saka, Hydro-geochemical processes in an area with saline groundwater in lower Shire River valley, Malawi: an integrated application of hierarchical cluster and principal component analyses, *Appl. Geochem.* 26 (8) (2011) 1399–1413.
- [54] A. Rahbar, M. Vadiati, M. Talkhabi, A.A. Nadiri, M. Nakhaei, M. Rahimian, A hydrogeochemical analysis of groundwater using hierarchical clustering analysis and fuzzy C-mean clustering methods in Arak plain, Iran, *Environ. Earth Sci.* 79 (2020) 1–17.
- [55] C.R. Das, S. Das, S. Panda, Groundwater quality monitoring by correlation, regression and hierarchical clustering analyses using WQI and PAST tools, *Groundw Sustain Dev* 16 (2022), 100708.
- [56] E. Eskandari, H. Mohammadzadeh, H. Nassery, M. Vadiati, A.M. Zadeh, O. Kisi, Delineation of isotopic and hydrochemical evolution of karstic aquifers with different cluster-based (HCA, KM, FCM and GKM) methods, *J. Hydrol. (Amst.)* 609 (2022), 127706.
- [57] O. Mohammadrezapour, O. Kisi, F. Pourahmad, Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality, *Neural Comput. Appl.* 32 (2020) 3763–3775.
- [58] H. Mao, et al., Deciphering spatial pattern of groundwater chemistry and nitrogen pollution in Poyang Lake Basin (eastern China) using self-organizing map and multivariate statistics, *J. Clean. Prod.* 329 (2021), 129697.
- [59] S. Hajji, N. Allouche, S. Bourri, A.M. Aljuaid, W. Hachicha, Assessment of seawater intrusion in coastal aquifers using multivariate statistical analyses and hydrochemical facies evolution-based model, *Int J Environ Res Public Health* 19 (1) (2021) 155.
- [60] G. Nasri, S. Hajji, W. Aydi, E. Boughariou, N. Allouche, S. Bourri, Water vulnerability of coastal aquifers using AHP and parametric models: methodological overview and a case study assessment, *Arabian J. Geosci.* 14 (2021) 1–19.
- [61] B. Sarkar, A. Islam, A. Majumder, Seawater intrusion into groundwater and its impact on irrigation and agriculture: evidence from the coastal region of West Bengal, India, *Reg Stud Mar Sci* 44 (2021), 101751.
- [62] E. Güler, C. Kaya, N. Kabay, M. Arda, Boron removal from seawater: state-of-the-art review, *Desalination* 356 (2015) 85–93.
- [63] USEPA, Drinking Water Standards and Health Advisories, 2012th, U.S. Environmental Protection Agency., 2012.
- [64] Who, Guidelines for Drinking-Water Quality, Fourth, 2011.

- [65] M. Rahman, M.A.N. Tushar, A. Zahid, K.M.U. Ahmed, M.A.M. Siddique, M.G. Mustafa, Spatiotemporal distribution of boron in the groundwater and human health risk assessment from the coastal region of Bangladesh, *Environ. Sci. Pollut. Control Ser.* 28 (2021) 21964–21977.
- [66] A. Mora, J. Mählknecht, R. Ledesma-Ruiz, W.E. Sanford, L.E. Lesser, Dynamics of major and trace elements during seawater intrusion in a coastal sedimentary aquifer impacted by anthropogenic activities, *J. Contam. Hydrol.* 232 (2020), 103653.
- [67] Z. Wang, Q. Su, S. Wang, Z. Gao, J. Liu, Spatial distribution and health risk assessment of dissolved heavy metals in groundwater of eastern China coastal zone, *Environ. Pollut.* 290 (2021), 118016.