RESEARCH PAPER
Homepage: https://cste.journals.umz.ac.ir/
Contrib. Sci. & Tech Eng, 2025, 2(1)
DOI: 10.22080/cste.2025.29082.1036

Estimating Groundwater Levels in Tehran Province Using Ensemble Learning Algorithms

Seyed Mojtaba Mousavimehr [1]*, Mohammad Reza Kavianpour [1]

[1] *Faculty of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran.*

Article Info
Received 03 April 2025
Accepted 04 May 2025
Available online 04 May 2025

Keywords:
Groundwater Level Estimation;
Machine Learning;
Ensemble Models;
Remote Sensing Data;
Water Resource Management.

Abstract:

The study of groundwater levels is of paramount importance due to its critical role in water resource management, agriculture, and ecosystem sustainability. This study uses machine learning algorithms to predict groundwater levels in observation wells across Tehran. A range of input parameters, including satellite-derived data from GRACE, GLDAS, and ERA5, were employed to train models for estimating groundwater level fluctuations. The primary aim was to evaluate and compare the performance of 12 different machine learning models, including Random Forest, AdaBoost, Support Vector Machine, and Artificial Neural Networks, among others, in terms of their prediction accuracy. The results indicated that ensemble-based models generally outperformed individual algorithms, achieving the highest coefficients of determination (R^2) and the lowest error metrics. Spatial analysis of the errors revealed that the northern part of the study area experienced higher prediction errors than the southern region, likely due to more significant groundwater level fluctuations, influenced by regional climatic conditions and topography. Furthermore, the study demonstrated that combining various input parameters, such as terrestrial water storage, total soil moisture, and precipitation, improved the accuracy of the groundwater level predictions. The models were evaluated using standard error metrics, including Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (R), with results showing strong agreement between predicted and observed data. The findings suggest that machine learning models, especially those leveraging high-resolution satellite and reanalysis data, can be highly effective for groundwater level prediction and management in regions with limited in-situ measurement data. This study provides valuable insights into the application of machine learning for groundwater monitoring, with promising results for future implementation in water resource management.

***Corresponding Author:** mmoosavimehr@mail.kntu.ac.ir

Supplementary information: Supplementary information for this article is available at https://cste.journals.umz.ac.ir/

Please cite this paper as: Mousavimehr, S. M., & Kavianpour, M. R. (2025). Estimating Groundwater Levels in Tehran Province Using Ensemble Learning Algorithms. Contributions of Science and Technology for Engineering, 2(1), 51-63. doi:10.22080/cste.2025.29082.1036.

1. Introduction

Groundwater (GWS) is one of the most important sources of freshwater in the world and plays a vital role in human life. Due to its stability, ease of extraction, widespread availability, and generally good quality, groundwater has been widely utilized by people. groundwater is considered the only reliable water source in many water-stressed regions—such as the semi-arid and arid areas of Asia, the Middle East, North Africa, and Mediterranean countries—as surface water bodies, whether seasonal or permanent, are often absent [1].

In these regions, where agricultural activities and food production predominantly drive water demand, groundwater accessibility is intrinsically linked to food security and, consequently, to national and regional socio-political stability. However, intensive and unsustainable groundwater abstraction has resulted in severe consequences, groundwater level declines, water quality degradation, and the manifestation of geohazards such as underground funnels and ground subsidence [2].

Despite these mounting challenges, groundwater remains a critical buffer against water scarcity, particularly under climatic variability and limited surface water availability. Therefore, accurate monitoring and effective groundwater management at regional scales are paramount. Traditionally, groundwater assessment and management have relied on observation wells, a method that is not only time-consuming but also constrained by economic limitations and insufficient spatial coverage.

In recent years, the emergence of advanced technologies—such as Interferometric Synthetic Aperture Radar (InSAR), Global Positioning System (GPS), and Gravity Recovery

ISSN 3060-6578

and Climate Experiment (GRACE) satellites—has revolutionized the way hydrological and climatic parameters are monitored. These remote sensing platforms enable high-resolution, continuous space-time observation of various environmental indicators, providing an unprecedented opportunity to monitor changes over both spatially extensive and data-scarce regions. The launch of the GRACE satellite mission on March 17, 2002, marked a significant turning point in global hydrological monitoring. It enabled the large-scale assessment of terrestrial water storage changes (ΔTWS), offering a powerful tool for detecting and analyzing regional and global groundwater depletion [3-5]. Furthermore, by integrating data from hydrological models, groundwater storage changes (ΔGWS) can be inferred from GRACE-based ΔTWS observations [6].

Accurate prediction and estimation of groundwater level fluctuations are critically important for effective water resource planning and management. Reliable forecasts enable informed decision-making and support the sustainable use of these valuable subsurface reserves. As a foundational step, sustainable groundwater management must be designed to allow for intelligent utilization of the resource while preserving ecological integrity. By anticipating patterns of groundwater level variation, it becomes significantly easier to strategically plan water use, optimize allocation, and prevent over-extraction, thereby reducing the risk of long-term environmental degradation.

Moreover, accurate predictions can play a key role in preventing future water crises. For instance, if it is known that a significant decline in groundwater resources will occur during a specific period, mitigation strategies such as improving water use efficiency in agriculture or utilizing alternative water resources can be designed. These proactive measures can help alleviate the economic and social impacts of such a crisis.

In addition, maintaining groundwater quality is essential. Excessive groundwater depletion can lead to the intrusion of saline water into aquifers, ultimately degrading the quality of water resources. This issue is especially critical in areas where agriculture heavily depends on groundwater. Therefore, forecasting these changes provides valuable information for engineers and resource managers, enabling them to take preventative actions to avoid contaminant or salinity intrusion into aquifers [7].

In long-term planning, predicting groundwater levels assists urban planners, farmers, and developers in making informed decisions. This includes the optimal selection of locations for residential developments, industrial expansion, or agricultural land use, ensuring that the balance between natural resources and human needs is maintained [8]. In this regard, researchers have made efforts to provide forecasts and estimations of groundwater level fluctuations using a variety of techniques, such as statistical methods, machine learning, and deep learning [9-12].

Amiri et al. (2023) conducted a study to predict groundwater level fluctuations by combining deep learning techniques with the GSM numerical model. After evaluating various machine learning models, they demonstrated that ensemble models outperformed traditional models in terms of performance [13].

Soltani and Azari (2022) utilized satellite data and machine learning algorithms to predict terrestrial water storage anomalies (TWSA) in the Urmia Lake Basin. Based on climatic change projections, the study's findings indicate that TWSA will decrease compared to the historical average during the periods 2021-2040 and 2041-2060. Additionally, due to changes in TWSA, groundwater availability (GWA) is expected to significantly decrease in the future [14].

Kardan et al. (2019) conducted a study to predict aquifer status using Bayesian networks and artificial neural networks (ANNs). The results of their study demonstrated that Bayesian network models outperformed both ANN and mathematical models. Furthermore, their findings indicate that Bayesian networks are effective tools for predicting groundwater levels [12].

In another study, Azizi et al. (2023) employed a hybrid modeling approach by integrating wavelet transform with machine learning algorithms to simulate and forecast groundwater levels in the Sahneh plain. The results demonstrated the hybrid model's high accuracy in predicting groundwater fluctuations, confirming its effectiveness for complex hydrogeological systems [15].

Improving the accuracy of groundwater level prediction remains a major challenge due to the system's complexity and data uncertainty. Employing advanced methods such as wavelet transform for data preprocessing and integrating system dynamics concepts into the training process of machine learning models may help enhance prediction performance and better reflect the real behavior of groundwater systems [16, 17].

In this study, machine learning and deep learning techniques were employed to predict the groundwater level using data from observation wells located within the study area in Tehran Province.

2. Case study

Tehran Province, covering an area of approximately 13,000 km² and home to over 15 million people, is located in the northern part of the Central Iranian Plateau. The province is bordered by the Alborz Mountain range to the north and surrounded by the Central Desert of Iran from the north and south, respectively. Geographically, it lies between 35°30′N and 35°42′N latitude and 50°55′E and 51°23′E longitude. The average annual precipitation is about 280 mm, while the annual evaporation exceeds 250 mm. The mean annual temperature is approximately 17°C, with recorded extremes ranging from -15°C to 43°C [18].

According to the latest data provided by the Iran Water Resources Management Company, by the end of 2017, more than 4,000 wells had been operating in Tehran Province, supplying water for irrigation, drinking, industrial, and other uses. Over the years of groundwater extraction, the total groundwater storage in the province has decreased by approximately 4 billion cubic meters [19].

This region is highly dependent on groundwater resources for domestic, agricultural, and industrial purposes. In fact, more than 50% of the drinking water supply for the city of Tehran is derived from groundwater sources. In recent years, continued drought, rapid population growth due to migration, and over-extraction of groundwater to meet increasing demands for livelihood and production have led to a water supply crisis and the emergence of land subsidence phenomena. The unsustainable exploitation of these resources has resulted in the designation of many plains in the region as overdrawn or restricted zones. Therefore, assessing and monitoring groundwater storage at the local scale in this region is of critical importance. The location of the study area and the observation wells are illustrated in Figure 1.

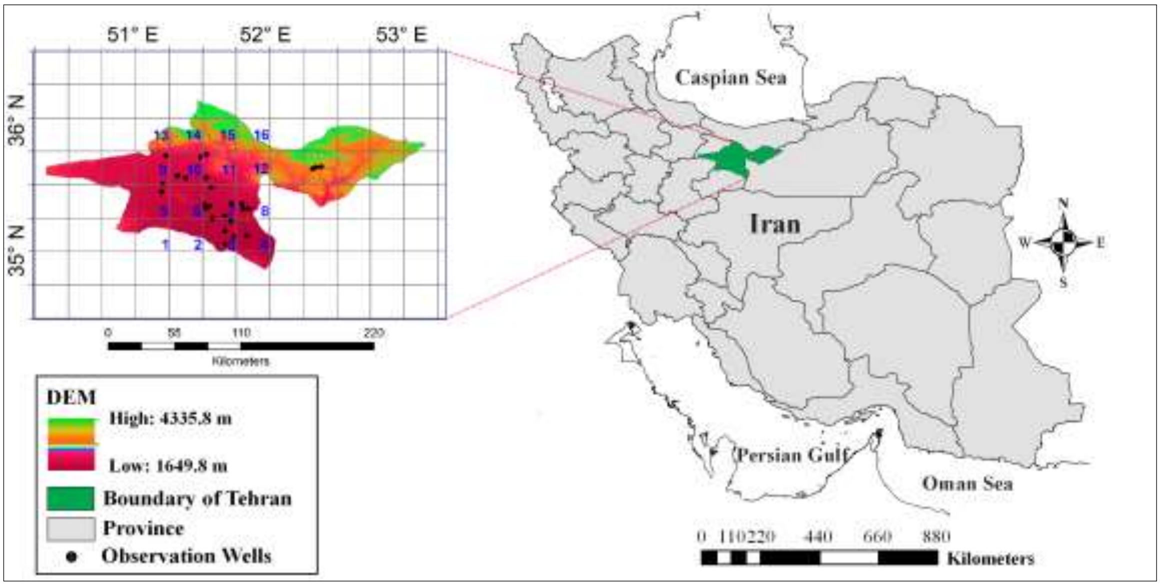


Figure 1. Study Area and Location of Observation Wells in Tehran Province, Iran

3. Data Collection and Analysis

This study aims to estimate groundwater level fluctuations across Tehran Province by integrating satellite-based observations with advanced machine learning and deep learning techniques. To achieve this objective, a comprehensive set of hydrological and remote sensing datasets was employed. These include monthly variations of Total Water Storage (TWS) derived from GRACE and GRACE Follow-On missions, along with surface and subsurface runoff parameters, soil moisture, canopy water content, snow water equivalent (SWE), evapotranspiration, and precipitation, sourced from various global and regional databases.

The predicted groundwater levels generated by the models were subsequently validated against in-situ measurements from observational wells to assess the accuracy and reliability of the estimations. The Table 1 summarizes the input variables used in the groundwater level estimation framework, along with their spatial resolution, measurement units, and respective data sources. Subsequent sections provide detailed descriptions of each data source and its relevance to the modelling process.

Table 1. Summary of the datasets used in the groundwater level estimation process

Variable	Symbol	Spatial & Temporal Resolution	Unit	Data Source	Time Span
Terrestrial Water Storage	TWS	1° × 1°, Monthly	cm	GRACE & GRACE-FO (https://www2.csr.utexas.edu/grace/)	2002–2023
Surface Runoff	Q_S	0.25° × 0.25°, Monthly	kg·m⁻²	GLDAS/NOAH v2.1 (https://disc.gsfc.nasa.gov/dataset)	2002–2023
Subsurface Runoff	Q_G	0.25° × 0.25°, Monthly	kg·m⁻²	GLDAS/NOAH v2.1 (https://disc.gsfc.nasa.gov/dataset)	2002–2023
Total Soil Moisture	SM_{Tot}	0.25° × 0.25°, Monthly	kg·m⁻²	GLDAS/NOAH v2.1 (https://disc.gsfc.nasa.gov/dataset)	2002–2023
Canopy Water	CW	0.25° × 0.25°, Monthly	kg·m⁻²	GLDAS/NOAH v2.1 (https://disc.gsfc.nasa.gov/dataset)	2002–2023
Snow Water Equivalent	SWE	0.25° × 0.25°, Monthly	kg·m⁻²	GLDAS/NOAH v2.1 (https://disc.gsfc.nasa.gov/dataset)	2002–2023
Evapotranspiration	E	0.25° × 0.25°, Monthly	m	ERA5 (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means)	2002–2023
Precipitation	Pr	0.25° × 0.25°, Monthly	m	ERA5 (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means)	2002–2023
Observational Wells	—	Monthly	m	IWRM (http://wrs.wrm.ir/amar/register.asp)	2001–2022

3.1. Groundwater Well Data

Monthly groundwater level measurements from observation wells were obtained from the official website of the Iran Water Resources Management Company (http://wrs.wrm.ir) [19]. These data span from 2001 to 2022 (corresponding to the Persian calendar years 1380 to 1401). In total, data are available for approximately 350 observation wells located across ten defined hydrological subregions: Eyvanaki, Namak Lake, Damavand, Firouzkouh, Saveh, Garmsar, Mobarakiyeh, Hoomand-Abasard, Varamin, and Tehran-Karaj.

Preliminary evaluation of the raw data revealed that, in some wells, measurements were available only for limited periods. Additionally, several wells exhibited large data gaps or missing values, rendering them unsuitable for robust time series analysis. In certain cases, monthly groundwater level fluctuations exceeded 2 meters, likely due to measurement or recording errors; such anomalies were either corrected or excluded from the analysis.

Following a rigorous quality control procedure—including filtering, outlier detection, and continuity checks implemented in MATLAB—only those wells with consistent and sufficiently long time series (spanning approximately two decades from 2002 to 2022) were retained. This time period was selected to ensure comparability with the temporal coverage of satellite-based GRACE and GRACE Follow-On datasets.

Ultimately, after applying various filters and conducting the Mann-Kendall trend test, observation wells exhibiting a significant increasing trend inconsistent with the general behavior of the dataset were excluded. As a result, the number of wells was reduced to 29. The spatial distribution of these wells is illustrated in Figure 1. The Z-values obtained from the Mann-Kendall test at a 5% significance level ($\alpha = 0.05$) for the observation wells are presented in Figure 2.

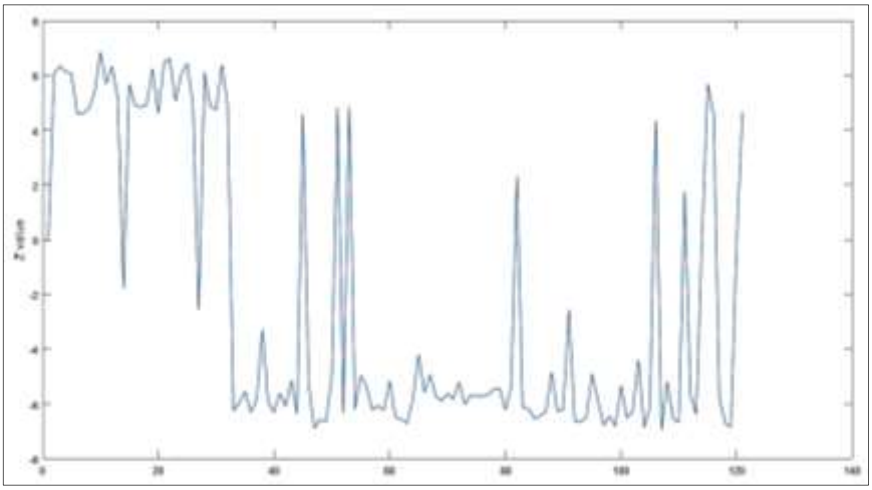


Figure 2. Variations in Mann-Kendall Z-values for different observation wells

According to Figure 2, the Mann-Kendall test reveals negative Z-values for the majority of observation wells, indicating a declining trend in groundwater levels across Tehran Province. Moreover, the significantly negative Z-values observed in a substantial number of wells reflect a strong downward trend.

Subsequently, cumulative and monthly variations in average groundwater level across all wells are illustrated in the following figure. Based on this figure, the cumulative monthly groundwater level change over the 20-year period from 2003 to 2023 exceeds 8 meters. A slight increase in the trend was observed between 2003 and 2007, followed by a noticeable and intensifying decline in subsequent years.

The range of monthly groundwater level fluctuations in Tehran Province is approximately ±0.8 meters, with wider variability observed at the beginning and end of the time series and narrower fluctuations in the mid-years. Additionally, the average groundwater level change for Tehran Province is estimated at approximately -0.03meters.

3.2. GRACE and GRACE-FO Data

GRACE data are provided at five processing levels: Level-0, Level-1A, Level-1B, Level-2, and Level-3. These datasets can be obtained from three main data centers: the Center for Space Research (CSR), the German Research Centre for Geosciences (GFZ), and the Jet Propulsion Laboratory (JPL). Each of these institutions employs different algorithms to convert GRACE raw data into Level-3 products. In this study, following a thorough evaluation, monthly Level-3 data from the Center for Space Research (CSR) at the University of Texas have been utilized to estimate groundwater storage variations derived from GRACE (Total Water Storage – TWS). The time series of satellite-derived data from GRACE and GRACE-FO is presented in Figure 4.

According to Figure 4, the temporal coverage of GRACE data spans from April 2002 to June 2017, while GRACE Follow-On (GRACE-FO) data cover the period from June 2018 to the present. As observed, there is an 11-month gap between the end of the GRACE mission and the beginning of GRACE-FO. A declining trend in Total Water Storage (TWS) is evident in both GRACE and GRACE-FO data, with a steeper decline observed in the post-2018 period covered by GRACE-FO.

Figure 5 illustrates the month-to-month variations in TWS, showing the differences between each month and its preceding month.

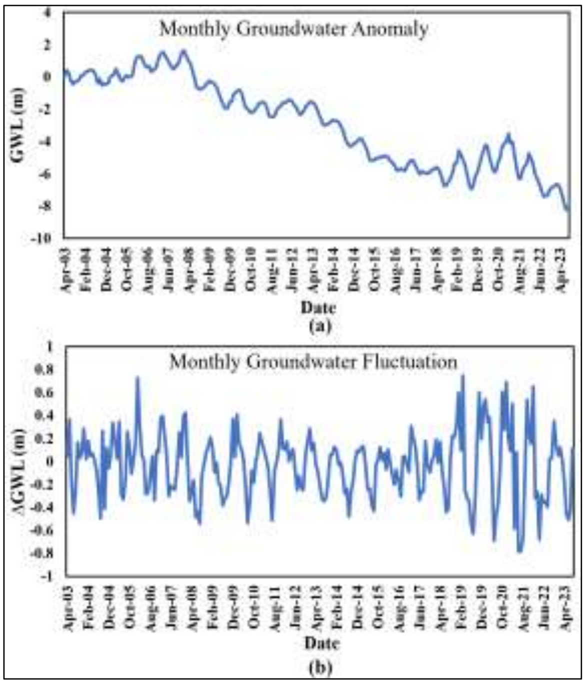


Figure 3. **(a) Cumulative changes in groundwater level of observation wells, and (b) Monthly groundwater level fluctuations in Tehran Province**

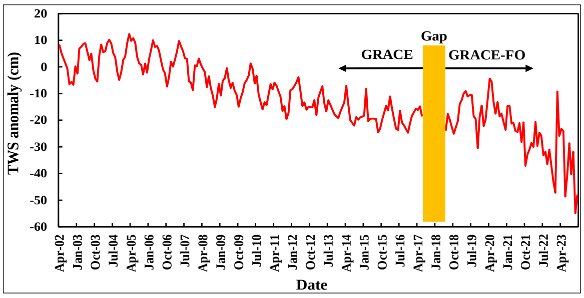


Figure 4. **Time series of monthly TWS variations derived from the GRACE satellite**

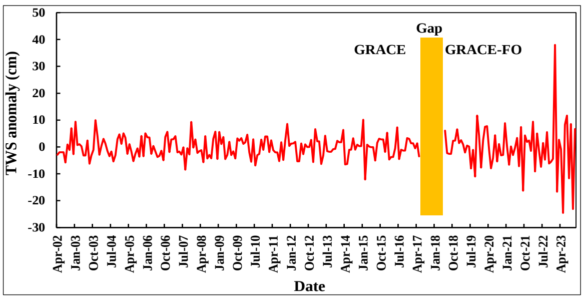


Figure 5. **Monthly variations in Total Water Storage (TWS) derived from the GRACE satellite.**

As observed, the monthly variations in TWS for the years prior to 2017 remain within a range of ±10 centimeters. However, in recent years, this variation has significantly increased, reaching up to more than 30 centimeters per month.

Subsequently, the annual variations in TWS, along with its cumulative values for the specified period, are presented below.

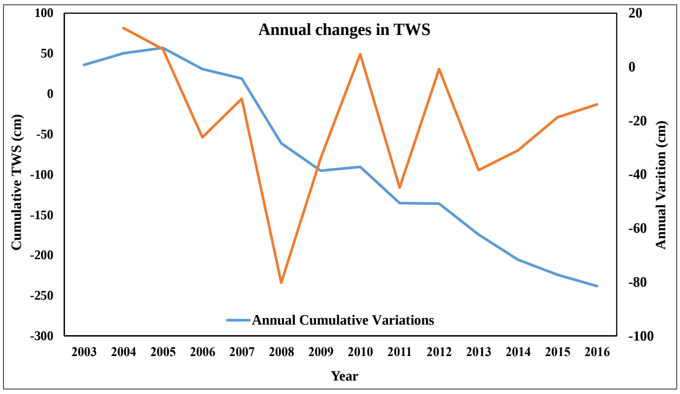


Figure 6. **Annual variations in Total Water Storage (TWS) derived from the GRACE satellite.**

3.3. GLDAS Data

The Global Land Data Assimilation System (GLDAS) was developed through collaboration between scientists from NASA, the Goddard Space Flight Center (GSFC), the National Oceanic and Atmospheric Administration (NOAA), and the National Centers for Environmental Prediction (NCEP) to generate land surface parameters [20]. GLDAS includes three land surface models (Mosaic, NOAH, and CLM), watershed land surface models, and hydrological (VIC) models.

This study used monthly data with a spatial resolution of 0.25° × 0.25° from the GLDAS/NOAH model to obtain soil moisture, snow water equivalent (SWE), canopy water (CW), and runoff data. These data are available from the Goddard Earth Sciences Data and Information Services Center (GES DISC) [21].

Version 2.1 of the GLDAS dataset includes various parameters, from which surface runoff (Qs), subsurface runoff (Qg), canopy water (CW), soil moisture (SM) at different depths (0–10, 10–40, 40–100, and 100–200 cm), and snow water equivalent (SWE) were used in this study. The data obtained from GLDAS 2.1 are measured in kilograms per square meter. To convert the GLDAS 2.1 data to centimeters of water, it should be multiplied by 0.1. Additionally, the runoff and groundwater flow data are provided as three-hour averages, and thus, these values should be multiplied by 8 × 30.

3.4. ERA5 Precipitation and Evaporation Data

Undoubtedly, precipitation and evaporation are key factors controlling the water balance, playing a significant role in streamflow, its conversion into groundwater flow, and the replenishment of groundwater aquifers. Given the importance of these parameters, it is crucial to examine their variations. One of the widely used data sources for studying spatial and temporal changes in precipitation and evaporation is the ERA5 database, which provides data with a spatial resolution of 0.25 degrees. The units of precipitation and evaporation are given in meters per day, and to convert these values to centimeters per month, they are multiplied by 30 × 100.

4. Methodology

This study aims to estimate the groundwater table levels of selected observation wells in cells 3, 4, and 10. These cells were chosen due to their extensive coverage from south to north. The first step involves gathering and analyzing input data, followed by developing a prediction model, which must be appropriately selected based on the dependence of the target variable. The target variable, or the variable to be predicted in this study, is the groundwater table level of the observation wells.

In general, the data considered as inputs for the machine learning-based models include Terrestrial water storage(TWS), cumulative surface runoff(Qs), subsurface runoff(Qg), snow water equivalent (SWE), total soil moisture at different depths from 0 to 200 cm(SM), canopy water(CW), precipitation(P), and evaporation(E).

Due to the expected limitations in model accuracy when relying solely on coarse-scale input data for groundwater level prediction, this study utilizes high-resolution datasets (0.25-degree spatial resolution) to improve modeling precision. The integration of finer spatial data enhances the reliability of the predictive models and supports more accurate groundwater monitoring, which is essential for effective water resource planning and sustainable management.

All input parameters used in this study have a spatial resolution of 0.25 degrees and a temporal resolution at the monthly scale. Since the original TWS data are provided at a coarser resolution of 1° × 1°, they were downscaled to 0.25-degree resolution to ensure consistency across all inputs. The downscaling procedure for TWS is described in detail by Mosavimehr and Kavianpour [22], and the results of their study were utilized in this research. Following a correlation analysis between groundwater levels in the selected wells and the input parameters, the optimal combination of predictors was selected for each well based on its spatial location. Table 2 presents the selected input variable combinations used to develop machine learning models for the three representative wells located in cells 3, 4, and 10. It is worth noting that input parameters with low correlation to the target variable were excluded from the modelling process.

Using this combination of input variables and 12 different methods, the monthly variations in groundwater table levels for the selected observation wells in cells 3, 4, and 10 were predicted. The methods employed in this study to predict the average changes in groundwater levels include: Linear Regression [23] (LR), Ridge Regression [24] (Ridg), Lasso Regression [25] (Lasso), Support Vector Machine [26] (SVM), Decision Tree [27] (DT), Artificial Neural Network [28] (ANN), Bagging Regressor [29] (BR), Random Forest [30] (RF), AdaBoost Algorithm [31] (Ada Boost), Gradient Boosting Algorithm [32] (GB), Stochastic Gradient Boosting [33] (SGB), and Extreme Gradient Boosting [34] (XGB). These methods are summarized with their abbreviations in the corresponding tables.

Table 2. Optimal combinations of input variables selected for machine learning model development, based on correlation analysis for three representative observation wells located in cells 3, 4, and 10.

Model No.	Input variables
Cell 10	TWS_{JPL}, SWE, SM_{Tot}
Cell 3	TWS_{JPL}, SM_{Tot}
Cell 4	TWS_{JPL}, SM_{Tot}

It is important to note that, the data were randomly split into training and test samples for modeling purposes, with 75% of the data used for training and the remaining 25% reserved for testing the models [35-37]. Also, it is worth noting that all machine learning models were coded using the Python Scikit-learn library [38-40].

To evaluate the performance of the developed models in predicting groundwater level, four statistical metrics were employed: Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (R). These metrics are defined as follows:

$$\mathrm{ME} = \frac{\sum_{i=1}^{n}(y_i' - y_i)}{n} \tag{1}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i' - y_i|}{n} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i' - y_i)^2}{n}} \tag{3}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(y_i' - \bar{y}')}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(y_i' - \bar{y}')^2}} \tag{4}$$

where y_i and $\bar{y}$ denote the i-th observed value and the mean of the observed values, respectively, $y_i'$and $\overline{y'}$ represent the i-th predicted value and the mean of the predicted values, respectively, N is the total number of samples.

In general, lower values of ME, MAE, and RMSE, and a correlation coefficient (R) closer to 1 indicate better model performance and higher predictive accuracy.

5. Results and Discussion

After model development, the performance of various machine learning models was evaluated for three selected wells using four statistical metrics: Coefficient of Determination (R^2), Mean Error (ME), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). These metrics were calculated under different scenarios and are presented in Tables 3, 4, 5, and 6, respectively. It is worth noting that all values presented in the tables are expressed in meters.

Table 3. Coefficient of Determination (R^2) Values for the Selected Wells and the 12 Prediction Methods

Cell	LR	Ridg	Lasso	SVR	DT	ANN	BR	RF	Adb	GB	SGB	XG
C10	0.820	0.820	0.820	0.383	0.838	0.828	0.856	0.852	0.868	0.869	0.842	0.847
C3	0.744	0.744	0.744	0.752	0.855	0.744	0.815	0.780	0.849	0.895	0.897	0.857
C4	0.844	0.844	0.845	0.705	0.848	0.340	0.857	0.852	0.793	0.849	0.852	0.807

Table 4. Mean Error (ME) Values for the Selected Wells and the 12 Prediction Methods

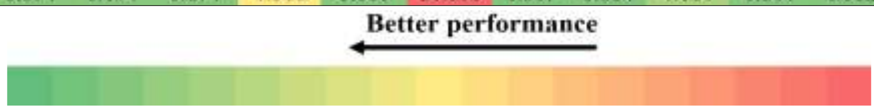
Cell	LR	Ridg	Lasso	SVR	DT	ANN	BR	RF	Adb	GB	SGB	XG
C10	0.564	0.564	0.564	5.641	0.456	0.626	0.507	0.526	2.128	1.141	1.212	1.361
C3	0.169	0.169	0.169	1.598	0.044	0.169	-0.046	0.003	-0.134	0.044	0.080	-0.045
C4	-0.035	-0.035	-0.034	0.148	-0.084	-1.524	-0.059	-0.060	-0.126	-0.071	-0.058	-0.082

Table 5. Mean Absolute Error (MAE) Values for the Selected Wells and the 12 Prediction Methods

Cell	LR	Ridg	Lasso	SVR	DT	ANN	BR	RF	Adb	GB	SGB	XG
C10	4.643	4.643	4.643	12.065	4.113	4.559	3.782	3.793	4.127	3.779	3.991	4.322
C3	1.380	1.380	1.380	2.526	1.081	1.380	1.201	1.321	1.109	0.956	0.965	1.015
C4	0.718	0.718	0.718	1.662	0.678	26.249	0.676	0.706	0.761	0.703	0.700	0.748

Table 6. Root Mean Square Error (RMSE) Values for the Selected Wells and the 12 Prediction Methods

Cell	LR	Ridg	Lasso	SVR	DT	ANN	BR	RF	Adb	GB	SGB	XG
C10	5.646	5.646	5.646	13.924	5.367	5.537	5.050	5.116	5.256	4.948	5.466	5.499
C3	1.860	1.860	1.862	3.739	1.384	1.860	1.575	1.721	1.427	1.201	1.196	1.381
C4	0.871	0.871	0.871	1.930	0.865	30.905	0.837	0.854	1.051	0.861	0.852	0.995



Based on the results presented in the above table, it can be observed that ensemble-based learning methods generally provide relatively better performance. The highest coefficient of determination (R^2) was obtained by the Stochastic Gradient Boosting (SGB) model. Additionally, this model also yielded the lowest error values across the evaluated metrics. Therefore, the SGB model is identified as the most effective model for monthly groundwater level estimation across different locations.

A comparison of the spatial distribution of errors indicates that the northern part of the study area exhibits higher error values compared to the southern part. This may be attributed to the more significant fluctuations in groundwater levels in that region, which are likely influenced by differences in precipitation patterns, geographical location, and elevation compared to the southern zones.

To facilitate a better understanding of the tabulated results, corresponding bar charts are also provided in Figures 7 to 10.

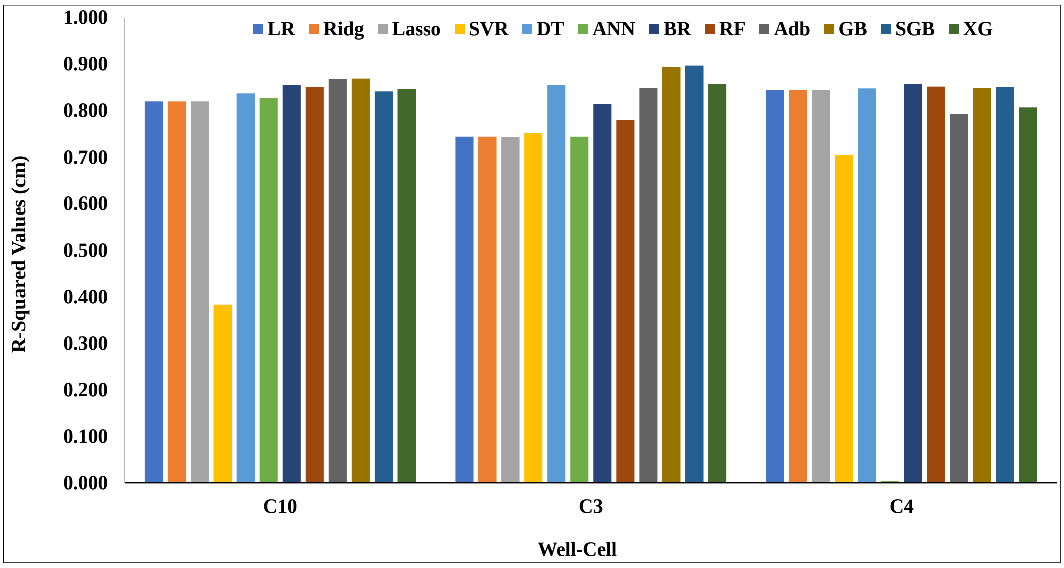


Figure 7. Bar Chart of R^2 Values Across Different Cells (Selected Wells) and the 12 Prediction Methods.

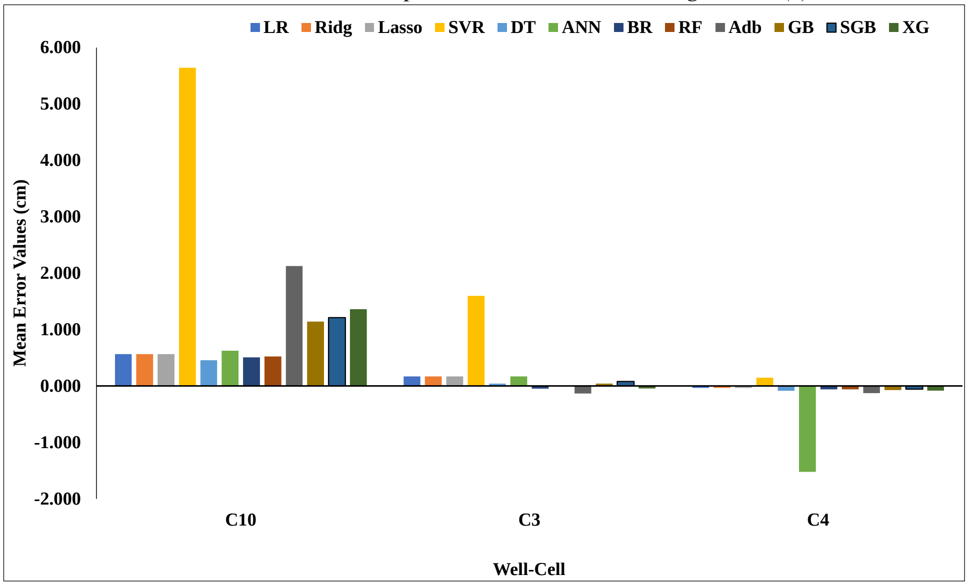


Figure 8. **Bar Chart of ME Values Across Different Cells (Selected Wells) and the 12 Prediction Methods.**

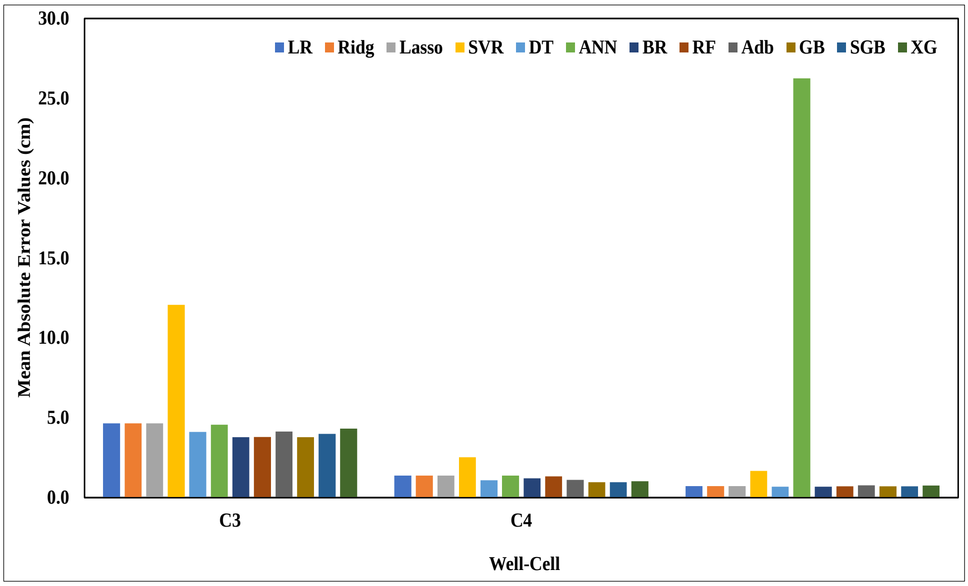


Figure 9. **Bar Chart of MAE Values Across Different Cells (Selected Wells) and the 12 Prediction Methods**

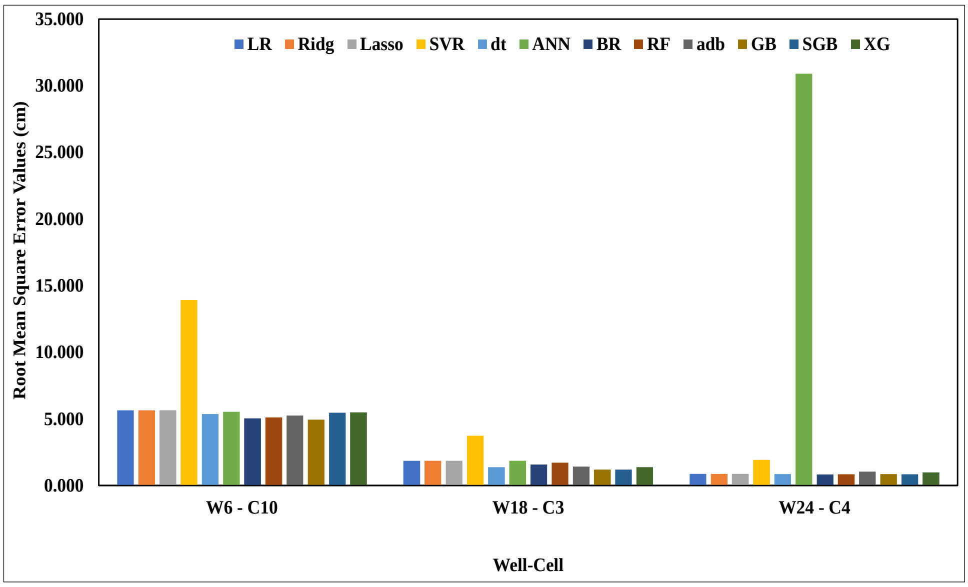


Figure 10. **Bar Chart of RMSE Values Across Different Cells (Selected Wells) and the 12 Prediction Methods**

As illustrated in the above figures, the SGB model demonstrated the best performance, with the lowest error values and the highest R^2. However, the performance difference between the SGB model and other ensemble-based models is not substantial.

A comparison between the Mean Error (ME) and the Mean Absolute Error (MAE) for the SGB and Decision Tree (DT) models reveals that the ME value for the SGB model is higher than that of the DT model, whereas the MAE of the SGB model is lower than that of the DT. This discrepancy suggests that the DT model exhibits a more symmetrical error distribution, while the SGB model tends to underestimate the actual values for the majority of the data points.

The performance comparison of various models across the selected cell indicates that there is no significant difference in R^2 values. However, the figure also shows relatively higher error values for Cell 10, located in the northern part of the study area.

5.1. Groundwater Level Prediction for the Selected Well in Cell 10

As shown in Tables 4 to 7, which presents the performance of various algorithms for this model, the AdaBoost (Ada) and Gradient Boosting (GB) algorithms exhibit the highest R^2 values and the lowest error values. Therefore, they provide the most accurate predictions of groundwater level fluctuations.

It is worth noting that the error values reported in the table are expressed in meters.

Table 7. Prediction Model Performance for Cell 10 During the Testing Period.

	ME	MAE	RMSE	R2
LR	0.5642043	4.6431212	5.6458621	0.8201
Ridg	0.5642061	4.643116	5.6458574	0.8201
Lasso	0.5639028	4.6425839	5.6456411	0.8201
SVR	5.6414544	12.065095	13.924478	0.3832
DT	0.4563499	4.112835	5.3667233	0.8375
ANN	0.6258802	4.5594345	5.5368978	0.8275
BR	0.506542	3.7816679	5.0503464	0.8556
RF	0.5255296	3.7932911	5.1161387	0.8517
Ada	2.1276266	4.1272769	5.255509	0.8677
GB	1.1412259	3.7785109	4.9477727	0.8694
SGB	1.2120369	3.9913322	5.4655153	0.8416
XG	1.361111	4.3218163	5.4988643	0.8465

Better performance

Furthermore, Figures 11 and 12 present the scatter plots of observed groundwater level fluctuations (from monitoring wells) versus the predicted values for the Ada and GB algorithms, respectively.

In addition, Figure 13 illustrates the time series of observed versus predicted groundwater levels during the testing period.

Overall, a strong correlation is observed between the predicted and actual groundwater levels in Cell 10, as indicated by the high R^2 values in the scatter plots. Additionally, Figure 13 demonstrates that the selected models have successfully captured the overall trend of groundwater level fluctuations with a satisfactory degree of accuracy. Notably, the models also perform reasonably well in predicting extreme values. However, as previously mentioned, these models tend to slightly underestimate the extreme values, which is a commonly observed behavior in machine learning-based prediction models.

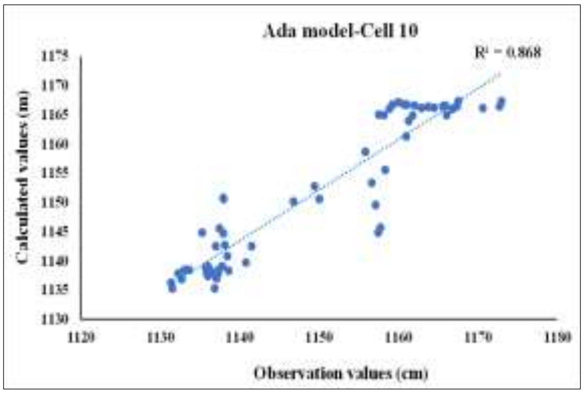


Figure 11. Scatter Plot of Observed vs. Predicted Groundwater Levels for the Ada Model

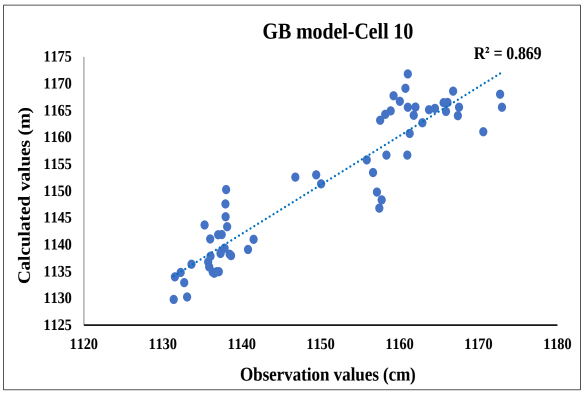


Figure 12. Scatter Plot of Observed vs. Predicted Groundwater Levels for the GB Model

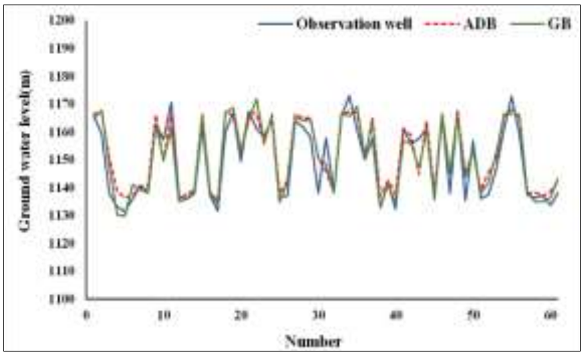


Figure 13. Time Series of Observed vs. Predicted Groundwater Levels for the Ada and GB Models

5.2. Groundwater Level Prediction for the Selected Well in Cell 3

As shown in Table 8, which presents the performance of various algorithms for this model, the AdaBoost (Ada) and Gradient Boosting (GB) algorithms achieved the highest R^2 values and the lowest error values. Therefore, they provide the most accurate predictions of groundwater level variations. It should be noted that the error values reported in the table are expressed in meters.

Table 8. Prediction Model Performance for Cell 3 During the Testing Period.

	ME	MAE	RMSE	R2
LR	0.169	1.380	1.860	0.744
Ridg	0.169	1.380	1.860	0.744
Lasso	0.169	1.380	1.862	0.744
SVR	1.598	2.526	3.739	0.752
DT	0.044	1.081	1.384	0.855
ANN	0.169	1.380	1.860	0.744
BR	-0.046	1.201	1.575	0.815
RF	0.003	1.321	1.721	0.780
Ada	-0.134	1.109	1.427	0.849
GB	0.044	0.956	1.201	0.895
SGB	0.080	0.965	1.196	0.897
XG	-0.045	1.015	1.381	0.857

Better performance

Subsequently, Figures 14 and 15 illustrate the scatter plots of observed groundwater level fluctuations (from monitoring wells) versus the predicted values for the SGB and GB algorithms, respectively. Figure 16 presents the time series of observed versus predicted groundwater levels for these two models.

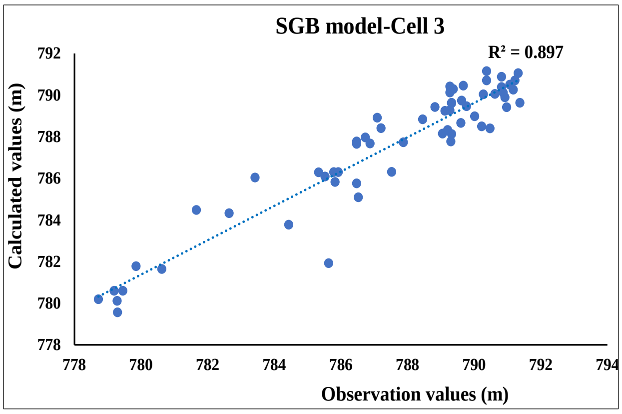


Figure 14. **Scatter Plot of Observed vs. Predicted Groundwater Levels for the SGB Model**

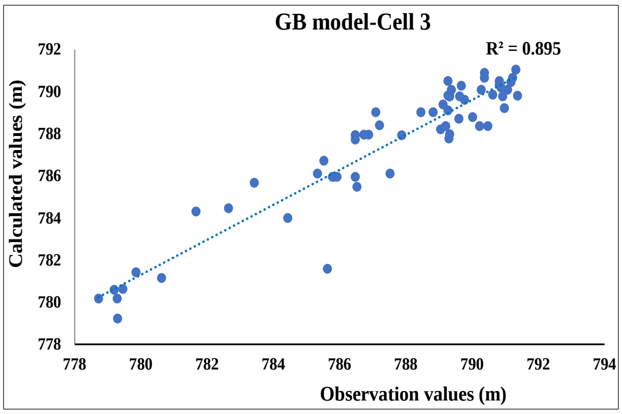


Figure 15. **Scatter Plot of Observed vs. Predicted Groundwater Levels for the GB Model**

The highest coefficient of determination among the different wells is observed for Cell 3, with a value of 0.895. According to the above figures, the selected models provide reliable estimates of the monthly groundwater level for the well in question. It is noteworthy that the models demonstrate a strong capability in predicting extreme values accurately.

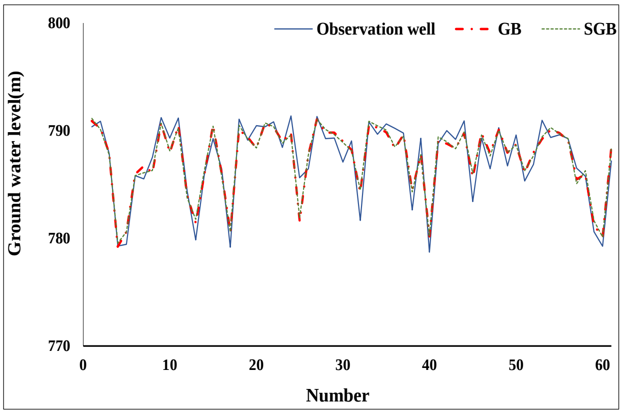


Figure 16. **Time Series of Observed vs. Predicted Groundwater Levels for the SGB and GB Models**

5.3. Groundwater Level Prediction for the Selected Well in Cell 4

As shown in Table 9, which presents the performance of various algorithms for this model, the AdaBoost (Ada) and Gradient Boosting (GB) algorithms achieved the highest R^2 values and the lowest error values. Therefore, they provide the most accurate predictions of groundwater level variations. It should be noted that the error values in the table are expressed in meters.

Table 9. Prediction Model Performance for Cell 4 During the Testing Period.

Model	ME	MAE	RMSE	R2
LR	-0.035	0.718	0.871	0.844
Ridg	-0.035	0.718	0.871	0.844
Lasso	-0.034	0.718	0.871	0.845
SVR	0.148	1.662	1.930	0.705
dt	-0.084	0.678	0.865	0.848
ANN	-1.524	26.249	30.905	0.340
BR	-0.059	0.676	0.837	0.857
RF	-0.060	0.706	0.854	0.852
Adb	-0.126	0.761	1.051	0.793
GB	-0.071	0.703	0.861	0.849
SGB	-0.058	0.700	0.852	0.852
XG	-0.082	0.748	0.995	0.807

Better performance

Subsequently, Figures 17 and 18 show the scatter plots of observed groundwater level fluctuations (from monitoring wells) versus the predicted values for the Bagging Regressor (BR) and Random Forest (RF) algorithms, respectively. Figure 19 displays the time series of observed values versus predicted groundwater levels for these two models.

The results for the well located in Cell 4 are similar to those of Cell 3. Although the BR and RF models performed best for this cell, the differences in performance among the ensemble-based models are relatively negligible and can be considered insignificant.

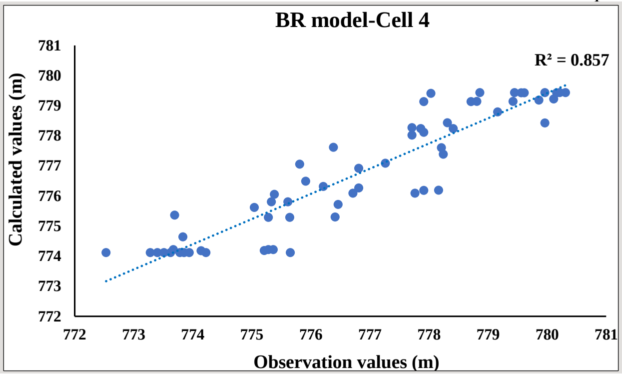


Figure 17. Scatter Plot of Observed vs. Predicted Groundwater Levels for the BR Model

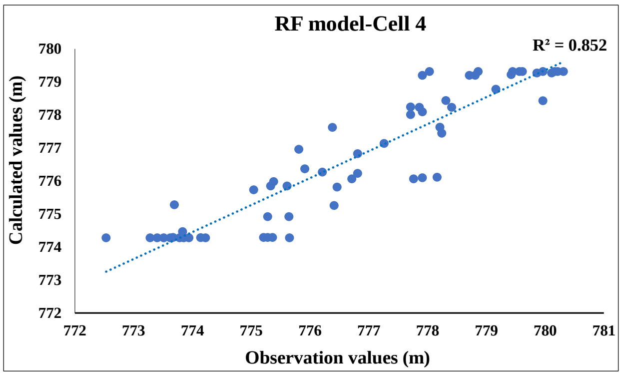


Figure 18. Scatter Plot of Observed vs. Predicted Groundwater Levels for the RF Model

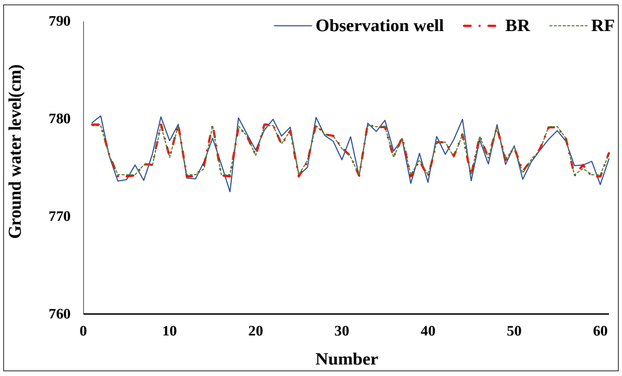


Figure 19. Time Series of Observed vs. Predicted Groundwater Levels for the BR and RF Models

6. Conclusions

Accurate prediction of groundwater level variations is crucial for sustainable water resource management, particularly in arid and water-stressed regions. Satellite-based observations, such as Total Water Storage (TWS) data from GRACE and GRACE-FO missions, offer valuable large-scale insights into terrestrial water dynamics, including subsurface changes. When integrated with advanced machine learning and deep learning algorithms, these datasets significantly improve the capability to model, predict, and track groundwater fluctuations over time. The synergy between remote sensing data and data-driven approaches enhances both spatial and temporal accuracy, thereby supporting evidence-based decision-making and long-term groundwater planning.

To this end, various machine learning-based models, including Random Forest, AdaBoost, and Convolutional Neural Networks (CNN), were employed to predict and estimate groundwater level fluctuations in observational wells across the study area in Tehran Province. The input data for these models consisted of aggregated climatic parameters obtained from GLDAS and ERA5 datasets, and the models' performance was evaluated to identify the most accurate predictive approach. In this study, different combinations of input variables were tested to enhance the accuracy of groundwater level estimations. Subsequently, the trained models were applied using high-resolution input data for final prediction and estimation of groundwater levels in the selected wells.

The main limitation of this study was the restricted availability of in-situ measurements, as the existing data were limited to a sparse and irregular network of observational wells. Nevertheless, the comparison between the predicted values and well observations showed a strong agreement, confirming the effectiveness of the proposed methodology. The results revealed a consistent declining groundwater level trend across the study area.

The study achieved reliable prediction performance by applying 12 different machine learning algorithms and testing various combinations of input parameters. The analysis indicated that models incorporating variables such as terrestrial water storage, total soil moisture, and snow water equivalent (in the northern areas), or precipitation (in the southern parts), yielded the most accurate groundwater level forecasts.

Comparison of different machine learning models for estimating groundwater levels in the observation wells revealed that ensemble-based approaches generally outperformed other methods. These models achieved the highest coefficients of determination (R^2) and the lowest error metrics among the evaluated techniques. Moreover, spatial analysis of model errors indicated that prediction errors were larger in the northern part of the study area compared to the southern region. This discrepancy is likely due to more pronounced groundwater level fluctuations in the north, influenced by distinct climatic conditions, higher precipitation variability, and notable differences in elevation and geographic setting relative to the southern zones.

As observed, the parameters of soil moisture and terrestrial water storage show the highest correlation with groundwater level for wells located in the southern regions. In contrast, in the northern regions, in addition to the aforementioned parameters, snow water equivalent also shows a significant correlation with monthly groundwater levels.

Ultimately, the evaluation of various models for direct groundwater level prediction at individual wells using satellite-derived and reanalysis input data demonstrates a satisfactory level of accuracy, highlighting the potential of these approaches for reliable groundwater monitoring.

7. Statements & Declarations

7.1. Funding

The author(s) received no specifc funding for this work

7.2. Author Contributions

Seyed Mojtaba Mousavimehr contributed to conceptualization, methodology, data acquisition and preparation, visualization and writing—review and editing, and provided software. Mohammad Reza Kavianpour was involved in supervision, investigation, conceptualization, methodology and writing—review and editing.

8. References

[1] Ashraf, S., Nazemi, A., & AghaKouchak, A. (2021). Anthropogenic drought dominates groundwater depletion in Iran. Scientific Reports, 11(1), 9135. doi:10.1038/s41598-021-88522-y.

[2] Pinder, G. F. (2011). Groundwater hydrology. In Groundwater Quantity and Quality Management. John Wiley & Sons, Hoboken, United States. doi:10.1029/eo070i008p00114-04.

[3] Frappart, F., Papa, F., Famiglietti, J. S., Prigent, C., Rossow, W. B., & Seyler, F. (2008). Interannual variations of river water storage from a multiple satellite approach: A case study for the Rio Negro River basin. Journal of Geophysical Research Atmospheres, 113(21). doi:10.1029/2007JD009438.

[4] Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. Water Resources Research, 44(2). doi:10.1029/2006WR005779.

[5] Velicogna, I., & Wahr, J. (2005). Greenland mass balance from GRACE. Geophysical Research Letters, 32(18), 1–4. doi:10.1029/2005GL023955.

[6] Sun, A. Y. (2013). Predicting groundwater level changes using GRACE data. Water Resources Research, 49(9), 5900–5912. doi:10.1002/wrcr.20421.

[7] Scibek, J., & Allen, D. M. (2006). Modeled impacts of predicted climate change on recharge and groundwater levels. Water Resources Research, 42(11). doi:10.1029/2005WR004742.

[8] Kinzelbach, W., Bauer, P., Siegfried, T., & Brunner, P. (2003). Sustainable groundwater management — problems and scientific tools. Episodes, 26(4), 279–284. doi:10.18814/epiiugs/2003/v26i4/002.

[9] Zhang, D., & Tong, J. (2023). Robust water level measurement method based on computer vision. Journal of Hydrology, 620. doi:10.1016/j.jhydrol.2023.129456.

[10] Adamowski, J., & Chan, H. F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 407(1–4), 28–40. doi:10.1016/j.jhydrol.2011.06.013.

[11] Barzegar, R., Fijani, E., Asghari Moghaddam, A., & Tziritis, E. (2017). Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Science of The Total Environment, 599–600, 20–31. doi:10.1016/j.scitotenv.2017.04.189.

[12] Moghaddam, H. K., Moghaddam, H. K., Kivi, Z. R., Bahreinimotlagh, M., & Alizadeh, M. J. (2019). Developing comparative mathematic models, BN and ANN for forecasting of groundwater levels. Groundwater for Sustainable Development, 9, 100237. doi:10.1016/j.gsd.2019.100237.

[13] Amiri, S., Rajabi, A., Shabanlou, S., Yosefvand, F., & Izadbakhsh, M. A. (2023). Prediction of groundwater level variations using deep learning methods and GMS numerical model. Earth Science Informatics, 16(4), 3227–3241. doi:10.1007/s12145-023-01052-1.

[14] Soltani, K., & Azari, A. (2022). Forecasting groundwater anomaly in the future using satellite information and machine learning. Journal of Hydrology, 612, 128052. doi:10.1016/j.jhydrol.2022.128052.

[15] Azizi, E., Yosefvand, F., Yaghoubi, B., Izadbakhsh, M. A., & Shabanlou, S. (2023). Modelling and prediction of groundwater level using wavelet transform and machine learning methods: A case study for the Sahneh Plain, Iran. Irrigation and Drainage, 72(3), 747–762. doi:10.1002/ird.2794.

[16] Alizadeh, M. J., Rajaee, T., & Motahari, M. (2017). Flow forecasting models using hydrologic and hydrometric data. Proceedings of the Institution of Civil Engineers: Water Management, 170(3), 150–162. doi:10.1680/jwama.14.00146.

[17] Motahari, M., Sotoodehnia, A., Nazari, B., & Yazdani, M. (2024). Development of a water resources utilization model and optimization of the patterns of rice genotypes with system dynamics approach. Applied Water Science, 14(11), 235. doi:10.1007/s13201-024-02295-z.

[18] Iran Meteorological Organization, Tehran, Iran. Available online: https://data.irimo.ir/ (accessed on May 2025).

[19] Iran Water Resources Management, Tehran, Iran. Available online: http://wrs.wrm.ir/amar/register.asp (accessed on may 2025).

[20] Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., & Toll, D. (2004). The Global Land Data Assimilation System. Bulletin of the American Meteorological Society, 85(3), 381–394. doi:10.1175/BAMS-85-3-381.

[21] Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, United States. Available online: https://disc.gsfc.nasa.gov/datasets (accessed on may 2025).

[22] Mousavimehr, S. M., & Kavianpour, M. R. (2025). A non-stationary downscaling and gap-filling approach for GRACE/GRACE-FO data under climatic and anthropogenic influences. Applied Water Science, 15(5), 1–15. doi:10.1007/s13201-025-02427-z.

[23] Seber, G. A., & Lee, A. J. (2003). Prediction and Model Selection. Linear Regression Analysis, John Wiley &

Sons, Hoboken, United States. doi:10.1002/9780471722199.ch12.

[24] Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67. doi:10.1080/00401706.1970.10488634.

[25] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, United States. doi:10.1007/978-0-387-84858-7.

[26] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. doi:10.1007/bf00994018.

[27] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. Chapman and Hall/CRC, New York, United States. doi:10.1201/9781315139470.

[28] Haykin, S. (2009). Neural networks and learning machines, 3/E. Pearson Education India, Chennai, India.

[29] Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. doi:10.1007/bf00058655.

[30] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. doi:10.1023/A:1010933404324.

[31] Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. Proceedings of the 13th International Conference on Machine Learning, 96, 148–156. doi:10.1.1.133.1040.

[32] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. doi:10.1214/aos/1013203451.

[33] Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics and Data Analysis, 38(4), 367–378. doi:10.1016/S0167-9473(01)00065-2.

[34] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785.

[35] Alizadeh, M. J., Kavianpour, M. R., Kisi, O., & Nourani, V. (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. Journal of Hydrology, 548, 588–597. doi:10.1016/j.jhydrol.2017.03.032.

[36] Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. PLoS ONE, 14(11), 224365. doi:10.1371/journal.pone.0224365.

[37] Tao, H., Al-Sulttani, A. O., Salih Ameen, A. M., Ali, Z. H., Al-Ansari, N., Salih, S. Q., & Mostafa, R. R. (2020). Training and Testing Data Division Influence on Hybrid Machine Learning Model Process: Application of River Flow Forecasting. Complexity, 2020(1), 8844367. doi:10.1155/2020/8844367.

[38] Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc, Sebastopol, United States.

[39] Mahesh, B. (2020). Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR), 9(1), 381–386. doi:10.21275/art20203995.

[40] Amanollah, H., Asghari, A., Mashayekhi, M., & Zahrai, S. M. (2023). Damage detection of structures based on wavelet analysis using improved AlexNet. Structures, 56, 105019. doi:10.1016/j.istruc.2023.105019.