Journal of Hydrology: Regional Studies 55 (2024) 101929

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Application of gravity recovery and climate experiment data and ensemble modeling to assess saltwater intrusion in the Miandoab coastal aquifer, Iran, under climate change

Vahid Nourani [a,b,d,*], Nardin Jabbarian Paknezhad [a], Zhang Wen [c], Sameh Ahmed Kantoush [d]

[a] Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, Iran
[b] College of Engineering, Information Technology and Environment, Charles Darwin University, Darwin, Australia
[c] School of Environmental Studies, China University of Geosciences, Wuhan, China
[d] Disaster Prevention Research Institute, Kyoto University, Goka-sho, Uji City, Kyoto 611-0011, Japan

ARTICLE INFO

Keywords:
Groundwater
Climate change
Clustering
Ensemble learning
Hydro-chemical facies evolution
Miandoab

ABSTRACT

Study region: The Miandoab aquifer, northwest of Iran, which is located in a sub basin of the Urmia Lake.
Study focus: To model the groundwater (GW) quantity and quality, shallow learning (Feed Forward Neural Network (FFNN), Adaptive neuro fuzzy inference system (ANFIS), Support Vector Regression (SVR)), their ensemble and deep learning models were applied. Projections by General Circulation Models (GCMs) for the Shared Socio-economic Pathways (SSP585) scenario, after bias correction, and changes of the model inputs including Normalized Difference Vegetation Index (NDVI), Gravity Recovery and Climate Experiment (GRACE), GRACE Follow-On (GRACE-FO) and GW level (GWL) obtained via Markov Chain model were employed for future climate change projections. To project GW quality (GWQ) parameters for future climate conditions, relationships between GWL and GWQ were established via the Fourier model.
New hydrological insights for the region: Results revealed that ensemble learning could outperform individual methods up to 23 %. The Hydro-chemical Facies Evolution (HFE) diagrams for 2050 and 2100 indicated that clusters near the shoreline may exhibit severe declining trend in GWL up to 1.53 m and potential intrusion of saltwater. In the higher altitude lands GWL may exhibit declining trend up to 11.74 m. In addition, HFE diagram indicated that the Ca-Cl water type will be more common in 2050.

Abbreviations: GW, Groundwater; FFNN, Feed Forward Neural Network; ANFIS, Adaptive neuro fuzzy inference system; SVR, Support Vector Regression; GCM, General Circulation Models; GRACE, Gravity Recovery and Climate Experiment; GRACE-Fo, GRACE Follow-On; NDVI, Normalized Difference Vegetation Index; GWL, GW level; SSP, Shared Socio-economic Pathways; GWQ, GW quality; HFE, Hydro-chemical Facies Evolution; AI, Artificial Intelligence; GRU, Gated Recurrent Units; DTW, Dynamic time warping; LSTM, Long short-term memory; FLDAS, Famine Early Warning Systems Network Land Data Assimilation System; CMIP6, Coupled Model Intercomparison Project Phase 6; SOM, Self-Organizing Map; FCM, Fuzzy c-means; CTS, Connected-triple-based similarity; SRS, SimRank-based similarity; ASRS, Approximate SimRank-based similarity; NSE, Nash-Sutcliffe efficiency; RMSE, Root mean square error; HVI, Hybrid Validity Index; SWSA, Surface water storage anomaly; GWSA, Groundwater storage anomalies; SMSA, Soil moisture storage anomaly; SWEA, Snow water equivalent anomaly.

* Corresponding author at: College of Engineering, Information Technology and Environment, Charles Darwin University, Darwin, Australia.
E-mail address: vahid.nourani@cdu.edu.au (V. Nourani).

https://doi.org/10.1016/j.ejrh.2024.101929
Received 9 March 2024; Received in revised form 31 July 2024; Accepted 2 August 2024
Available online 10 August 2024
2214-5818/

1. Introduction

Groundwater (GW), as the critical natural resource, plays fundamental role in supporting human societies and ecosystems. GW serves as the primary source of drinking water, irrigation and industry. Moreover, it sustains ecosystems by replenishing surface water bodies, supporting aquatic habitats, and maintaining ecological balance. Recognizing the immense importance of GW, it is imperative to understand its characteristics, monitor its quality and quantity, and implement sustainable management strategies to ensure its long-term availability for present and future generations. Evaluating GW quality (GWQ) is essential to ensure the safety and suitability of water for human consumption and to protect ecosystems from potential contamination.

In coastal aquifers, the intrusion of seawater into freshwater GW is a significant concern. This saltwater intrusion occurs when the interface between the saline seawater and the freshwater aquifer shifts landward. It is primarily driven by natural processes such as tidal fluctuations, as well as human activities such as excessive GW pumping and coastal development. The intrusion of seawater can have detrimental effects on the quality and availability of freshwater resources in the coastal areas. As seawater infiltrates the freshwater aquifers, it contaminates the previously potable GW, rendering it unsuitable for various uses, including drinking water supply and irrigation. The increased salinity levels can also have adverse impacts on ecosystems and agricultural productivity. Understanding the availability, recharge rates, and extraction rates of GW helps prevent overexploitation and depletion of this valuable resource. Accurate assessments of both GWQ and quantity are effective for managing water resource, ensuring the long-term availability of clean and sufficient water for various uses.

The Gravity Recovery and Climate Experiment (GRACE) satellite presents unique opportunity for accurately measuring spatial variations in GW storage at large basin scale with high precision at monthly resolutions. Utilizing GRACE data allows continuous monitoring of GW storage at regular intervals, surpassing the limitations posed by the scarcity and uneven distribution of piezometers in the study area. The quantification of changes in total water supplies by GRACE data proves crucial in hydrological studies, particularly in investigating drought conditions. Water within large watershed plays pivotal role in the hydrological cycle and exerts significant influence on the local climate system. Previous studies have successfully employed GRACE data to calculate GW anomalies, highlighting the applicability and effectiveness of this dataset (Foroumandi et al., 2023; Sabzehee et al., 2023; Wang et al., 2023).

Like all hydrological process, GW quantity and quality is affected under climate change. The hydrological cycle is strongly influenced by climate change, as any alteration in climate has substantial implication for the hydrological cycle. Furthermore, climate change impacts the accessibility of water by changing water sources (Sivarajan et al., 2019). Various methodologies have been used to assess the impact of climate change on GW recharge and its repercussions on GW level (GWL) (e.g., see Nourani et al., 2022a). To predict subsurface water region, prevalent approach involves utilizing data obtained from general circulation models (GCMs). Multiple studies applied GCMs for climate change impacts assessments in hydrology (e.g., see, Baghanam et al., 2019; Nourani et al., 2018; Nourani, et al., 2019).

Due to the complexity of GW modeling, the Artificial Intelligence (AI)-based methods both of Shallow learning (Nourani et al., 2018) and deep learning (Emmert-Streib et al., 2020; Nourani et al., 2022b) were successfully applied for GWL modeling via employing different inputs. AI methods have ability to establish non-linear connections between input parameters and targets. These methods do not necessitate any preconceived assumptions about the underlying physical system being modeled. Hence, these statistical techniques are effective in uncovering associations within the data. According to the provided literature review, the majority of previous studies focusing on GWL modeling overlooked the influence of anthropogenic activities and neglected the prediction of future GWL and GWQ conditions. To address this research gap, the current study utilized a combination of remote sensing and GCMs datasets to model and project GWL and GWQ. This approach enabled a more comprehensive analysis, considering the impact of human activities and providing insights into the future states of GWL and GWQ.

In this study, the GWL and GWQ in Miandoab aquifer, which play vital role in supporting agricultural activities in the region were modeled and predicted under climate change scenarios. For this purpose, shallow, ensemble and Gated Recurrent Units (GRU) AI methods were used for GWL modeling. Based on the authors' knowledge, this is the first study that used GCM, GRACE, and Normalized Difference Vegetation Index (NDVI) (as representative of anthropogenic activities) for prediction of the GWQ. On the other hand, the abundance of piezometers within the aquifer makes it an arduous task to assess each one individually. To simplify this process, clustering techniques could be employed to group piezometers with similar patterns together, allowing for the selection of a representative piezometer (centroid member) from each cluster. This approach facilitates more efficient and manageable assessment of the piezometers, ensuring that their characteristics and behaviors are adequately captured. Thus, clustering ensemble (Link-Clue) was also used for clustering GW piezometers. This study aimed at following 3 main objectives:

i. GWL modeling and projection for future via application of the GCMs, GRACE and NDVI data,
ii. Using clustering ensemble for clustering the piezometers and the aquifer zoning,
iii. Projection of GWQ parameters and the Hydro-chemical Facies Evolution (HFE) diagram for investigation of seawater intrusion into the aquifer.

2. Materials and methods

2.1. Proposed methodology

The overall applied methodology in this paper for modeling and prediction of GWL and GWQ is depicted in Fig. 1. At the first step,

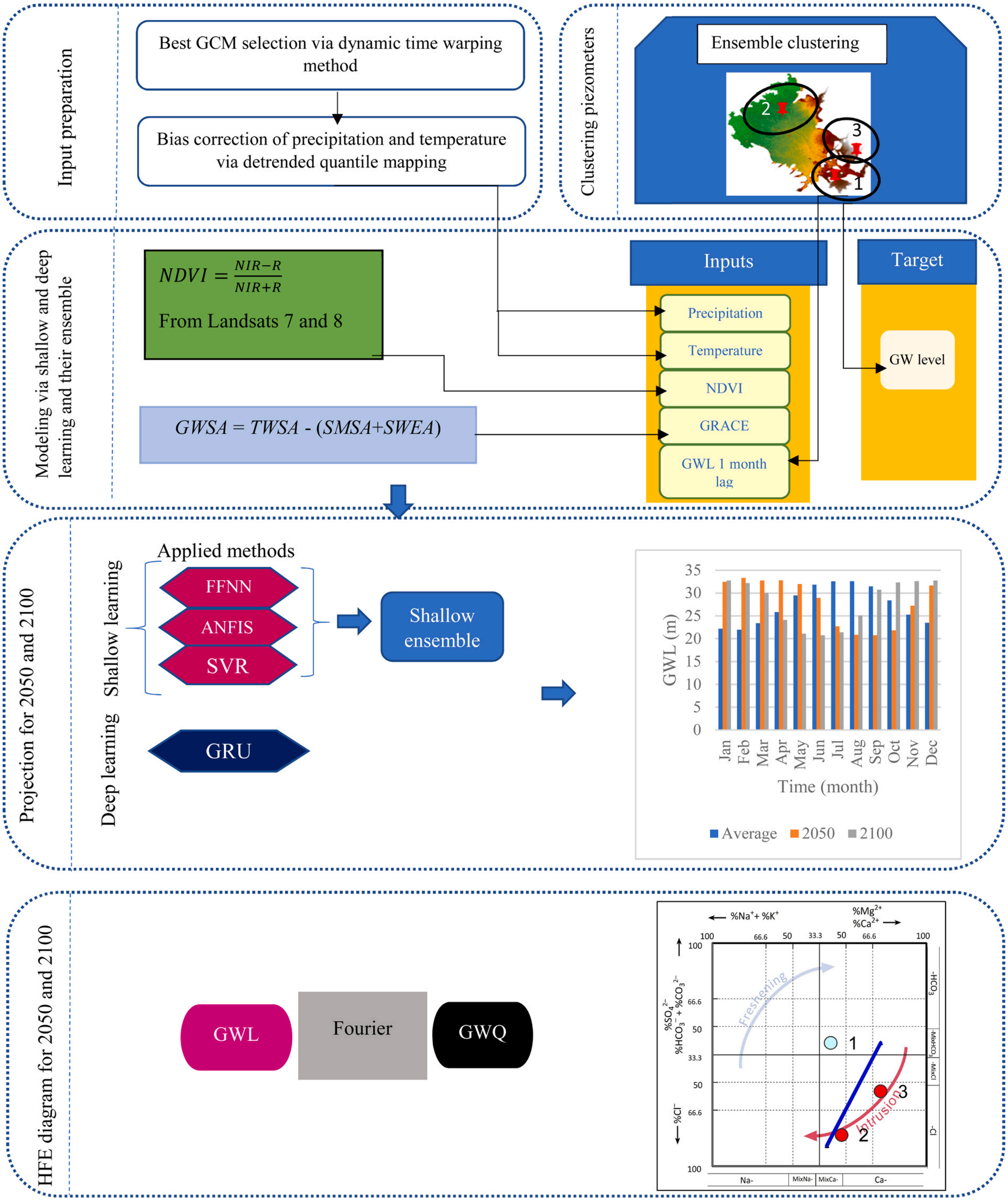
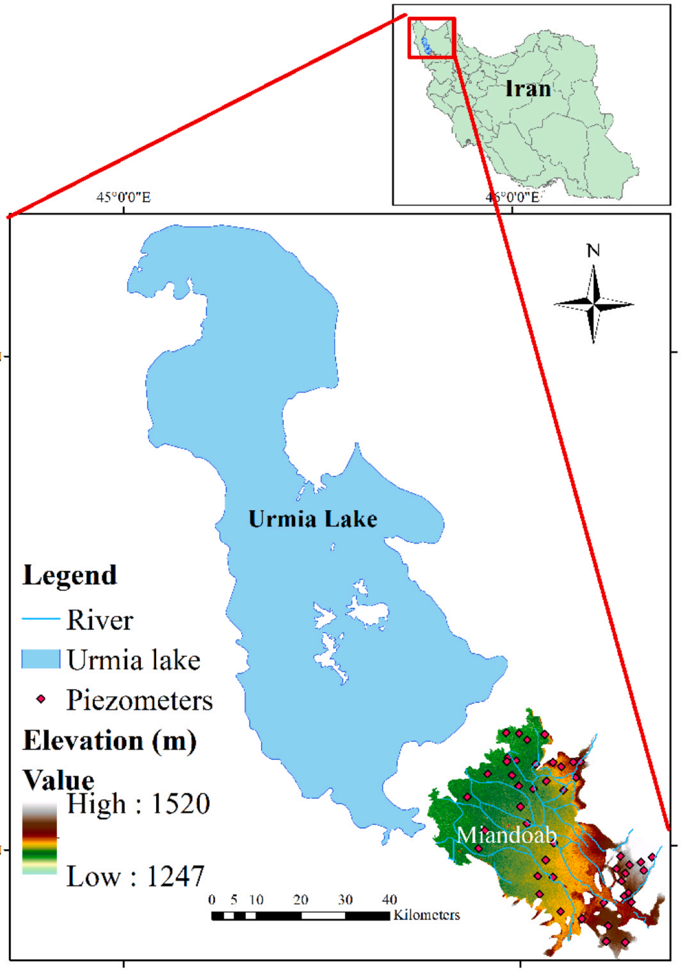
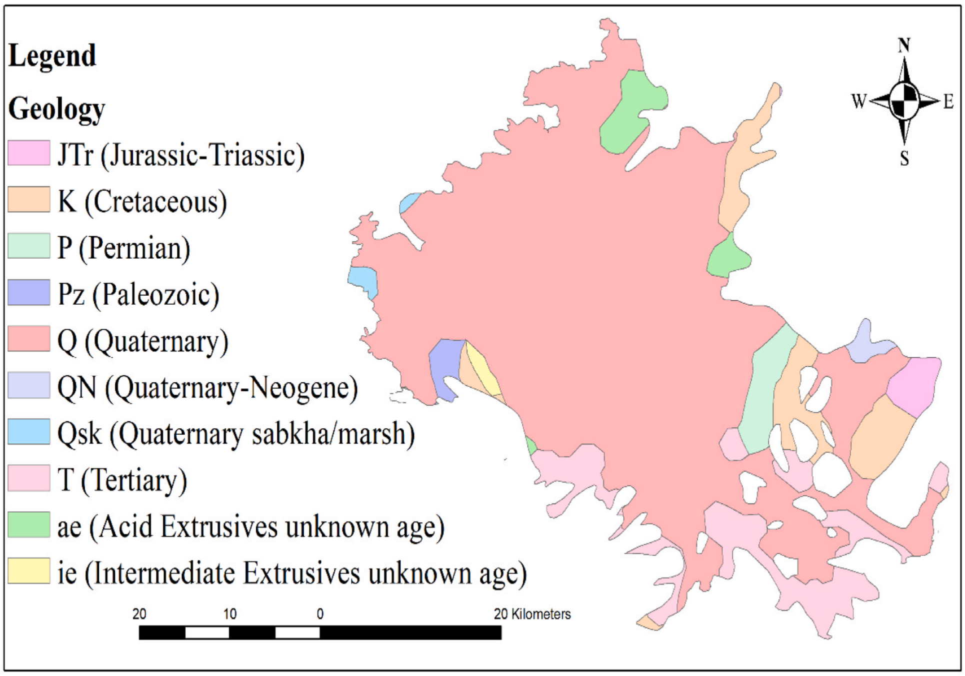


Fig. 1. The overall applied methodology. At first step the appropriate GCM was selected based on DTW. In the second step, piezometers were clustered via ensemble of clustering method. Then, shallow and deep learning methods (Feed Forward Neural Network (FFNN), Adaptive Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR) and their ensemble were considered for GWL modeling and projection for future. At the final step, Fourier model was used to estimate GWQ for 2050 and 2100 and to plot HFE diagram.

a



b



(caption on next page)

Fig. 2. a) The map of the case study b) Geological map of the Miandoab.

Table 1
The applied data for GWL modeling.

Applied data	Resolution	Time period	Source*
GCM	$(1.3^{\circ}-2.8^{\circ})\times(1.5^{\circ}-2.8^{\circ})$	2000–2014	https:// esgf- node. llnl. gov/ search/ cmip6/
GRACE	111 km	2002-present	https://grace.jpl.nasa.gov
FLDAS (SWE_inst, SoilMoi100_200cm_tavg)	11132 m	2000-present	https://developers.google.com/earth-engine/datasets/catalog/NASA_FLDAS_NOAH01_C_GL_M_V001#description
Landsat	30 m	2000-present	https://developers.google.com/earthengine/datasets/catalog/LANDSAT_LE07_C02_T1_L2#description https://developers.google.com/earthengine/datasets/catalog/LANDSAT_LC08_C02_T1_L2#description
GWL, GWQ	-	2000-present	Regional water company of Iran
Hydraulic conductivity	-	-	Regional water company of Iran

* The links were accessed at date 06/1/2023

the appropriate GCM is selected based on Dynamic Time Warping (DTW). In the second step, for determination of the representative piezometers of the aquifer, all piezometers are clustered via the proposed ensemble clustering method. The bias corrected precipitation and temperature of the GCM, and also, GRACE and NDVI data are considered as inputs of the modeling. These data are then projected for future via the Markov chain method. Then, shallow, ensemble and deep learning methods are applied for GWL modeling and projection for the future under climate change scenarios. At the final step, Fourier equation is used to estimate GWQ for 2050 and 2100 and to plot the HFE diagram.

2.2. 2.2 Case study

The Miandoab basin, situated in the northwestern part of Iran, is a sub basin of the Urmia Lake (refer to Fig. 2a). Urmia Lake is located in the northwestern region of Iran, with its latitude ranging from 37°4′ to 38°17′ and longitude spanning from 45°13′ to 46°. In recent years, the water level of the Urmia Lake has experienced a significant decrease, as well as soil degradation and desertification of its watershed. These environmental crises have been attributed to both climate change and anthropogenic activities, which have played significant roles in exacerbating these issues. According to Javadzadeh et al. (2020), there exists a reverse relationship between the GWL of the Miandoab aquifer and the water level of the lake. This suggests that as the lake level increases, the GWL in the aquifer tends to decrease.

The Miandoab area features diverse geological formation. The climate can be characterized as semi-arid to arid. The yearly average of rainfall in the region ranges from 200 to 500 mm, with the majority occurring during the winter season. The arid climate in the area contributes to high rates of evaporation, intensifying the challenge of water scarcity. The Miandoab basin holds great agricultural importance, with the cultivation of crops such as wheat, barley, and fruits (Es' haghi et al., 2022). Unfortunately, excessive exploitation of GW and inadequate implementation of irrigation techniques have created to challenges in Miandoab. In Fig. 2b the geological map of study area is presented. The study area encompasses different geological formations described completely by Norouzi and Moghaddam, (2020).

2.3. Applied data

The GRACE data used for investigation of the GWL fluctuation impact; different GCMs applied for climate change impact assessment on GWL, and Landsat data were used as representative of anthropogenic activities in the modeling process (see Table 1).

The GRACE mission employs two satellites for calculation of the gravitational properties of the Earth and explore its water storage. GRACE provides valuable data on variations of terrestrial water storage anomaly in monthly scale. GRACE was launched in 2002 and operated until its end of mission in 2017. Following its launch in 2018, GRACE-FO serves as a successor mission aimed at sustaining the important data collection efforts initiated by GRACE. GRACE-FO guarantees the uninterrupted continuity of the dataset, enabling the monitoring of long-term trends and changes over time. The vertical water model is used to extract the GW storage anomalies (GWSA), which is influenced from SWSA (surface water storage anomaly), SMSA (soil moisture storage anomaly), SWEA (snow water equivalent anomaly) based on Eq. 1:

$$TWSA = GWSA+SWSA+SMSA+SWEA \quad (1)$$

According to Foroumandi et al. (2023), surface water storage anomalies are negligible in arid and semiarid regions because the TWSA is significantly controlled by GWSA (Eq. 2) and SMSA:

$$GWSA = TWSA - (SMSA+SWEA) \quad (2)$$

The auxiliary data used to calculate *GWSA* for *SMSA* and *SWEA* in this study were collected from the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) dataset in monthly scale (see Table 1).

To facilitate the incorporation of climate change effects, GCMs could be employed in GWL and GWQ modeling. GCMs are essential tools used in climate science to simulate and understand the Earth's climate system. The Coupled Model Intercomparison Project Phase 6 (CMIP6) represents the latest generation of GCMs developed by the international climate modeling community. CMIP6 aims to provide comprehensive assessment of past, present, and future climate conditions by simulating various components of the Earth system, including the atmosphere, oceans, land surface, and cryosphere. CMIP6 GCMs simulate important climate variables such as temperature, precipitation, winds, and sea ice extent, enabling researchers to study climate change impacts, understand underlying mechanisms, and inform policy and decision-makers. The CMIP6 project facilitates model intercomparison and evaluation, allowing for the assessment of uncertainties and improving our understanding of climate projections, ultimately contributing to more accurate and robust climate predictions. In this paper, the historical and Shared Socio-economic Pathways (SSP585) data were used for modeling and projection of GWL. The historical dataset in CMIP6 represents the observed or simulated climate conditions from the past. The SSP scenarios are alternative pathways that describe potential socioeconomic developments and their associated greenhouse gas emissions and climate outcomes. SSP585 presents Fossil-fueled Development - Pathway with high population growth, rapid economic growth driven by fossil fuel use, and limited environmental policies.

For consideration of anthropogenic activities in the modeling, NDVI was employed in this study. NDVI is employed as a quantitative measure of vegetation greenness, providing valuable insights into vegetation density and facilitating the assessment of changes in the plant health. The calculation of NDVI involves determining the ratio between the values of red and near infrared spectral bands, for Landsat 7 and Landsat 8.

2.4. DTW method

The selection of an appropriate GCM (compatible with observed hydro-climatological data of case study) is crucial to achieve accurate and reliable model performance in regional climate studies. The DTW technique is specifically designed to assess the similarity between two sequences of data, such as time-series data. By employing dynamic programming approach, DTW identifies the optimal alignment of two signals. DTW aims to match the patterns exhibited by two signals via identifying the best path through grid mapping, connecting elements of one signal to elements of the other in one-to-many fashion, resulting in minimum distance metric. As a result, DTW provides detailed information regarding differences in signal amplitudes and temporal variations. This method utilizes the Euclidean distance as measure of similarity between the two temporal sequences. Given two time series sequences $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_m\}$ with lengths n and m, a n-by-m matrix, $D_{base} = (d_{base}(u_i, v_j))_{n\times m}$ stores the Euclidean distances between $u_i \in U$ $\forall i = 1, 2, \ldots, n$ and $v_j \in V \forall j = 1, 2, \ldots, m$ as (Belgiu et al., 2020):

$$d_{base(i,j)} = |u_i - v_j| \tag{3}$$

The $d_{i,j}$ is calculated as:

$$d_{i,j} = d_{base(i,j)} + \min\{d_{i-1,j}, d_{i-1,j-1}, \quad d_{i,j-1}\} \tag{4}$$

A warping path, W, is a contiguous set of matrix elements that defines a mapping between U and V. The k^{th} element of W is defined as $w_k = (i,j)_k$, $W = w_1, w_2, \ldots, w_k, \ldots, w_K$. *DTW* is as follow:

$$DTW(U, V) = \min\left\{\frac{1}{K}\sqrt{\sum_{k=1}^{K} w_k}\right\} \max(m, n) \le K < m + n - 1 \tag{5}$$

the path which minimizes the warping cost (Eq. 5) is considered in the modeling process.

2.5. Markov chain

In this study to use NDVI and GRACE data, Markov chain was employed for projection of future values. Markov chain has the ability to forecast future time series data and through analyzing the current state of a system and its transition probabilities, Markov chain can make probabilistic predictions about the future states of the system (Stepchenko and Chizhov, 2015). The Markov chain assumes sequence of discrete states, allowing for the calculation of transition probabilities between these states. It is random process that undergoes transitions within state space, characterized by the property of "memorylessness". This property implies that the probability distribution of the next state solely depends on the current state and is independent of the past sequence of events. A stochastic process $X = \{X_n; n = 0, 1, \ldots\}$ with discrete state space S is considered first-order discrete-time Markov chain if the following condition holds for every $j \in S$ and $n = 0, 1, \ldots, N$:

$$\Pr\{X_{n+1} = j| \tag{6}$$

For any given set of states $i_0, \ldots, i_n$ within the state space, the values of X_i are derived from a finite or countable set S known as the chain's state space. Moreover, Markov chain is said to possess stationary transition probabilities if:

$$\Pr\{X_1 = j| \tag{7}$$

In stationary probability scenario, the process can be fully described by the initial conditions, specifically the one-step transition probabilities. To represent these probabilities, square matrix is utilized and typically denoted by P as follow:

$$P = \begin{vmatrix} P(1,1) & P(1,2) & \dots & P(1,r) \\ P(2,1) & P(2,2) & \dots & P(2,r) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ P(r,1) & P(r,2) & \dots & P(r,r) \end{vmatrix} \tag{8}$$

Where

$$P(i,j) = \Pr\{X_1 = j | X_0 = i\} \tag{9}$$

The number of states, denoted as r, determines the dimensions of matrix P. As P represents probabilities, it is always nonnegative, and the sum of the elements in each row is equal to one. A matrix that satisfies these conditions, where all entries are nonnegative and the row sums are equal to one, is referred to as Markov matrix.

2.6. Clustering

Due to the presence of numerous piezometers in the aquifer, considering all piezometers into the modeling process can be exhausting work. To address data redundancy, clustering techniques can be employed to identify a representative piezometer for each region. This study utilized well-known clustering methods, including K-means (Boongoen and Iam-On, 2018), Self-Organizing Map (SOM) (Kohonen, 1984), Fuzzy c-means (FCM) (Bezdek et al., 1984), and hierarchical clustering (Murtagh and Contreras, 2012). These methods were combined using the LinkCluE clustering approach. Sharghi et al. (2022) highlighted that employing an ensemble clustering could lead to precise outcomes compared to using a single clustering method. Although the mentioned classic clustering methods can be reasonably effective, they may demonstrate varying performance levels when applied to the identical data. Consequently, ensembling could improve the overall performance of clustering. Proposed ensemble method introduced by Iam-on and Garrett (2010) was utilized in this study. The initial stage of the cluster ensemble involves creating a cluster ensemble. $\Pi = \{\pi_1, \dots, \pi_M\}$, which is a collection of M base clustering results. The output E produced from this function is an N × M matrix of cluster labels for N data points from M base clusters. The objective of this method is to enhance the effectiveness of the traditional pairwise similarity approach by incorporating link-based similarity metrics. By doing so, it aims to improve the evaluation of similarity measures between various data points. The LinkCluE method provides three functions for creating such similarity matrix: the connected-triple-based similarity (CTS), the SimRank-based similarity (SRS) and the approximate SimRank-based similarity (ASRS). Their main input argument is a matrix of cluster ensemble E. A similarity matrix S, acquired as the output of CTS, SRS or ASRS, is used together with any similarity-based clustering algorithm to generate the final result.

The computation of the CTS involves (Iam-on and Garrett, 2010):

$$CTS(x_i, x_j) = \frac{1}{M} \sum_{m=1}^{M} S_m(x_i, x_j) \tag{10}$$

where $S_m(x_i, x_j)$ is the similarity between data points x_i, $x_j \in X$. For any ensemble member, $\pi_m \in \Pi$, $m = 1 \dots M$, $S_m(x_i, x_j)$ is calculated as:

$$S_m(x_i, x_j) = \begin{cases} \quad if C(x_i) = C(x_j) \quad 1 \\ otherwise \quad Sim^{WCT}(C(x_i), C(x_j)) \times DC \end{cases} \tag{11}$$

where DC ∈ (0; 1] is a constant decay factor. Sim^{WCT} is similarity between clusters C_i and C_j.

The computation of the SRS involves:

$$SRS(a,b) = \frac{DC}{|N_a||N_b|} \sum_{a' \in N_a} \sum_{b' \in N_b} SRS(a', b') \tag{12}$$

where DC is constant decay factor within the interval (0, 1], *SRS(a, b)* is the entry in the SRS matrix, $N_x \subset V$ denotes the set of vertices connecting to $x \in V$. Accordingly, the similarity between data points x_i and x_j is the average similarity between the clusters to which they belong, and likewise, the similarity between clusters is the average similarity between their members. ASRS matrix is obtained via Eq. 13:

$$ASRS(a,b) = \frac{1}{|N_a||N_b|} \sum_{a' \in N_a} \sum_{b' \in N_b} sim^{clus}(a', b') \tag{13}$$

where N_x denotes the set of vertices connecting to data point $x \in V$ (i.e., a set of clusters to which x belongs) and sim^{clus} (y; z) is a similarity value between clusters y and z, which can be obtained using the weighted SimRank algorithm (Iam-on and Garrett, 2010).

2.7. Modeling via AI models

The AI methods are the commonly used tools for establishing relationship between inputs and target and creating model based on training data, which is able to forecast future condition based on the training data. FFNN, ANFIS, SVR as shallow learning methods and their ensemble were applied in the modeling. The performance of each shallow learning model can vary across different conditions and case studies. In this study, the individual performance of each shallow learning model was examined. Additionally, an ensemble model was employed to enhance overall performance.

FFNN is a widely used and accessible AI modeling tool, while ANFIS is capable of managing uncertainty, and SVR helps mitigate the probability of overtraining. FFNN has gained popularity due to its capability to address problems that are not linearly separable. In practical applications, a three-layer FFNN is often sufficient. The first layer represents the input variables, the second layer serves as the hidden layer, and the third layer functions as the output layer. Connections between layers are established through weights, and each unit calculates the sum of its inputs, incorporating a bias or threshold term, followed by a non-linear activation function that transforms the sum into an output (Bebis and Georgiopoulos, 1994).

ANFIS, initially introduced by Jang (1993), is a hybrid model that combines elements of fuzzy logic and artificial neural networks. This model employs a feed-forward network architecture in which the nodes in different layers handle fuzzy parameters, similar to distributed parameter fuzzy inference systems. ANFIS-based modeling commences by initializing a model using a dataset comprised of input-output pairs. The ANFIS system utilizes the entered membership function and input-output data to optimize membership functions through hybrid and backpropagation algorithms. Once the ANFIS system generates a model, it undergoes validation against predefined criteria.

SVR is an enhanced version of Support Vector Machines specifically designed for regression tasks (Cheng and Shiu, 2014). Unlike traditional regression algorithms that aim to minimize error rates, SVR focuses on fitting the errors within a defined threshold. SVR can be employed with both linear and nonlinear regression models. It introduces a region known as a "tube" around the function being optimized, aiming to identify the tube that provides the best approximation of the continuous-valued function while minimizing prediction errors. SVR utilizes an ε-insensitive loss function that penalizes predictions that deviate beyond a certain ε threshold from the desired output.

In addition to shallow learning methods, recently deep learning models gained attention in modeling hydrological process, but higher number of parameters in their structure is their disadvantages. GRU is one of the deep learning methods, that used in this study. GRU is a type of recurrent neural network. GRU is developed to address challenges such as long-term memory and gradients in backpropagation (Fu et al., 2022). The memory block of GRU is the key feature, which consists of two gates (reset gate and update gate) and a memory cell. The cell state, represented by a horizontal line on top of the block, acts like a conveyor belt that controls the transfer of information to the next moment through the following formulas (Fu et al., 2022):

$$r_t = \sigma(W_r.[h_{t-1}, X_t]) \tag{14}$$

$$z_t = \sigma(W_z.[h_{t-1}, X_t]) \tag{15}$$

$$h_t = \quad = \sigma(W_z.[h_{t-1}, X_t]) \tag{16}$$

Where σ is the sigmoid function. W_r, W_z are weight matrices. z_t is the update gate at time step t, which determines how much of the new candidate activation should be incorporated into the hidden state. X_t is the input at time step t, r_t is the reset gate at time step t,

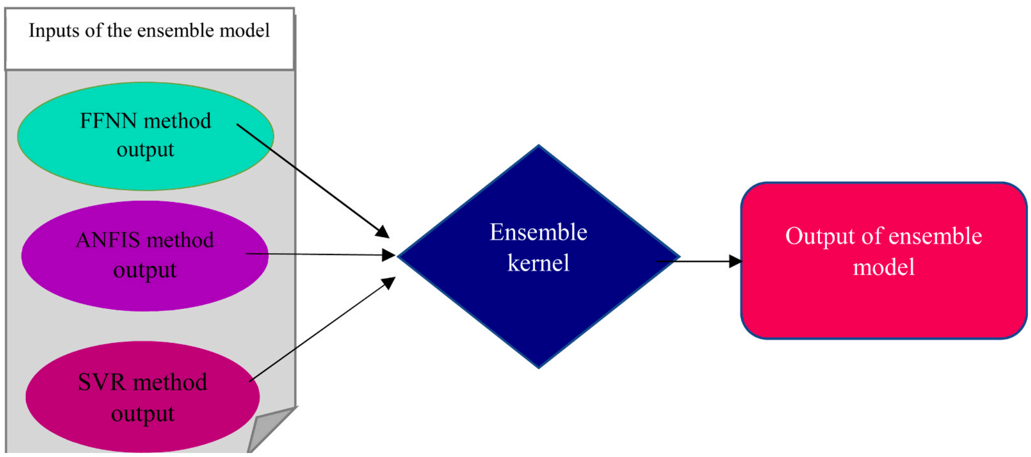


Fig. 3. The diagram illustrating the ensemble method depicts that the inputs to the ensemble consist of the outputs from multiple models.

which determines how much of the previous hidden state should be forgotten. GRU includes reset and update gates in its architecture. These gates take the current time step input x and the previous time step's hidden state h_{t-1} as inputs, and their output is calculated by a fully connected layer with a sigmoid activation function.

By using the sigmoid function, the values of the elements in the reset gate and update gate are constrained between 0 and 1. As a result, each element in the reset gate and update gate has a value within the range of [0,1]. The GRU model calculates candidate hidden states to assist with later hidden state calculations. Specifically, the candidate hidden state $\widetilde{h}_t$ at time step t and hyperbolic tangent function (tanh) is computed as:

$$\widetilde{h}_t = \tanh(W_{\widetilde{h}}.[r_t.h_{t-1}, X_t]) \tag{17}$$

The reset gate determines how the hidden state from the previous time step is incorporated into the candidate hidden state of the current time step. The hidden state at the last time step may contain all the historical information of the time series up to that point. $W_{\widetilde{h}}$, and W_O are weight matrices. Thus, the reset gate can be used to discard irrelevant historical information. The hidden state h_t at time step t is expressed as (Fu et al., 2022):

$$h_t = (1 - z_t).h_{t-1} + z_t.\widetilde{h}_t \tag{18}$$

$$y_t = \sigma(W_O.(h_t)) \tag{19}$$

where $\widetilde{h}_t$is the candidate activation at time step t, which is a combination of the previous hidden state and the new input. The final output is y_t, which is the result of the σ activation on the hidden layer states h_t. GRU networks have a simpler architecture compared to Long short-term memory (LSTM). They also have memory cells, but each cell contains only two gates: reset gate and update gate. Due to its simpler architecture with fewer gates and parameters, GRU is generally computationally more efficient compared to LSTM. GRU networks can be quicker to train and require fewer computational resources, making them suitable in scenarios with limited computational power.

Given the well-established effectiveness of modeling ensembles in previous studies, an opportunity arises to enhance the regionalization process by applying ensemble learning to shallow networks, enabling a parallel learning approach. By combining the outputs of single models, along with an ensemble kernel (any AI model can be used as kernel), it will be feasible to develop a precise and adaptable model that demands less data. In this regard, the outputs of the single models are utilized as inputs for a new network, depicted in Fig. 3, with the GWL as the target variable.

Since deep learning has many parameters, it was not considered in ensemble learning. Ensemble works as parallel processing, with lower parameters, because at first the models are calibrated and then are used in the ensemble technique, but deep learning models have numerous parameters must be trained by larger size of data set.

2.8. Detrended quantile mapping

Detrended quantile mapping is a statistical technique used in climate and hydrological modeling to correct the biases of the

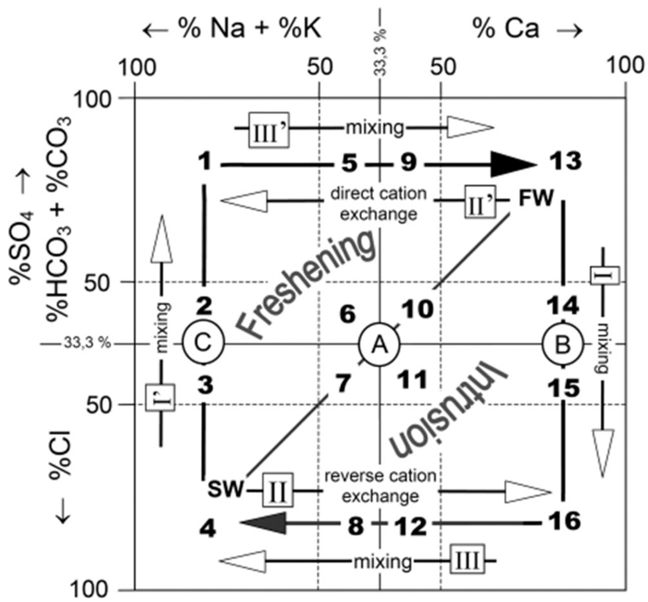


Fig. 4. The HFE diagram (Giménez-Forcada, 2010).

variables, such as temperature or precipitation (Cannon et al., 2015). It is particularly useful when comparing and reconciling climate model outputs with historical or observed data. The main goal of detrended quantile mapping is to adjust the shape and spread of the distribution of the modeled variable to match that of the observed or reference data, while preserving any existing trend in the data. It is an extension of the traditional quantile mapping method, which is commonly used for bias correction. Quantile mapping equates cumulative distribution functions $F_{o,h}$ and $F_{m,h}$, respectively, observed data $x_{o,h}$, denoted by the subscript o, and modeled data $x_{m,h}$, denoted by the subscript m, in a historical period, denoted by the subscript h as follow:

$$\widehat{x}_{m,p}(t) = F_{o,h}^{-1}\left\{F_{m,h}\left[\frac{\overline{x}_{m,h}x_{m,p}(t)}{\overline{x}_{m,p}(t)}\right]\right\}\frac{\overline{x}_{m,p}(t)}{\overline{x}_{m,h}} \tag{20}$$

Where $\overline{x}_{m,h}$ and $\overline{x}_{m,p}(t)$, respectively, involve estimations of the long-term output average, based on historical data projections in the future period p for time t.

2.9. Hydro-chemical facies evolution (HFE) diagram

The HFE diagram serves as an effective tool in GWQ assessment, as it provides visual representation of the complex relationships between different hydro-chemical parameters. It aids in identifying potential water quality factors, such as salinization, pollution, or changes in GW recharge patterns. By tracking HFE, scientists and water resource managers can make informed decisions regarding GW management, protection, and remediation strategies. Gimenez Forcada (2010) proposed HFE as convenient method for assessing the status of coastal aquifer in terms of intrusion or freshening phases occurring over time. These phases can be identified by analyzing the distribution of anion and cation percentages in the square diagram (Fig. 4). The HFE diagram focuses on the major cations (Ca^{2+} and Na^{+}) and anions (HCO_3^-, SO_4^{2-}, and Cl^-) percentages, which are the key factors determining the dynamics of saline/saltwater intrusion. During the intrusion stage, two almost simultaneous processes affect the aquifer. Firstly, there is an increase in salinity (line I), leading to reverse exchange reactions (line II). This results in the characteristic Ca-Cl facies. Subsequently, the composition of GW evolves (line III) towards saline/saltwater facies (Na-Cl). In the freshening process, the introduction of freshwater recharge induces direct exchange reactions (lines I' and II'), leading to the formation of a Na-HCO_3^- facies. Finally, the water evolves along line III' towards the composition of freshwater, allowing for the recovery of the aquifer.

The diagram consists of three sections: (A) Mixing between end members, specifically freshwater and seawater; (B) HFE evolution during the intrusion period; and (C) HFE evolution during the freshening period. There are also labels (I) and (I′) representing the initial process of mixing in the intrusion and freshening stages, (II) and (II') representing reverse and direct base exchange reactions, and (III) and (III') representing mixing during the later stages until reaching equilibrium with the dominant flow's chemical facies.

2.10. Modeling evaluation metrics

The Nash-Sutcliffe efficiency (NSE) and the root mean square error (RMSE) were applied in this study as modeling evaluation metrics. Hybrid Validity Index (HVI) was considered for assessment of the clustering performance.

NSE metric is applied to assess the performance of models and it is between $-\infty$ to 1. It quantifies the agreement between observed and modeled values, with values closer to 1 indicating a better fit and values below 0 indicating poor model performance (Nourani et al., 2019):

$$NSE = 1 - \frac{\sum_{i=1}^{N}(R_i - Z_i)^2}{\sum_{i=1}^{N}(R_i - \overline{R})^2} \tag{21}$$

where R_i, Z_i and $\overline{R}$ are the actual values, modeled values and mean of the observed values, respectively. N is the number of the observations.

RMSE measures the standard deviation of the residuals, which are the prediction errors, serves as an indicator of the dispersion. *RMSE* is calculated as follow:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(R_i - Z_i)^2}{N}} \tag{22}$$

Choosing the most appropriate clustering method can be challenging because of presence of numerous cluster assessment criteria

Table 2
Performance of different methods of clustering.

Method	HVI
Kmeans	58.58
SOM	64.29
Hirarchical	70.77
FCM	51.37
LinkCluE (ensemble clustering)	73.54

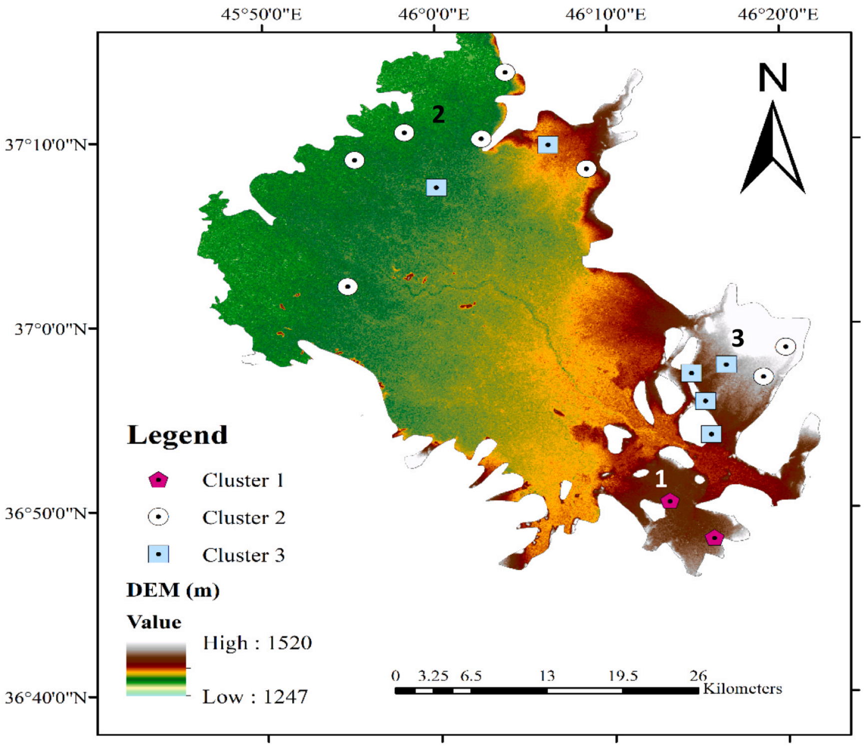


Fig. 5. Spatial distribution of the clusters in the Miandoab aquifer.

Table 3
The value of DTW between GCMs precipitation and observed precipitation values (the links for GCMs are presented in Table 1).

GCM	DTW	Institution	Atmosphere lat/lon grid(°)
BCC-ESM2	9.26	Beijing Climate Center	2.8 × 2.8
KACE−1	9.10	Nat. Inst. of Meteorological Sciences/Korea Meteorological Admin	2.2 × 2.2
IPSL-CM6A-LR	10.12	Institute Pierre Simon Laplace, France	1.3 × 2.5
MCM-UA1	10.86	National Institute for Environmental Studies, Japan	4.5 × 4.5
MIROC-ESM2	9.90	Nanjing University of Information Science and Technology, Nanjing	1.9 × 1.9
NESM2	10.69	Commonwealth Scientific and Industrial	1.2 × 1.8
ACESS-CM2	9.75	Chinese Academy of Sciences, China	2.3 × 2.0
FGOALS	9.90	Norwegian Climate Center, Norway	1.9 × 2.5
NORCPM	13.20	Russia	2 × 1.5
CSIRO	11.99	NASA Goddard Institute for Space Studies, New York, USA	2× 2.5
GISS-E2	9.5	Beijing Climate Center	2.8 × 2.8

that utilize diverse methodologies to evaluate the model efficiency. To streamline the investigation, this study utilized the HVI. The HVI is based on combination of multiple clustering assessment criteria, providing a comprehensive and efficient assessment of clustering performance (Sharghi et al., 2022) as:

$$HVI = \frac{0.5CH + 0.5SI}{DB} \tag{23}$$

Where CH, SI and DB are respectively Calinski-Harabasz (Caliński and Harabasz, 1974), Silhouette Index (Rousseeuw, 1987) and Davies- Bouldin (Davies and Bouldin, 1979) metrics. Superior clustering performance is indicated by greater CH and SI, as well as lower DB metric. Consequently, the overall clustering performance can be quantitatively assessed by evaluating the HVI, where a higher HVI value signifies better clustering performance.

3. Results and discussion

3.1. Results of clustering

Given the existence of multiple piezometers within study aquifer, including all in the modeling process can be a challenging

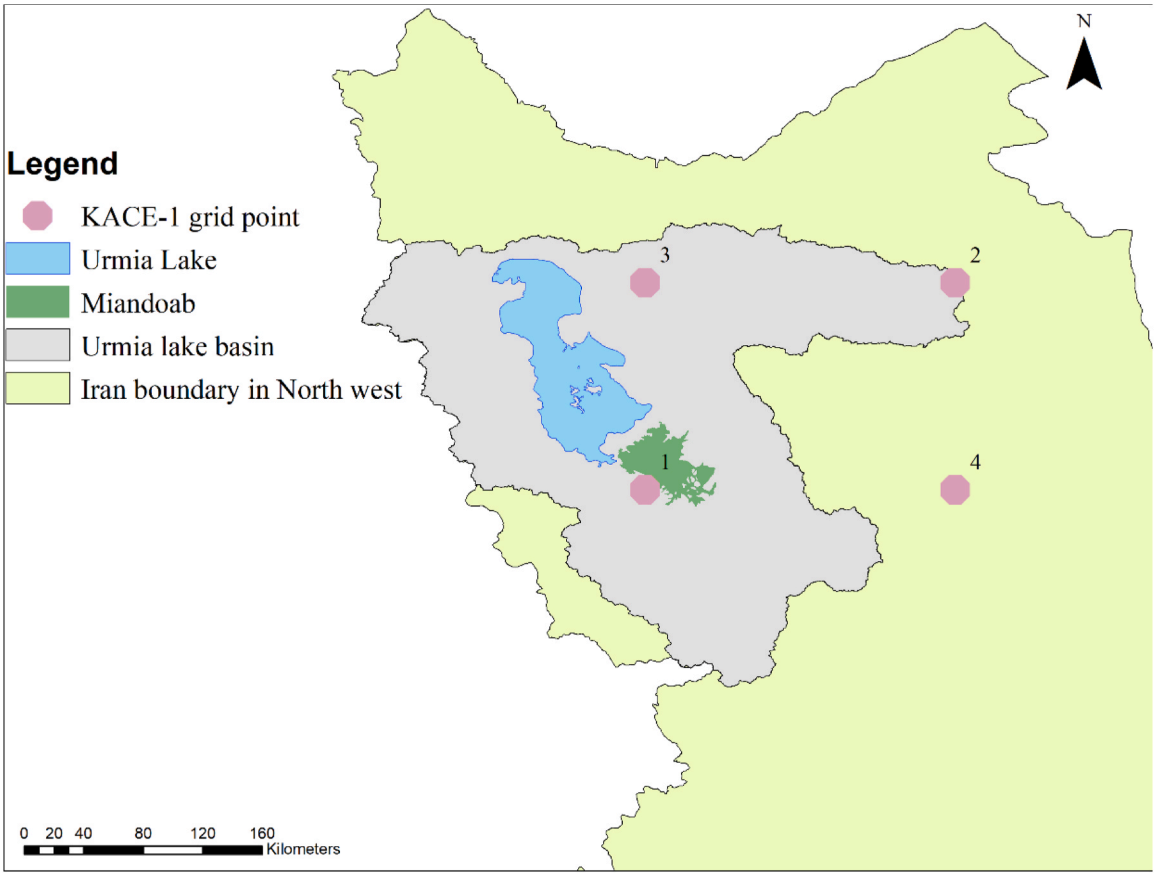


Fig. 6. The grid points of the KACE-1 GCM. The grid point 1 is in the proximity of the case study and its precipitation and temperature were used for modeling.

endeavor. To overcome data redundancy, clustering was employed to identify representative piezometers. In this study, among 153 GWL piezometers, the data from

50 piezometers that cover longer time period were considered. For clustering purposes at first, single clustering methods were employed independently then ensemble of clusters was used for final clusters identification.

In the k-means and FCM methods, respectively five and three clusters were considered by examination of different values. In the FCM method, the iterations were set up to 100, and a minimum improvement of $1e^{-5}$ in the objective function between consecutive iterations was considered. For the SOM method, a 2×2 grid was used to define the dimensions. In hierarchical clustering, the weighted

Table 4
Performance of the modeling based on NSE and RMSE metrics. The input of modeling consists of precipitation, temperature, GRACE, NDVI and GWL with one month lag.

Region	Method	NSE train	NSE verification	RMSE train (m)	RMSE verification (m)
piezometer 1	FFNN	0.95	0.90	0.12	1.00
	ANFIS	0.97	0.92	0.11	0.95
	SVR	0.95	0.85	1.28	1.83
	Ensemble	0.98	0.96	0.10	0.15
	GRU	0.86	0.61	1.38	3.13
piezometer 2	FFNN	0.94	0.82	0.58	0.94
	ANFIS	0.95	0.88	0.55	0.80
	SVR	0.88	0.74	0.65	0.83
	Ensemble	0.98	0.97	0.05	0.10
	GRU	0.65	0.65	1.87	1.45
piezometer 3	FFNN	0.90	0.75	0.17	0.20
	ANFIS	0.98	0.85	0.02	0.07
	SVR	0.96	0.71	0.21	0.30
	Ensemble	0.97	0.93	0.02	0.05
	GRU	0.85	0.70	0.37	0.33

linkage method was chosen to calculate the distance between clusters, and the Euclidean distance metric was employed to measure the dissimilarity between data points. The final number of clusters obtained by k-means, hierarchical, SOM, and FCM methods were 5, 3, 5, and 3, respectively. Consequently, a total of 17 piezometers, representing multiple clusters, were considered by removing duplicate piezometers and excluding clustering methods with subpar performance. These representative piezometers were utilized as inputs for the proposed approach. The output of ensemble resulted in three clusters, with 1, 8, and 6 piezometers assigned to clusters 1, 2, and 3, respectively. Cluster 2 housed the majority of the piezometers, indicating a higher concentration in that particular cluster. The evaluation of clustering methods was conducted using predefined metrics, and the corresponding results are summarized in Table 2. According to the evaluation metrics presented in Table 2, the ensemble method LinkCluE exhibited superior performance compared to the individual techniques. The LinkCluE approach takes into account the collective merits of multiple clustering methods, concurrently. The LinkCluE approach effectively mitigates the influence of biases and errors inherent in individual models, resulting in a more comprehensive and accurate representation of cluster structures. The location of three clusters is depicted in Fig. 5. The piezometers in clusters 1, 2 and 3 are located in Southern, North and East parts of the Miandoab aquifer, respectively.

Cluster 1 located utmost distance from the Urmia Lake (see Fig. 2) and also in higher elevation compared to the piezometers in clusters 2 and 3. The piezometers in cluster 2 are in the proximity of the coastline and in lower elevation of the aquifer. The majority of the piezometers in cluster 3 located at the east part of the aquifer.

3.2. Results of the GCM selection

In order to assess the impact of climate change on GWL and GWQ, GCMs were considered as input of the modeling. From multiple GCMs, in order to select the appropriate GCM, DWT was applied in the modeling. Since precipitation is the most influential parameter in GW modeling, the DTW values were calculated for precipitation data from different GCMs (precipitation of the nearest grid point to the case study was used in analysis) and the observed time series of precipitation in the case study and the results are tabulated in Table 3.

The GCM with lower DTW distance from the observed precipitation was selected for further analysis and applied in the modeling process. In this way, the DWT was lowest for KACE-1, so its data were used for the subsequent modeling process across all clusters. The parameters of GCM from the nearest grid point to the case study were considered as inputs of the models (see Fig. 6, grid point 1).

Since GCMs are subject to uncertainties and potential sources of error due to various factors like the inherent complexity of the Earth's climate system, computational constraints, thus the bias correction of GCMs is essential before to be used in the modeling. In this study, detrended quantile mapping was used for bias correction (Cannon et al., 2015). Then bias corrected precipitation and temperature were then used for GWL modeling.

3.3. Results of modeling via AI models

The importance of input selection in AI models cannot be overstated, as it directly impacts the performance, efficiency, and interpretability of the modeling. The correct selection of relevant and representative input features is crucial for capturing the essential

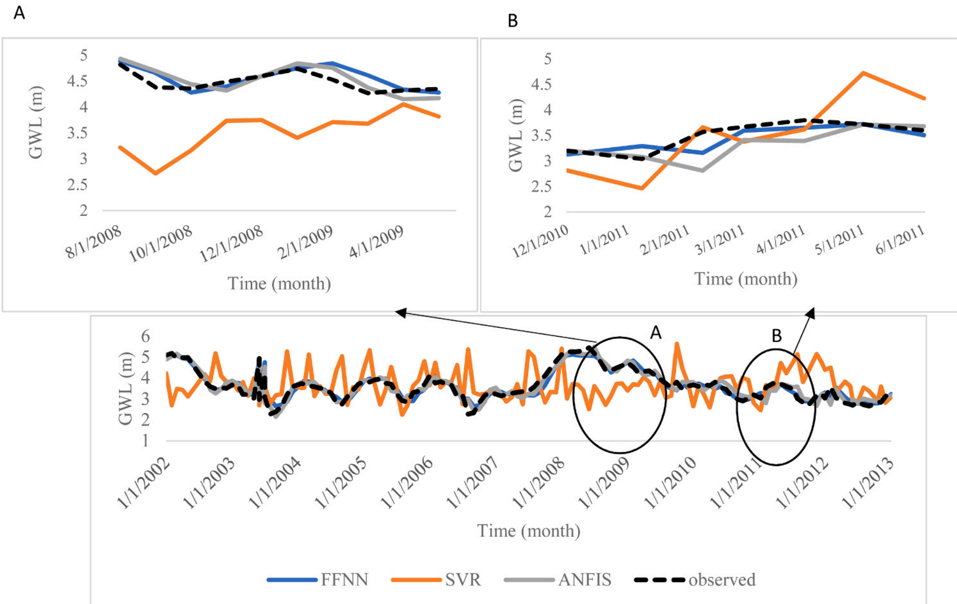


Fig. 7. The observed and output of the individual methods. In the detail of sections, A and B, it is obvious that in section A, ANFIS outperformed the other methods, while in section B the SVR outperformed the other methods.

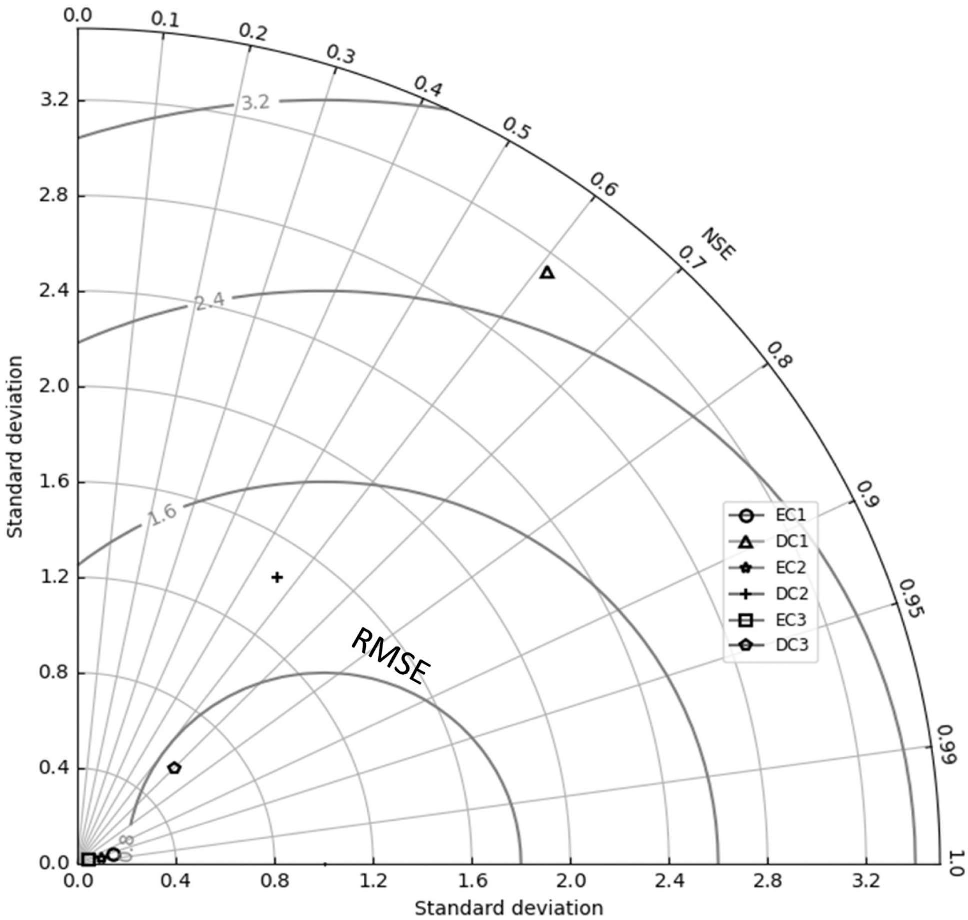


Fig. 8. The Taylor diagram of the ensemble and deep learning models for clusters 1,2 and 3. The shallow learning methods are closer to coordinate origin and also x axis, which indicates their more accurate performance. Cn, n=1, 2 and 3 indicate to the cluster number; E and D present ensemble of shallow learning models and deep learning model, respectively.

patterns and relationships in the data, reducing noise and irrelevant information, and improving the model's ability to be generalized to the unseen data. The precipitation and temperature affect the evapotranspiration; thus, they have significant impact on GW and multiple studies have used these parameters as inputs for GWL modeling (e.g. see, Foroumandi et al., 2023). Therefore, these two parameters from KACE-1 were used in input layer of AI models after bias correction. GRACE and GRACE-FO data as the representative of the GW storage anomaly and GWL with 1-month lag time were the other inputs.

Incorporating a one-month lag time in the modeling of GWL has demonstrated the potential to enhance modeling accuracy. Moreover, the time series of GWL with a one-month lag exhibits a higher correlation with the primary GWL time series. This approach has been previously employed in studies such as Javadinejad et al. (2020), where a one-month lag was considered in GWL modeling. Furthermore, Tao et al. (2022) emphasized that fluctuations in GWL serve as a direct indicator of the impact of GW development and provide valuable insights into aquifer dynamics within GWL time-series data. Consequently, there is a high possibility to accurately forecast future GWL from its previous data.

Moreover, NDVI inputted to the model as the representative of the anthropogenic activities. During the modeling process, 70 % of the data were assigned for training purposes, while the remaining 30 % were allocated for the verification. The optimal epoch number was determined through iterative examination. The applied data were from 2002 to 2012 for the modeling as the KACE-1 GCM was available up to 2012 and for the recent years (as the observed values were available) the model was reconfirmed. The Levenberg–Marquardt and Tansig were considered as learning algorithm and active function, respectively (Nourani et al., 2019). The kernel function of the SVR was radial basis. The layer and unit number in the GRU method set as one. This adjustment was found to be effective in optimizing the model's performance. Deep learning is generally more suitable for time series data with longer lengths, and minimum layers and units may enhance the efficiency when dealing with shorter datasets. The sigmoid function was determined to be the most suitable activation function in training GRU. The mean squared error was employed as the loss function, and the Adam optimizer algorithm was utilized as the learning algorithm. Fine-tuning of the hyperparameters within the GRU model was performed to determine the optimum parameters. Parameters such as the number of GRU layers, the number of hidden units in each layer, the learning rate, batch size, dropout rate, and number of epochs, as well as the window size, were investigated, with values ranging up to

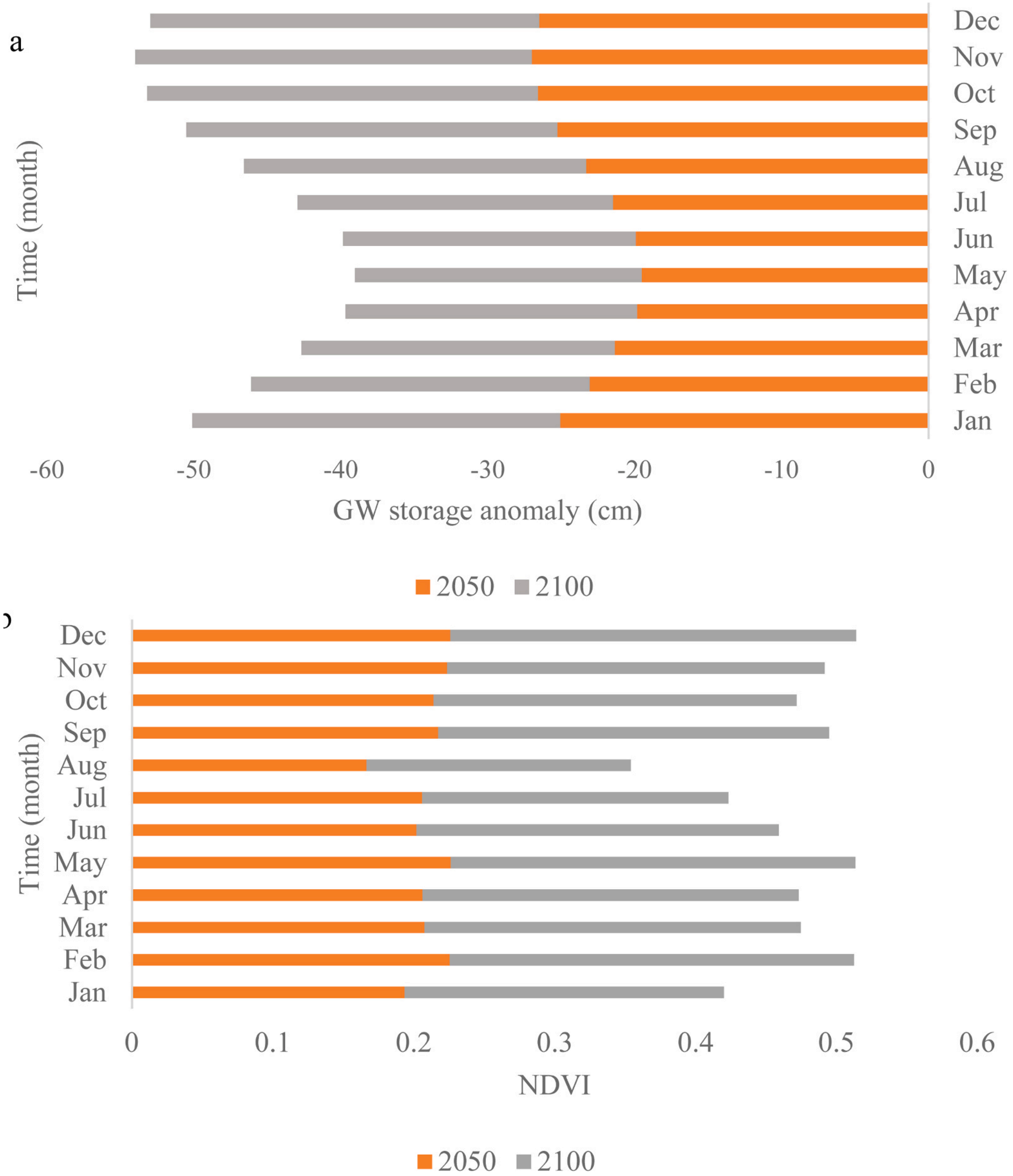


Fig. 9. Predicted monthly values via the Markov chain for a) GRACE data b) NDVI. The monthly anomalies of GW storage are computed with regard to the time-mean baseline that includes all months.

5, 10, 5, 5, 5, 1000, and 4, respectively. The outcomes of the applied methods and the ensemble approach are presented in Table 4.

Investigation of the results shows that for among shallow learning models, in all cases, the performance of ANFIS was more accurate (up to 6 %) than the performance of the FFNN. ANFIS demonstrating its superior performance compared to FFNN due to its more versatile and adaptable structure, which considers the strengths of neural networks and fuzzy logic. ANFIS flexible structure allows to adapt its architecture and parameters to accommodate a broader range of problems by leveraging the input data. Fig. 7 illustrates the output of the individual models. The analysis of Fig. 7 reveals that ANFIS displayed superior performance in certain instances, while SVR exhibited better results in others. This observation underscores the importance of employing ensemble learning techniques, as they enable the final model to leverage the strengths of all three individual models. Consequently, an ensemble of models was incorporated into the modeling process to enhance the effectiveness and efficiency of the shallow learning models. The parallel processing approach was used to ensemble the results of the applied individual models and the outputs of the individual models were regarded as inputs for modeling of ensemble model. For training the ensemble method, ANFIS was utilized as the kernel due to its satisfactory performance compared to other individual methods.

It should be noted that, alternative models can be employed as the kernel for the ensemble. The findings demonstrated that the ensemble of models could significantly enhance the efficiency of individual models by up to 35 %. The collective power of an ensemble

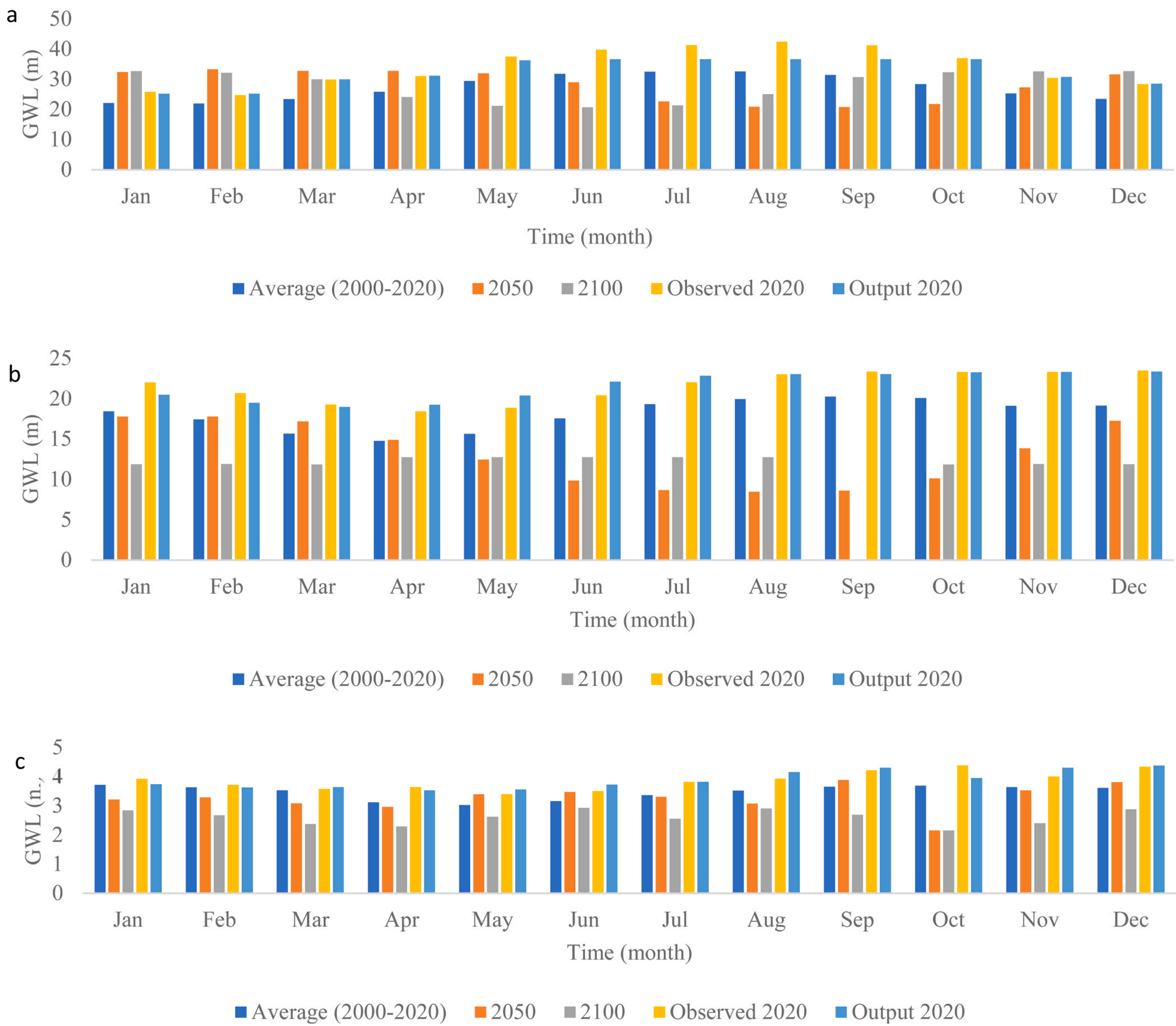


Fig. 10. The GWL monthly average values, observed values in 2020, outputs in 2020, 2050 and 2100 for representative piezometers of a) cluster 1 b) cluster 2c) cluster 3. The results indicated that GWL in 2050 and 2100 will have decreasing trend for clusters 2 and 3. For cluster 1 there is declining trend in dry season but increasing in wet season.

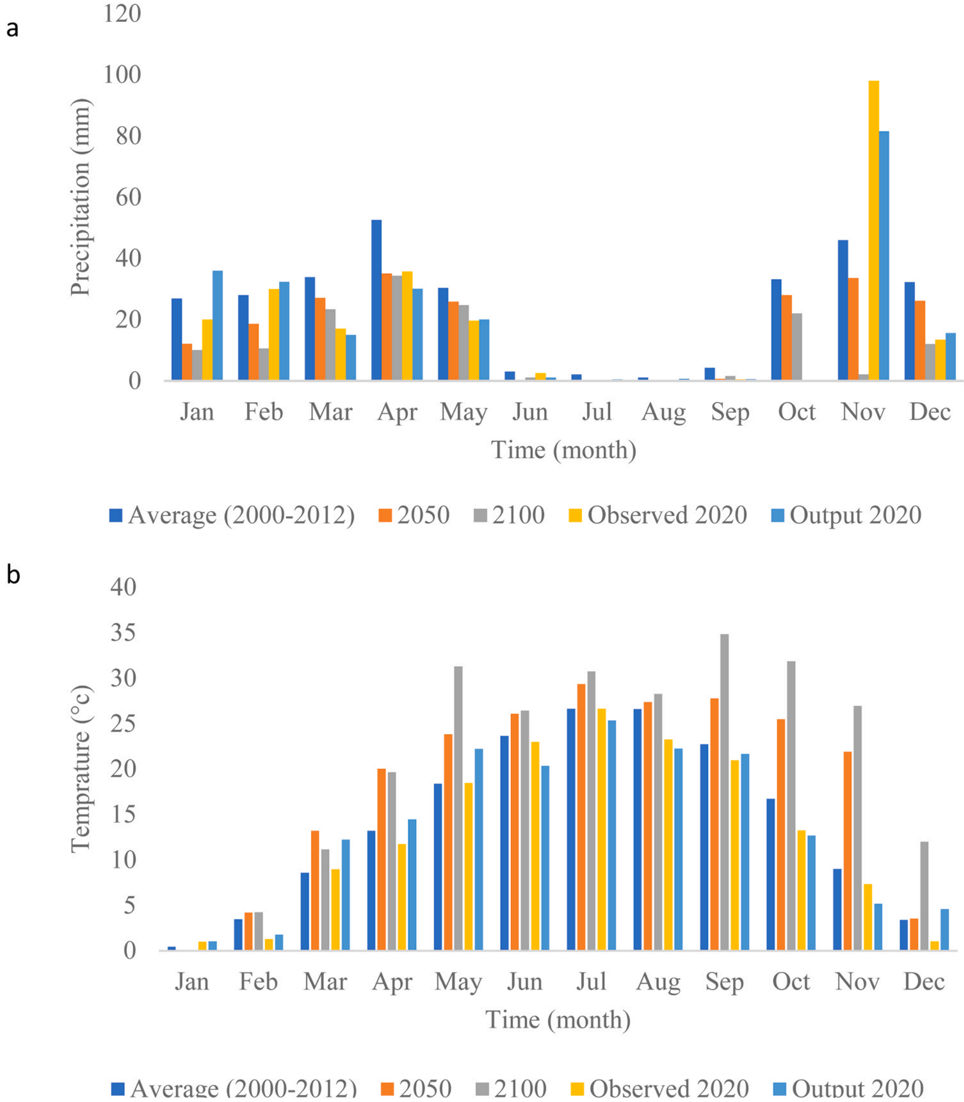


Fig. 11. The monthly average of data for historical period (2000–2012), observed 2020 and output in 2020, 2050 and 2100 for a) precipitation b) temperature.

lies in its ability to capture a wide range of distinctive data characteristics that may not be able to effectively detected by the individual models. Also, overfitting can be mitigated through ensembling, resulting in improved performance. Furthermore, ensembling contributes to the enhanced generality of the model by leveraging the advantages of multiple individual models. Every individual model within the ensemble possesses the capability to detect distinctive patterns, and when combined, patterns are integrated effectively to enhance the final outcome. In Fig. 8, the Taylor diagram of the applied methods is depicted. As shown in Fig. 8, the ensemble of shallow learning models is closer to the coordinate origin, which shows better performance. Moreover, evaluation of the modeling performance indicated that the ensemble learning outperformed deep learning up to 41 % based on the NSE metric. The modeling conducted in this study was on a monthly scale, and given that the historical data from GCM are only available up to 2012, the time series were relatively short. Consequently, deep learning methods did not exhibit optimal performance as they typically require extensive data to effectively calibrate their numerous parameters. On the other hand, ensemble, which is a parallel learning process and it is independent of large dataset, models individually and the obtained results are more accurate.

3.4. Projection of GWL and quality for future

Once the model (ensemble of shallow learning) and GCM (KACE-1) with optimal performance were identified, they were utilized to forecast the future state of GWL and GWQ. To validate the constructed model, it was applied to simulate the states for 2020 and consequently, and compare to the presently accessible observed values of 2020. To forecast the future state, bias corrected

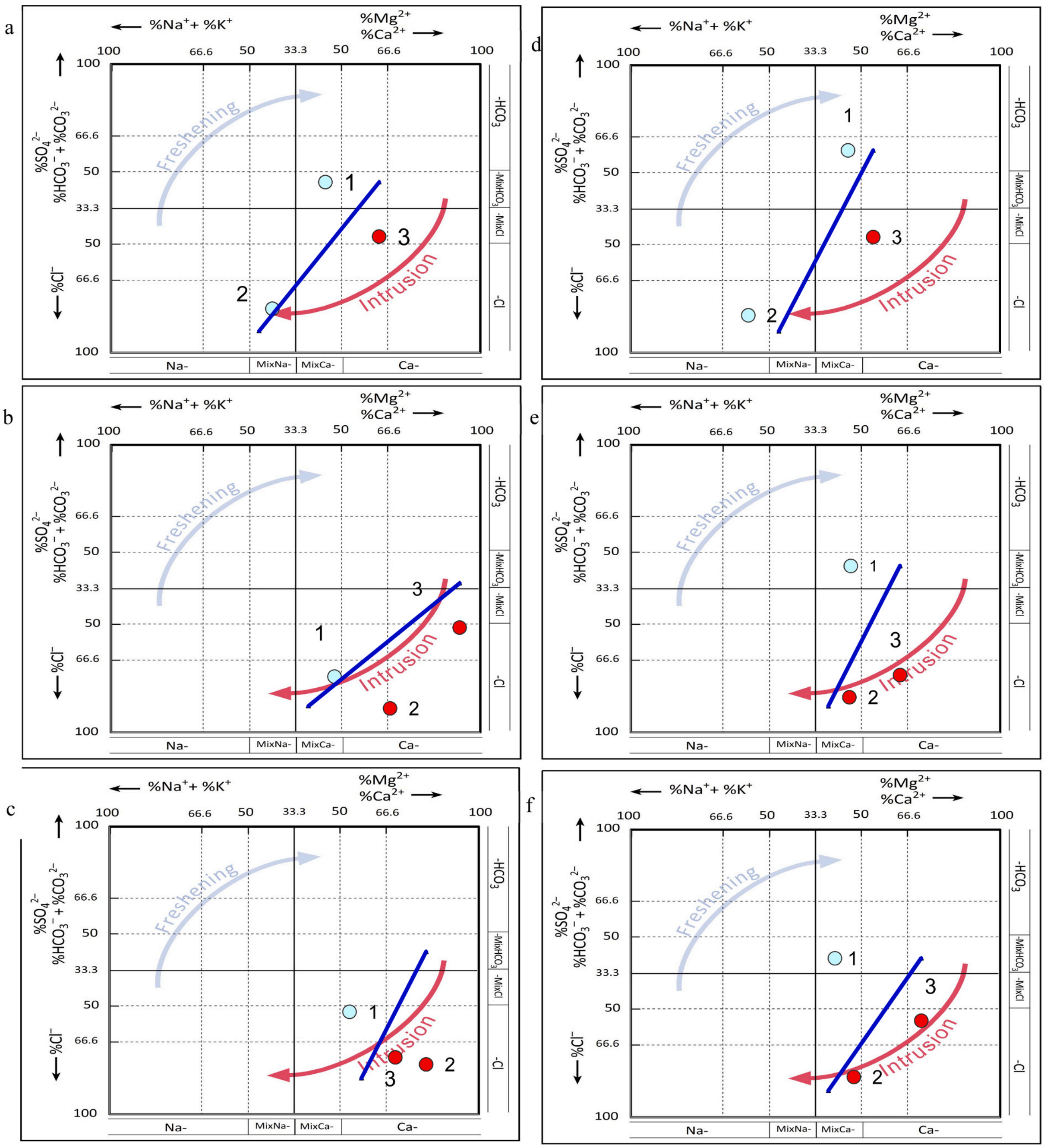


Fig. 12. The HFE diagram for three clusters of region a) average (2000–2012) dry seasons b) 2050 dry season c) 2100 dry season d) average (2000–2012) wet seasons e) 2050 wet season f) 2100 wet season. There are intrusion facies for clusters 2 and 3 in 2050 and 2100. For cluster 1 in dry seasons representative point of cluster 1 becomes closer to the mixing line.

precipitation and temperature parameters of the GCM's SSP585 that represent high-emissions scenario were used to project GWL for 2050 and 2100. To find out the NDVI and GRACE data values for 2050 and 2100, the Markov chain method was used and its results imposed as inputs of the projection model. The predicted monthly values of NDVI and GRACE data for future via the Markov chain are shown in Fig. 9.

The time series of the GWL for the average of the modeling period (2000–2012), and predicted values of the GWL for 2020, 2050 and 2100 are depicted in Fig. 10. For revalidation of the constructed model, the model was performed for 2020 and the results was compared with the observed values in 2020, which are available now (see, Fig. 10). According to Fig. 10, for central piezometers of clusters 1, 2 and 3, the GWL may decrease down to 11.74 m, 11.64 m and 1.53 in 2050 and 8.23 m., 8.59 m and 1.19 m in 2100, respectively; however, there may be an increase in GWL for cluster 1 in wet seasons. The value of the precipitation and temperature for

Table 5
Hydro-chemical facies of GW based on HFE diagram for clusters in dry and wet seasons.

Season	Cluster	Water type in average of past period	HFE facies	Water type in 2050	HFE facies	Water type in 2100	HFE facies
Dry	1	MixCa-MixHCO3	Fresh.	MixCa-Cl	Fresh.	Ca-Cl	Fresh.
	2	MixNa-Cl	Fresh.	Mg-Cl	Intrus.	Ca-Cl	Intrus.
	3	Ca-MixCl	Intrus.	Ca-Cl	Intrus.	Ca-Cl	Intrus.
Wet	1	MixCa-MixHCO3	Fresh.	MixMg-MixSO4	Fresh.	MixMg- MixSO4	Fresh.
	2	MixNa-Cl	Fresh.	MixMg- Cl	Intrus.	MixMg- Cl	Intrus.
	3	Ca-MixCl	Intrus.	Ca- Cl	Intrus.	Ca- Cl	Intrus.

average of

historical period (2000–2012), 2050 and 2100 are presented in Fig. 11. It is worthy to mention that projected values are based on high-emissions scenario, moreover the GCMs data consist of

uncertainty and in this study only bias correction was applied and downscaling was not used for the GCMs data.

Furthermore, the scenarios used for future projections involve inherent uncertainty. The applied scenarios rely on extreme and pessimistic assumptions, which means that the resulting future temperature and precipitation values may appear to be exaggerated. Changes in temperature and precipitation patterns can have impact on GWL. Higher temperature and lower precipitation can lead to increased evaporation rates, which can cause decrease in the GWL. When temperatures rise, evaporation rates increase, resulting in greater water loss from the aquifer. Additionally, reduced precipitation means less recharge to the aquifer, further contributing to decline in GWL. The results showed decreasing and increasing trends for precipitation and temperature, respectively. Moreover, Nourani et al. (2019) conducted downscaling of precipitation and temperature data for the years 2050 and 2100, leading to the conclusion that there will be a decreasing trend in precipitation and an increasing trend in temperature in the future for the area.

Dehghanipour et al. (2019) expressed that GW head in the Miandoab aquifer has dropped over time due to a reduction of precipitation as well as an increase of GW withdrawal for agricultural purposes.

The positioning of the piezometers within cluster 1 indicates a considerable distance from Urmia Lake. The Urmia Lake (refer to Fig. 2) has a mitigated impact on the piezometers within its vicinity, resulting in lower variation in response to the lake level fluctuations. Additionally, cluster 1 is situated in the higher altitude lands compared to the remaining clusters, which can influence the wells within cluster 1 to exhibit increase in GWL. Furthermore, in elevated regions, agricultural activities and the presence of wells are usually limited. Cluster 1 consisted of only two piezometers, whereas most of the wells were distributed across clusters 2 and 3, which exhibited a declining GWL. This observation leads to the conclusion that the study aquifer is experiencing a significant concern regarding the decline of GWL. In addition, Javadzadeh et al. (2020) mentioned that the water level in the observation wells close to the lake, especially in the eastern and southern coasts of the lake (e.g. Miandoab aquifer) experience lower GWL than the ecological water level. The aquifer type throughout the entire region is consistent, being unconfined. Therefore, it seems that the variation in GWL was not influenced by the aquifer type.

In order to investigate GWQ, HFE diagram was plotted for the average parameters during the modeling period (2000–2012) as well as for predicted parameters for future (2050 and 2100). GWQ parameters were predicted by establishing correlation between GWL and GWQ parameters. Correlation analysis was used here instead of modeling via the AI models because GWQ data were reported only

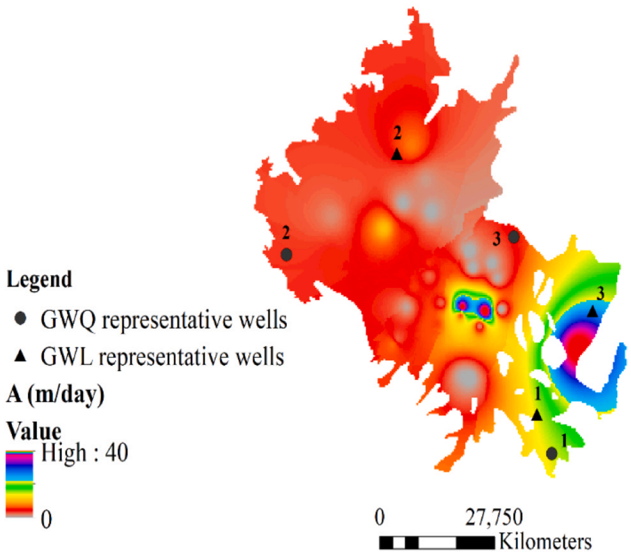


Fig. 13. Hydraulic conductivity map of the case study and location of the representative GWL and GWQ piezometers of clusters 1,2 and 3.

twice a year, therefore the length of GWQ time series were too short to be modeled via the AI models. The GWL time series were in monthly scale while GWQ time series were 6 monthly data, so the GWL time series were adjusted to match the 6-month interval. To establish the GWL-GWQ relationship using historical data, Curve fit toolbox was used for the modeling and among multiple equations of the Curve fit toolbox, the Fourier model outperformed the other equations. The GWQ modeling parameters was calibrated and verified via Fourier model. The better performance of the Fourier equation could be due to its sin and cos kernel functions, proves to be more effective in modeling hydrologic processes exhibiting seasonality. Subsequently, the predicted GWL values were used to estimate GWQ parameters.

Additionally, to consider the Urmia Lake water quality parameters, based on assumption that the current condition of the lake would be stable, linear regression was employed to project the data related to the quality parameters of the Lake for 2050 and 2100. The HFE diagrams for average of the modeling period (2000–2012), 2050 and 2100 are presented in Fig. 12 and the facies are tabulated in Table 5. Based on Fig. 12, the cluster 1 for average of the historical data (2000–2012) in both wet and dry seasons has refreshing facies (located in part 10 in Fig. 12). This may be due to the far distance of the cluster 1 from the shoreline. The extent of saltwater intrusion is directly correlated with the perpendicular distance from the coast. It is evident that the maximum level of saltwater intrusion occurs when the aquifer is in close proximity to the shore and favorable hydrogeological conditions exist for transmission. However, the hydraulic conductivity is high in this part of the case study, but its more distance and higher GWL prevent seawater intrusion. The hydraulic conductivity of the aquifer plays critical role in determining the vulnerability of GW to seawater intrusion, as it governs the ease with which saltwater can infiltrate and contaminate the freshwater aquifer. As tabulated in Table 5, in wet seasons there is Mixed Ca–Cl water type, which is compatible with study of Amiri et al. (2016) expressing that in wet period most of the samples show water type of Mixed Ca–Mg–Cl.

The hydraulic conductivity map of the case study is presented in Fig. 13. The extent of the saline waterfront's spread under constant hydraulic pressure is directly associated with the hydraulic conductivity of the GW system.

Due to the aquifer's high hydraulic conductivity, the cone of depression during pumping expands, leading to an increased likelihood of saltwater up-coning. The representative point of cluster 2 with Mg-Cl facies indicates simple binary mixing with little or no intervention of base exchange reactions. However, it is near the coastline and the lower hydraulic conductivity of this part decreases the seawater intrusion effect. The representative point of cluster 3 placed in the right and beneath line A and below the horizontal line at 33.3 %, which is initial process of mixing in intrusion. During the phase of seawater intrusion, there is an initial increase in salinity and rapid and marked reverse exchange of $Na^{+}\backslash Ca^{2+}$, which is recognized by the characteristic Ca-Cl facies. This might be due to the high hydraulic conductivity of the area of cluster 3 (see, Fig. 13).

GWL below sea level holds significant importance as it increases the vulnerability to seawater intrusion. The GWL plays critical role in maintaining hydraulic pressure along the coast, which helps prevent the intrusion of saltwater. Water level in representative well of cluster 2 is the highest and for cluster 3 it is the lowest. As the freshwater level decreases, the intrusion of the saltwater front extends further into the aquifer, thereby increases the risk of aquifer vulnerability. The likelihood of saltwater intrusion is low in an aquifer where the freshwater level is significantly higher than sea level. In the HFE diagram of 2050 for dry season, representative well of cluster 1 became near the mixing line (part 12 in Fig. 12); this might be due to the decline of GWL in which discharge is minimal in dry season in both 2050 and 2100 as presented in Fig. 10. The position of representative well of cluster 1 doesn't change in wet season and the Fig. 10 shows increase in GWL for wet season. For representative points of clusters 2 and 3, both points moved to right and left parts of the mixing line, which indicates the intrusion in both dry and wet seasons. Moreover, the GWL in these two clusters in both 2050 and 2100 shows decreasing trend. Therefore, it could be concluded from the obtained results that there is a chance to the saltwater intrusion into the aquifer under the influence of climate change and anthropogenic activities. The obtained results in this study are compatible with the results of Jeihouni et al. (2018) which stressed that high drop rate in the water table leads to reduction of freshwater pressure in the transition zone, and accordingly the saltwater penetrates into the surrounding plains aquifers in the Urmia Lake area. Regarding the obtained results of this study, the GW table has a downward trend at surrounding aquifers of the Urmia Lake. However, the water level of Urmia Lake has a descending trend during the study period, but the downward trend of water table is more expressed than the one for the water level of the Lake. This pressure change caused saltwater intrusion from saltwater aquifers beneath the lake to surrounding plains freshwater aquifers.

The importance of the results of this research is about prediction of the GWL and GWQ for future under consideration of the climate change impacts and scenarios and also anthropogenic activities. The findings of this study provide valuable information and insights to policymakers and decision-makers, enabling them to effectively manage GW resources and regulate their utilization. By considering intrusion probabilities in the aquifer, appropriate measures can be implemented to control and safeguard GW resources.

4. Conclusions

In this study, GWL and GWQ were modeled and predicted for future under climate change effects. For this purpose, the shallow learning methods and their ensemble, and deep learning model of GRU were applied. Among different GCMs DTW method selected KACE-1 as the most appropriate GCM for GWL modeling. Clustering ensemble method of Link-Clue was used for clustering the piezometers of the case study leading to three clusters. The precipitation, temperature of KACE-1 GCM, NDVI, GRACE data and GWL with a lag were used as inputs of modeling GWL and GWQ for representative (centroid) piezometers of the clusters (one for each cluster).

Evaluation of the modeling performance indicated that ensemble learning outperformed the deep learning up to 35 % based on the NSE metric. In addition, it was concluded that the ensemble model improved the performance of individual models up to 23 % in this study. For future projection, NDVI and GRACE data were predicted for 2050 and 2100 via the Markov chain. Projection of the GWL under the SSP585 showed that in 2050 and 2100, there will be decline in GWL down to 20.2 m, especially for the piezometers of clusters 2 and 3. The trained model was used for prediction of GWL for future. Then the Fourier model was used to make relationship between the GWL and GWQ and HFE diagram was plotted for quality parameters in 2050 and 2100. Results indicated that in 2050 and 2100, GW may have intrusion facies especially near the shoreline. Comparison of the obtained results of this study with the results of other researchers showed that there was agreement about the saltwater intrusion in the regions near the Urmia Lake and also GWL decreasing trend in Miandoab aquifer due to reduction of precipitation and anthropogenic activities. In addition, in wet seasons there is Mixed Ca–Cl water type which is compatible with previous studies.

As suggestion for future studies, for improvement of the modeling performance under climate change, future studies can fill the gap in GCMs data. Currently, GCMs provide data only until 2012, resulting in 10-year gap in historical data and the training process. In addition, the uncertainty associated with the modeling of GWQ and GWL can be quantified via different methods of uncertainty quantification. Future studies can investigate the factors influencing GWQ, such as land use practices, industrial activities, and agricultural inputs, and develop models to predict changes in GWQ under different scenarios.

CRediT authorship contribution statement

Nardin Jabbarian Paknezhad: Software, Investigation, Formal analysis. **Vahid Nourani:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization. **Sameh Ahmed Kantoush:** Project administration, Funding acquisition. **Zhang Wen:** Resources, Data curation.

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.

Data Availability

Data will be made available on request.

Acknowledgements

This study was supported by the JSPS Core-to-Core Program (Grant number: JPJSCCB20220004). The MEXT/JSPS fellowship number is L24526.

References

Amiri, V., Nakhaei, M., Lak, R., Kholghi, M., 2016. Assessment of seasonal groundwater quality and potential saltwater intrusion: a study case in Urmia coastal aquifer (NW Iran) using the groundwater quality index (GQI) and hydrochemical facies evolution diagram (HFE-D). Stoch. Environ. Res. Risk Assess. *30*, 1473–1484.
Baghanam, A.H., Nourani, V., Keynejad, M.-A., Taghipour, H., Alami, M.-T., 2019. Conjunction of wavelet-entropy and SOM clustering for multi-GCM statistical downscaling. Hydrol. Res. *50* (1), 1–23.
Bebis, G., Georgiopoulos, M., 1994. Feed-forward neural networks. IEEE Potentials *13* (4), 27–31.
Belgiu, M., Zhou, Y., Marshall, M., Stein, A., 2020. Dynamic time warping for crops mapping. The International Archives of the Photogrammetry. Rem. Sens. Spatial Inf. Sci. *43*, 947–951.
Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. *10* (2-3), 191–203.
Boongoen, T., Iam-On, N., 2018. Cluster ensembles: a survey of approaches with recent extensions and applications. Comput. Sci. Rev. *28*, 1–25.
Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. Theory Methods *3* (1), 1–27.
Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28 (17), 6938–6959.
Cheng, C.-H., Shiu, H.-Y., 2014. A novel GA-SVR time series model based on selected indicators method for forecasting stock price. 2014 Int. Conf. Inf. Sci. Electron. Electr. Eng.
Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 2, 224–227.
Dehghanipour, A.H., Zahabiyoun, B., Schoups, G., Babazadeh, H., 2019. A WEAP-MODFLOW surface water-groundwater model for the irrigated Miyandoab plain, Urmia lake basin, Iran: multi-objective calibration and quantification of historical drought impacts. Agric. Water Manag. *223*, 105704.
Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An introductory review of deep learning for prediction models with big data. Front. Artif. Intell. *3*, 4.
Es' haghi, S.R., Karimi, H., Rezaei, A., Ataei, P., 2022. Content Analysis of the Problems and Challenges of Agricultural Water Use: A Case Study of Lake Urmia Basin at Miandoab, Iran. In: SAGE Open, *12*.
Foroumandi, E., Nourani, V., Huang, J.J., Moradkhani, H., 2023. Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: a deep learning approach. J. Hydrol. *616*, 128838.
Fu, K., Li, H., Deng, P., 2022. Chaotic time series prediction using DTIGNet based on improved temporal-inception and GRU. Chaos Solitons & Fractals *159*, 112183.
Giménez-Forcada, E., 2010. Dynamic of sea water interface using hydrochemical facies evolution diagram. Groundwater *48* (2), 212–216.
Iam-on, N., Garrett, S., 2010. Linkclue: a matlab package for link-based cluster ensembles. J. Stat. Softw. *36*, 1–36.
Jang, J.-S., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. *23* (3), 665–685.
Javadinejad, S., Dara, R., Jafary, F., 2020. Modelling groundwater level fluctuation in an Indian coastal aquifer. Water SA *46* (4), 665–671.
Javadzadeh, H., Ataie-Ashtiani, B., Hosseini, S.M., Simmons, C.T., 2020. Interaction of lake-groundwater levels using cross-correlation analysis: a case study of Lake Urmia Basin, Iran. Sci. Total Environ. *729*, 138822.
Jeihouni, M., Toomanian, A., Alavipanah, S.K., Hamzeh, S., Pilesjö, P., 2018. Long term groundwater balance and water quality monitoring in the eastern plains of Urmia Lake, Iran: a novel GIS based low cost approach. J. Afr. Earth Sci. *147*, 11–19.
Kohonen, T. (1984). Self-Organization and Associative Memory Springer. *New York Berlin Heideberg*.
Murtagh, F., Contreras, P., 2012. Algorithms for hierarchical clustering: an overview. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. *2* (1), 86–97.

Norouzi, H., Moghaddam, A.A., 2020. Groundwater quality assessment using random forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). Arab. J. Geosci. *13*, 1–13.

Nourani, V., Baghanam, A.H., Gokcekus, H., 2018. Data-driven ensemble model to statistically downscale rainfall using nonlinear predictor screening approach. J. Hydrol. *565*, 538–551.

Nourani, V., Khodkar, K., Paknezhad, N.J., Laux, P., 2022b. Deep learning-based uncertainty quantification of groundwater level predictions. Stoch. Environ. Res. Risk Assess. *36* (10), 3081–3107.

Nourani, V., Paknezhad, N.J., Huang, J.J., 2022a. Application of PPIE method to assess the uncertainty and accuracy of multi-climate model-based temperature and precipitation downscaling. Theor. Appl. Climatol. *147* (3-4), 1327–1343.

Nourani, V., Paknezhad, N.J., Sharghi, E., Khosravi, A., 2019. Estimation of prediction interval in ANN-based multi-GCMs downscaling of hydro-climatologic parameters. J. Hydrol. *579*, 124226.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65.

Sabzehee, F., Amiri-Simkooei, A., Iran-Pour, S., Vishwakarma, B., Kerachian, R., 2023. Enhancing spatial resolution of GRACE-derived groundwater storage anomalies in Urmia catchment using machine learning downscaling methods. J. Environ. Manag. *330*, 117180.

Sharghi, E., Nourani, V., Zhang, Y., Ghaneei, P., 2022. Conjunction of cluster ensemble-model ensemble techniques for spatiotemporal assessment of groundwater depletion in semi-arid plains. J. Hydrol. *610*, 127984.

Sivarajan, N., Mishra, A.K., Rafiq, M., Nagraju, V., Chandra, S., 2019. Examining climate change impact on the variability of ground water level: a case study of Ahmednagar district, India. J. Earth Syst. Sci. *128* (5), 1–7.

Stepchenko, A., Chizhov, J., 2015. Applying markov chains for NDVI time series forecasting of Latvian regions. Inf. Technol. Manag. Sci. *18* (1), 57–61.

Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., 2022. Groundwater level prediction using machine learning models: a comprehensive review. Neurocomputing *489*, 271–308.

Wang, J., Xu, D., Li, H., 2023. Constructing GRACE-Based 1 km resolution groundwater storage anomalies in arid regions using an improved machine learning downscaling method: a case study in Alxa League, China. Remote Sens. *15* (11), 2913.