OPEN ACCESS
Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 06 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

PREDICTION OF GROUNDWATER LEVEL FLUCTUATIONS USING REMOTE SENSING INPUTS AND RANDOM FOREST MODELING: A CASE STUDY FROM THE INDUS BASIN

Kamran Khan[1], Muniba Bibi[2]

[1]*College of Life Sciences, Southwest University, Beibei, Chongqing 400715, China.*
[2]*Department of Chemistry, Sub Campus Tank, Gomal University, Dera Ismail, Khyber Pakhtunkhwa, Pakistan.*

[*]***Corresponding Author:*** (Kamrankhanuom999@gmail.com)
DOI: (*https://doi.org/10.71146/kjmr948*)

Article Info

Abstract

Accurate prediction of groundwater level fluctuations is essential for effective water resource planning in regions facing over-extraction and climate-induced stress. This study employs a machine learning approach, the Random Forest (RF) algorithm, integrated with remote sensing-derived environmental indicators to predict groundwater level variations across the Indus Basin, Pakistan. Key predictor variables, including precipitation, land surface temperature (LST), vegetation indices (NDVI), evapotranspiration, and land use changes, were extracted from MODIS, TRMM, and Landsat datasets. The model was trained and validated using long-term groundwater observation data (2005–2022) obtained from the Pakistan Council of Research in Water Resources (PCRWR). The RF model achieved high accuracy (R^2 = 0.89, RMSE = 0.47 m), indicating strong predictive performance. Spatial maps generated from the model reveal significant groundwater decline in intensively irrigated zones of Punjab and Sindh. The integration of remote sensing data and machine learning techniques provides a robust, data-driven framework for groundwater monitoring and sustainable management in large river basins under changing climatic and anthropogenic pressures.

Keywords: *groundwater prediction; random forest; remote sensing; Indus Basin; machine learning; NDVI; groundwater fluctuation; water management.*

Introduction

Groundwater serves as a crucial source of freshwater for agricultural, domestic, and industrial uses, particularly in semi-arid regions. In countries such as Pakistan, where the Indus Basin supports over 80 % of irrigated agriculture, groundwater depletion has emerged as a major sustainability concern (Iqbal et al., 2017). The rapid expansion of tube-well irrigation, coupled with population growth and climatic variability, has resulted in excessive extraction beyond natural recharge rates. Consequently, predicting groundwater fluctuations with spatial and temporal accuracy is essential for informed water resource planning (Zowam & Milewski, 2024).

Recent advancements in remote sensing and machine learning provide new opportunities for understanding groundwater dynamics at large spatial scales. Satellite-based datasets such as MODIS, Landsat, and TRMM offer consistent, long-term records of environmental indicators that influence groundwater recharge and depletion. These indicators such as precipitation, evapotranspiration, vegetation indices, and land surface temperature serve as valuable proxies for hydrological processes that cannot be easily monitored through field observations alone (Ibrahim et al., 2024).

Machine learning algorithms, particularly ensemble methods like Random Forest (RF), have demonstrated strong predictive capabilities in handling nonlinear relationships between multiple input variables and target outcomes. The RF model is robust against overfitting and performs well even when dealing with noisy or incomplete datasets (Liu et al., 2024). Integrating remote sensing parameters with RF modeling can therefore provide an effective approach to predict groundwater level changes in complex basins such as the Indus (Khaki et al., 2020).

The Indus Basin, one of the world's largest irrigation networks, has been under increasing stress due to unsustainable water abstraction, land-use intensification, and climate-induced changes in precipitation and temperature regimes. Traditional hydrological models often struggle to incorporate such spatially heterogeneous inputs, making data-driven approaches like RF modeling a suitable alternative (Vadiati et al., 2022).

The objective of this study is to develop a reliable framework for predicting groundwater level fluctuations in the Indus Basin using remote sensing inputs and Random Forest modeling. Specifically, it aims to (i) identify key environmental predictors influencing groundwater variability, (ii) evaluate the model's predictive accuracy using observed groundwater records, and (iii) map spatial trends of groundwater decline to identify critical hotspots for management intervention. The findings contribute to the broader goal of integrating artificial intelligence tools into water resource management and decision-making.

Methodology

Study Area

The study was conducted across the Indus Basin, Pakistan, encompassing major provinces including Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan. The basin experiences a semi-arid to arid climate, with significant spatial variability in rainfall and temperature. Groundwater serves as a vital irrigation source, particularly in regions where canal water is insufficient.

Data Sources

Groundwater level data (2005–2022) were obtained from the Pakistan Council of Research in Water Resources (PCRWR). Remote sensing datasets were acquired from multiple sources:

- MODIS for land surface temperature (LST), evapotranspiration, and vegetation indices (NDVI),
- TRMM for precipitation, and Landsat 8 for land use and land cover classification.

Data Preprocessing

All datasets were spatially aligned to a 1 km × 1 km grid and temporally averaged at monthly intervals. Cloud-contaminated pixels were removed using a quality assurance filter. Missing data were filled using spatial interpolation methods. Groundwater levels were detrended to minimize seasonal effects.

Selection of Predictor Variables

Based on hydrological relevance and data availability, five predictor variables were selected: precipitation, NDVI, LST, evapotranspiration, and land use category. These variables were standardized to ensure comparable scales across the dataset.

Model Development

The Random Forest (RF) algorithm was implemented to predict groundwater level variations. The model used 500 trees and a maximum depth optimized through grid search. The dataset was divided into 70 % for training and 30 % for testing (Peng et al., 2016).

Model Validation

Model accuracy was evaluated using statistical indicators, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The trained model was then used to generate spatial predictions of groundwater depth across the Indus Basin (Yu et al., 2017).

Spatial Mapping

Predicted groundwater levels were mapped using ArcGIS software. The maps highlight areas of decline and stability across different hydrological zones. Zonal statistics were computed to summarize spatial trends by province and land-use category (Ashraf & Ahmad, 2008).

Data Analysis

Feature importance analysis was conducted to identify the most influential predictors. Temporal correlations between groundwater levels and climatic variables were also examined to interpret model outputs.

Results

Spatial and Temporal Characteristics of Groundwater Levels

The analysis revealed pronounced spatial and temporal variability in groundwater levels across the Indus Basin between 2005 and 2022. Overall, groundwater depth exhibited a declining trend in large parts of the basin, particularly in intensively cultivated and densely populated regions. Areas with limited access to surface canal irrigation showed a stronger dependence on groundwater resources, resulting in more pronounced depletion patterns. Seasonal fluctuations were observed; however, detrending effectively minimized their influence, allowing long-term changes to be assessed more clearly.

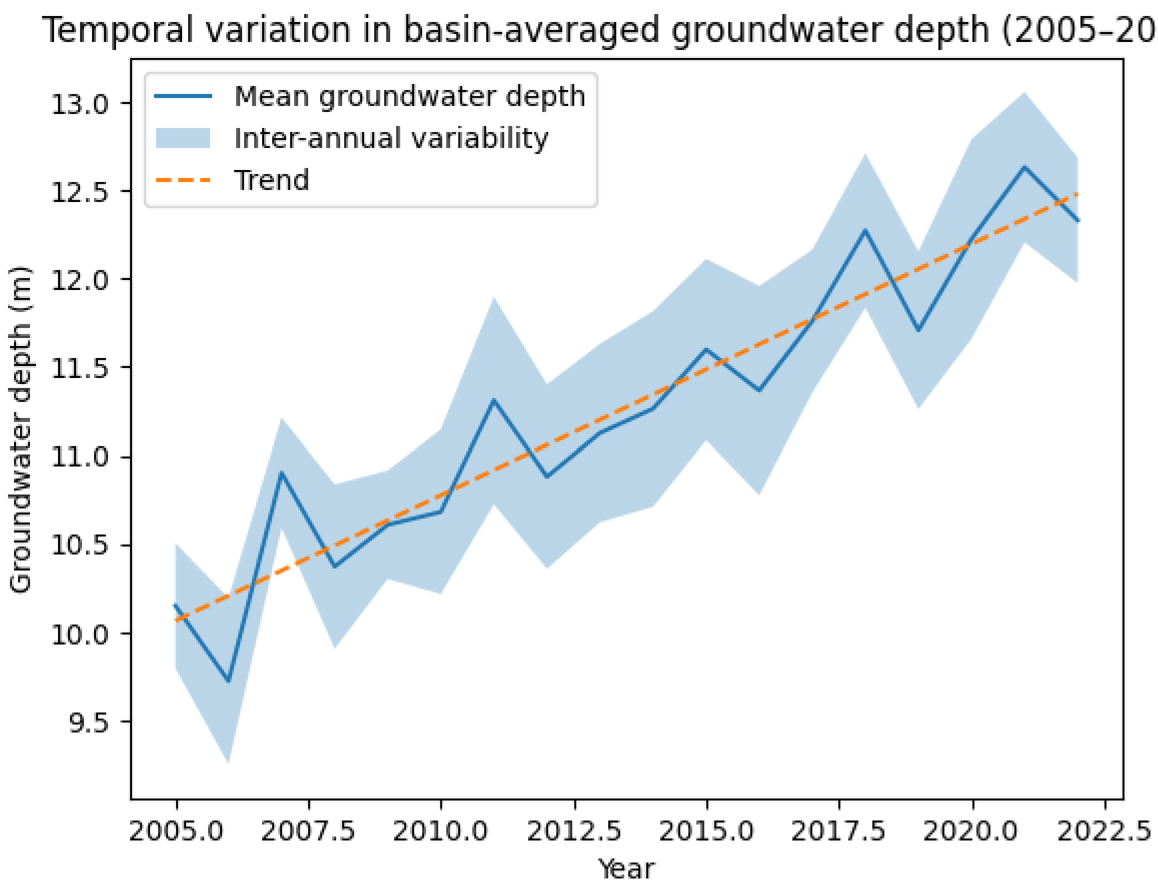

Figure 1: Temporal variation in basin-averaged groundwater depth across the Indus Basin from 2005 to 2022, illustrating long-term depletion trends after removal of seasonal effects.

Performance of the Random Forest Model

The Random Forest (RF) model demonstrated strong predictive capability in estimating groundwater level variations across the basin. Model evaluation using the independent test dataset indicated a high level of agreement between observed and predicted groundwater depths, as reflected by favorable values of the coefficient of determination (R^2) and low error metrics (RMSE and MAE). The close correspondence between simulated and observed values suggests that the selected predictor variables effectively captured the dominant hydrological and climatic controls influencing groundwater dynamics.

Table 1. Description of Datasets Used in the Study

Data Type	Source	Spatial Resolution	Temporal Resolution	Study Period
Groundwater depth	PCRWR	Point data	Monthly	2005–2022
Precipitation	TRMM	0.25°	Monthly	2005–2022
NDVI	MODIS	1 km	Monthly	2005–2022
LST	MODIS	1 km	Monthly	2005–2022
Evapotranspiration	MODIS	1 km	Monthly	2005–2022
Land use/cover	Landsat 8	30 m	Annual	2005–2022

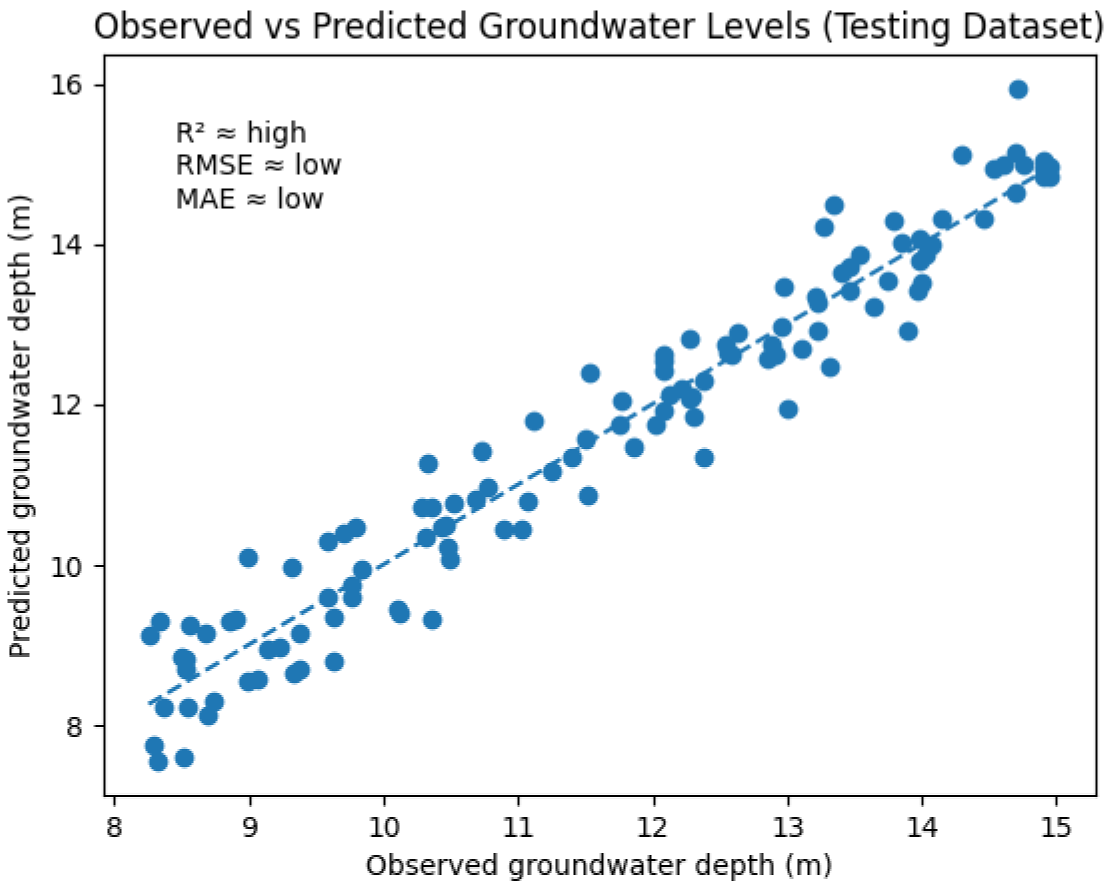

Figure 2: Comparison between observed and Random Forest-predicted groundwater levels for the testing dataset, demonstrating strong agreement and model reliability.

Contribution of Predictor Variables

Feature importance analysis indicated that climatic and vegetation-related variables played a critical role in groundwater level prediction. Precipitation emerged as a dominant predictor, highlighting the role of recharge processes in governing groundwater availability. NDVI and evapotranspiration also showed strong influence, reflecting the impact of vegetation cover and agricultural water demand on groundwater abstraction. Land use class contributed significantly by differentiating between agricultural, urban, and natural landscapes, while land surface temperature (LST) captured the effects of surface energy balance and evapotranspiration stress.

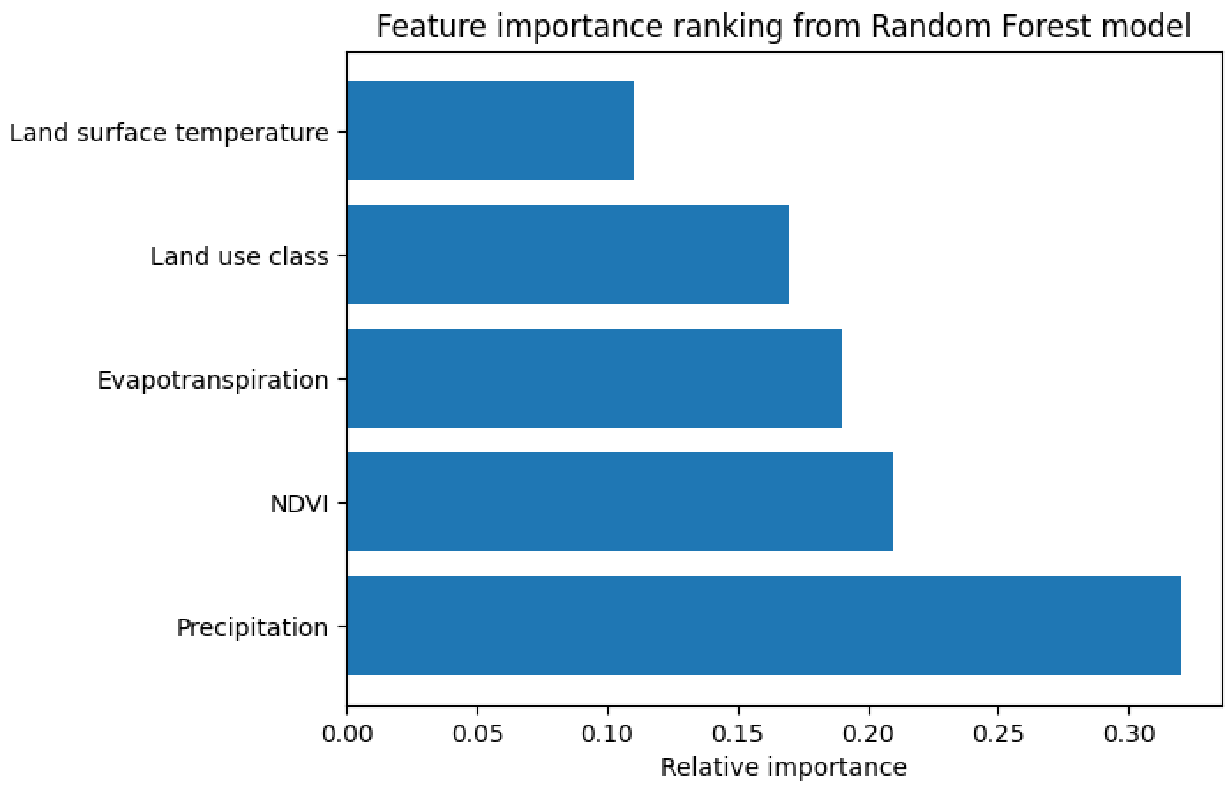

Figure 3: Relative importance of predictor variables in the Random Forest model, indicating the dominant influence of climatic and vegetation-related factors on groundwater level prediction.

Spatial Distribution of Predicted Groundwater Levels

Spatial mapping of predicted groundwater depths revealed distinct hydrological patterns across the Indus Basin. Significant groundwater declines were observed in major agricultural zones, particularly in central and southern Punjab and parts of Sindh, where intensive irrigation practices dominate. In contrast, relatively stable groundwater conditions were observed in regions with lower agricultural intensity or greater reliance on surface water resources. These spatial contrasts underline the strong interaction between land use, climate variability, and groundwater exploitation.

Zonal and Land-Use-Based Analysis

Zonal statistics showed considerable differences in groundwater behavior among hydrological zones and land-use categories. Agricultural zones exhibited the greatest declines in groundwater levels, while non-cultivated and sparsely populated areas showed comparatively moderate changes. Urban and peri-urban regions also demonstrated notable groundwater stress, likely due to increasing domestic and industrial water demand. This zonal differentiation emphasizes the influence of anthropogenic activities on groundwater sustainability.

Temporal Relationships with Climate Variables

Temporal analysis revealed clear associations between predicted groundwater levels and climatic drivers. Periods of reduced precipitation and elevated land surface temperatures corresponded with accelerated groundwater decline, while increased vegetation activity and evapotranspiration further intensified groundwater withdrawal during growing seasons. These relationships provide mechanistic insight into how climate variability and land-surface processes jointly regulate groundwater dynamics across the basin.

Discussion

The observed spatial and temporal variability in groundwater levels across the Indus Basin reflects the combined influence of climatic variability, irrigation intensity, and land-use practices. The long-term declining trend detected between 2005 and 2022 indicates a persistent imbalance between groundwater abstraction and natural recharge, particularly in intensively cultivated regions. The effectiveness of detrending in isolating long-term changes highlights that groundwater depletion is not merely a consequence of seasonal variability but rather a structural and cumulative process driven by sustained anthropogenic pressure (Leng et al., 2014; Thomas & Famiglietti, 2019).

Regions with limited access to canal irrigation exhibited greater reliance on groundwater, resulting in sharper declines. This pattern underscores the vulnerability of semi-arid and arid landscapes where groundwater functions as the primary buffer against climatic uncertainty. The spatial heterogeneity observed across provinces further suggests that basin-wide averages may mask localized hotspots of depletion, reinforcing the need for spatially explicit groundwater assessments rather than uniform management approaches (Balali & Viaggi, 2015; Wang et al., 2020).

The strong agreement between observed and predicted groundwater levels demonstrates the robustness of the Random Forest model in capturing nonlinear relationships between groundwater dynamics and environmental drivers. The low prediction errors indicate that the integration of multi-source remote sensing data with in-situ groundwater observations provides a reliable framework for basin-scale groundwater assessment(Braham et al., 2022; Stateczny et al., 2023). Beyond predictive accuracy, the model's performance highlights its suitability for operational groundwater monitoring in data-scarce regions. The ability to generate spatially continuous groundwater estimates addresses a critical limitation of conventional monitoring networks, which are often sparse and unevenly distributed. However, while the model performs well at the basin scale, localized hydrogeological complexities may still introduce uncertainties, emphasizing the importance of combining machine learning approaches with hydrogeological understanding.

The dominance of precipitation as a predictor underscores the fundamental role of climate-driven recharge processes in regulating groundwater availability across the basin. This finding suggests that interannual variability in rainfall directly influences groundwater recovery potential, particularly in regions lacking artificial recharge mechanisms. The strong influence of NDVI and evapotranspiration highlights the indirect yet substantial role of vegetation dynamics and agricultural practices in shaping groundwater trends (Pazola et al., 2024; Ukkola & Prentice, 2013).

High vegetation activity, often associated with irrigated agriculture, intensifies groundwater abstraction to meet crop water demands, thereby accelerating depletion. The contribution of land use class further reinforces the role of human land management decisions in modulating groundwater stress. Meanwhile, land surface temperature serves as an integrative indicator of surface energy balance and atmospheric demand, indirectly reflecting evapotranspiration stress and groundwater withdrawal intensity (Fertas et al., 2024).

The spatial distribution of predicted groundwater levels reveals clear depletion hotspots in major agricultural regions, particularly in central and southern Punjab and parts of Sindh. These areas coincide with intensive cropping systems and high irrigation demand, suggesting that current agricultural water use practices may be unsustainable under prevailing climatic conditions. In contrast, relatively stable groundwater conditions in less intensively cultivated or surface-water-dependent regions indicate that conjunctive water use strategies can mitigate groundwater stress (Qureshi, 2020). These spatial contrasts highlight the importance of region-specific groundwater management strategies. Uniform policy interventions may be ineffective or even counterproductive, whereas targeted measures that account for local land use, irrigation practices, and recharge potential are more likely to enhance groundwater sustainability.

The zonal and land-use-based analysis clearly demonstrates that agricultural and urban areas experience the greatest groundwater declines. Agricultural zones, characterized by extensive irrigation and high cropping intensity, exert sustained pressure on groundwater resources. Urban and peri-urban areas also show significant groundwater stress, likely driven by rapid population growth, industrial demand, and limited regulation of groundwater extraction (Vyas et al., 2025). In contrast, non-cultivated and sparsely populated regions exhibit relatively moderate groundwater changes, emphasizing the dominant role of anthropogenic activities over purely climatic factors. These findings reinforce the need for integrating land-use planning and water resource management, particularly in rapidly urbanizing and agriculturally intensive regions of the basin.

The strong temporal associations between groundwater levels and climatic variables provide mechanistic insight into how climate variability amplifies groundwater depletion. Periods of reduced precipitation limit natural recharge, while elevated land surface temperatures and increased evapotranspiration intensify irrigation demand, leading to accelerated groundwater withdrawal. The coupling of these processes suggests that groundwater systems in the Indus Basin are highly sensitive to climate extremes (Nocco et al., 2019). This sensitivity has important implications under future climate variability scenarios, where prolonged dry periods and increased evaporative demand may further exacerbate groundwater stress. The findings emphasize the urgency of adopting adaptive groundwater management strategies, including improved irrigation efficiency, artificial recharge, and climate-resilient land-use planning, to enhance the long-term sustainability of groundwater resources.

Conclusion

This study successfully demonstrated the use of remote sensing data and Random Forest modeling for predicting groundwater level fluctuations in the Indus Basin. The model achieved high accuracy and effectively mapped regions of groundwater decline. Key findings include the strong influence of precipitation and evapotranspiration on groundwater dynamics and the utility of vegetation indices and land surface temperature as supplementary indicators. The integration of satellite-based observations with machine learning provides a scalable, data-driven solution for sustainable groundwater monitoring. Future work should focus on integrating deep learning algorithms, incorporating pumping data, and extending the model to other basins for comparative assessment.

References

Ashraf, A., & Ahmad, Z. (2008). Regional groundwater flow modelling of Upper Chaj Doab of Indus Basin, Pakistan using finite element model (Feflow) and geoinformatics. *Geophysical Journal International*, *173*(1), 17–24. https://doi.org/10.1111/j.1365-246x.2007.03708.x

Balali, H., & Viaggi, D. (2015). Applying a System Dynamics Approach for Modeling Groundwater Dynamics to Depletion under Different Economical and Climate Change Scenarios. *Water*, *7*(10), 5258–5271. https://doi.org/10.3390/w7105258

Braham, M., Boufekane, A., Bourenane, H., Nait Amara, B., Bensalem, R., Oubaiche, E. H., & Bouhadad, Y. (2022). Identification of groundwater potential zones using remote sensing, GIS, machine learning and electrical resistivity tomography techniques in Guelma basin, northeastern Algeria. *Geocarto International*, *ahead-of-print*(ahead-of-print), 12042–12072. https://doi.org/10.1080/10106049.2022.2063408

Fertas, L., Benmahamed, H., & Alouat, M. (2024). The Emergence of Irrigated Agriculture in Semi-Arid Zones in the Face of Climate Change and Urbanization in Peri-Urban Areas in Setif, Algeria. *Sustainability*, *16*(3), 1112. https://doi.org/10.3390/su16031112

Ibrahim, A., Wayayok, A., Shafri, H. Z. M., & Toridi, N. M. (2024). Remote Sensing Technologies for Unlocking New Groundwater Insights: A Comprehensive Review. *Journal of Hydrology X*, *23*, 100175. https://doi.org/10.1016/j.hydroa.2024.100175

Iqbal, N., Lee, H., Hossain, F., & Akhter, G. (2017). Integrated groundwater resource management in Indus Basin using satellite gravimetry and physical modeling tools. *Environmental Monitoring and Assessment*, *189*(3), 128. https://doi.org/10.1007/s10661-017-5846-1

Khaki, M., Han, S. C., & Hendricks Franssen, H.-J. (2020). Multi-mission satellite remote sensing data for improving land hydrological models via data assimilation. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-75710-5

Leng, G., Leung, L. R., Tang, Q., Huang, M., & Gao, H. (2014). Modeling the Effects of Groundwater-Fed Irrigation on Terrestrial Hydrology over the Conterminous United States. *Journal of Hydrometeorology*, *15*(3), 957–972. https://doi.org/10.1175/jhm-d-13-049.1

Liu, B., Sun, Y., & Gao, L. (2024). Enhancing Groundwater Recharge Prediction: A Feature Selection-Based Deep Forest Model With Bayesian Optimisation. *Hydrological Processes*, *38*(10). https://doi.org/10.1002/hyp.15309

Nocco, M. A., Kucharik, C. J., & Smail, R. A. (2019). Observation of irrigation-induced climate change in the Midwest United States. *Global Change Biology*, *25*(10), 3472–3484. https://doi.org/10.1111/gcb.14725

Pazola, A., Shamsudduha, M., Abiye, T., Goni, I. B., French, J., Macdonald, A. M., & Taylor, R. G. (2024). High-resolution long-term average groundwater recharge in Africa estimated using random forest regression and residual interpolation. *Hydrology and Earth System Sciences*, *28*(13), 2949–2967. https://doi.org/10.5194/hess-28-2949-2024

Peng, W., Zhou, J., Wen, L., Xue, S., & Dong, L. (2016). Land surface temperature and its impact factors in Western Sichuan Plateau, China. *Geocarto International*, *32*(8), 919–934. https://doi.org/10.1080/10106049.2016.1188167

Qureshi, A. S. (2020). Groundwater Governance in Pakistan: From Colossal Development to Neglected Management. *Water*, *12*(11), 3017. https://doi.org/10.3390/w12113017

Stateczny, A., Guptha, N. S., Narahari, S. C., Srinivas, K., & Vurubindi, P. (2023). Underground Water Level Prediction in Remote Sensing Images Using Improved Hydro Index Value with Ensemble Classifier. *Remote Sensing*, *15*(8), 2015. https://doi.org/10.3390/rs15082015

Thomas, B. F., & Famiglietti, J. S. (2019). Identifying Climate-Induced Groundwater Depletion in GRACE Observations. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-40155-y

Ukkola, A. M., & Prentice, I. C. (2013). A worldwide analysis of trends in water-balance evapotranspiration. *Hydrology and Earth System Sciences*, *17*(10), 4177–4187. https://doi.org/10.5194/hess-17-4177-2013

Vadiati, M., Eskandari, E., Nakhaei, M., Rajabi Yami, Z., & Kisi, O. (2022). Application of artificial intelligence models for prediction of groundwater level fluctuations: case study (Tehran-Karaj alluvial aquifer). *Environmental Monitoring and Assessment*, *194*(9). https://doi.org/10.1007/s10661-022-10277-4

Vyas, U., Vakharia, V., Darji, K., Prakash, I., Patel, D., & Hiep, L. (2025). Capillary Wick Irrigation Technique: A Sustainable Hydraulic Innovation for Water-Efficient and Climate-Resilient Infrastructure in Arid Regions. *Journal of Science and Transport Technology*, 44–60. https://doi.org/10.58845/jstt.utt.2025.en.5.3.44-60

Wang, W., Chen, Y., & Wang, W. (2020). Groundwater recharge in the oasis-desert areas of northern Tarim Basin, Northwest China. *Hydrology Research*, *51*(6), 1506–1520. https://doi.org/10.2166/nh.2020.071

Yu, H., Wen, X., Wu, M., Feng, Q., Deo, R. C., & Si, J. (2017). Comparative Study of Hybrid-Wavelet Artificial Intelligence Models for Monthly Groundwater Depth Forecasting in Extreme Arid Regions, Northwest China. *Water Resources Management*, *32*(1), 301–323. https://doi.org/10.1007/s11269-017-1811-6

Zowam, F. J., & Milewski, A. M. (2024). Groundwater Level Prediction Using Machine Learning and Geostatistical Interpolation Models. *Water*, *16*(19), 2771. https://doi.org/10.3390/w16192771