AI-Driven Spatial Data Analysis of Groundwater Level and Gravimetric Data in Roorkee Region, India

Himangshu Sarkar[1], Chandra Shekhar Prasad Ojha[1], Sanjay Kumar Ghosh[1]

[1]Department of Civil Engineering, Indian Institute of Technology Roorkee, India – hsarkar@ce.iitr.ac.in; c.ojha@ce.iitr.ac.in; sanjay.ghosh@ce.iitr.ac.in

Keywords: Groundwater level prediction, Hydro-gravimetry, Machine Learning, XG-Boost, Random Forest, Sustainable Development

Abstract

Excessive consumption of groundwater can lead to a significant imbalance between groundwater recharge rates and water demand. This disparity underscores the importance of accurately estimating future groundwater storage to ensure global water and food security, in line with sustainable development goals (SDGs) related to clean water and sanitation and sustainable cities and communities. However, traditional methods face challenges in predicting groundwater storage due to their inherent complexity. To address this gap and align with SDGs, this study aims to develop a regression-based machine learning model for spatially varying groundwater level prediction. The primary goal is to improve local water resource management and encourage responsible water usage. The study evaluates the use of K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), XG Boost and Polynomial regression models, using two groups of input parameters. The results show that the XG-Boost model establishes a strong relationship between input and output parameters. The developed KNN model can be reliably used for local groundwater level prediction and can also contribute to sustainable urban development, ultimately aligning with the SDGs.

1. Introduction

Groundwater is a source of fresh water that fulfils approximately one-third of the world's water demands, making it essential for various purposes. It is closely linked to socio-economic development (Robins and Fergusson, 2014). Groundwater serves as a significant supplier of domestic freshwater (36%), water for agriculture (42%), and industrial water demand (27%) (Rulli and D'Odorico, 2013; Taniguchi and Hiyama, 2014; Döll et al., 2012). Despite the projected increase in global water demand, recent studies have raised concerns about declining groundwater levels in different parts of the world (Castellazzi et al., 2016; MacDonald et al., 2016; Richey et al., 2015). The primary causes of this decline are attributed to human activities and the adverse impacts of climate change (Haas and Birk, 2017; Graaf et al., 2017). Additionally, natural factors like evapotranspiration, hydraulic properties, and other events contribute to seasonal fluctuations in water tables (Rathay et al., 2018). Reduced precipitation and higher temperatures during dry periods also play a role in diminishing groundwater levels (Stoll et al., 2011). Furthermore, the increasing reliance on groundwater, along with spatial-temporal variations and disparities in groundwater resources, have worsened the situation (Yu et al., 2018). India, in particular, faces critical challenges due to hydro-climatic variability and droughts, posing significant difficulties for scientists and water resource management.

Advancements in computer modeling, computing power, and information processing have led to the development of practical tools for understanding complex natural systems. In the field of hydrology, researchers have focused on the applicability of machine learning methods to improve groundwater studies (Shortridge et al., 2015). Machine Learning is the subset of artificial Intelligence which develops computer algorithms and statistical models (Kenda et al., 2018). It is well known that the behavior of ML models changes as data structure changes. So far supervised learning techniques seem better for prediction purposes, but both cluster and ensemble learning models also useful subject to availability of good quality of data sets.

Not all machine learning (ML) techniques are universally suitable for groundwater problems due to variations in data quality and availability (Kasiviswanathan et al., 2016). Some methods may underperform with limited, sparse, or noisy data (Brajard et al., 2020; Nguyen et al., 2019). The K-Nearest Neighbor (KNN) method is widely used for both regression and classification due to its simplicity and robustness, making it effective for ML-based prediction and forecasting in noisy data environments (Navot et al., 2005; Nguyen et al., 2019). Random Forest, developed by Leo Breiman in 2001, uses multiple regression trees based on bootstrap resampling, averaging their outputs to improve accuracy and reduce high bias (Breiman 2001; Dietterich 2000; Mohammadi 2019). Support Vector Machines (SVM), proposed by Cortes and Vapnik (1995), aim to find an optimal hyperplane to separate classes, often outperforming earlier classification methods (Farzin et al., 2021). Polynomial regression fits data with a polynomial function, offering moderate flexibility and computational efficiency, and is still considered linear in statistical estimation (Peckov 2012). XG-Boost, a popular boosting algorithm, constructs shallow decision trees, minimizing errors and preventing overfitting through regularization; it has become widely used in data mining due to its adaptability in hyperparameter tuning (Lu and Ma, 2020; Chen and Guestrin, 2016; Bhagat et al., 2020). Each method offers unique advantages depending on data characteristics and problem context.

This study aims to address this challenge by evaluating the effectiveness of five specific machine learning models, namely K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine, Polynomial regression and XG-boost in predicting groundwater levels using limited site-data. The primary focus is on assessing their capacity to handle sparse and noisy samples. To achieve this objective, the study assesses the potential of the ML models in dealing with limited data and has

https://doi.org/10.5194/isprs-annals-X-G-2025-763-2025 |

been explored for its likely contribution to sustainable groundwater management.

2. Material and Methods

2.1 Study Area

This study area is shown in Fig. 1 and lies in the Roorkee block of state Uttarakhand, India, situated between 77° 50' 30.012" E to 77° 56' 30.012" E longitude and 29° 48' 15.012" N to 29° 55' 30" N latitude, covering a total area of 99.401 km² (refer to Fig.1). The area is characterized by flat terrain, with an elevation range of 254 to 279 meters above mean sea level (msl). Land use is primarily agricultural (57%), with residential zones occupying 39% of the area. Groundwater levels at the specified wells range from 5.30 to 9.00 meters below ground level (bgl), with seasonal fluctuations between 0.47 and 3.65 meters. Subsurface layers reveal a cyclic succession of grey micaceous sands, silt, clay/brownish-grey clay, sand, and gravel, interspersed with occasional pebbles and boulders. These deposits are part of the quaternary alluvium associated with terrace, fan, and channel formations (Mishra et al., 2024).

Roorkee experiences a highly unpredictable continental climate, heavily influenced by its proximity to the towering Himalayas (Sudha et al., 2010). The city experiences four distinct seasons. Summer begins in March and lasts until July, with temperatures around 28°C. Following this, the monsoon season brings significant changes in weather, marked by heavy rainfall due to monsoon clouds blocked by the Himalayas. This rainy season typically continues until October. After the monsoons, the post-monsoon period offers mild weather, with temperatures ranging from 15°C to 21°C. Roorkee receives an average annual rainfall of about 260 mm.

The winter season spans from December to February, bringing frigid conditions and occasional cold waves due to katabatic winds descending from the Himalayas. Groundwater is the primary water resource for the city, although the Ganga Canal traverses the center of Roorkee, and the perennial Solani River flows through the area. Both water bodies support agriculture, which significantly contributes to the economy of Roorkee, as well as to the district and state. These water sources also play an essential role in recharging groundwater levels.

The site was selected to study groundwater (GW) level fluctuations around Roorkee. A total of 14 bore wells were identified to monitor temporal groundwater levels through gravity measurements. Additional details about the study area can be found in Fig.1.

2.1.1 Data

A relative gravimeter (ZLS-B-96) with a precision of 0.001 mGal was used to record gravity observations near the observation wells. Since the relative gravimeter measures only relative gravity values, these readings were converted to absolute gravity values using a reference absolute gravity station on the IIT Roorkee campus. Simultaneously, the depth to the water table was measured using a water level indicator, which has a precision of 5 mm. The field data were collected on a weekly basis. Gravity and GWL data collection procedure is presented in Fig.2. Geographical positioning (latitude, longitude, and orthometric height) of the observation wells essential for gravity observations, a dual-frequency GPS receiver was used for this purpose. meteorological data have been collected from the meteorological observatory located at the hydrology department of IIT Roorkee. A total of 165 samples were collected for the gravity-based model and meteorological-based model run. Observation wells have been selected in the area such that to form a network. The gravity data obtained from the observation well has been transformed into absolute gravity values measured in Gal, while the water levels have been converted to orthometric height expressed in meters. Additionally, the time of observation has been converted to seconds.

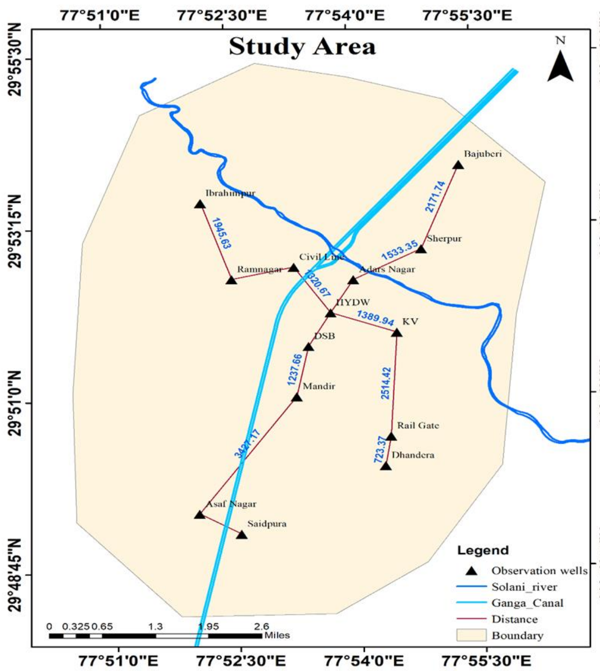


Figure 1. Observation stations are located in the study area.

Figure 2. Gravity and Water level observation taking process at Asaf Nagar site.

2.2 Relation Between Temporal Gravity and GWL

Gravity is the force that the Earth exerts on a unit mass, acting perpendicular to the equipotential surface. According to Newton's law of gravitation, this force depends on the distribution of mass and its spatial position relative to the Earth's surface. At any given point, variations in groundwater levels cause an increase or decrease in the surrounding water mass, directly affecting the gravitational force at that location. This change in mass, reflecting fluctuations in groundwater storage, can be determined by observing changes in the water level. Thus, as per Newton's law, an increase or decrease in mass around a station leads to a corresponding rise or drop in gravity readings. The change in gravity, Δg, associated with variations in total groundwater storage Δh, can be calculated accordingly (Jacob et al., 2010). The equation can be given by

$$\overline{\Delta g} \propto \Delta h \tag{1}$$

2.2.1 Approximation Analysis Between Gravity Variation and GWL Fluctuation

According to the Jacob approximation discussed in section 2.2, the relationship between changes in the value of acceleration due to gravity (Δg) and the altitude difference (Δh) in a time lag, changes in the gravity value have been suggested to vary linearly with the altitude (GWL) difference (Eqn. 1). The above approximations have been analyzed here.

This analysis focuses on examining the applicability of the above relationship over the field dataset recorded for this study. To carry out this study, the gravity values were recorded at different altitudes in the vicinity of the wells to study their influence on groundwater extraction. These readings were used to obtain Δg and Δh values for different step sizes or time lags, i.e., studying the relationship between these two parameters by taking differences between the consecutive values, then increasing the step size and obtaining the values with a step size of two, then increasing the step size to three, and so on. For instance, if the recorded gravity values are:

$$g_1, g_2, g_3, g_4, \dots\dots\dots\dots\dots\dots\dots, g_{(n-1)}, g_n$$

and the respective GWL altitude be:

$$h_1, h_2, h_3, h_4, \dots\dots\dots\dots\dots\dots\dots, h_{(n-1)}, h_n$$

For the first case, keeping the step size at 1:

$$\Delta g_1 = g_2 - g_1, \Delta g_2 = g_3 - g_2,$$
$$\Delta g_3 = g_4 - g_3, \dots\dots\dots \Delta g_{(n-1)} = g_n - g_{(n-1)}$$

and

$$\Delta h_1 = h_2 - h_1, \Delta h_2 = h_3 - h_2, \ \Delta h_3 = h_4 - h_3,$$
$$\dots\dots\dots \Delta h_{(n-1)} = h_n - h_{(n-1)}$$

For the second case, keeping the step size at 2:

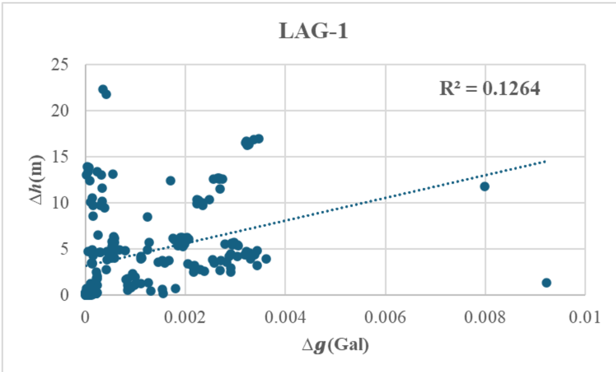
$$\Delta g_1 = g_3 - g_1, \Delta g_2 = g_4 - g_2, \Delta g_3 = g_5 - g_3,$$
$$\dots\dots\dots \Delta g_{(n-2)} = g_n - g_{(n-2)}$$

and

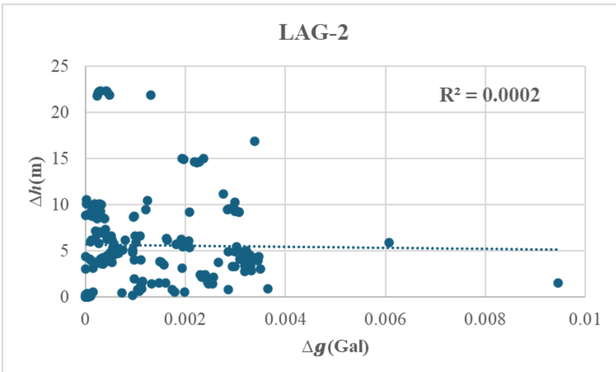
$$\Delta h_1 = h_3 - h_1, \Delta h_2 = h_4 - h_2, \Delta h_3 = h_5 - h_3,$$
$$\dots\dots\dots \Delta h_{(n-2)} = h_n - h_{(n-2)}$$

2.2.2 Visualization of the approximation

Similarly, higher step sizes can be analyzed to study the effect of altitude differences on gravity values over large values. For every step size Δg was plotted against Δh and the relationship between the two parameters was observed and compared with the linear relationship between the two, as stated in section 2.2.1. This analysis has been performed here for a single-well observation case. The approximation has been realized through scatter plots drawn between Δg vs Δh as described in time lag (1- 2) and shown in Fig. 3.



(a)



(b)

Figure 3. Scatter plots between $\Delta g \ vs \ \Delta h$ at different time lags- a. at Lag 1, b. at Lag 2.

Upon comparison, the observation of the plots leads to the conclusion that for every step size, one obtains a nonlinear relationship between the two parameters under study, with significantly low correlation values for all the cases. To encounter this nonlinearity in the relationships between the two parameters, the approximation is being realized by evaluating artificial AI-based machine learning algorithms applied over the dataset to optimize the values towards linearity.

2.3 Development of ML models

In this study, the models were developed using two different sets of parameters. The first set consisted of gravimetric parameters, and the second set was combined with gravity and meteorological parameters shown in Table 1. The dependent variable used in the models was the groundwater level from Mean Sea Level.

Table 1. Model Group

Sr. no	Model Group Names	Parameters	Total no of Samples
1	Gravimetric Group	Gravity, time, Location of wells (latitude, longitude)	165
2	Combined gravity and meteorology group	Gravity, time, Location of wells (latitude and longitude), Mean temperature and precipitation.	165

A comparison was made between these two parameter groups. Prior to training the models with the available data, an Exploratory Data Analysis (EDA) was conducted to understand the data structure, pattern identification, and the selection of models. Subsequently, the data was divided into training and testing datasets in a ratio of 85:15. Hyper-parameter tuning of the selected models was performed using the grid search cross-validation method, and evaluation metrics such as Root Mean Square Error (RMSE) and R^2 were employed. The models were compared with each other and also within their respective parameter groups to assess their performance. If the testing error of the model was deemed high, the initial process was repeated to refine the model. Conversely, if the error fell within an acceptable range, the developed model was compared to other models for further analysis and evaluation. The workflow of the entire process can be visualized in the accompanying Fig. 4. The aim was to analyze the effectiveness of the gravimetric parameters versus the meteorological parameters in predicting the groundwater level.

This section focused on the development of machine learning models (KNN, Random Forest, SVM, XG-Boost, and Polynomial regression) that have been created as part of current research to predict groundwater levels. These prediction models have been developed using the open-source software Jupyter Notebook. Machine learning models applied in practice can be influenced by varying input variable ranges. To put it differently, when input variables possess differing ranges, the model's calculations may be skewed, as it might assign greater importance to input variables with larger ranges, regardless of the actual significance of input variables with smaller ranges in predicting the target variable. Consequently, it is imperative to normalize the data before constructing these predictive models. In the case of the datasets used in the KNN, Random Forest (RF), XG-Boost, SVM, and polynomial regression models, normalization has been executed in such a manner that the values fall within the range of zero to one. This normalization technique employs the minimum and maximum normalization methods, which have been demonstrated to enhance model performance by reducing computation time and minimizing errors during the model execution process.

$$X_n = \frac{(X - X_{min})}{(X_{max} - X_{min})} \tag{2}$$

Where Xmin and Xmax are the minimum and maximum values, respectively, and Xn is the normalized value.

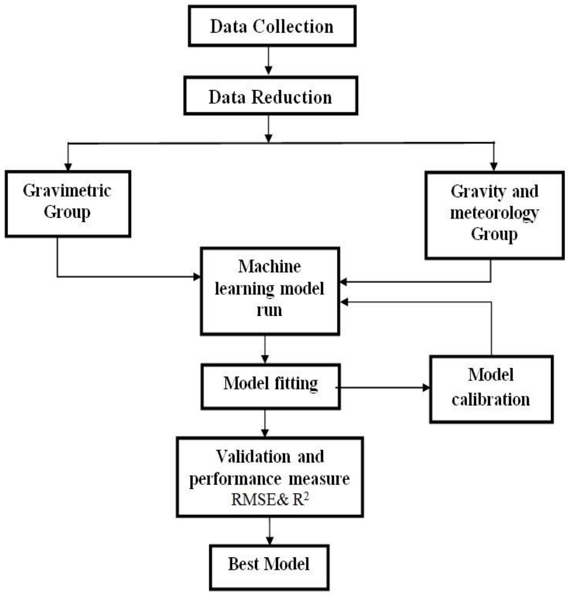


Figure 4. Flowchart of methodology

The process of splitting the normalized input dataset into two segments, namely the training and validation sets. Although there is no universally prescribed ratio for dataset splitting during model calibration and validation, it is generally recommended that the size of the validation dataset should fall within the range of 10% to 40% of the total dataset. In the case of the models applied in our study, the specific data partitioning ratios were determined through a series of trials involving variations in the splitting ratios ranging from 70:30 to 85:15. Ultimately, the choice of the splitting ratio was made based on the criterion of achieving the lowest error, as measured by the Root Mean Square Error (RMSE), during the model calibration process.

2.3.1 Model Calibration

The calibration of machine learning models stands as a crucial phase in the development and implementation of predictive systems. The calibration process encompasses fine-tuning the model by optimizing the training algorithm to minimize the cost function. The model learns its weights from the provided training dataset. However, a common challenge that arises at this stage is overfitting, where the model becomes overly sensitive to noise, resulting in adverse training effects. Hyperparameters, while not directly learned from the training data, introduce complexity and play a crucial role in achieving the optimal model architecture. Consequently, one can engage in hyperparameter tuning for all models as part of the calibration process to ensure that they are finely adjusted and perform at their best.

Calibration also serves as a corrective measure, rectifying any systematic biases or uncertainties inherent in a model's predictions. This not only enhances the model's reliability but also makes it more interpretable and actionable for decision-makers, ultimately leading to more informed and effective choices in various domains, from finance to healthcare to environmental science.

2.3.2 Model Validation

Validation of machine learning models is a critical step in the development and evaluation of predictive models. It involves assessing the model's performance to ensure its accuracy and reliability in predictions. The model's performance during training and testing is assessed using the following metrics: root mean square error (RMSE) and coefficient of determination (R2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_{actual} - y_{predicted})^2} \quad (3)$$

$$R^2 = 1 - \frac{(\sum_{i=1}^{n}(y_{predicted} - y_{mean})^2)}{(\sum_{i=1}^{n}(y_{actual} - y_{mean})^2)} \quad (4)$$

where ‘ypredict’ is predicted output, ‘yactual’ - actual output, ‘ymean’ - mean from actual prediction, and ‘n’ - total number of the dataset.

3. Results and Discussion

For both model groups total 166 numbers of samples were collected from 14 different observation wells. There were a total of 14 wells, of which the water level depth, gravity data, and all the required data have been collected. All five models were trained in both model groups using 85% of the total data, and the rest of the 15% was used for testing. Using metrics such as RMSE and R^2 the model evaluation was done. Additionally, techniques like K-fold cross-validation have been employed to further enhance validation by partitioning the data into multiple subsets and iteratively training and testing the model. As the training samples were less, cross-validation method has been used for model evaluation using five folds in the dataset. Testing samples have been used for testing the models and also to know how the model is predicting the new data. To check the prediction capability of the models, 23 samples have been used to test data sets.

Hyperparameter tuning was carried out using the Grid search cross-validation method in which three cross-validation (CV) folds are used for every model. For linear regression there was no hyperparameter tuning was done. For polynomial regression, grid search CV has been used to identify the correct polynomial degree for the combined gravity and meteorological models. For KNN, the value of k is determined, for the gravimetric model, the value of k is 5, and for the meteorological model, the value of k is 3. For the SVM model, the radial basis function (RBF) kernel has been used as a hyperparameter for both models. Then for both Random forest and XG boost, no of estimates i.e. no of trees for the model and maximum depth of each is determined using hyper parameter tuning. Details about hyper-parameter tuning are shown in Table 2.

In gravimetric parameter model run, among all models XG-Boost model has provided best training and testing RMSE of 0.123 and 0.65 and model has R^2 of 0.985. Although, RF also has training and testing RMSE of 0.182 and 0.678 and R^2 of 0.961. Performance of all models run has been shown in Table 3. As tested the prediction capability of all models, it has been observed that XG-Boost and RF have similar prediction accuracy of 3.17 m and 3.11m. The correlation output of all models prediction capability against the actual presented through the Taylor diagrams in Fig. 5 and Fig. 6.

Table 2. Best parameters for each model

Model	**Hydro-Gravimetric parameters**	**Meteorological Parameter**
KNN	K=5	K = 3
SVM-RBF	kernel='rbf'	kernel='rbf'
Random Forest	max_depth=5, n_estimators= 100	max_depth=3, n_estimators= 100
XG-Boost	max_depth=7, n_estimators=100	max_depth=5, n_estimators= 200
Polynomial regression	Degree = 2	degree = 2

In gravimetric parameter model run, among all models XG Boost model has provided best training and testing RMSE of 0.123 and 0.65 and model has R2 of 0.985. Although, RF also having training and testing RMSE of 0.182 and 0.678 and R2 of 0.961. Performance of all models run has been shown in table 3. As tested the prediction capability of all models, it has been observed that XG boost and RF have similar prediction accuracy of 3.17 m and 3.11m. The correlation output of all model's prediction capability against the actual presented through the taylor diagrams in Fig. 5 and Fig. 6.

Table 3. Training and testing outcomes of all groups of models run.

Model	With Gravimetric Parameters			With Hydro-Gravimetric Parameters		
	RMSE		R^2	RMSE		R^2
	Training	Testing		Training	Testing	
KNN	0.379	0.497	0.867	3.754	4.052	0.211
Random Forest	0.182	0.678	0.961	3.423	3.211	0.343
SVM-RBF	0.392	0.478	0.842	4.457	3.873	0.113
Polynomial-regression	0.761	0.90	0.404	4.144	3.687	0.038
XG Boost	0.123	0.65	0.985	3.339	3.149	0.376

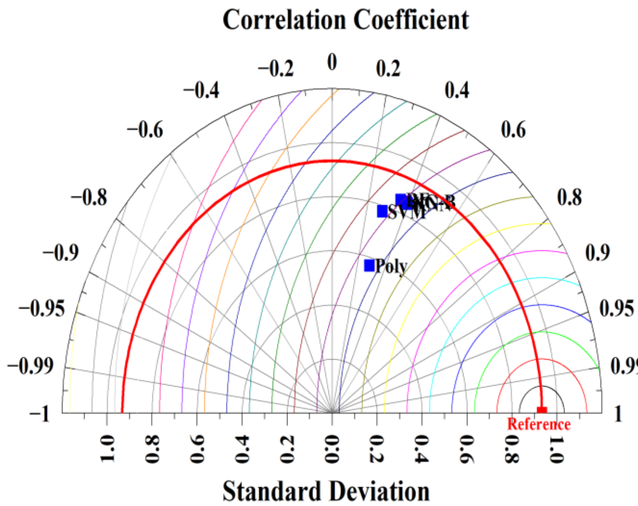


Figure 5. Taylor diagram representing the correlation between observed and predicted GWL during model development in the gravimetric group.

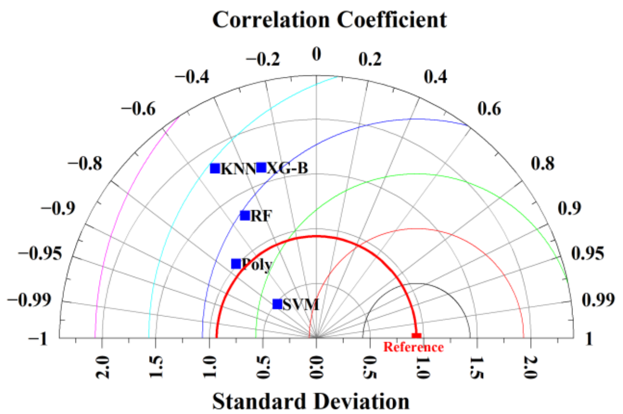


Figure 6. Taylor diagram representing the correlation between observed and predicted GWL during model development in the Hydro-gravimetric group.

In the model run with gravity and meteorological parameters, XG Boost also performs best with training and testing RMSE of 3.339 and 3.149, and the model has R^2 of 0.376. XG Boost provided prediction accuracy of 3.29 m. From the remaining four models, RF and KNN also have shown relatively good performance with training and testing RMSE and R^2. Training and testing performance of all models is shown with detailed information in Table 3. Output of all model's prediction capability against the actual is shown in Fig. 6 and Fig. 7.

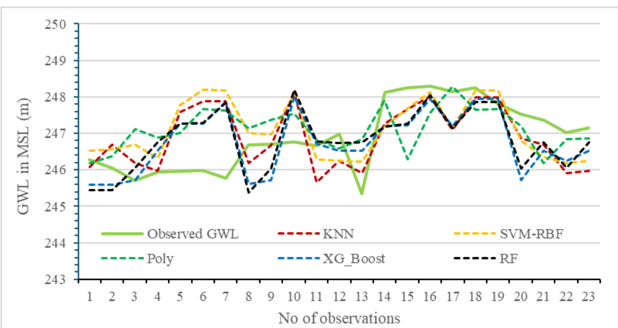


Figure 7. Prediction outcomes of all models with respect to actual data for gravimetric group.

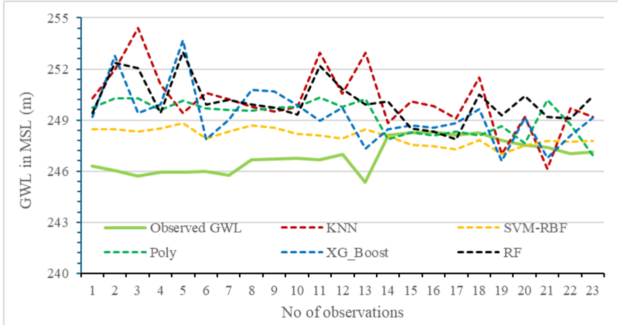


Figure 8. Prediction outcomes of all models with respect to actual data for gravity and meteorological combined group.

In both model groups, it has been observed that SVM has not performed well during training and testing. This denotes that the model has not generated a proper relation between independent variable and dependent variable. But interestingly, the model gives feasible R^2 and also less prediction error. Among all of the models, outputs of polynomial regression tend to over fitting as the training error is very high as compare to the testing error and R^2. In this study the performance of the model categorized as excellent model fit (RMSE $\leq$ 0.50 or $R^2 > 0.75$), good model fit (0.50< RMSE or 0.50 < $R^2 \leq 0.75$), fair model fit (0.60<RMSE 0.70 or 0.25 < $R^2 \leq 0.50$) and poor model fit (RMSE > 0.70 or $R^2 \leq 0.25$).

To assess and visualize differences in model performance, absolute error was calculated for both model groups and displayed through error box diagrams in Fig. 9 and Fig. 10. These diagrams indicate that XG-Boost generally exhibits lower bias with non-linear data structures, achieving a more accurate fit. However, this model may also show high variance due to its sensitivity to specific data points and localized patterns within the training set. This trade-off between bias and variance can lead to decreased performance on new, unseen data, as reflected in the elevated RMSE and absolute error observed in the hydro-gravimetric group, indicating potential overfitting issues.

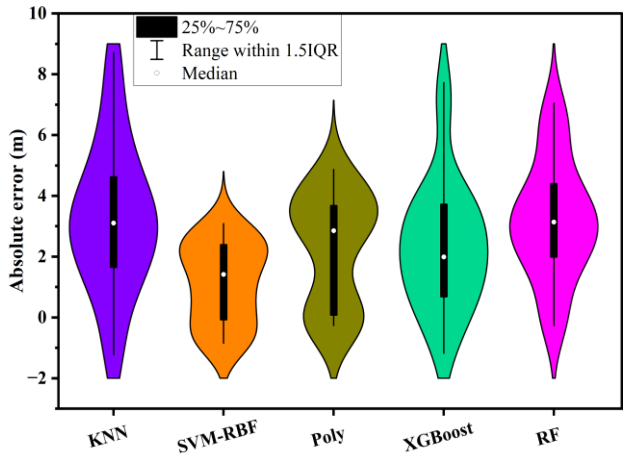


Figure 9. Violin box diagram representing the absolute error between observed and predicted GWL during model development in the gravimetric group.

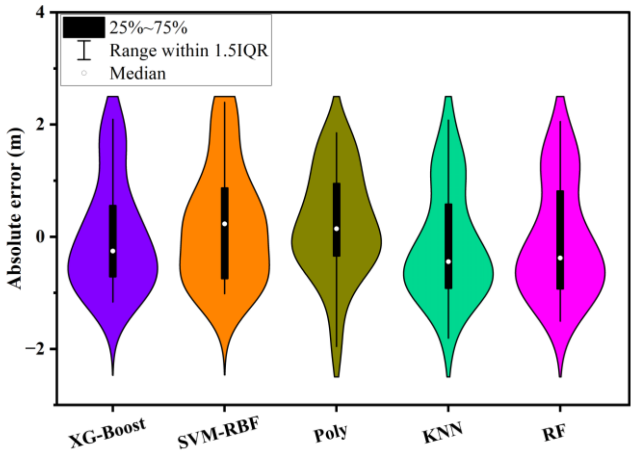


Figure 10. Violin box diagram representing the absolute error between observed and predicted GWL during model development in the Hydro-gravimetric group.

For comparative evaluation between the model groups, it has been observed that the gravimetric model group has been giving good results as compared to meteorological parameter group. Hyper parameters tuning is done to tune the model according to the data set for having good training and testing results. Among the various ML algorithms, XG boost has demonstrated favorable performance when compared to other models and meteorological group. This is evident from the evaluation metrics, such as RMSE and R^2 values, which indicate the accuracy and explanatory power of the models. With limited data source, regression models have shown promising results, indicating their potential for accurate groundwater level estimation. Gravity parameters have shown comparable performance to meteorological parameters, and the inclusion of additional observation data can lead to further improvements.

4. Conclusions

This study has undertaken a comprehensive analysis of five groundwater level prediction models, utilizing data from 14 observation wells and two distinct sets of parameters: gravimetric and meteorological. The primary objective was to evaluate the performance of various machine learning algorithms in predicting groundwater levels, with a focus on model accuracy and generalization of parameters. Several key findings and insights have emerged from this study:

Model Performance and Evaluation:

i. In the both model group, both XG Boost and Random Forest (RF) models demonstrated outstanding performance. XG Boost achieved a remarkable training and testing RMSE. These models displayed a high degree of precision in predicting groundwater levels.

ii. The study employed hyperparameter tuning techniques, including Grid Search Cross-Validation, to optimize model performance. These efforts ensured that each model was fine-tuned to the dataset, enhancing their predictive capabilities.

Comparative Evaluation and Insights:

i. A comparative analysis between the gravimetric and meteorological model groups revealed that the gravimetric models generally outperformed their meteorological counterparts. This underscores the significance of gravimetric data in groundwater level prediction.

ii. Among the various machine learning algorithms assessed, XG Boost consistently demonstrated superior performance, highlighting its effectiveness in handling limited data sources and producing accurate groundwater level estimates.

iii. The study's findings indicate that even with a relatively small dataset, regression models can provide promising results for groundwater level estimation. This suggests the potential for leveraging machine learning approaches in real-world applications related to water resource management.

iv. The study also emphasized the valuable role of gravity parameters in predicting groundwater levels, showing their comparable performance to meteorological parameters. The inclusion of diverse observation data can lead to further improvements in model accuracy.

In conclusion, this study provides valuable insights into the predictive modelling of groundwater levels, offering a framework for data-driven decision-making in water resource management. The demonstrated accuracy and generalization of the models, particularly XG Boost and Random Forest, highlight their potential utility in addressing groundwater-related challenges, which are crucial for sustainable water resource management and urban planning.

References

Bhagat, S.K., Tiyasha, T., Tung, T.M., Mostafa, R.R. and Yaseen, Z.M., 2020: Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicology and Environmental Safety*, *204*, 111059. https://doi.org/10.1016/j.ecoenv.2020.111059.

Brajard, J., Carrassi, A., Bocquet, M. and Bertino, L., 2020: Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of computational science*, *44*, 101171. https://doi.org/10.1016/j.jocs.2020.101171.

Breiman, L., 2001: Random forests. *Machine learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324.

Castellazzi, P., Martel, R., Galloway, D.L., Longuevergne, L. and Rivera, A., 2016: Assessing groundwater depletion and dynamics using GRACE and InSAR: Potential and limitations. *Groundwater*, *54*(6), 768-780. https://doi.org/10.1111/gwat.12453.

Chen, T. and Guestrin, C., 2016: Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794. https://doi.org/10.1145/2939672.2939785.

Cortes, C. and Vapnik, V., 1995: Support vector machine. *Machine learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018.

de Graaf, I.E., van Beek, R.L., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E.H. and Bierkens, M.F., 2017: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. *Advances in water Resources*, *102*, 53-67. https://doi.org/10.1016/j.advwatres.2017.01.011.

Dietterich, T.G., 2000: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, *40*, 139-157. https://doi.org/10.1023/A:1007607513941.

Döll, P., Hoffmann-Dobrev, H., Portmann, F.T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G. and Scanlon, B.R., 2012: Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, *59*, 143-156. https://doi.org/10.1016/j.jog.2011.05.001.

Farzin, M., Avand, M., Ahmadzadeh, H., Zelenakova, M. and Tiefenbacher, J.P., 2021: Assessment of ensemble models for groundwater potential modeling and prediction in a karst watershed. *Water*, *13*(18), 2540. https://doi.org/10.3390/w13182540.

Haas, J.C. and Birk, S., 2017: Characterizing the spatiotemporal variability of groundwater levels of alluvial aquifers in different settings using drought indices. *Hydrology and Earth System Sciences*, *21*(5), 2421-2448. https://doi.org/10.5194/hess-21-2421-2017.

Jacob, T., Bayer, R., Chery, J. and Le Moigne, N., 2010: Time-lapse microgravity surveys reveal water storage heterogeneity of a karst aquifer. *Journal of Geophysical Research: Solid Earth*, *115*(B6). https://doi.org/10.1029/2009JB006616.

Kasiviswanathan, K.S., Saravanan, S., Balamurugan, M. and Saravanan, K., 2016: Genetic programming based monthly groundwater level forecast models with uncertainty quantification. *Modeling Earth Systems and Environment*, *2*, 1-11. https://doi.org/10.1007/s40808-016-0083-0.

Kenda, K., Čerin, M., Bogataj, M., Senožetnik, M., Klemen, K., Pergar, P., Laspidou, C. and Mladenić, D., 2018: August. Groundwater modeling with machine learning techniques: Ljubljana polje aquifer. In *Proceedings* (Vol. 2, No. 11, 697). MDPI. https://doi.org/10.3390/proceedings2110697.

Lu, H. and Ma, X., 2020: Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, *249*, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169.

MacDonald, A.M., Bonsor, H.C., Ahmed, K.M., Burgess, W.G., Basharat, M., Calow, R.C., Dixit, A., Foster, S.S.D., Gopal, K., Lapworth, D.J. and Lark, R.M., 2016: Groundwater quality and depletion in the Indo-Gangetic Basin mapped from in situ observations. *Nature Geoscience*, *9*(10), 762-766. https://doi.org/10.1038/ngeo2791.

Mishra, R., Kumar, S., Sarkar, H. and Ojha, C.S.P., 2024: Utility of Certain AI Models in Climate-Induced Disasters. *World*, *5*(4), 865-900.

Mohammadi, B., 2019: Predicting total phosphorus levels as indicators for shallow lake management. *Ecological Indicators*, *107*, 105664. https://doi.org/10.1016/j.ecolind.2019.105664.

Navot, A., Shpigelman, L., Tishby, N. and Vaadia, E., 2005: Nearest neighbor based feature selection for regression and its application to neural activity. *Advances in neural information processing systems*, *18*.

Nguyen, D., Ouala, S., Drumetz, L. and Fablet, R., 2019: Em-like learning chaotic dynamics from noisy and partial observations. *arXiv preprint arXiv:1903.10335*.https://doi.org/10.48550/arXiv.1903.10335.

Peckov, A., 2012: A machine learning approach to polynomial regression. *Ljubljana, Slovenia, URL: http://kt. ijs. si/theses/phd_aleksandar_peckov. pdf*.

Rathay, S.Y., Allen, D.M. and Kirste, D., 2018: Response of a fractured bedrock aquifer to recharge from heavy rainfall events. *Journal of Hydrology*, *561*, 1048-1062. https://doi.org/10.1016/j.jhydrol.2017.07.042.

Richey, A.S., Thomas, B.F., Lo, M.H., Reager, J.T., Famiglietti, J.S., Voss, K., Swenson, S. and Rodell, M., 2015: Quantifying renewable groundwater stress with GRACE. *Water resources research*, *51*(7), 5217-5238. https://doi.org/10.1002/2015WR017349.

Robins, N.S. and Fergusson, J., 2014: Groundwater scarcity and conflict–managing hotspots. *Earth perspectives*, *1*, 1-9. https://doi.org/10.1186/2194-6434-1-6.

Rulli, M.C. and D'Odorico, P., 2013: The water footprint of land grabbing. *Geophysical Research Letters*, *40*(23), 6130-6135. https://doi.org/10.1002/2013GL058281.

Shortridge, J.E., Guikema, S.D. and Zaitchik, B.F., 2015: Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds. *Hydrology & Earth System Sciences Discussions*, *12*(10). doi:10.5194/hessd-12-11083-2015.

Stoll, S., Hendricks Franssen, H.J., Butts, M. and Kinzelbach, W., 2011: Analysis of the impact of climate change on groundwater related hydrological fluxes: a multi-model approach including different downscaling methods. *Hydrology and Earth System Sciences*, *15*(1), 21-38. https://doi.org/10.5194/hess-15-21-2011.

Sudha, Tezkan, B., Israil, M., Singhal, D.C. and Rai, J., 2010: Geoelectrical mapping of aquifer contamination: a case study from Roorkee, India. *Near Surface Geophysics*, *8*(1), 33-42. https://doi.org/10.3997/1873-0604.2009049.

Taniguchi, M. and Hiyama, T. eds., 2014. *Groundwater as a key for adaptation to changing climate and society* (pp. 145-pp). Springer. ISBN : 978-4-431-54967-3. (Book)

Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J. and Wu, M., 2018: Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. *Water resources management*, *32*, 301-323. https://doi.org/10.1007/s11269-017-1811-6.