Theme 11

APPLICATION OF AI, ML, IoT, CLOUD COMPUTING AND OTHER ADVANCED TECHNIQUES IN GROUNDWATER

COMPLEXITY AND CONNECTIVITY IN GROUNDWATER SYSTEMS: ROLE OF CHAOS THEORY AND COMPLEX NETWORK THEORY

Bellie Sivakumar* and Deepthi B

Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, India
**Corresponding author e-mail: b.sivakumar@iitb.ac.in*

Groundwater constitutes more than a quarter of the world's available freshwater resources. Over the past few decades, population growth, improved living standards, diminishing per capita availability of surface water of required quality, and advancements in technology have, in one way or another, significantly increased the exploitation of groundwater resources. It is important to note that groundwater serves as the only reliable source of safe drinking water and domestic supply in many countries and is also a primary or sole water source for irrigation in many regions. As a result, there are already serious concerns about the exploitation and sustainability of groundwater resources. Climate change and environmental degradation are expected to exacerbate this problem in future. Ensuring sustainable planning and management of groundwater resources, including conjunctive use, is vital for our future water, food, and energy security. To achieve this, a thorough understanding of groundwater flow and transport dynamics is one of the essential requirements.

Over the past several decades, a wide range of methodologies and mathematical models have been developed to study groundwater flow and transport. Despite significant advancements in such theoretical developments and their practical applications, many challenges in groundwater modelling still persist, especially in representing the complexity and connections in groundwater systems. In recent times, there have been some significant theoretical developments in the field of complex systems science, to study the complexity and connections in the systems. Among these, chaos theory and complex network theory have been gaining significant attention in hydroclimatic studies, including identification of the system complexity, prediction, catchment classification, propagation of droughts and other extreme hydroclimatic events, performance assessment of General Circulation Models (GCMs), and downscaling of outputs from GCMs. These theories offer unique perspectives on the complexity, connectivity, and nonlinearity of hydroclimatic systems. Despite their relevance and usefulness in hydro-climatology, these theories have not gained much attraction in studying groundwater systems, especially when compared with the volume of studies on surface water hydrologic systems.

The present study explores the application of chaos theory and complex network theory to examine the complexity and connectivity in groundwater systems. In particular, the study focuses on the temporal dynamics of groundwater levels at various locations across a region. For implementation, a groundwater level monitoring network of 200 wells across the United States has been considered. These 200 wells are spread across California, Arkansas, Nebraska, Idaho, and Texas, which collectively account for 46% of the total fresh groundwater withdrawals across all categories and have significant groundwater usage for irrigation. Daily groundwater level data observed over the period 2010–2020, obtained from the US Geological Survey (USGS), have been studied.

For studying the temporal dynamics of groundwater levels, each well has been considered as part of the network. Chaos theory-based methods have been coupled with complex network theory concepts to construct the groundwater level network. In particular, phase-space reconstruction and the False Nearest Neighbor (FNN) algorithm, two popular chaos theory-based methods, have been employed. The concept of phase space reconstruction has been applied to transform the single-variable groundwater level time series into a multi-dimensional phase space using an embedding technique. The optimal embedding dimension for reconstruction has been determined using the FNN algorithm. During the network construction process, each reconstructed vector in the phase space is treated as a node, while the connections between these nodes, established based on a distance threshold for the reconstructed vectors, are defined as links.

Five complex networks-based measures—clustering coefficient, degree centrality, closeness centrality, betweenness centrality, and shortest path length—have been used to analyze the properties of the groundwater level network. The clustering coefficient measures the extent to which nodes in the network form tightly connected groups. In groundwater systems, high clustering indicates regions with similar hydrologic behavior, reflecting localized patterns of recharge or pumping. Degree centrality identifies the number of connections a node has, highlighting its influence within the network. Wells with high degree centrality are critical as they may impact groundwater flow across broader areas, making them essential for monitoring and management. Closeness centrality evaluates the efficiency of information transfer across the network. Wells with high closeness centrality represent areas where groundwater dynamics are highly interconnected. Betweenness centrality measures the extent to which a node acts as a bridge in the network. Wells with high betweenness centrality are pivotal for maintaining network connectivity and can play a crucial role in transferring water or contaminants. Finally, shortest path length quantifies the efficiency of a network in transmitting information between the nodes in the network, indicating how quickly changes, such as recharge or contamination, propagate through the wells.

The results obtained from the present analysis have been interpreted to provide significant insights into the characteristics of the groundwater levels. For example, the FNN dimensions obtained for the groundwater level time series from the 200 wells have been interpreted to describe the complexity of groundwater level dynamics. The results from the five complex networks-based measures have been used to identify the extent of temporal connections in groundwater level dynamics and type of network. These results further help in identifying the appropriate model for studying the groundwater monitoring network. Based on the results from the present study, the important role of coupling chaos theory and complex network theory in advancing future research on groundwater systems has also been emphasized.

Keywords: *Groundwater systems, nonlinearity, chaos theory, complex networks*

PRIORITIZATION OF SUB-WATERSHED OF PHALGU RIVER BASIN (PRB), MIDDLE GANGA PLAIN, INDIA USING INTEGRATED PCA AND MACHINE LEARNING APPROACH

Bibhukalyan Mohapatra* and Prafull Singh

Department of Geology, Central University of South Bihar, Gaya (Bihar), India
**Corresponding author e-mail: geobibhu2015@gmail.com*

Watershed is defined as an area where the incoming rainwater is drained as surface runoff into a water body. Morphometric analysis of watershed aids in managing water crisis, soil erosion and prioritizing watershed with respect to its vulnerability. The traditional method of watershed prioritization includes activities like digitizing stream networks, contours in the toposheet, manually measuring stream length or extensive field survey. The manual survey, in particular, is labour intensive, time-consuming and expensive. Modern day tools like Remote sensing (RS), GIS and Machine learning (ML) have reduced the expenses and enhanced the accuracy in morphometric analysis for prioritizing the sub-watersheds for sustainable water and land resources management. This study highlights the effectiveness of the satellite data and machine learning approach to prioritize the sub-watersheds (SW) of the Phalgu river basin using morphometric parameters, PCA-WSA (Principal component analysis – weighted sum approach) and linear discriminant analysis (LDA).

The Phalgu River Basin which covers the major districts of Bihar and Jharkhand states in India. The Phalgu river basin is located between the geographic coordinates 24°00’00"N to 25°30'00"N latitude and 84°30’00" E and 85°30’00"E longitude and covers total area of around 6286 km^2. The river Falgu is a tributary of river Ganga and is a major river system of the south Bihar. The river originates in the Chotanagpur plateau by the confluence of the two rivers namely Lilajan (also known as Niranjan or Nilanjan) and Mohana at Bodhgaya. In this study, most of the morphometric parameters have been extracted and analyzed with the help of Shuttle Radar Topography Mission (SRTM-DEM) which have 30 m spatial resolution. The river basin elevation ranges from 46 to 762 m above mean sea level (amsl). The watershed has been delineated using the SRTM-DEM obtained from USGS Earth Explorer website along with utilization of Python toolkit in the Jupyter integrated development environment (IDE) for statistical and machine learning approaches (PCA and LDA). Sub-watersheds (SW1 to SW9) have been classified using the spatial analyst tool of ArcGIS 10.8.2. The morphometric parameters studied and classified to understand the river morphology are linear, relief and areal or shape parameters. A total 22 morphometric parameters have been selected and analysed. The rank of each parameter in each sub-watershed is determined after all morphometric values for individual sub-watersheds have been calculated. The linear and relief parameters are directly related to soil erosion. The sub-watershed with the highest value in these parameters is ranked 1st, and so on. On the other hand, soil erosion is indirectly related to shape parameters. The sub-watershed with the lowest value in relief parameter is ranked 1st, and so on. Following the determination of all ranks for each parameter in an individual sub-watershed, the compound parameter (C_p) value for the individual sub-watershed has been determined. When all of the ranks of SW1 are combined and divided by 22, C_p comes out to be 6. The same process has been performed for all the other sub-watersheds. Based on the C_p values, the sub-watersheds were classified as high, medium, and low.

The 22 morphometric parameters have been further reduced to 7 important components for analysis of rotated component matrix. The rotated component matrix shows that each component considers some most highly correlated parameters. The scree plot also includes the component loading matrix, which offers a visual representation of the loadings for each component. The top five components account for 95.1% of the overall variance in the actual information and are clearly considerable. After rotated component matrix, the primary morphometric parameters derived from PCA are lemniscate ratio, hypsometric integral, drainage texture, stream frequency, form factor, basin relief and circulatory ratio. Subsequently, these factors have been used to prioritize the 9 sub-watersheds of the Phalgu river watershed. Further, the WSA was used along with preliminary priority ranking obtained from PCA derived parameters, to obtain the compound factors to categorize them into three groups namely high, medium, and low.

LDA is a dimensionality reduction approach that keeps as much information as possible while reducing the number of dimensions or parameters in a dataset. LDA determines the eigen value/vector of the product of scatter matrices and it has been found that the first two eigen values corresponding to bifurcation ratio, and drainage texture have significant. The linear discriminants are then obtained by multiplying the standardized data with the eigen vector matrix. The two linear discriminants represent the 22 parameters data to prioritize the sub-watersheds. The important morphometric parameters derived from LDA are the bifurcation ratio and drainage texture. The LDA method produces two highly correlated parameters which have been ranked and prioritized as per the standard method.

The study has used three methods: morphometric analysis, PCA and LDA to prioritize Phalgu sub-watersheds. Using the morphometric analysis, higher priorities have been designated for SW5, SW8 and SW7 sub-watersheds. Using PCA, higher priorities have been assigned to SW8, SW7 and SW6 sub-watersheds. Utilizing LDA, higher priorities have been designated for SW7, SW8 and SW9 sub-watersheds. Thus, it has been observed that different methods may yield varying priority levels. However, when two out of the three methods consistently indicate a similar priority for a particular sub-watershed, it is reasonable to consider that priority. By identifying the overlapping results and considering the consistency among multiple methods, a reliable assessment of sub-watershed priority can be achieved. Accordingly, based on the results obtained in the study, sub-watersheds SW7 and SW8 have been determined to have a high priority as both SW7 and SW8 sub-watershed consistently exhibit similar high priorities across all the three distinct methods. SW3, SW5, SW6 and SW9 sub-watersheds have been assigned a moderate priority, while SW1, SW2 and SW4 sub-watersheds have been categorized as having low priorities. Thus, based on the prioritization analysis carried out under the study, sub-watersheds SW7 and SW8 are recommended for focused conservation efforts and effective land management practices. These high priority sub-watersheds are of utmost importance for decision makers and management authorities because it will aid in best conservation and land management practices against soil erosion and flooding. It has also been observed that because of the limited data availability, a deep learning model couldn't be implemented in the present study. As a way forward, there is a plan to expand the data collection efforts to include more micro-level watersheds, which will help in continuation of the classification study using deep learning models, which would further improve the results.

Keywords: *Watershed prioritization, morphometric parameters, PCA-WSA, LDA*

ADVANCING GROUNDWATER LEVEL PREDICTION USING PHYSICS-INFORMED NEURAL NETWORKS AND MACHINE LEARNING FOR TRANSIENT HEAD VARIATION

Mayank Bajpai[1]*, Narendra[1], Shreyansh Mishra[1], Shishir Gaur[1] and Kamal Singh[2]

[1]Indian Institute of Technology, Banaras Hindu University, Varanasi, India
[2]Telecom Saint Etienne / Université Jean Monnet, Saint-Étienne, France
[*]Corresponding author e-mail: mayankbajpai.rs.civ22@iitbhu.ac.in

Groundwater (GW) management is a critical component of sustainable water resources planning, particularly in regions facing water scarcity. Accurate prediction of GW head is essential for effective monitoring and management. Traditional methods often struggle with poor resolutions, necessitating advanced approaches. This study explores the potential of ML models, including XGBoost, Bayesian regression, and ANN, alongside a Physics-Informed Neural Network (PINN) to predict transient GW head variations. A hypothetical MODFLOW model was developed to simulate the dynamic behavior of GW heads under varying boundary conditions and pumping scenarios, providing a dataset for model training and evaluation. The study employed a conceptual MODFLOW model, designed to simulate transient GW head variations with 8 pumping wells. The boundary conditions were specified as fluxes, with an inflow of 20 m^3/d at the left boundary and an outflow of -30 m^3/d at the right. The generated test dataset covered a range of time steps to capture the dynamic nature of the system. The data was used to train three ML models: XGBoost, Bayesian regression, and ANN. Additionally, a PINN was trained, integrating the governing physical equations of groundwater flow into the neural network architecture to improve prediction accuracy. For the ML models, the objective was to minimize the mean squared error (MSE) during training, using a train-test split for model validation. The PINN was designed to incorporate the physical equation of transient groundwater flow. This equation was embedded into the PINN's loss function, combining data-based and physics-based components. The optimization process aimed to balance the physics-based loss with the data-fitting loss, making PINN particularly suitable for data-scarce conditions.

The performance of each model was evaluated using error metrics including Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The results indicated that the PINN outperformed the traditional ML models in capturing the spatial and temporal variations of GW head. The performance evaluation of the different models based on RMSE, R^2, AIC, and BIC reveals significant variations in their predictive capabilities. The PINN achieved the lowest RMSE value and an R^2 indicating its superior accuracy in predicting groundwater head. Its ability to integrate physical laws into the learning process is further reflected in the lowest AIC and BIC values, which suggest a better balance between model complexity and fit. Among the ML models, XGBoost also demonstrated strong performance with an RMSE and an R^2, albeit slightly less effective than the PINN in terms of AIC and BIC. The ANN provided relatively good results with an RMSE and an R^2, though its AIC and BIC values were higher than those of the PINN and XGBoost. In contrast, the Bayesian Regressor struggled to achieve comparable accuracy, with a notably higher RMSE and a lower R^2, alongside higher AIC and BIC values. These results underscore the advantages of using PINN for capturing complex spatial and temporal variations in groundwater head, while XGBoost and ANN serve as competitive alternatives in less complex scenarios.

The study demonstrates that PINNs offer a highly effective and innovative approach for predicting transient groundwater head variations, significantly outperforming traditional ML models. PINNs leverage the integration of physical governing equations with data-driven learning, enabling them to achieve exceptional predictive accuracy while maintaining consistency with hydrogeological principles. This dual advantage addresses the limitations of purely data-driven models, particularly when dealing with sparse or unevenly distributed datasets, which are common in groundwater studies. The PINN model achieved the lowest RMSE and the highest R^2 indicating superior accuracy in capturing both spatial heterogeneities and temporal dynamics of groundwater heads. Its superior performance is further validated through model selection criteria such as the AIC and the BIC. These results reflect an optimal balance between model complexity and goodness-of-fit, confirming the robustness of the PINN approach in complex groundwater systems. Comparatively, XGBoost also performed well, achieving a competitive R^2 of 0.9999. However, its slightly higher RMSE and AIC suggest that while XGBoost excels in predictive accuracy, it may not fully capture the non-linear and transient variations inherent in groundwater flow dynamics as effectively as PINNs. This limitation arises from XGBoost's reliance on statistical learning alone, which lacks the physical constraints embedded in the PINN framework. The ANN provided reasonable predictive accuracy but lagged in model efficiency and stability, likely due to its dependency on large training datasets and the absence of physical constraints. Similarly, the Bayesian Regressor struggled to capture the complex variations of groundwater heads, reflected in its higher error metrics and suboptimal AIC and BIC values, indicating poorer fit and higher model complexity. These findings underscore the transformative potential of PINNs for groundwater modeling, particularly in regions with limited or unreliable observational data. By embedding physical laws within the learning process, PINNs not only reduce data dependency but also ensure physically realistic outputs, making them highly applicable for hydrogeological systems with varying spatial and temporal complexities. Moreover, the ability of PINNs to integrate observed and simulated data makes them a powerful tool for reducing uncertainties in groundwater head predictions under changing hydrological or anthropogenic stress conditions. From a water resource management perspective, the accuracy and reliability of PINN-based predictions can significantly enhance decision-making processes. For instance, more precise estimations of transient groundwater head variations allow water managers to design optimal pumping strategies, identify critical recharge zones, and develop sustainable groundwater extraction policies. In data-scarce regions, where traditional models often fail due to insufficient data, PINNs can provide reliable insights, enabling proactive management to mitigate groundwater depletion and ensure long-term resource sustainability. The future research directions should focus on scalability and generalizability of PINNs across diverse hydrogeological settings, including heterogeneous aquifers, karst systems & regions with strong river-aquifer interactions. Additionally, coupling PINNs with real-time data acquisition systems, such as IoT-based sensors, can further enhance predictive capabilities by integrating live observational data into the learning process. This would facilitate the development of dynamic groundwater management frameworks, adaptable to changing climatic & anthropogenic conditions, thereby ensuring the sustainable use and resilience of groundwater resources.

Keywords: *Groundwater modeling, physics-informed neural networks, machine learning*

GROUNDWATER POTENTIAL ZONES MAPPING USING ARTIFICIAL INTELLIGENCE TECHNIQUES FOR SURAT DISTRICT, INDIA

Prajakta Jadhav[1*], Vivek Manekar[2] and Jayantilal Patel[2]

[1]Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra, India
[2]Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India
[*]Corresponding author e-mail: ppjadhav@sggs.ac.in

Groundwater is a vital component of the hydrologic cycle and its depletion rate is the serious concern. Monitoring and controlling groundwater are crucial in many regions, requires advanced technology and methods. Global population expansion, urbanization, industrialization, and agricultural operations have resulted in an increase in groundwater consumption. Approximately half of India's total water requirements are met by the groundwater used in agriculture. About 60% of the water utilized for irrigation is extracted from groundwater to meet the requirements. The groundwater level has been declining in most parts of India in recent decades due to population growth at an alarming rate and subsequent significantly higher extraction rate from the groundwater than recharging rate. For the sustainable use of groundwater water resources, groundwater management is essential. The Surat district relies heavily on agriculture, which has led to the extraction of massive amounts of groundwater for drinking and irrigation. There are 35 monitoring stations in the district. The depth to groundwater level in the district is between 0 and 31.5 m. A water level of 5 m to 10 m below ground level covers the central area of the district. Factors related to topography, hydrology, geology, and atmosphere affect the availability and movement of groundwater. Geographic information systems (GIS) and Remote sensing (RS) applications including frequency ratio, multi-criteria decision analysis, weight of evidence, and analytical hierarchy processing, have been used in numerous studies for groundwater potential mapping. However, the efficiency of the groundwater potential evaluation under these studies was insufficiently precise and was subjective because the models used in these researches are based on conventional weighted techniques or expert opinion. Using the Analytical Hierarchy Process, weights are often assigned manually. The results of past studies show that artificial intelligence (AI) techniques integrating with RS and GIS for groundwater potential zone mapping are faster and more accurate than traditional multi-criteria decision analysis and analytical hierarchy processing techniques. The primary goal of this study is to create groundwater potential maps for the Surat area of India applying AI techniques such as support vector machine (SVM) and boosted regression tree (BRT). Surat district situating between 20°30’: 21°35’N and 72°35’:74°20’ E has a total geographical area of 4,414 km2. The district is located in the south of the Gujarat state. It is bounded by Bharuch district in the north, the Arabian sea in the west, the Tapi district in the east and the Valsad district in the south. Groundwater level data were obtained from Central Groundwater Board (CGWB) and Rainfall data were procured from the India Meteorological Department (IMD) for May, August and November, (pre-monsoon, peak monsoon, and post-monsoon respectively) for the year 2023. Thematic layers of the groundwater potential determining factors were created using Arc GIS software after gathering information regarding the groundwater potential determining factors, including rainfall, slope, elevation, aspect, topographic wetness index, and distance from streams. BRT is one of the most popular methods for improving a single model's performance and prediction accuracy. It can fit numerous models and combine them for prediction. The SVM model employs kernel functions to identify a hyperplane in an infinite-dimensional space that optimizes the distance between datasets using data pattern and regression recognition. In this study, predication of groundwater potential maps has been done by the cell analysis of thematic maps. The observed groundwater potential maps, which were produced using a spatial interpolation approach such as the inverse distance weighted technique, were used to validate the predicted groundwater potential maps. The resilience and accuracy of the groundwater potential maps were assessed using performance assessment measures such as root mean square error (RMSE), coefficient of determination (R2) and mean square error (MSE). The equal break categorization system was used to divide groundwater potential maps into five categories. Predicted groundwater potential zone mappings showed that the SVM model was superior for pre-monsoon with performance evaluation measures of 0.57, 0.32 and 0.97 for RMSE, MSE and R2 respectively and for peak-monsoon with the performance evaluation measures of 0.42, 0.17 and 0.93 for RMSE, MSE and R2 and respectively. The BRT model was superior for post-monsoon with performance evaluation measures of 0.57, 0.32 and 0.94 for RMSE, MSE and R2 respectively. The study concludes that the developed SVM and BRT models produce good predictions of groundwater potential zone mappings. The study findings indicate that, for the prediction of groundwater potential zone mappings, the SVM model performs better for pre-monsoon and peak-monsoon months while the BRT model performs better for post-monsoon months.

Keywords: *Artificial intelligence techniques, groundwater potential maps, SVM, BRT, remote sensing, GIS*

GROUNDWATER QUALITY PREDICTION USING MACHINE LEARNING

Devanand Guleri*, D.C. Jhariyaand and Mridu Sahu
National Institute of Technology, Raipur (Chhattisgarh), India
**Corresponding author e-mail: devanandguleri@gmail.com*

Groundwater, an essential natural resource is highly susceptible to contamination from agricultural runoff, industrial waste, and natural geochemical processes. Monitoring and predicting groundwater quality is essential to avoid public health issues arising from waterborne contaminants such as heavy metals, harmful chemicals, and microbial agents. Recent advancements in machine learning have demonstrated the potential to leverage data-driven models for environmental monitoring, which could revolutionize traditional approaches by reducing reliance on costly laboratory-based testing. Our study aims to develop a machine learning model capable of predicting groundwater potability with high accuracy using readily available water quality indicators, thus, offering a valuable tool for resource management and decision-making.

The dataset for this study was obtained from an open-source platform, Kaggle, and comprises a total of 3,276 samples of groundwater quality measurements. Each sample includes nine physico-chemical parameters: pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, and Turbidity. These parameters provide an extensive profile of water chemistry and are commonly used indicators of water quality. The dataset also includes a target variable, *Potability*, which labels each sample as potable (drinkable) or non-potable based on these chemical attributes. Our approach began with data pre-processing to prepare the dataset for analysis and modelling. One key challenge was handling missing values, particularly in parameters such as pH and Sulfate. We addressed this issue by using statistical imputation techniques to estimate missing values based on the distribution of existing data, thereby preserving the dataset's integrity and improving model accuracy. Furthermore, to account for varying measurement scales across different features, we applied feature scaling, normalizing each parameter to ensure that no single attribute disproportionately influenced the model outcomes.

Exploratory data analysis (EDA) was conducted to gain insights into the data distribution and identify any patterns or correlations among features. Visualizations, including histograms and scatter plots, helped highlight the variations in parameters like pH and Turbidity, which play significant roles in determining water quality. Potability distribution was also assessed to ensure the data was not imbalanced, as this could affect model training. Statistical tests were applied to verify the significance of correlations between potability and the nine features, assisting in feature selection by narrowing down the attributes most influential in predicting groundwater quality.

After pre-processing and feature selection, we implemented a range of machine learning classifiers to evaluate their suitability for groundwater quality prediction. These included: (i) Random Forest Classifier: An ensemble learning method that operates by constructing multiple decision trees and combining their outputs for improved accuracy. Random Forest was selected for its robustness in handling noisy datasets and complex data relationships, (ii) Gradient Boosting Classifier: This boosting method sequentially builds models to correct errors from previous iterations, often resulting in high predictive performance but at the cost of longer training times, (iii) Decision Tree Classifier: Known for its interpretability, the

Decision Tree algorithm builds a model based on feature splits, which is particularly useful for identifying key decision points in potability determination, (iv) K-Nearest Neighbors (KNN): A simple, instance-based method that classifies samples based on the majority label among the nearest neighbors in feature space, and (v) Logistic Regression: Although not as powerful with nonlinear data, Logistic Regression was included as a baseline to evaluate performance against more complex algorithms.

Model training and evaluation were performed using a split of 80% for training and 20% for testing. Additionally, we applied grid search and cross-validation to optimize hyperparameters for each classifier, ensuring maximum model accuracy and generalizability. For model assessment, accuracy, precision, recall, and F1-score metrics were used, providing a comprehensive evaluation of each algorithm's performance. Among the classifiers, the Random Forest model achieved the highest predictive accuracy, with an accuracy score of approximately 65.10%, outperforming other models in both sensitivity and specificity for potability prediction. The ensemble nature of Random Forest allowed it to manage the high-dimensionality of the dataset effectively, particularly in distinguishing potable from non-potable samples based on nuanced feature interactions. The Gradient Boosting model also demonstrated competitive performance, albeit with a slightly lower accuracy than Random Forest. However, it exhibited excellent precision and recall scores, indicating a strong ability to generalize across diverse groundwater samples. On the other hand, the K-Nearest Neighbors model displayed lower accuracy, likely due to sensitivity to feature scaling and noise in the dataset. Logistic Regression and Decision Tree classifiers provided reasonable baseline results, yet they were less robust compared to the ensemble methods.

This study demonstrates that machine learning models, particularly ensemble methods such as Random Forest, can be effective tools for predicting groundwater quality. By leveraging key water quality indicators, these models can classify water samples as potable or non-potable with high accuracy, providing a valuable resource for environmental monitoring agencies. Our findings underscore the potential of data-driven approaches in groundwater quality assessment, enabling scalable and rapid evaluations that can complement traditional water quality testing. The implementation of machine learning for groundwater quality prediction has broader implications for water resource management, particularly in areas with limited access to laboratory testing facilities. Future work will involve expanding this model to include additional chemical and biological parameters, enhancing its predictive capabilities. Additionally, the model's application could be extended to real-time water quality monitoring systems, facilitating proactive responses to water contamination events.

Keywords: *Groundwater quality, machine learning, Random Forest, water potability, environmental monitoring*

INTEGRATING REMOTE SENSING, GIS, AND MACHINE LEARNING FOR GROUNDWATER LOGGING DETECTION AND FORECASTING IN AGRICULTURE-DOMINATED REGIONS

Dolon Banerjee* **and Sayantan Ganguly**

Indian Institute of Technology, Ropar, India
**Corresponding author e-mail: dolon.20cez0004@iitrpr.ac.in*

Waterlogging is a persistent and critical environmental issue, especially in agrarian regions like Punjab, India. Excessive water accumulation in the soil leads to saturation and reduces oxygen availability for plant roots, which can result in crop damage and long-term soil degradation. In the Punjab state of India, the irrigation intensity and the region's seasonal rainfall make it particularly prone to waterlogging. Addressing this issue requires a robust monitoring and forecasting system supporting proactive water management strategies. Traditional ground-based methods, though valuable, have limitations in scalability and efficiency. Therefore, this study aims to evaluate the effectiveness of integrating Geographic Information Systems (GIS) and remote sensing techniques with advanced machine learning (ML) methods to monitor and forecast waterlogging in Punjab.

Combining GIS-based analysis and ML modelling, this study seeks to identify waterlogged regions and forecast future occurrences based on environmental and agricultural factors. Historical data from the Central Ground Water Board (CGWB) is used here, alongside satellite-derived data, to establish a reliable methodology. The study's main objectives are two-fold: first, to establish the accuracy of GIS-based techniques for waterlogging detection, and second, to develop and validate an ML-based predictive model that forecasts waterlogging under various environmental conditions. This framework is intended to serve as a tool for managing waterlogging risks, ensuring agricultural productivity, and guiding policy decisions in Punjab and similar regions facing waterlogging challenges. The primary dataset for observed waterlogged areas (from 2001 to 2015) for pre- and post-monsoon was obtained from CGWB records, which serve as a reliable ground truth reference. In addition to CGWB data, LANDSAT satellite imagery is utilized to analyze and quantify the waterlogging across Punjab. This dataset is complemented by environmental and agricultural factors, including precipitation records, groundwater abstraction, irrigation demands, soil types, and slope data for a more detailed understanding. Historical precipitation and temperature data were acquired from the Indian Meteorological Department (IMD; https://mausam.imd.gov.in/responsive/rainfallinformation_swd.php), while soil characteristics and slope data were obtained from the National Aeronautics and Space Administration Earthdata (https://www.earthdata.nasa.gov/). This analysis includes precipitation to understand its role in waterlogging, especially during monsoon heavy rainfall periods, when soil properties such as texture and infiltration rate can affect water retention. Irrigation demand data reflect agricultural practices that may enhance waterlogging. Integrating these data sources provides a comprehensive view, enabling the ML model to capture complex relationships between contributing factors and waterlogging.

The GIS analysis begins with standardizing LANDSAT images for consistent temporal and spatial analysis. Using spectral indices like the Normalized Difference Water Index (NDWI) and Modified NDWI (MNDWI), waterlogged regions are identified and delineated based on significant changes in water content. These indices clearly distinguish water bodies and saturated soil areas from surrounding dry land clearly, making them suitable for waterlogging detection. To ensure the robustness of this detection method, GIS-detected waterlogged areas were cross-validated with CGWB's data. Microwave satellite data is also analysed for temporal validation, especially from 2016 to 2024. Microwave data, known for penetrating vegetation and cloud cover, is beneficial for detecting waterlogged areas during the monsoon season or in densely vegetated regions. The resulting GIS-based waterlogging maps are compared with CGWB data to establish the method's accuracy, laying the foundation for further analyses.

To forecast waterlogging, we develop an ML model here using data from 2001 to 2014 sourced from CGWB. The model considered factors like precipitation, irrigation demand, soil type, and slope influencing water retention and drainage. Precipitation, especially during monsoon seasons, directly affects soil saturation. High irrigation demand in rice-growing areas can exacerbate waterlogging. Similarly, soil type for low-permeability soils and slope, which influences drainage, also play significant roles. By incorporating these factors, the model can effectively predict waterlogging risk. After preprocessing, including normalization and handling of missing values, we evaluated several ML algorithms for predictive accuracy, including Random Forests, Decision Trees, Gradient Boosting Machines, Extreme Gradient Boosting Machines and Long-Sort Term Memory. Model performances were assessed using metrics such as Root Mean Square Error (RMSE) and R^2 values. Feature importance analysis was also conducted to identify the most influential variables, providing insights into key drivers of waterlogging in the region.

After model validation, the best ML model was employed to forecast waterlogging patterns from 2015 to 2024, relying on projected values for environmental and agricultural factors. This prediction simulated various future scenarios and assessed the model's generalizability. Simultaneously, a GIS analysis using microwave satellite data was used to monitor waterlogging across Punjab independently for the same period. Comparing the GIS results with the ML predictions validates the robustness of the ML model in real-world applications, offering a dual-layered validation framework for future forecasting accuracy. The study yielded two key outputs: (1) a set of GIS-derived waterlogging maps from 2001 to 2024 and (2) an ML-based model capable of forecasting waterlogged regions. The remote sensing and GIS-based analysis revealed waterlog-prone areas, particularly in regions with high irrigation demands and certain soil types, confirming the feasibility of such waterlogging detection techniques. The cross-validation with CGWB data quantified the GIS method's accuracy, providing a basis for reliable spatial detection. The model performance metrics depicted the best model with corresponding accuracy in identifying waterlogged areas based on environmental predictors, particularly during monsoon seasons. The variables such as precipitation intensity and irrigation levels emerged as primary factors driving waterlogging. The forecasting capability of the model, validated by microwave data comparisons, provided an insightful prediction framework, enabling proactive water management strategies. By evaluating model performance against real-world data from 2016 to 2024, we established the model's reliability for practical forecasting applications.

This research established a comprehensive methodology combining remote sensing, GIS, and ML to detect and forecast waterlogging in Punjab, India. The GIS-based technique, using LANDSAT and microwave data, demonstrated a high spatial accuracy, validating its usage for large-scale waterlogging detection. The cross-validation with CGWB data underscored its reliability, and temporal validation using microwave data supported its suitability over the years. The ML model was a robust predictive tool for waterlogging forecasting, with potential applications in various environmental and agricultural scenarios. The model provided a detailed understanding of the conditions under which waterlogging occurs by quantifying the influence of factors such as precipitation, irrigation, soil type, and slope. Its forecasting capability will allow policymakers and water resource managers to take preemptive actions, mitigating the impact of waterlogging on agriculture and the environment. In conclusion, combining GIS-based mapping with ML predictions offered a powerful and scalable solution for waterlogging management. The results from this study supported the utility of GIS in detecting waterlogged areas and affirmed the predictive power of ML models when integrated with comprehensive environmental data. While applied in Punjab, this framework is adaptable and could be implemented in other waterlogged regions facing similar challenges, offering a versatile and data-driven tool for sustainable water management and planning.

Keywords: *Groundwater, water logging, remote sensing, GIS, machine learning*

DATA-DRIVEN GROUNDWATER MANAGEMENT IN WATER STRESS AREA OF CENTRAL INDIA: A MACHINE LEARNING ALGORITHM BASED ANALYSIS OF AQUIFER DYNAMICS AND OVER-EXTRACTION RISKS USING HIGH FREQUENCY DATA

Vaishnavi Parihar*, Lata Udsaiya, Pradip Roy and Ashok Kumar Biswal

Central Ground Water Board, North Central Region, Bhopal, Madhya Pradesh, India
**Corresponding author e-mail: vaishnavi.p-cgwb@gov.in*

This study investigates groundwater (GW) dynamics in the over-exploited Ratlam district of Madhya Pradesh, where GW extraction has reached a critical 130% as of the GWRE 2024 report. The region's reliance on groundwater for agriculture and limited surface water resources make it highly vulnerable to water scarcity, especially during the dry season when wells and hand pumps often dry up. Residents frequently resort to tanked water from distant areas and drill deep bore wells, sometimes reaching 363 m. The area's black cotton soil, highly suitable for agriculture, further drives groundwater demand. In addition, surface water from rivers is occasionally used for irrigation, but during the non-monsoon months, bore wells become essential. This study employs high-frequency data collected via Digital Water Level Recorders (DWLRs) installed in 14 strategically chosen wells. These DWLRs capturing water levels four times daily from 2023 to 2024, provided a comprehensive dataset for examining temporal and spatial variations in groundwater levels. Monthly rainfall data from the IMD was also paired with groundwater data to assess correlations between rainfall and GW levels. This is one of the first studies in central India to use such high-resolution, machine learning (ML)-enhanced methods, marking a significant contribution to sustainable GW management in a semi-arid region facing extreme extraction pressures.

The DWLR data was aggregated daily and averaged monthly to create structured time series data for each well, allowing for the assessment of both seasonal and short-term fluctuations in GW levels. Applying a rolling mean to the GW level data provided a clearer view of long-term trends by smoothing out short-term variability. This comparison between the rolling mean and the raw data helped us identify deviations, potentially indicating over-extraction or climate impacts, while enhancing trend clarity. The Augmented Dickey-Fuller (ADF) test was used to check stationarity in the data, distinguishing between seasonal fluctuations and persistent trends. Autocorrelation analysis was also conducted, revealing the time-dependent relationships within the GW levels, such as patterns driven by monsoon cycles and lagged recharge responses. Several analyses were performed to study the relationship between groundwater levels and rainfall. Scatter plots allowed for a visual inspection of linear and non-linear relationships, while heat maps provided an overview of rainfall-GW correlations across various wells and time periods. Cross-correlation analysis helped determine the lagged response of groundwater levels to rainfall, showing the time, it takes for rainfall events to impact the aquifer system in specific areas. Granger causality tests were applied to assess whether changes in rainfall could predict GW level fluctuations, helping to distinguish causative relationships from mere correlations. This method allowed us to identify wells where GW levels are highly sensitive to rainfall, providing insight into which areas are most vulnerable to GW depletion under changing rainfall patterns. Python programme was used to analyze the data, employing libraries such as Pandas for data processing, NumPy for computations, Matplotlib and Seaborn for visualizations, and Statsmodels for advanced statistical testing. SciPy supported autocorrelation and other statistical analyses.

In Ratlam, the pattern of GW decline is notable, particularly in deeper wells used for irrigation. This pattern involves a sharp decrease in water levels after October, peaking in January, and then stabilizing at a deeper level. The average GW decline in January is on an average 0.039 m mbgl per day assessed from all wells. The maximum extraction is as high as 0.13 mbgl/day in some areas. Prolonged, intense pumping leads to depths of up to 38 mbgl. High GW fluctuations, ranging from less than 1 m to over 30 m annually, suggest a complex aquifer response influenced by seasonal rainfall, extraction rates, and soil type. The study found that deeper aquifers exhibit more pronounced GW level fluctuations, especially during winter cropping than the shallower piezometers. During January, for e.g., deeper wells experienced declines of up to 0.1 mbgl/day, compared to 0.0009 mbgl/day in shallower wells. This stark difference implies that agricultural extraction targets deeper aquifers, which face prolonged recovery times, thereby creating GW sustainability issues.

Spatial analysis revealed that recharge mainly occurs between August and September, few months after the monsoon onset, across all wells. Recharge rates were faster in certain wells but almost all wells showed prolonged periods of drawdown after recharge in monsoon due to agricultural use, especially in rural, agriculture-dominated areas. These areas displayed more pronounced declines relative to semi-urban areas where surface water supply decreases the groundwater demand. Our correlation analysis of GW levels and rainfall demonstrated that over 50% of wells had a strong positive correlation with rainfall, some wells show perfect correlation up to 0.9, indicating the region's reliance on rainfall for aquifer recharge. Cross-correlation analysis provided insight into delayed recharge response times, varying across wells and revealing that certain areas are more dependent on rainfall and hence, more susceptible to seasonal GW depletion. Spatial correlation analysis indicated that wells in certain zones showed synchronous GW level fluctuations, suggesting interconnected aquifer sections, while others exhibited isolated behaviour indicating localized depletion. These findings help in determining critical regions in Ratlam region where the aquifer system is either highly responsive to natural recharge or vulnerable to prolonged drawdowns.

The integration of ML techniques has enabled a comprehensive understanding of district Ratlam's aquifer dynamics, revealing intricate relationships between rainfall, GW extraction, and aquifer recharge. The study's high-frequency monitoring and data-centric analysis highlight critical trends and zones of fluctuation, essential for addressing GW management in semi-arid, water-stressed areas. Our analysis suggests that the aquifer's response to recharge is highly variable, with deeper confined aquifers showing slower recuperation rates, especially under heavy agricultural extraction. Given the district's reliance on rainfall, our findings stress the need for proactive water management strategies to mitigate extraction pressures and develop resilient GW systems. We have tried to identify high-risk zones based on extraction rates, recharge patterns, and also periods of high GW depletion. The high-resolution data coupled with ML-based analysis provides valuable insights into central India's groundwater dynamics, advancing the scientific understanding of aquifer behaviour in over-exploited regions. The approach is scalable, offering a framework for similar studies in regions facing GW depletion challenges, essential for sustainable resource management due to rising demands and climate uncertainties.

Keywords: *Over-extraction, groundwater, data-driven management, machine learning*

DEVELOPMENT OF USER INTERFACE FOR SUSTAINABLE GROUNDWATER MANAGEMENT BY INTEGRATING THE AI, ML, IOT, CLOUD COMPUTING AND OTHER ADVANCED TECHNIQUES

Suresh Thonte[1*], Prasad I.N.[1], Thiyageshwaran N.[1] and Sunitha Veerla[2]

[1]Department of Civil Engineering, Indian Institute of Science, Bangalore
[4]International Centre of Excellence for Dams, Indian Institute of Technology, Roorkee, India
[*]Corresponding author e-mail: tsureshiisc@gmail.com

Groundwater is a vital resource for billions of people worldwide, providing a significant portion of the freshwater used for drinking, agriculture, and industry. However, over-extraction, contamination, and the impacts of climate change have created an urgent need for more effective groundwater management. Traditional methods of managing groundwater are increasingly inadequate in addressing these challenges, necessitating the adoption of advanced technological solutions. Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), the Internet of Things (IoT), Deep Learning (DL), Artificial Neural Networks (ANN), Fuzzy Logic Models, and Cloud Computing have opened new avenues for improving groundwater management. These technologies enable real-time monitoring, predictive analysis, and data-driven decision-making, offering significant potential to optimize groundwater use, detect contamination, and enhance sustainability. Despite the promise of these technologies, their full potential remains untapped due to fragmented systems and a lack of integration across platforms. The objective of this paper is to present and explore the significance of the Development of a User Interface for sustainable groundwater management by integrating AI, ML, IoT, Cloud Computing and other advanced techniques. By creating an intuitive platform, the research aims to enable stakeholders from farmers to policymakers to make data-driven decisions that enhance groundwater sustainability. The user interface will combine real-time data collection from IoT sensors, AI-driven predictive models, and Cloud Computing to process and analyze large datasets, allowing for better decision-making and resource management. By articulating this ideology, the paper seeks to provoke critical reflection and discussion, ultimately contributing to Sustainable Development Goals (SDGs).

The development of the user interface for sustainable groundwater management involved several key stages: system design, data integration, algorithm development, and user interface (UI) design. These stages were informed by existing literature, case studies, and expert input. The system should be designed to integrate multiple technologies, including IoT sensors for real-time groundwater monitoring, Cloud Computing for scalable data storage and processing, and AI/ML algorithms for predictive modeling and decision support. The architecture will include IoT-based groundwater sensors that measure parameters such as water levels, temperature, and water quality indicators (e.g., pH, turbidity). These sensors transmit data to a cloud platform, where the data is processed and analyzed using AI and ML models to forecast water availability and quality trends. In data integration and analysis, the system should be designed to integrate historical groundwater data, weather patterns, land usage data, and sensor data into a centralized Cloud database. The AI and ML algorithms, including regression models and neural networks, should be employed to analyze the data and predict future groundwater trends, identify potential contamination risks, and optimize resource management. In the User Interface (UI) design, the user interface should be developed to be accessible to a wide range of stakeholders, including water resource

managers/Engineers, farmers, and local governments. The UI should have features of interactive dashboards, real-time monitoring tools, trend analysis visualizations, and predictive analytics. Users can view current groundwater conditions, forecast trends, and receive alerts for potential issues (e.g., over-extraction or contamination).

In the testing and validation, the system should be tested using data from several regions with varying groundwater conditions. User feedback was collected to refine the UI and improve system functionality. Performance metrics such as system accuracy, response time, and user satisfaction should be evaluated. The data requirement is essentially consisting of groundwater sensor data mainly Real-time data from IoT sensors installed in boreholes and wells in different regions. Parameters include groundwater levels, temperature, electrical conductivity (EC), pH, and turbidity and historical groundwater data on past groundwater trends, including historical water levels and usage patterns, sourced from regional water management authorities and weather and climate data consisting of precipitation, temperature, and evaporation rates from meteorological sources, used to inform the system's predictive models. Finally, the geospatial data that includes geographic data on land use, population density, and water demand, is integrated into the Cloud system to provide context for groundwater management decisions. The development of the user interface for sustainable groundwater management should result in a functional prototype integrating IoT sensors, AI/ML-based predictive modeling, and Cloud data storage. Key results should include a) Real-time Monitoring: The system should capture accurate groundwater data, displayed via interactive maps and dashboards. Alerts were triggered when water levels or quality parameters (e.g., pH, turbidity) exceeded predefined limits. b) Predictive Modeling: AI and ML algorithms should accurately forecast groundwater trends (weekly, monthly, seasonal) based on historical and real-time data, identifying drivers of depletion and contamination, such as over-extraction and agricultural runoff. c) User Experience: Groundwater professionals should find the UI intuitive and actionable, enabling informed decision-making for water allocation and resource conservation. The integration of multiple technologies should improve data accessibility and interpretation. d) System Performance: Cloud-based architecture should provide scalable data storage and quick processing of large datasets, with fast response times for real-time monitoring and predictive analysis.

The development of a user interface integrating AI, ML, IoT, and Cloud Computing for sustainable groundwater management will be a successful approach to addressing the challenges of groundwater depletion and contamination. The system should demonstrate the ability to monitor groundwater resources in real time, predict future trends, and provide stakeholders with actionable insights to guide resource management decisions. This paper presents ideology by highlighting the potential of advanced technologies to transform groundwater management practices by developing a user interface. However, further studies are needed to refine the system's algorithms, enhance data accuracy, and explore additional features, such as decision-support tools and scenario modeling. Adopting such integrated systems can play a crucial role in ensuring the sustainable use of groundwater resources in the face of growing demand and environmental challenges. Future research should focus on expanding the system's applicability to different geographical regions, integrating additional data sources, and improving user customization features to meet the specific needs of various stakeholders.

Keywords: *Artificial intelligence, machine learning, IoT, deep learning*

REAL-TIME GROUNDWATER PREDICTION MODELS USING RNN AND LSTM TECHNIQUES

Ashish Gupta[1*] and BG Rajeev Gandhi[2]

Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India
[*]Corresponding author e-mail: ashishgupta02515@gmail.com

Accurate real-time groundwater level predictions are essential for proactive water resource management, especially in regions facing acute water scarcity and heightened vulnerability under climate variability. Groundwater depletion and fluctuating recharge rates create urgent challenges for sustainable management, particularly in semi-arid regions like Madhya Pradesh. This state experiences varied geology, ranging from ancient Archean rocks to Deccan volcanic formations, which greatly influence its groundwater dynamics. State relies heavily on groundwater, and with over 1,600 observation wells monitored by the CGWB, data from this region provides a crucial foundation for understanding groundwater dyanamics. This study presents an AI-powered, real-time groundwater prediction model that incorporates advanced neural network architectures specifically, recurrent neural networks (RNN) combined with long short-term memory (LSTM) techniques. These architectures are well-suited to the sequential nature of groundwater data, allowing the model to provide high spatial and temporal resolution predictions. By synthesizing inputs from satellite-based observations, historical records from national databases, and localized monitoring stations, this model delivers accurate localized and real-time groundwater predictions. Our research focuses on three key challenges: (1) achieving model adaptability across diverse climatic and geological regions, (2) ensuring data accuracy through multi-source integration, and (3) improving spatial data interpretation in regions where observational data is limited. The model incorporates an error-correction mechanism and spatial data interpolation techniques, which are particularly beneficial for data-scarce areas. This integrated approach enhances the model's robustness and potential utility in groundwater management, especially in regions where conventional data may be sparse or inconsistent.

Groundwater serves as a primary source of drinking water and irrigation in many regions, yet its availability is increasingly compromised by climate-induced variability and over-extraction. The Madhya Pradesh exemplifies the challenges associated with groundwater management, given its geological complexity and regional disparities in recharge rates and extraction levels. From the Archean rocks in the northern and southeastern parts to Deccan Trap formations in the west, the geological diversity affects groundwater storage and movement. The India-WRIS provides extensive data on rainfall, soil characteristics, and other critical factors, which when integrated with satellite observations and CGWB well data allows for a comprehensive approach to groundwater prediction. However, traditional monitoring and management methods often lack the resolution and immediacy needed to respond effectively to these challenges. By leveraging AI-driven neural networks, this study aims to address these limitations and offer a predictive model adaptable to the diverse hydrogeological contexts found within State and beyond.

The methodology centers around a robust framework designed to incorporate and process multi-dimensional groundwater data. These include comprehensive data integration, model design and adaptive mechanisms to refine predictions over time. The architecture utilizes RNN-LSTM layers for processing time-series data effectively, enabling a nuanced approach to groundwater prediction. The model draws on multi-source data inputs. This data is enriched by satellite-based precipitation and terrestrial water storage data (e.g., GRACE) and historical groundwater records from national databases. Additional data on rainfall, temperature, and humidity from India WRIS allow the model to incorporate essential climatic factors impacting groundwater recharge and demand. The model's architecture is rooted in RNN, enhanced by LSTM layers that allow it to capture both short-term and long-term dependencies in groundwater data. RNN layers process sequential data effectively, while the LSTM component manages prolonged dependencies influenced by seasonal and annual climatic patterns. Backpropagation through time (BPTT) optimizes model weights, ensuring that each layer accurately learns from the historical groundwater data patterns. This design enables the model to accommodate cyclical fluctuations in groundwater levels and adapt to climatic changes. Recognizing that certain regions have limited observational data; the model employs an error-correction mechanism and spatial data interpolation. These mechanisms adjust predictions dynamically, compensating for data gaps by leveraging geological and hydrological data from nearby locations. For instance, in areas like the Deccan Traps, where data is scarce, spatial interpolation based on geological similarities improves prediction accuracy. The error-correction component continually refines model outputs, adapting over time to reduce prediction errors and enhance reliability, especially in data-limited settings. Multi-source integration is vital to the model's functionality, blending real-time data from satellite observations, well data, and climatic inputs to improve overall accuracy. A cloud-based processing system manages these data streams, ensuring continuous adaptability to environmental changes. By harmonizing data from diverse sources, the model captures groundwater level fluctuations in response to both local and broader climatic events, providing a realistic and timely representation of groundwater conditions across Madhya Pradesh's varied landscapes. The model's design allows for scalability and adaptability across regions with diverse geological features. This adaptability is tested within Madhya Pradesh's hydrogeological contexts, such as the hard rock regions of the north and alluvial plains in the central areas. By integrating parameters specific to local geology and climate, the model dynamically adjusts its calculations, enabling deployment across regions with differing groundwater behaviors. This scalability is crucial for extending the model's utility to other regions facing similar groundwater challenge.

The real-time groundwater prediction model developed in this study has critical implications for groundwater management and policy-making. By delivering localized, high-resolution predictions, this framework equips decision-makers with actionable insights to address regional groundwater stress proactively. In water-scarce regions like Madhya Pradesh, where demand often outpaces recharge, predictive insights can inform decisions on sustainable extraction limits, recharge initiatives, and water conservation strategies. Additionally, the model's capacity to identify high-risk extraction zones and potential recharge areas is invaluable for targeted interventions, such as implementing rainwater harvesting in vulnerable regions. As climate change intensifies hydrological variability, real-time prediction models provide the flexibility needed to adapt groundwater management practices to fluctuating conditions. This adaptability not only supports immediate decision-making but also contributes to long-term resource sustainability. By integrating AI-driven analysis with multi-source data, this model presents a scalable approach to groundwater management that can be applied in regions facing similar water security challenges globally.

Keywords: *Real-time groundwater prediction, RNN-LSTM models, AI, water management*

A HYBRID APPROACH TO MODELLING ETO AND GROUNDWATER RECHARGE

Manita[1*], Deepak Khare[1] and Manish Kumar Nema[2]

[1]*Indian Institute of Technology, Roorkee, India*
[2]*National Institute of Hydrology, Roorkee, India*
[*]*Corresponding author e-mail: manita@wr.iitr.ac.in*

This research focuses on developing a hybrid predictive model for estimating reference evapotranspiration (ETo) and understanding its relationship with groundwater levels using cloud computing and machine learning techniques. ETo is a critical component of the hydrological cycle and plays an important role in water resource management, irrigation practices, and climate adaptation strategies. The accurate estimation of ETo is crucial in cities, where it is influenced by several climatic factors such as temperature, humidity, wind speed, and solar radiation. This study seeks to address this challenge by leveraging hybrid machine learning models for improved ETo predictions, while using cloud-based data processing to handle large-scale computations and provide real-time predictions for surface and ground water management. The research relies on NASA Power's climate data as the primary source for estimating ETo. NASA's Power API provides global climate data, which includes key variables like temperature, humidity, wind speed, and solar radiation. These variables are essential for estimating ETo, particularly when using the Penman-Monteith equation. However, the traditional Penman-Monteith approach does not fully capture the complexities of urban climates, and hence, a hybrid machine learning approach is used to improve the accuracy of the predictions. The hybrid model combines the strengths of multiple machine learning algorithms to provide a more robust and accurate prediction of ETo under varying climatic conditions. This hybrid model is designed to handle non-linear relationships and adapt to the variable and complex nature of urban environments.

In this study, the hybrid prediction model integrates multiple machine learning algorithms to generate more reliable estimates of ETo. While basic machine learning models like Random Forest (RF) and Support Vector Machines (SVM) are effective in capturing non-linear relationships, the hybrid approach combines these algorithms with additional techniques to improve prediction accuracy and adaptability. The hybrid model's ability to combine different methods ensures better generalization across different urban environments with varying climates and urbanization levels. This approach is particularly useful when using climate data from NASA Power's API, which offers valuable global climate information but requires sophisticated models to account for the complexity of urban settings. The use of API-based data extraction is central to the methodology of this study. NASA Power's API provides accessible, from 1990s to near real-time climate data that can be used directly for calculating ETo. The API allows for easy retrieval of necessary climate variables, such as temperature, wind speed, humidity, and solar radiation, which are critical inputs for the hybrid model. Through an automated process, data for specific urban locations can be extracted directly from the cloud-based API, ensuring accurate and timely access to the most up-to-date climate information. This method eliminates the need for manual data collection, making the process more efficient and scalable. It also allows for dynamic data retrieval, which is essential for continuous monitoring of ETo and groundwater levels over time.

To process and analyse the large datasets generated from NASA Power's API, the research utilizes cloud computing platforms. These platforms enable the execution of complex calculations for ETo estimation and hybrid model predictions in a scalable and cost-effective manner. Cloud computing provides the computational power necessary to run the hybrid machine learning models without the limitations of local hardware resources. By leveraging cloud infrastructure, the study can process large volumes of climate data, perform real-time calculations, and deliver quick predictions. This cloud-based approach allows the model to be scalable, flexible, and adaptable to different urban locations, making it an efficient tool for urban water management and climate adaptation strategies. The study also explores the relationship between ETo and groundwater levels. Groundwater plays a significant role in urban water systems, and its levels are affected by multiple factors, including precipitation, soil characteristics, and evapotranspiration. As ETo increases, water is lost through evaporation and transpiration, which can reduce the amount of water available for groundwater recharge. Conversely, lower ETo can facilitate increased infiltration, promoting groundwater replenishment. By integrating groundwater data with the hybrid ETo model, this research aims to provide a better understanding of how variations in ETo influence groundwater recharge and water resources. The relationship between ETo and groundwater is dynamic and location-specific, making it crucial to use flexible, data-driven models for accurate predictions.

To evaluate the accuracy of the hybrid model, the study uses root mean square error (RMSE) and mean absolute error (MAE) metrics to quantify prediction accuracy. These metrics are used to compare the ETo estimates generated by the hybrid machine learning model with the values calculated using traditional methods, such as the Penman-Monteith equation. By validating the model with real-world data, the study aims to demonstrate the effectiveness of the hybrid approach in urban settings. Preliminary evaluations suggest that the hybrid model provides more accurate predictions compared to traditional methods, as it is better equipped to account for the non-linear interactions between climatic variables. This improved accuracy is particularly important for surface and ground water management, where precise ETo estimates are crucial for irrigation and groundwater conservation. While the results of the study are still under analysis, initial findings indicate that the hybrid machine learning model, when combined with cloud computing and API-based data extraction, significantly enhances the accuracy and flexibility of ETo predictions in urban areas. The integration of real-time climate data from NASA Power's API allows for continuous updates and predictions, making it a powerful tool for decision-makers involved in urban water resource management. The study also demonstrates the potential of cloud-based models for scaling predictions across diverse urban environments, offering a valuable tool for sustainable urban development. In conclusion, this research highlights the potential of using hybrid machine learning models combined with cloud computing to estimate ETo and analyse its impact on groundwater recharge. By leveraging NASA Power's climate data through API access, the study provides a scalable and flexible approach for estimating ETo in real-time, which is crucial for urban water management. This work represents a significant step toward using data-driven approaches for sustainable urban water resource management and climate adaptation.

Keywords: *Evapotranspiration, hybrid modelling, groundwater recharge, API, climate data*

COMPARATIVE PERFORMANCE ANALYSIS OF LINEAR REGRESSION, KNN, AND RANDOM FOREST FOR MODELING TOTAL DISSOLVED SOLIDS IN GROUNDWATER: A CASE STUDY

A.R. Malavika and V. Santhosh*

MES Ponnani College, Ponnani, Kerala, India
**Corresponding author e-mail: santhoshviswapal@gmail.com*

Total Dissolved Solids (TDS) refers to the whole concentration of dissolved substances in groundwater, encompassing minerals, salts, and organic material. The amounts of Total Dissolved Solids significantly influence water quality, contingent upon its use in residential, agricultural, and industrial contexts. Excessively high or low TDS negatively impacts the potability of drinking water, irrigation practices, and aquatic ecosystems. Therefore, it necessitates the implementation of suitable monitoring and management strategies for groundwater resources. Conventional TDS measurement methods are laborious and costly. Consequently, the advancement of more efficient methodologies is imperative. Machine learning model-based techniques offer a promising approach to forecasting groundwater quality indicators, including TDS, by including several chemical and physical water characteristics. The performance of three different machine-learning algorithms: Linear Regression, K-Nearest Neighbours (KNN), and Random Forest, is compared in the context of modelling TDS levels in groundwater. All these models have pros and cons related to complexity, interpretation, and effectiveness of forecasts. Linear Regression is one of the most elementary statistical techniques that assumes a relationship to be linear between variables and the target. Hence, it is both easy and powerful. KNN is an instance-based non-parametric learner that is versatile and capable of fitting any kind of non-linear interactions. It is, however, sensitive to data scaling. Random Forest, on the other hand, is an ensemble learning technique using several decision trees to enhance predictive accuracy and prevent overfitting. This study seeks to identify which model most accurately predicts total dissolved solids in groundwater, hence enhancing water quality management.

Data was collected from 60 groundwater sample locations for pre-monsoon season for this investigation. Each sample provides a collection of water quality characteristics, including pH, Salinity, Electrical Conductivity (EC), Total Alkalinity (TA), Total Hardness (TH), Sodium, Potassium, Magnesium, Calcium, Chloride, Nitrate, Sulphate and Fluoride. Prior to the implementation of machine learning models, the dataset underwent multiple preprocessing stages. All missing values were replaced with the median value of each parameter to preserve data integrity. All input variables were standardized to a uniform scale to ensure that all features contribute equally to the model, particularly for algorithms such as KNN, which depend on distance metrics. Feature selection approaches were employed to determine the most relevant factors for predicting TDS and to minimize noise. Linear Regression, KNN, and Random Forest models were developed and trained using the pre-processed dataset.

The linear regression model assumes a linear relationship between water quality parameters and TDS. Ordinary least squares were used to decrease sum of squared differences between expected and actual data. Simple linear regression shows how each parameter influences TDS. K-Nearest Neighbours (KNN) is non-parametric method prediction based on training set data similarity. KNN may describe non-linear interactions, however it is computationally expensive for large datasets and requires accurate distance measurements. Random Forest learning method combines decision trees to avoid overfitting and increase accuracy. In order to ensure diversity, each decision tree is trained on a random sample and has random feature selection at each split. To predict, all tree projections were averaged. Random Forest can express complex, non-linear relationships and manage large datasets without much change. Metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) were plotted to assess performance of each model. A correlation matrix was also built to assess the relationship between all measures of water quality and TDS, elucidating the contribution of each component to TDS prediction.

The Linear Regression model showed outstanding performance, as evidenced by a significantly high R^2 value (0.9987), suggesting that the model accounted for almost all the variance in TDS. The MSE and MAE values were comparatively low, indicating that the model accurately predicted the TDS values. This suggests that the total dissolved solids in the specified groundwater dataset exhibit a linear trend. Random Forest had superior performance, although lacked the accuracy of Linear Regression. The R^2 value (0.9715) indicates that the model accounted for almost 97% of the variance in TDS, therefore, representing a favourable outcome. Nonetheless, the MSE and MAE values were greater, signifying that the prediction errors of the model exceeded those of Linear Regression. Typically, Random Forest exhibits more resilience to complex and non-linear data; but, in this instance, it performed inferiorly than Linear Regression. The overall performance of the three models was overshadowed by the other two, since KNN exhibited the poorest performance. The MSE and MAE values were significantly elevated in comparison to the Linear Regression and Random Forest models, indicating greater prediction errors. The R^2 score is 0.8475, indicating that the model accounts for around 85% of the variance in TDS, which is acceptable but significantly lower than the other models.

The performance comparison of Linear Regression, Random Forest, and KNN in predicting TDS in groundwater reveals that Linear Regression is the superior model, exhibiting the lowest MSE and MAE, along with the highest R^2 value. The straightforwardness of adopting the linear relationships between input factors and TDS contributes significantly to its performance in this context. Robust and adaptable, Random Forest exhibited marginally inferior performance compared to Linear Regression, however yielded satisfactory results. KNN had the highest prediction errors and the lowest R^2 value, perhaps due to its sensitivity to hyperparameter adjustments and the data's complexity. This study advocates for the utilization of Linear Regression as the model for predicting TDS in groundwater due to its accuracy and interpretability. Future work may involve refining the KNN model, or exploring alternative machine learning techniques such as support vector machines or neural networks to enhance the prediction of groundwater quality.

Keywords: *Ground water quality, TDS prediction, machine learning, KNN, linear regression, random forest*

UTILITY OF OPEN AND SATELLITE-BASED DATA FOR GROUNDWATER RESOURCE ESTIMATION USING WATER ACCOUNTING PLUS (WA+) FRAMEWORK FOR AN INDIAN PENINSULAR BASIN

Prabhash Kumar Mishra[1*], Rajat Agarwal[2], Nitesh Patidar[3] and Ranjan Kumar Ray[4]

[1]National Institute of Hydrology, Roorkee, India
[2]Central Ground Water Board, New Delhi, India
[*]Corresponding author e-mail: erprabhash@gmail.com

Water resources, particularly groundwater, play a critical role in drinking, agriculture, and forestry. Although, Central Groundwater Board (CGWB) is the central nodal agency, the onus of groundwater development and management is mainly a state subject, and lies with respective Groundwater divisions of State Governments. Unless a resource is properly quantified, its management and conservation is a far cry. In case of groundwater, its estimation is a tedious, time consuming and expensive activity. However, with technological advancements, high computational capabilities, use of open and satellite-based data for estimation of groundwater resource is gaining momentum these days. Although, these estimates will never replace the field-based observation accuracy, but provide considerable information on the groundwater resource of a region with comparable accuracy. The python-based water accounting plus (WA+) framework, developed by IHE-Delft, The Netherlands is a step in this direction. The framework utilizes open and satellite-based data, and accounts the water resources of a basin utilizing the outputs from specially developed 'WaterPix' model with the concept of blue and green water pixels. The blue and green water pixels indicate the type of water used in a region. Precipitation and soil moisture comes under green water, whereas surface water and deep groundwater which are basically a form of delayed precipitation, comes under blue water. This concept of green and blue water also applicable to rainfed and irrigated area. Yield from rainfed area is only contributed by green water, whereas, an additional supply as irrigation (blue water) along with the green water contribute the yield in case of irrigated area. In the present study, WaterPix model was set-up for the Mahanadi basin. The Mahanadi basin is one of the major multi-states (Chhattisgarh and Odisha), east flowing peninsular river basin with a geographical area of 1,41,589 km^2 (Chhattisgarh – 75136 km^2, Odisha – 65580 km^2). The major significant tributaries of Mahanadi River are Pairi, Seonath, Hasdeo, Mand, Kalma and Jonk, which joined above the Hirakud dam, whereas the Ong and Tel rivers join below the Hirakud dam. The basin has a subtropical climate, with the average temperature during the summer months being about 29°C and in winters about 21°C. The average annual rainfall of 106 years (1901–2006) is about 1400 mm, out of this about 90% rainfall received during the monsoon period (June–September). As the waters of the Mahanadi basin are not yet allocated or abstracted completely, the basin provides ample scope for evolving an integrated approach for equitable, democratic and sustainable management of its resources. However, rapid changes in the Mahanadi basin, especially during the last two decades, suggest that not only is the water use increasing, but new problems are emerging in the inter-sectoral allocation and also between the two major riparian states, Odisha and Chhattisgarh. One of the root causes of riparian disputes is trust deficit in the methods and criteria adopted in the estimation of water availability particularly for basins spread of multiple states and transboundary in nature. In a situation like this, a well-defined, unbiased scientific approach which mostly utilizes the remote sensing based open data for accounting the basin's water resources is handy in addressing conflicts. Initially, the input data such as precipitation, evapotranspiration, soil moisture, net dry matter (NDM), leaf area index (LAI), GRACE (Gravity Recovery and Climate Experiment), etc. were downloaded from open sources at daily and monthly scales. Then, the hydrological model WaterPix was set-up and simulated for the period 2004-2020 to assess soil moisture and water balance in the basin. The model estimates the surface runoff, total runoff, changes in water storage, percolation, base flow and the supply of green and blue water, based on soil moisture water balance. Prior to that, Globcover and NRSC LU, Population density, MIRCA irrigated, MIRCA rainfed, WDPA, and crop data from IWMI were used to develop the water accounting-based land use (WALU). It is one of the essential inputs before setting up the WA+ model. Finally, WA+ sheets were generated for utilized flow, surface water, groundwater, and resource base. In the paper, the factsheet on groundwater (Sheet 6) is discussed in detailed. The groundwater sheet estimates vertical recharge both natural and man-made as well as GW abstraction, return flows, changes in storage, and baseflow contributions. The basin's average annual recharge, groundwater abstraction, change in aquifer storage, and baseflow were computed as 110.47 BCM/year, 53.67 BCM/year, 19.02 BCM/year, and 73.8 BCM/year, respectively. The assessment supports the development of sustainable groundwater management plans, such as safeguarding against excessive drawdown, issuing extraction permits, and advancing managed aquifer recharge (MAR) techniques to enhance long-term water security. Similarly, the mean change in water storage fluctuated between −52 mm and 26 mm across the basin, with negative values indicating a decrease in water storage. This decline was most prominent in the northern and southeast parts of the basin, largely attributed to the imbalance between water withdrawals and return flows (such as groundwater recharge). Percolation rates also showed significant variation, ranging from 744 mm to 2,157 mm throughout the basin. Overall, the analysis suggests that WA+ is an effective tool for assessing water availability and usage of surface water and groundwater in river basins. The study helps the decision-makers particularly the basin authority to enhance resource planning, mitigate risks, and optimize water use for long-term sustainability and resilience in the face of growing water challenges.

Keywords: *Groundwater, WA+, WaterPix, Mahanadi basin, green and blue water*

PREDICTION OF GROUNDWATER LEVEL USING MACHINE LEARNING TECHNIQUES OF THE WESTERN ARAVALI RANGE, RAJASTHAN, INDIA

Sushindra Kumar Gupta*, Satendra Kumar and Anupma Sharma

National Institute of Hydrology Roorkee, India
**Corresponding author e-mail: skgupta.nihr@gov.in*

Groundwater is of paramount importance and serves as a primary water source for arid and semi-arid regions. Groundwater level (GWL) prediction is crucial in arid and semi-arid region due to the uneven distribution of rainfall and unsystematic utilization of groundwater at both global and regional scales. Over the past few decades, groundwater resources have experienced substantial decline due to over irrigation practices, infrastructure development, day-by-day increase of population, increase of industry area etc. The GWL is direct and simple estimation of groundwater availability and accessibility. Having a proper understanding of past, present and future scenario of GWL can provide proper planning of water resources, formulation of new policy of central/state level and practitioners in hydrology sectors with better insight of and perception to develop strategies for the water resources planning, development and management, to ensure sustainable socio-economic development. Although the western Aravali region has been identified as a significant for groundwater level changes. The primary goals of the current investigation are (1) to predict the GWL based on standalone and ensemble machine learning algorithms for the Ajmer city, Rajasthan, (2) to compared the proposed and existing models over the study area, and (3) to compared between the observed and predicted GWL over the study area.

GWL has been predicted using the suggested AdaBoost, Classification and Regression Tree (CART), Random Forest (RF), and REP tree models. An ensemble modelling method for GWL prediction is the AdaBoost model. It is pointing to a classification tree boosting strategy that is adaptable. It is a non-parametric approach that effectively separates outliers without requiring the learners to be defined. AdaBoost begins by giving the data set equal weights in order to create an initial decision tree (DT) for training. The fitted model provides the whole training in the following stage. The larger weights are labelled as misclassified, whereas the weight of the accurate predictors is marked as fixed. With respectable results, CART has been used extensively for groundwater modelling, including GWL prediction. Since the CART is regarded as an unstable model, we can enhance the model's performance by using the bagging strategy. An ensemble approach known as the RF model is used to model streamflow, rainfall-runoff, and groundwater. In order to improve model accuracy and prevent training overfitting issues, the suggested RF model is built using the random subspace method to create a multiple decision tree with controlled variance. Regression and classification are the primary uses for the Reduce Error Pruning tree (REPTree) model. For groundwater modelling, particularly GWL prediction, CART has been utilised extensively with respectable results. We can enhance the model's performance by using the bagging strategy, since the CART is regarded as an unstable model. An ensemble algorithm for streamflow, groundwater, and rainfall-runoff modelling is the RF model. The suggested RF model is built using the random subspace approach, which creates a multiple decision tree with controlled variance to improve model accuracy and prevent overfitting during training. Cut Down on Error the Pruning Tree (REPTree) model is mostly utilised for regression and classification. The REPTree model is an extremely effective methodology for quick learning that creates decision/regression trees using reduced error pruning strategies. The model

efficiently prunes the tree by employing back-fitting with reduced error, which lessens instability and mitigates mistakes brought on by overfitting.

Its primary objective is to create a set of guidelines for making decisions based on predicted factors. One of the main methods for building decision trees is the REPTree model process, which builds regression trees based on variance data by using condensed error pruning. By creating numerical ranges inside the model, this technique makes predictions that are more reliable and accurate. The water resources department of Rajasthan State in India provided the GWL data for this study, which was gathered between 2011 and 2023. In this investigation, data from 252 wells mostly drilled and piezometric wells were included. Both the Central Groundwater Board (CGWB) and the State Groundwater Department (SGWD) of the planned study region measure the GWL data. Similarly, the NASA power platform has been used to gather the pre-monsoon and post-monsoon datasets for average rainfall, average temperature, maximum temperature, and lowest temperature. The models' ideal input combination was assessed using a number of different methodologies, including the average rainfall, average temperature, maximum temperature, and minimum temperature before and after the monsoon. The suggested models have been trained, validated, and tested using a total of 3984 data sets. As a result, 70% of the dataset was used for model training, 20% for model validation, and 10% for testing the suggested models.

The model's performance was evaluated against that of alternative methods using four statistical indices: correlation validation coefficient (CC), Nash-Sutcliffe Efficiency (NSE), root mean square error (RMSE), normalised root means square error (nRMSE), and four visual representations: scatter plot, violin plot, radar diagram, and Taylor diagram, in that order. According to the GWL prediction results, the boosting models (AdaBoost & REPTree) had a lower NSE (0.98) and CC (0.97) than the bagging models (RF & CART). In general, the RF model performs better than the CART, AdaBoost, and REPTree models. In the areas of watershed management and aquifer management, researchers, academics, and policymakers can all benefit from the prediction of GWL in this study by preserving an optimal harvest from these valuable natural resources.

Keywords: *GWL, bagging technique, boosting technique, Aravalli range*

DESIGN, IMPLEMENTATION AND VALIDATION OF AN AUTOMATED INFILTROMETER

Kritika Bansal*

Department of Civil Engineering, Indian Institute of Technology Kanpur, India
**Corresponding author e-mail: kritikab20@iitk.ac.in*

This study introduces the design, development, and calibration of an automated infiltrometer primarily focusing on modifying a conventional Double-Ring Infiltrometer system using sensors and a data acquisition system. Infiltration is a vital hydrological parameter with applications in rainwater harvesting and contaminant transport analysis. Traditional infiltrometer systems often require labor-intensive manual operations and lack precision for diverse soil and environmental conditions. The automated system retrofits a double-ring infiltrometer with sensors, a microcontroller, and a data acquisition configuration to enable real-time monitoring and precise measurement of water loss. Systematic calibration involving rigorous laboratory testing to establish baseline performance ensures measurement accuracy and reliability. Preliminary results demonstrate the potential of this automated infiltrometer to deliver high reliability and accuracy compared to traditional methods, enhancing its utility in applications such as water resource management, agricultural planning, and environmental monitoring. Its modular, cost-efficient design facilitates widespread adoption and scalability, particularly in resource-constrained settings. The system's compatibility with existing hydrological models further extends its utility, allowing seamless integration into watershed simulations and climate resilience studies. This work represents a significant advancement in hydrological instrumentation by addressing critical challenges in infiltration measurement. Future efforts will focus on integrating predictive algorithms for real-time data analysis and refining the system's portability for broader field applications. The automated infiltrometer offers a practical and efficient solution to modern water resource challenges, contributing to sustainable management practices in the face of increasing climate variability.

***Keywords**: Double ring infiltrometer, soil infiltration measurement, real time data collection, hydrological applications, data acquisition system*

GEOSPATIAL INTELLIGENCE FOR GROUNDWATER MANAGEMENT

Prateek Khanna*

Amity University, Ranchi, India
**Corresponding author e-mail: prateekkhanna.bea@gmail.com*

Groundwater, the vital elixir of sustainability, is increasingly an important resource for drinking water, irrigation, and industrial processes. However, unsustainable practices and a lack of information has led to its overexploitation and contamination. This is where geospatial intelligence (GEOINT) emerges as a game-changer for effective groundwater management. High-resolution satellite imagery, aerial photography, and Light Detection and Ranging (LiDAR) data can be used to create detailed maps of aquifers, including their extent, depth, and potential yield. By integrating geospatial data with rainfall records, evaporation data, and geological information, hydrologists can build models to simulate groundwater flow and predict recharge areas. Repeated satellite radar measurements can detect changes in ground elevation with high precision, allowing us to monitor groundwater depletion and identify areas with declining water tables. Remote sensing techniques can help identify potential sources of groundwater contamination, such as agricultural runoff or industrial waste disposal sites. GEOINT provides objective data and insights, enabling informed decisions on well placement, pumping rates, and artificial recharge projects. Locating productive aquifers and optimizing well locations can reduce drilling costs and energy consumption associated with groundwater extraction. Monitoring groundwater levels and quality helps ensure sustainable use of this precious resource and prevent long-term depletion or contamination. By tracking changes in groundwater levels and identifying potential contamination sources, GEOINT can help trigger early warnings to prevent water crises.

This study employed a comprehensive approach to demonstrate the application of GEOINT in groundwater management. The research began with a thorough review of existing literature on GEOINT and groundwater management to identify best practices and challenges. Three case studies from India, California (USA), and Africa were then analyzed to demonstrate the successful application of GEOINT in groundwater management. The first case study relates to India where geospatial data is being used to map and manage groundwater resources in drought-prone regions, enabling targeted interventions like rainwater harvesting and well construction. The CGWB utilizes geospatial data to create aquifer maps, monitor groundwater levels, and assess groundwater quality across the country. The National Remote Sensing Centre (NRSC) has developed various geospatial technologies for groundwater management, including satellite-based drought monitoring and groundwater potential zonation. Several state governments are actively using geospatial data to tackle drought and water scarcity challenges. For example, the Maharashtra government's Jal Yukti Shivar Abhiyan leverages geospatial data to identify and prioritize areas for rainwater harvesting structures. However, given the enormity of the problem with rising population and fast depleting resource, much more needs to be accomplished in this domain. The next case study is about the Central Valley which is California's agricultural heartland, but it also relies heavily on groundwater for irrigation. Decades of over-pumping have led to declining water tables, causing land subsidence (sinking) and impacting agricultural productivity. Climate change with less precipitation and higher temperatures further exacerbates the situation. California's Central Valley relies on geospatial modeling to track groundwater levels and implement sustainable water management practices in the face of prolonged drought. The

California Department of Water Resources (DWR) utilizes the California Central Valley Groundwater Model (CVM) to assess groundwater availability, evaluate drought impacts, and inform sustainable water management practices. The CVM helps water managers understand how changes in groundwater pumping, surface water deliveries, and climate can affect groundwater levels throughout the Central Valley.

The third case study discusses about several African countries which are utilizing GEOINT to identify suitable locations for drilling wells in remote areas, improving access to clean water for rural communities. For example, The United Nations International Children's Emergency Fund (UNICEF) has used GEOINT in Ethiopia and Madagascar to identify drilling locations for water wells. By pinpointing areas with higher likelihood of groundwater, GEOINT significantly increases the success rate of well drilling projects, saving time, resources, and money. Focusing drilling efforts on promising locations minimizes environmental impact and avoids disruption to pristine areas. The study yielded several key findings that demonstrate the effectiveness of GEOINT in groundwater management.

The analysis of geospatial data revealed that GEOINT can accurately map and monitor groundwater resources, including aquifer extent, depth, and potential yield. For instance, the case study in India demonstrated that GEOINT-based mapping enabled the identification of potential groundwater recharge zones, which informed the development of effective groundwater management strategies. The study found that GEOINT can optimize well placement and pumping rates, reducing drilling costs and energy consumption. The case study in California showed that GEOINT-based analysis enabled the optimization of well placement, resulting in a significant reduction in drilling costs. The analysis revealed that GEOINT can establish early warning systems for groundwater depletion and contamination. The case study in Africa demonstrated that GEOINT-based monitoring enabled the detection of groundwater contamination sources, allowing for proactive measures to prevent further contamination. The study also emphasized the importance of capacity building and data accessibility in ensuring the effective use of GEOINT in groundwater management. The case studies highlighted the need for training and capacity building programs to enhance the skills of water management professionals, policymakers, and local communities in interpreting and utilizing geospatial data.

In a nutshell, the study demonstrates the potential of GEOINT in supporting effective groundwater management. The application of GEOINT enables improved mapping and monitoring of groundwater resources, allowing for better decision-making. By adopting a GEOINT-based approach, groundwater management can be transformed, ensuring the sustainable use of this vital resource for future generations. Future advancements in remote sensing technologies will provide even more detailed information about groundwater recharge zones. Making geospatial data readily available and accessible can empower local communities, researchers, and entrepreneurs to participate in developing innovative solutions for groundwater management. Equipping water management professionals, policymakers, and local communities with the skills to interpret and utilize geospatial data is essential for maximizing the impact of GEOINT initiatives. GEOINT can support the integration of surface water, groundwater, and treated wastewater resources, optimizing water use efficiency and ensuring long-term sustainability.

Keywords: *GEOINT, groundwater management, SAR, remote sensing*

ASSESSING GROUNDWATER RESOURCES IN SIKKIM, INDIA USING REMOTE SENSING, GIS AND GOOGLE EARTH ENGINE

Manish Subba[1*], Brijesh Kumar Yadav[1] and Aditya Kapoor[2]

[1]Department of Hydrology, Indian Institute of Technology, Roorkee, India
[2]University of Nebraska-Lincoln, Nebraska, United States of America
[*]Corresponding author e-mail: manish_s@hy.iitr.ac.in

Groundwater is a critical component of the hydrologic cycle. It represents a vital component of Earth’s freshwater resource as it accounts for 99% of the total unfrozen freshwater on earth. It’s significant role in supporting ecosystems, agricultural, industrial and domestic water supply warrants its sustainable management, especially in fragile ecosystems. Therefore, accurate, periodic, and low latency observation is paramount for its effective sustainable management. However, due to lack of ground observations, particularly in remote or challenging terrains, quantifying spatio-temporal changes in groundwater resource are difficult by employing conventional methods that require in-situ data collections. In such cases, remote sensing data can be beneficial for providing frequent estimates even in inaccessible or remote areas. Traditional GIS and Remote Sensing (RS) platforms work on local computer systems that are prone to machine failure and computations are time consuming. These drawbacks can be circumvented by using cloud-based platform such as Google Earth Engine (GEE). GEE has more than 600 datasets with many products potentially instrumental in performing water resources assessment. Additionally, its parallel processing capabilities results in faster computations facilitating quicker decision making. This study employs a water balance approach to estimate changes in groundwater storage for the north eastern Indian state Sikkim, where the components of the water balance have been computed using remote sensing products from GEE catalogue. To assess the stress on groundwater resources, the Stage of Groundwater Development (SGD) has also been computed and potential remedial measures have been suggested.

Sikkim is a north eastern state of India and is a part of eastern Himalaya with an approximate area of 7100 km^2. The entire state exhibits slope mostly greater than 20% where the conventional groundwater estimation methods often prove to be inadequate, making it an ideal study area of employing the remote sensing and GIS based water resources assessment methodology. Sikkim is a part of Teesta catchment that provides runoff to Teesta River that flows from North-west to South-eastern direction. The majority of the area (>57%) is covered by trees and 12% of the area is covered by snow. Water bodies and built-up area cover around 0.6% and 0.2% of the study area. Analysis of the collected lithological data exhibits large spatial variability with schist, phyllites, clays and pebbles constituting the majority of formations. Akin to the majority of India, Sikkim receives majority of its annual rainfall (>75%) in monsoon season that spans from June to September and rest of the rainfall is received in non-monsoon season. Apart from the domestic and agricultural usages there are currently 10 pharmaceutical industries in the south-eastern part of Sikkim that relies on groundwater for their industrial consumption. Additionally, the entire population of Sikkim relies heavily on springs as a source of freshwater, and for their domestic use. This over-reliance, and limited groundwater recharge due to the topography, and hydrogeology of the area poses significant risks of aquifer depletion, water quality issues, and ecological and hydrological deterioration. The depletion of the groundwater resource is evident in the decline in the observed water tables. Proposed methodology employs an integrated approach that amalgamates RS, GIS and GEE computational abilities to generate estimates of groundwater storage and subsequent computation of SGD. This integrated approach provides a more scalable and low latency framework for groundwater assessment in regions with complex terrain and scarce availability of field data. The proposed methodology is based on a water balance equation to compute changes in groundwater storage. The components of the water balance were computed using the Climate Hazards Center InfraRed Precipitation with Station Data (CHIRPS-Pentad) dataset, TerraClimate dataset, Soil Moisture Active Passive (SMAP L-4) dataset, Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) dataset, and Global Surface Water (GSW) dataset. Net groundwater inflows and out flows of the area is not considered as it was considered to be significantly small for the study area relative to the other components of the water balance. The SGD was computed using the proposed approach came out to be 12.12%, which according to the GEC falls under the safe category. Albeit the SGD in the study area was found to be quite low, uncontrolled use and mismanagement of groundwater resources coupled with climate change may lead to several adverse impacts on the overall environment and sustainability of the surrounding ecosystems. The study recommends multiple approach in addressing these challenges. Rainwater harvesting structures can be constructed to facilitate the groundwater recharge in valleys having high water demands. The collected water may be used directly or can be injected into the groundwater via injection well or infiltration basins. Small sized check dams and other water conservation practices can be implemented around the Industrial setups requiring large water demands. These dams may be constructed along the vast network of streams prevalent in the area to enhance the infiltration of surface water into the groundwater via stream bed seepages, but only after proper planning considering the factors spanning from soil parameters to logistical ones. Additionally, awareness campaigns and community engagement programs are emphasized for fostering sustainable water use practices at the base level. However, one must be mindful of the fact that remote sensing approach relies on pixel values to make inferences, which may fail to capture all the hydrogeological complexities and could be unrepresentative of the area under scrutiny, and estimates computed with remote sensing will almost always be less accurate compared to those obtained from ground observations.

This study implies that high-resolution remote sensing datasets aboard Google Earth Engine (GEE) can overcome data and computational limitations of traditional methodologies to a large extent. This study’s framework can be applied to similar mountainous terrains, and data scarce situations for guiding sustainable groundwater management. Furthermore, the research provides a foundation for future studies assessing the relationships between climate change and groundwater dynamics. This study can be further expanded/ modified to incorporate climate projections like precipitation, temperature, and evapotranspiration which allows for scenario-based planning, resilient and adaptive management strategies. With advancement in the field of remote sensing, and as the quality of the remote sensing datasets are continuously getting better, and accessibility is getting easier, the results using this approach are expected to improve over time. In conclusion, this study underlines the importance of advance remote sensing and GIS methodologies in addressing the complexities of groundwater resource assessment and management, especially in regions with complex terrains like Sikkim where conventional methodologies may be infeasible to provide adequate insights.

Keywords: *Groundwater management, remote sensing, GIS, Google Earth Engine, groundwater resource assessment, water balance, hydrology*

DIGITAL TWIN FOR SUSTAINABLE WATER MANAGEMENT FOR A VILLAGE IN INDIA

Dinesh Kumar Singh* and Vishnu Sharma

Dept. of Computer Science and Engineering, FCE, Poornima University, Jaipur, India
Corresponding author e-mail: 2022phdodddinesh12942@poornima.edu.in

Approximately 75% of rural households in India lack access to safe tap water, with extreme water stress predicted by 2050 due to over-extraction of groundwater. Groundwater is crucial for irrigation, especially in areas with low rainfall and for better livelihood and socio-economic development. This is counterproductive for the vision of Viksit Bharat 2047 and UN Sustainable Development Goal-6 (SDG-6) which aims to ensure everyone has access to safe and affordable drinking water. Among multiple reasons for the worsening groundwater situation in India, such as, access to cheap pumps, subsidised electricity, changing crop patterns and increasing population, lack of sound scientific understanding among stakeholders is also a major reason. Groundwater is not directly visible; hence, it is difficult for the local communities such as farmers, gram panchayat (GP), water user association, and other water stakeholders to visualize the complex hydrogeological aspects and estimate the volume of water available using equations, and modelling. There is a need for village communities to learn about groundwater science to make informed decisions and improve their management practices. This requires community members to take responsibility for gaining this knowledge and manage this scare resource sustainably. But this requires a technology-based method and system which is simple to understand and use considering the competencies of the village level stakeholders.

The digital twin (DT) concept is emerging as a helpful tool for water management. A digital twin network acts as a virtual model of a physical network, critical for analysing and managing elements or nodes of water resources in real-time. It can simulate water supply and demand, predict water quality, and detect possible contamination, thereby supporting a transparent and informed decision-making which is the need of the hour. Randullabad village in India, which has limited access to clean water for its population of 400 homes, exemplifies this issue. Hydrology of Randullabad covers key aspects of rainfall, groundwater, streams, and water management. Challenges include the hydrogeological complexity of basaltic aquifers, which have uneven recharge capacities, and the impact of climate change, leading to rainfall variability. Between 1-1.5°C increase in temperature over 30 years has raised evaporation rates, worsening water scarcity conditions. Annual rainfall in Randullabad ranges from 500-600 mm, primarily during the monsoon season from June to September. High runoff and low infiltration result in limited groundwater recharge. The area is characterized by basaltic aquifers with low permeability and poor storage capacity, where groundwater resides in fractures and weathered zones. The average annual rainfall has seen reductions, alongside a rise in drought occurrences, intensifying dependency on groundwater for irrigation, leading to over-extraction that harms groundwater quality. The village struggles with an inadequate water supply network, resulting in scarcity, competition, and poor water quality that threatens socio-economic growth, social conflicts, and public health, especially for children. The lack of real-time monitoring and data-driven decision-making worsens these problems. Currently, Randullabad's water management relies on sporadic manual data collection and a few sensors, making it hard to get a real-time understanding of water availability. This leads to slow responses and poor decisions, causing water wastage

and conflicts over resources. To overcome these challenges, a new, integrated approach is needed for sustainable water management.

This paper proposes using a digital twin for water management (DToWM) in Randullabad village to tackle existing shortcomings. The proposed architecture utilizes real-time data from sensors, IoT, satellite remote sensing, historical data, and machine learning to create a virtual model of the village's water supply network especially better understanding the recharge and storage potential of the aquifer. This can lead to the creation, maintenance, and monitoring the manage aquifer recharge interventions such as check dams, tanks, ponds. Implementing this approach could significantly improve water sustainability and public health in Randullabad. Groundwater contamination is another concern, stemming from agricultural runoff, geogenic sources, and improper wastewater disposal. High nitrate levels are found in many wells, primarily due to fertilizer use, and fluoride contamination poses health risks. Remediation strategies include improved agricultural practices, established treatment systems for waste, groundwater recharge initiatives, and community awareness projects for sustainable farming methods. DToWM can enhance Randullabad's participatory groundwater management for periodic groundwater assessment. Remote sensing and geophysical surveys can identify recharge zones and aquifer characteristics, while hadrochemical analysis and isotopic studies help track contaminant sources and flow patterns. Moreover, IoT-based monitoring systems can be used for real-time data collection on groundwater levels and quality and alerting the users. Recommendations for improvement involve using advanced modeling tools for better predictions and integrating traditional knowledge with modern practices for effective groundwater management. In conclusion, advanced techniques and community participation are vital in improving groundwater assessment and management in Randullabad, helping to address water scarcity and ensure sustainability in semi-arid regions.

Keywords: *Water management, digital twin, groundwater management, sustainable development goals*

MACHINE LEARNING-BASED GROUNDWATER LEVEL PREDICTION FOR SUSTAINABLE WATER RESOURCE MANAGEMENT IN AN INDUSTRIAL CATCHMENT

Ratnakar Swain* and Mani Haarika Nadella

[1]*Department of Civil Engineering, National Institute of Technology, Rourkela, Odisha, India*
**Corresponding author e-mail: swainrk@nitrkl.ac.in*

Groundwater is a vital source of freshwater, playing a crucial role in agriculture and economic development. Accurate groundwater level (GWL) prediction is essential for sustainable water resource management. As soft computational data-driven technologies have advanced in recent years, numerous machine learning models have been created and are being used for GWL forecasting. ML models are more desirable than physically based and numerical methods because they can simulate and predict GWL without needing a thorough understanding of the underlying topographical and hydro-geophysical characteristics. An efficient substitute that does not call for precise and particular physical factors and attributes are machine learning (ML) techniques. Some of the widely used ML models in GWL Prediction are Support Vector Machine (SVM), Artificial Neural Networks (ANN), Extreme learning machine (ELM), Fuzzy logic, Adaptive Network-based Fuzzy Inference System (ANFIS). For many hydro-meteorological applications, hybrid machine learning and genetic models have been attempted and tested in numerous researches. The literature evaluation indicates that these hybrid machine learning models perform better in terms of prediction than many solo models. The ability of hybrid models to identify intricate mathematical nonlinear interactions between the independent and dependent parameters is one of its primary benefits. The objective of this study is to downscale Monthly GWL from quarterly GWL data using the Kalman Filter algorithm; To predict GWL using rainfall, evapotranspiration, mean temperature, and infiltration data using three ML methods – Decision tree, Random Forest and Bagged Tree; To compare and analyze the performance of the three models.

Talcher is one of the 4 sub-divisions of Angul district in the state of Odisha. The latitude and longitude of Talcher are N 20° 56' 57.372", E 85° 14' 0.744". It has a savanna climate which is known as tropical wet and dry climate with the Classification: Aw. The quantity of rainfall during summers surpasses that of winters. The area's yearly temperature is 30.53ºC (or) 86.95ºF and is 4.56% higher than India's average. Talcher typically receives about 131.22 mm of precipitation and has 134.21 rainy days annually that is36.77% of the time. Brahmani River flows through Talcher. The Brahmani River and its major tributaries such as Nandira Jhor, Singhara Jhor, Tikra Jhor, Samakoi, Nigra, Gambhira, etc. run through the Angul district. For the training stage the mean value of GWL is 3.90 m, with a maximum value of 7.45 m and minimum value of 0.22m. From the graphs and the performance metrics, it can be observed that for the training stage the Random Forest model is best among the three models used with r2= 0.82, RMSE = 0.62, MSE = 0.38 and NSE = 81.77. Decision tree models, with r2 around 0.68, has less accurate results than other models. Cross validation seems to not increase accuracy in prediction for the decision tree model during the training stage. The Bagged Tree model has an r2 value of 0.7, RMSE value of 0.85, MSE value of 0.76 and NSE value of 63.34. From the observed GWL versus predicted GWL graph of Bagged tree, it can be observed that around half of the values are over estimated and half of the values are underestimated. Smaller GWL values are over estimated and larger GWL are under estimated

by the Bagged Tree model. The mean value of GWL for the testing stage is 3.94 m with maximal and minimal values of GWL for the testing stage being 8.83 m and 1.22 m, respectively. It can be observed that for the testing stage, from the graphs and the performance metrics, the Random Forest model is best among the three models used with r2= 0.80, RMSE = 0.61, MSE = 0.37 and NSE = 94.53. For decision tree, the cross validated models have shown an increase in accuracy. The 10-fold cross validated decision tree model has an r2 value of 0.70, RMSE value of 0.72, MSE value of 0.53 and NSE value of 73.70. The performance of Decision tree and Bagged tree models increased in the testing stage from the training stage. But the performance of Random Forest decreased slightly in the testing stage. Similar to the training stage values, the Bagged tree has overestimated smaller values of GWL and under estimated larger values of GWL. As in the training stage, here too the ensemble models have better performance than Decision tree.

In this study, multiple models were built for groundwater level prediction and compared to explore potential knowledge of GWL predictions. Forecasting of GWL is essential for sustainable management of ground water resources. GWL datasets from 2003 -2020 from the Talcher region, Odisha were collected and used to train and test three ML models for long-term prediction. Rainfall, temperature, calculated infiltration and calculated evapotranspiration values were used for the prediction of GWL. The Evapotranspiration values were calculated using the FAO-24 Radiation equation. Kalman filter was used to downscale GWL data. The performance of three machine learning models - Decision Tree, Random Forest and Bagged Tree in predicting GWL was evaluated. Each model is evaluated using the mean square error (MSE), root mean square error (RMSE), coefficient of determination (R^2), and Nash–Sutcliffe efficiency (NSE). The Random Forest model is seen to have best accuracy in both the training (r^2 = 0.82) and testing stages (r^2 = 0.80).

Keywords: *Decision tree, groundwater level, random forest, bagged trees, water resources management*

PARTICIPATORY GIS ANALYSIS AND MAPPING OF DRINKING WATER QUALITY CHALLENGES IN TANZANIA

Julian Ijumulana[1,3*], Joseph Mtamba[2], Felix Mtalo[2] and Prosun Bhattacharya[1]

[1]*KTH-Royal Institute of Technology, International Groundwater Arsenic Research Group, Department of Sustainable Development, Environmental Science and Engineering, Teknikringen, Sweden*

[2]*Department of Transportation and Geotechnical Engineering, College of Engineering and Technology, University of Dat es Salaam, Tanzania*

[3]*Department of Water Resources Engineering, College of Engineering and Technology, University of Dar es Salaam, Tanzania*

[*]*Corresponding author e-mail: julian@udsm.ac.tz; prosun@kth.se*

Increasing contamination and scarcity of critical water sources in arid and semi-arid regions of the globe are two challenges compromising the efforts to provide clean and safe water services for various purposes including drinking. This study investigated the practice of clean and safe water provision in Tanzania between 1960 and 2013. The study used community perspectives on the drinking water quality of critical water sources (around 65,000) captured during the national water point mapping in 2013. The critical water sources included both groundwater (Hand DTW, Machine DBH, SW, and Spring) and surface water (Dam, Lake, and River) resources. The methods used throughout the study combined classical statistics and Geographic Information Systems (GIS) tools to analyze trends in water supply services and critical water sources' quality in Tanzania. For the past five decades, groundwater resources contributed around 77% of the water demand while surface water resources contributed around 23%. However, the water users reported quality problems experienced during drinking and included, in ascending order, fluoride, milky, colour and salinity (salty). The salinity problem was reported across the country with Dar es Salam, Mara, Manyara, Dodoma, and Rukwa regions being the top five regions with severely affected critical water sources. In descending order, the most contaminated critical water sources in these regions were machine dug-boreholes (Machine DBH), shallow wells, and hand-dug tube wells (Hand DTW). Likewise, the colour and milky problems were reported across the country. The top five regions with severely affected critical water sources were Ruvuma, Mbeya, Kagera, Kigoma and Morogoro for colour problems and Shinyanga, Kigoma, Kagera, Lindi and Kilimanjaro for milky problems. The two drinking water problems were reported in all critical water sources except lakes. On the other hand, the fluoride problem was reported in fourteen (14) regions with the top five being, in descending order, Arusha, Mbeya, Manyara, Kilimanjaro, and Morogoro which are connected to the East African Rift Valley. Other regions where fluoride is a water quality problem included Kagera, Simiyu, Mtwara, Dar es Salaam, Mwanza, Mara, Pwani and Kigoma. The most affected critical water sources in these regions included rivers, hand-dug tube wells, machine-dug boreholes, shallow wells, and springs. The findings of this study highlight the quality of the critical water sources supporting the drinking water supply schemes across the country. The study further demonstrates the importance of participatory GIS in water supply services and management as it uses the feedback of the water users to map the water quality challenges at various critical water sources.

Keywords: *Critical water sources, water quality, exploratory data analysis, geospatial technology, GIS*

COMPARATIVE EVALUATION OF IMD AND NASA DATASETS FOR HYDROLOGICAL MODELLING

Shyam Sundar Bhardwaj and Madan Kumar Jha*

Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, India
**Corresponding author e-mail: madan@agfe.iitkgp.ac.in*

Water is a critical resource with social, economic, and environmental values essential for sustainable development. Intricate hydrological processes characterize the watersheds in the Himalayan Glacier System (also referred to as the “Water Towers” of Asia), and the significance of glacial meltwater in the overall water budget of mountain rivers is not well-established. Streamflow modeling in snow-dominated catchments is of utmost importance as it helps to manage water resources, predict and mitigate floods, optimize hydropower generation, preserve ecosystems, monitor water quality, adapt to climate change, plan infrastructure, support agriculture, and facilitate recreational and tourism activities. SWAT is a physically based, semi-distributed hydrological model and a potential tool for rainfall-runoff modeling at a basin scale as well as for estimating how climate change will affect water supplies, irrigation management, estimation of evapotranspiration, sediment analysis, snowfall, and snowmelt estimation, integrated surface water groundwater modeling. The objective of this study was to compare the performance of the SWAT model for streamflow simulation using NASA and IMD datasets in the hilly catchment of Uttarakhand, India.

The Bhagirathi River Basin, situated in the Himalayan region of India was selected as a case study to evaluate the efficacy of IMD and NASA datasets for basin-scale hydrological modeling. This basin lies in the Uttarakhand state and covers an area of 4663.81 km^2. The DEM data with a resolution of 30 m × 30 m was sourced from USGS Earth Explorer, while soil maps at a 1:5000 k scale were obtained from the FAO DSMW database. Land Use/Land Cover (LULC) data, with a resolution of 100 m × 100 m, was acquired from the Copernicus Global Land Service. Climate data, including precipitation, wind speed, relative humidity, solar radiation, and temperature for the period 2000–2020, were retrieved from NASA's POWER Data Access Viewer. Daily streamflow data (2006–2014) were collected from the Central Water Commission (CWC), Uttarakhand. Additionally, precipitation and temperature data for the study area were extracted from the IMD website. These datasets provided the inputs for setting up the SWAT model and analyzing its performance for hydrological modeling. SWAT simulates the hydrologic cycle based on the water balance equation. Based on the literature review, 25 surface, subsurface, and snow parameters were selected, and pre-calibration sensitivity analysis was performed using the SUFI-2 algorithm to identify influential parameters. The ArcSWAT model was calibrated using daily streamflow data for the years 2006−2010 and 2011−2014 were used for validation at the watershed outlet.

The assessment of the SWAT model using IMD and NASA datasets indicates significant disparities in statistical metrics between the calibration (2006–2010) and validation (2011–2014) periods for daily and monthly streamflow simulations. The Nash-Sutcliffe Efficiency (NSE) for IMD data increases from 0.58 in the calibration phase to 0.69 in the validation phase for daily data and from 0.74 to 0.88 for monthly data. The Kling-Gupta Efficiency (KGE) values are consistent, measuring 0.78 and 0.76 for daily data and higher values of 0.87 and 0.88 for monthly data, reflecting superior model performance at the monthly scale. The Root Mean Square Error (RMSE) is much higher for daily simulations (87.11–87.39

m³/s) than for monthly simulations (62.62–47.06 m³/s), indicating the model's superior accuracy at aggregated temporal scales. Percent Bias (PBIAS) values show a slight rise from 2.82% to 7.49% (daily) and from 2.66% to 7.74% (monthly) during the calibration and validation period, respectively, indicating underestimation of streamflow based on IMD data. The correlation coefficient (r) increases from 0.79 to 0.84 for daily data and from 0.87 to 0.94 for monthly data, indicating enhanced concordance between observed and simulated values over broader scales. The model performance using NASA data is significantly superior, especially at the monthly scale. The NSE values surpass those of the IMD data, rising from 0.79 during calibration to 0.74 during validation for daily data and to 0.92 and 0.93, respectively, for monthly data. KGE values are consistently elevated, measuring 0.78 and 0.80 for daily data and 0.86 and 0.91 for monthly data, signifying strong model simulations utilizing NASA data. RMSE values are significantly lower than those of IMD data, ranging from 61.95 to 79.75 m³/s for daily data and 34.19 to 37.39 m³/s for monthly data, demonstrating superior prediction accuracy. PBIAS results are promising, with near-zero figures (0.96% and -5.58%) for daily data and 0.95 and -5.56% for monthly data during calibration and validation, respectively, indicating slight bias. The correlation coefficient (r) demonstrates robustness, increasing from 0.89 to 0.87 for daily data and from 0.96 to 0.97 for monthly data, indicating very good concordance. NASA data outperforms IMD data across most statistical metrics, particularly in terms of lower RMSE and higher NSE, KGE, and r values. Both datasets showed reliable performance, with better predictive accuracy at the monthly scale. Our results will promote the use of gridded data for simulating hydrological processes in ungauged catchments or catchments with very few gauging stations, for instance, hilly catchments. Combining insights from multiple datasets could potentially reduce uncertainties and improve the robustness of hydrological modeling in data-scarce regions. Future research should focus on integrating multiple datasets to address discrepancies and enhance the reliability of water balance predictions for sustainable water and land resource management.

Keywords: *Bhagirathi River Basin, SWAT, hilly catchment, streamflow simulation, climate change, water balance components*

ET-BASED IRRIGATION PERFORMANCE ASSESSMENT OF A COMMAND AREA USING GOOGLE EARTH ENGINE

Harsh Upadhyay[1*], Sukant Jain[1] and Silvan Ragettli[2]

[1]*National Institute of Hydrology, Roorkee, India*
[2]*Hydrosolutions Ltd, Zurich, Switzerland*
[*]*Corresponding author e-mail: harshup.nihr@gov.in*

Effective utilization of water resources and ensuring food security have become pressing challenges due to increasing water demand and the simultaneous threats to water scarcity posed by climate change, pollution and overexploitation. Agriculture, as the largest consumer of global freshwater resources, plays a key role in this equation, with irrigation efficiency and water productivity emerging as the critical factors for optimizing water use and increasing crop yields. Multiple international organizations have adopted sustainable approaches with the aim of enhancing crop water productivity and have poured substantial amounts of investment into such ventures. With the advancements in remote sensing technologies and advent of various geospatial platforms, large-scale, rapid and cost-effective analyses of irrigation schemes have become a possibility. The primary determinants of water use efficiency include evapotranspiration (ET), which signifies crop water consumption (input) and the crop yield (output). A variety of earth observation datasets provide ET and biomass products at a global scale with varying spatial and temporal resolutions for assessment and incorporation into further analyses. However, these datasets only showcase the variability in water consumption and biomass generation and do not translate directly into any indicators of irrigation performance of a scheme.

In this study, eight indicators: 1) Cropping intensity, 2) Scheme utilization, 3) Less water-intensive crops, 4) Adequacy, 5) Equitable delivery, 6) Crop water productivity, 7) Reliability and, 8) Yield gap due to water stress were formulated to quantify the performance of an irrigation scheme incorporating data derived from remote sensing products. Each of these indicators are also associated with a set of critical and target values which aid in determining whether the performance based on a singular indicator is acceptable or not. The entire analysis was carried out using a tool developed in Google Earth Engine platform by the World Bank under the umbrella of Central Water Commission. The tool essentially consisted of two components viz., Cropmapper and Irrigation Performance Assessment indicators sheet. The Cropmapper was used to create a web-based application for the purpose of visualization of various geospatial layers such as the crops grown in the study area along with their seasonal distribution, ET-green, ET-blue, biomass generated and vegetation index map. The irrigation performance indicators sheet forms the crux of the tool output and contains the estimates of the aforementioned indicators as well as the relevant supporting information pertaining to the estimation such as season-wise and crop-wise rainfall, ET-green, ET-blue, volume of irrigation water requirement and total biomass generated. In order to evaluate the performance of irrigation schemes, the tool employs a score-based criterion which requires determination of scores for each indicator as per their individual performance computations and assignment of certain weights to each indicator based on their importance and influence upon the irrigation performance. Thus, the overall score represents a weighted score based on the performance scores of individual indicators. The major input datasets for the tool include the shapefile for the command area boundary, map for onset dates of monsoon from IMD, remote sensing layers viz., land cover maps from Sentinel-1 and Sentinel-2, actual

evapotranspiration and total biomass production maps from FAO as well as CHIRPS rainfall data. The overall methodology of the tool operation in Google Earth Engine can be sectioned into five parts – 1.) Data preprocessing, 2.) Crop type mapping, 3.) Irrigated area mapping, 4.) Creation of web application and, 5.) Generation of irrigation performance assessment indicators sheet, with each of these parts consisting of multiple subsequent steps.

In this study, the irrigation performance of a scheme implemented in command area of Ong catchment namely, Ong Diversion Weir Irrigation Project, was evaluated. All the remote sensing layers required as input for this study were acquired through the earth engine data catalog. The land cover layer was subjected to unsupervised classification to segregate the land cover into 10 classes and each of these classes were manually provided with labels belonging to a specific crop or crop rotation based on the agricultural statistics, cropping calendar and vegetation index information of the area. The tool also supports supervised classification using geo-tagged in-situ crop type samples obtained from field survey which can be digitized into the earth engine database as reference points. Separate crop maps were generated for the years from 2018 to 2022 using transfer learning, a deep learning technique, on the basis of an input labelled crop map for a single year, i.e., 2022. The irrigation performance of the canal for the entire agricultural year (June-May) of 2022 was evaluated based on the indicator sheet generated. The cropping intensity was found to be 87% which was less than the critical value thus, the indicator score was only 1. The irrigation utilization was 49% which was in between its critical and target values of 25% and 80%, respectively, thus, earning an indicator score of 5. Paddy was the major crop in the area which is a high water-intensive crop. The fraction of low water-intensive crops was only 37% resulting in an indicator score of 4. The adequacy and equity of the scheme were computed to be 0.2 and 0.04, respectively, both being within their critical and target values, with scores of 6 and 9, respectively. The water productivity for the paddy crop was found to be 1.1 kg/m^3 which was well beyond the target value of 1 kg/m^3 thus, earning a score of 10. The reliability and yield gap due to water stress were calculated to be 27% and 24% with scores of 4 and 5, respectively. The overall weighted average score for the irrigation scheme was 5.7 which was relatively low. The results of this irrigation performance assessment showcased that the irrigation scheme was not able to achieve optimal water use conditions due to discrepancies in some of the indicators especially cropping intensity. The indicators with low scores can act as waypoints to pinpoint the lacunae in irrigation management and may offer insights to address the detected inefficiencies. These findings highlight the potential of the tool to assess irrigation performance at larger scale and finer resolutions.

***Keywords**: Google Earth Engine, irrigation performance assessment, irrigation efficiency, ET-green, ET-blue*

APPLICATION OF MACHINE LEARNING TECHNIQUES FOR THE PREDICTION OF GROUNDWATER LEVELS FOR NASHIK DISTRICT IN MAHARASHTRA, INDIA

Yuvraj Nanasaheb Dhivar[1]*, Shiv Kumar[2] and Madan Kumar Jha[1]

[1]Indian Institute of Technology, Kharagpur, West Bengal, India
[2]G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India
[*]Corresponding author e-mail: yuvrajdhivar97@gmail.com

A severe groundwater crisis currently exists in India due to over-extraction and contamination of groundwater, which covers nearly 60% of all the districts in India and poses a threat to the drinking water security of the human population. Despite the national scenario that the availability of groundwater is favorable, there are pockets in certain areas of the country that face water scarcity because of non-uniform poor groundwater development, and over-exploitation leading to falling water levels. Therefore, understanding the response of groundwater levels (depths to water table) to climatic variables and pumping is necessary for sustainable groundwater management, specifically in meeting increasing groundwater demands in the agriculture sector. Since groundwater is a hidden water source with high spatio-temporal variability, quantifying its availability and the long-term impact of climate patterns on it is a highly complex task. To model the groundwater response to climate drivers, water demand, and surface water hydrology, it is necessary to find the relationship between them. There are conceptual and process-based modeling techniques to simulate the complex groundwater process, which requires a large number of hydrogeological data. Acquiring such enormous data is difficult and expensive. Consequently, it is desirable to explore data-driven and machine-learning techniques built on nonlinear interdependencies that could predict fluctuations in groundwater levels without an extensive understanding of the fundamental physical parameters. Among the several machine-learning algorithms available, artificial neural network (ANN) and support vector machine (SVM) were employed in this study. Nashik district of Maharashtra state was selected for the study, as groundwater has special significance for agricultural development, and groundwater development in some parts of the district has reached a critical stage.

In the study, components of groundwater recharge and groundwater discharge were estimated using input parameters such as recharge due to rainfall, recharge due to return flow of irrigation, recharge due to seepage from canals, draft due to irrigation use, industrial use, domestic use, and livestock use. Depths to water table for the period of 20 years, from 1998 to 2018, were used to develop four groundwater models using different combination sets of input variables. The data of 181 hydrograph stations were collected from GSDA, Nashik. Data of rainfall, canals, human population, livestock population, tanks and ponds, minor irrigation structures, and pumping wells were collected from the statistical department of the Nashik district.

While two models used annual data, the other two models used seasonal data as input variables to predict the pre-monsoon and post-monsoon depths of the water table. Several ANN structures were employed using the input variables of the models and different hyper-parameters were tuned in the case of SVM. The best combination of neurons/layers and activation function in the case of ANN and hyper-parameters in the case of SVM were selected based on the hit and trial method. Performance metrics such as correlation coefficient (r), coefficient of determination (R^2), Nash-Sutcliff efficiency, mean absolute error, root mean square error, mean absolute percentage error, root mean square percentage error, and relative absolute error were used to compare the performance of the models. Based on the global ranking, out of four ANN and four SVM models, the best models were selected for the prediction of pre-post-monsoon depths to the water table. The performance of the ANN and SVM models were also compared.

Based on the global ranking, it was observed that that out of four models, annual data-based models performed better than seasonal data-based models for both ANN and SVM algorithms. Hence, only annual data-based models of ANN and SVM were selected for the comparison. Out of 181 nodes (hydrograph stations), 63 nodes were selected for 5 years; hence, 63×5 = 315 predicted depths to water table for both models were compared. On comparing the results, it was found that 167 (53.01%) predicted pre-monsoon depths were within 10 percent deviation for ANN models while it was 172 (54.60%) for the SVM model, which indicated that the SVM model performed better than the ANN to predict pre-monsoon depth. On the other hand, the ANN model with 68 depths (21.58%) out of 315 predicted depths performed better than the SVM model with 52 values (16.50%) in forecasting post-monsoon depths to the water table.

The study concludes that both the models, i.e., artificial neural network (ANN) and support vector machine (SVM), perform well in predicting pre- and post-depths to the water table. On comparing the performance of both the models, the SVM model performs better in predicting the pre-monsoon depths to the water table, whereas the ANN model performs well in predicting the post-monsoon depths to the water table. The predicted depths of water table from both ANN and SVM models follow the trend of the observed water table precisely and accurately in both seasons. The results reveal that the performance of annual data-based models is better than the performance of the seasonal data-based models, which implies that with the increase in number of input variables the performance of the models improves.

Keywords: *Groundwater level prediction, depth to water table, Nashik district, seasonal depth of groundwater, machine learning, artificial neural network, support vector machine*

A HYBRID APPROACH USING DEEP LEARNING AND HYDROLOGICAL MODELING TO IDENTIFY AQUIFER STRESS ZONES IN THE BARAKAR RIVER BASIN, JHARKHAND

Amit Bera[1*], Sanjit Kumar Pal[1] and Bhabani Prasad Mukhopadhyay[2]

[1]*Department of Applied Geophysics, Indian Institute of Technology (Indian School of Mines) Dhanbad, Dhanbad, Jharkhand, India*

[2]*Department of Earth Sciences, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India*

[*]*Corresponding author e-mail: amitbera12312@yahoo.com*

Aquifer stress assessment in semiarid hard rock terrains, such as the Barakar River Basin in Jharkhand, is crucial for managing the region's groundwater resources. In these areas, aquifers are often characterized by low porosity and limited recharge capacity, making them particularly vulnerable to over-extraction and climate variability. The harsh climatic conditions, coupled with irregular rainfall patterns, further exacerbate the risk of groundwater depletion. Therefore, a comprehensive assessment of aquifer stress, including factors like groundwater recharge, lateral flow, and base flow, is essential to monitor aquifer health and ensure sustainable water management. Understanding aquifer stress in these hard rock terrains is vital for developing strategies to mitigate water scarcity and support agricultural and domestic water needs in the region. This study presents a novel hybrid approach that integrates deep learning techniques, specifically Convolutional Neural Networks (CNN), with hydrological modeling using the Soil and Water Assessment Tool (SWAT) to assess and identify aquifer stress zones in the Barakar River Basin. The research aims to enhance the accuracy and efficiency of aquifer stress assessments by employing advanced machine learning models in conjunction with traditional hydrological models. This area remains underexplored in aquifer health studies.

The methodology of this study combines data-driven deep learning techniques with the SWAT model, leveraging 16 influential factors that encompass hydrological, physiographical, and socioeconomic variables. These factors include precipitation, land use, soil texture, slope, groundwater recharge, water extraction rates, and socioeconomic activities, collectively contributing to the aquifer system's stress. The deep learning component, based on CNNs, classifies the aquifer stress into four distinct categories: Minimal Stress, Moderate Stress, High Stress, and Critical Stress. These classifications are crucial for understanding the severity of the aquifer's stress and informing future management strategies. The integration of these datasets was critical in providing a holistic view of the factors contributing to aquifer stress.

The results of the study showed that approximately 32% of the Barakar River Basin was categorized as Minimal Stress, indicating areas where groundwater levels are stable, and recharge rates are adequate. On the other hand, Critical Stress zones, which accounted for 24% of the basin, were identified in areas with severe depletion of groundwater resources, where extraction rates exceed natural recharge, leading to significant stress on the aquifer system. These findings are consistent with observed groundwater depletion trends and highlight certain areas' vulnerability to long-term water scarcity. Integrating hydrological modeling through the SWAT model provided valuable insights into the hydrological budget components of the Barakar River Basin. By coupling the hydrological model's predictions of

runoff and recharge potential with the CNN's ability to analyze spatial patterns in the data, the study offers a comprehensive and integrated framework for aquifer stress assessment. The use of SWAT enabled a deeper understanding of how climatic and physiographic factors influence groundwater availability. In contrast, the deep learning approach captured the complex spatial relationships that influence stress levels on the aquifer system. By providing an accurate and efficient method for aquifer stress assessment, the approach can be used by policymakers and water resource managers to identify vulnerable areas and prioritize intervention strategies. The integration of deep learning with traditional hydrological models represents a promising advancement in the field of aquifer stress assessment. It offers a scalable and adaptable solution for assessing aquifer stress in regions with limited data availability. Model validation demonstrated the robustness of the CNN model, with a precision score of 0.92, an AUC-ROC score of 0.91, and an F1 score of 0.84. This indicates that CNNs, with their ability to model complex, non-linear relationships in data, are well-suited for the task of aquifer stress classification, where the interplay of multiple environmental and socio-economic factors creates a highly intricate system. In conclusion, this research demonstrates the potential of hybrid deep learning and hydrological modeling for aquifer stress zone identification, providing a powerful tool for sustainable groundwater management. The successful application of CNNs in conjunction with the SWAT model in the Barakar River Basin serves as a benchmark for future studies and offers valuable insights for managing aquifer stress in hard rock terrains. This integrated approach improves the accuracy of aquifer stress assessments and provides a framework that can be adapted to other regions facing similar groundwater challenges. The study ultimately contributes to the global efforts in sustainable aquifer management, offering a pathway toward better resource allocation, conservation, and long-term water security.

Keywords: *Hybrid approach, SWAT, deep learning, hydrological modeling, aquifer stress*

CLEANER TECHNOLOGY FOR REDUCING FRESH WATER CONSUMPTION IN THE ELECTROPLATING SECTOR OF MORADABAD COUNTER CURRENT MECHANISM (CCM)

Siddhartha Baidya[1], Umesh Chandra[1], Suresh Babu SV[1], Vinod Tare[2,3] and Nitin Kaushal[1]

[1]*WWF, India, New Delhi, India*
[2]*Indian Institute of Technology, Kanpur, Uttar Pradesh, India*
[3]*Centre for Ganga River Basin and Studies (cGanga), IIT Kanpur, Uttar Pradesh, India*
[*]*Corresponding author e-mail: sbaidya@wwfindia.net*

The ecological integrity of rivers and sustainable use of water resources are critical for social and economic progress of human society. The continuous compounding of stress on groundwater is contributed by the agriculture, domestic and industrial sectors. The metalware cluster of Moradabad city is known across the world for brass art metalware and has carved a niche for itself in the handicraft industry. The city of Moradabad is the epicenter of metalware artwork and is popularly known as the brass city. There are about 1200 registered metalware industries in the city and thousands of unregistered household units. The metalware cluster is entirely dependent on groundwater. The status of groundwater table of Moradabad has a critical ecological context because the 'River Ramganga' is an aquifer-fed river in this region, therefore the flows in the river directly depend on the status of groundwater in the district. Technological improvements have been a key in the success story of Moradabad's metalware sector. However, the consequence of this technological progress or the gaps in the current technologies on the environment in general and water use in particular has received relatively less attention. The metalware industry uses large volume of groundwater with very limited measures to reduce water use across industrial processes. The consumption of fresh water varies remarkably across big export units and household-level units. Fresh water is mainly consumed either during the rinsing of electroplated articles or after Ultrasonic Cleaning of articles before rinsing. Rinsing in the majority of units is still an inefficient system, which includes continuously and simultaneously operating multiple fresh water taps thus leading to extraordinary high use of fresh water. It is estimated that about seventy percent of the total freshwater used by the metalware industry is consumed in the rinsing process.

This paper documents the impact of WWF-India's interventions from 2013 to 2024 around collective action industries, on the mainstreaming of cleaner technologies in Moradabad's metalware cluster. WWF-India in collaboration with Indian Institute of Technology (IIT), Kanpur initiated exploring cleaner technology interventions in 2013-14, which primarily consisted of investigating and field testing a technical solution to reduce fresh water use during the rinse process. However, it is critical to state that this journey of exploring a technological solution was weaved around the concept of '*multistakeholder approach*'. WWF-India and IIT-K methodically studied the uptake of technology by the metalware sector and later developed a technology solution to reduce fresh water consumption in the rinsing process. This partnership led to the development of a new manual rinsing mechanism known as the Counter Current Mechanism (CCM). This rinsing system essentially comprised of a composite iron tank, which had multiple partitions and discontinuation of all freshwater taps except one. This manual CCM was demonstrated in nine pilot industries during 2014-2016 and the impact of the manual CCM was studied in partnership with industries and academia. The stakeholders actively worked on improving the design of the manual CCM and by 2017-18 a new design of the CCM was developed, which made the CCM as a completely automatic mechanism. An automatic CCM comprised of a composite Poly Propylene (PP) tank with three partitions and holes drilled on walls separating each chamber, which allowed water to flow from one partition to the other. A TDS controller mechanism is incorporated in the CCM to disconnect the flow of fresh water whenever the TDS is under the threshold (a pre-defined benchmark value). The consumption of water during rinsing at this particular CCM is measured by installing a dedicated water flow meter. The CCM is supplied with fresh water from one tap only unlike from two to three taps in the conventional rinsing system. There will be no supply of fresh water from the tap whenever until the TDS value in the most polluted chamber is below the threshold TDS as this would imply that the TDS of water is within range and fit for rinsing. However, the moment the TDS value of the most polluted chamber goes above the threshold TDS value the solenoid valve would be activated and that will lead to supply for fresh water from the tap. The supply of fresh water from the tap will stop the moment the TDS value in the polluted chamber falls below the threshold TDS value due to dilution.

The collective action by WWF-India and other stakeholders has led to the adoption of 162 automatic CCMs by forty industries during 2010 and till now. The cumulative share of these forty industries in the total export of metalware products is estimated to be in the range of 40 to 57 percent. Therefore, these forty industries are a good sample to capture the economic brevity of Moradabad’s metalware cluster. The impact of automatic CCM on reduction of freshwater consumption in rinsing was studied by recording water consumption and number of articles rinsed in the CCM being studied on a daily-basis. A similar dataset was collected during the baseline study, i.e. before converting or replacing the traditional rinsing facility into an automatic CCM. It was estimated based on the data set that the average reduction in freshwater consumption was in the range of twenty-five percent to forty-seven percent after installation of automatic CCM. This sharp reduction in freshwater use further reduced the running cost of the Effluent Treatment Plants (ETP) thus enabling the industry to acquire double gain. The study based on stratified random sampling also estimated a total reduction of about 22 million liters from June 2020 to June 2024. The assessment methodology and the dataset used were also subjected to a creditable third-party auditor and the audit report validated WWF-India’s methodology and its accuracy. This study has also assisted the district administration in encouraging the cluster for wider adoption of automatic CCM, which subsequently followed the formation of a working group and a letter issued by the district administration to representatives of the metalware sector. The stakeholders involved in the initiative equivocally recognize the necessary policy push that is needed to scale up the adoption of automatic CCM by the positive consequence of this technological solution.

Keywords: *Cleaner technologies, metalware sector, counter current mechanism, automatic, freshwater reduction*

DEVELOPMENT OF A SEDIMENTATION INFORMATION SYSTEM FOR RESERVOIRS IN THE KANDI REGION, PUNJAB, INDIA

Mahesh Chand Singh[1*], Jaswinder Singh[2] and Vishnu Prasad[3]

[1,2]*Department of Soil and Water Engineering, Punjab Agricultural University, Ludhiana, Punjab, India*
[3]*Water Technology Centre, ICAR-Indian Agricultural Research Institute, New Delhi, India*
[*]*Corresponding author e-mail: msrawat@pau.edu*

Reservoir sedimentation is a persistent issue in water resource management, particularly in regions with sediment-prone catchments such as the *Kandi* region of Punjab, India. To address the challenges posed by sediment accumulation and capacity loss, this study presents the development of a Reservoir Sedimentation Information System specifically tailored for reservoirs in the *Kandi* region, with a focus on the Chohal and Damsal reservoirs. Leveraging the capabilities of remote sensing and GIS, this research successfully estimates key metrics including water spread area, reservoir capacity, and sedimentation rates using data from Landsat 7&8 and Sentinel-2 satellite. The developed system integrates these metrics into a user-friendly platform, providing a powerful tool for policymakers, water resource managers, and local stakeholders to make informed decisions regarding reservoir management and sedimentation control. The primary objective of this study was to develop a reservoir sedimentation monitoring system that could offer insights into the sedimentation patterns of the key reservoirs in the *Kandi* region. This study captured sedimentation dynamics and their impact on water spread area and capacity. The results reveal that Chohal and Damsal reservoirs display a positive correlation between water spread area and increasing elevation, underscoring the precision of satellite-based monitoring in assessing reservoir capacity changes over time. In particular, the study's dual use of Landsat and Sentinel datasets allowed for a comparative analysis, highlighting the improved accuracy achieved with higher-resolution Sentinel-2 imagery in estimating reservoir metrics. For the Chohal Dam reservoir, analysis of Landsat data reveals a capacity loss of approximately 30.2% over a 36-year period, with an annual sedimentation rate of 0.84%. Comparatively, the Sentinel-2 data indicate a reduced capacity loss of about 25.2% over the same timeframe, corresponding to an annual sedimentation rate of 0.70%. These results suggest that the higher spatial resolution available from Sentinel-2 yields more accurate capacity estimates and highlights the variations in sedimentation rates based on the data source. Similarly, for the Damsal Dam reservoir, Landsat imagery indicates a capacity reduction of 33.6% over 34 years, with an annual sedimentation rate of 0.99%. Sentinel-2 data, on the other hand, show a capacity loss of approximately 27.8%, with an annual sedimentation rate of 0.82%, again illustrating the precision gained with higher-resolution satellite imagery. These findings validate the use of remote sensing in monitoring long-term changes in reservoir capacity and sedimentation rates, which are crucial for effective reservoir management. The development of the information system represents a significant advancement in reservoir sedimentation analysis by integrating spatial datasets, elevation-area-capacity curves, and sedimentation rates into a single GIS-based platform. This system was designed to offer an intuitive interface and interactive features that empower stakeholders to engage directly with the data. A key feature of this system is the development of volume calculators for both reservoirs. These calculators are built to utilize elevation-area-capacity relationships derived from satellite data and historical records. By incorporating both the original and updated elevation-area-capacity curves, the developed system enables users to observe capacity trends over time, enhancing

their ability to project future sedimentation impacts and capacity reductions. Beyond the capacity calculator, this system consolidates additional data on soil characteristics, water quality, and sediment composition, offering a comprehensive view of reservoir health. Soil characteristics within the catchment areas, for instance, provide insights into erosion sources and sediment quality, informing conservation strategies. Water quality data allows stakeholders to evaluate the ecological status of the reservoirs and assess potential impacts of sedimentation on water quality for downstream uses including agricultural irrigation. By encompassing these variables, the system supports an integrative approach to reservoir management. One of the key achievements is demonstrating the value of such systems for regions with limited access to ground-based survey data. In the *Kandi* region of Punjab, physical monitoring of reservoirs can be challenging due to rugged terrain and resource constraints. This system offers a practical, efficient alternative by capitalizing on remote sensing technology to deliver timely updates on reservoir status, allowing for proactive sedimentation management. Furthermore, by automating data analysis and providing user-friendly outputs, the system reduces reliance on specialized technical skills, enabling local officials, resource managers, and other stakeholders to interpret and utilize the data effectively. The outcomes of this study underscore the critical role of ongoing monitoring and adaptive management practices to ensure the resilience of water resources in sediment-prone areas. Reservoir sedimentation, if left unmanaged, can severely diminish storage capacity, affecting water availability for agriculture, drinking water, and industrial uses. By quantifying capacity losses and sedimentation rates, this system provides stakeholders with the knowledge needed to implement timely interventions, such as desilting operations, catchment area conservation, and sustainable water use strategies. Additionally, the system's ability to project future sedimentation trends facilitates long-term planning, helping policymakers make informed decisions to secure water resources for future generations.

This study establishes the developed system as a vital tool for the sustainable management of the Chohal and Damsal reservoirs in the *Kandi* region of Punjab. By incorporating Landsat and Sentinel-2 satellite data, the system provides accurate, real-time insights into sedimentation rates, reservoir capacity, and other essential metrics. This GIS-based platform not only improves the accuracy of sedimentation assessments but also offers a comprehensive decision-support system that integrates multiple dimensions of reservoir health. The findings highlight the significant potential of remote sensing and GIS in enhancing reservoir management practices, particularly in sediment-prone regions like the *Kandi* area, where traditional monitoring approaches are often impractical. This system also underscores the importance of satellite-based methodologies in adaptive water management and ecological conservation. As sedimentation continues to impact water bodies globally, such systems offer a scalable, replicable solution for monitoring and managing reservoir sedimentation in real time. By enabling proactive reservoir management and promoting sustainable resource utilization, the developed system aligns with broader goals of water security and environmental resilience in Punjab and similar regions facing sedimentation challenges. This study demonstrates the transformative potential of combining GIS and remote sensing to address complex environmental challenges and provides a foundation for future innovations in reservoir management systems.

Keywords: *Reservoir sedimentation, GIS, remote sensing, reservoir capacity, Kandi region*

ADVANCING GROUNDWATER LEVEL PREDICTION USING PHYSICS-INFORMED NEURAL NETWORKS AND MACHINE LEARNING FOR TRANSIENT HEAD VARIATION

Shreyansh Mishra, Mayank Bajpai*, Adwitiya Roy, Shishir Gaur and Anurag Ohri

Indian Institute of Technology, Banaras Hindu University, Varanasi, India
**Corresponding author e-mail: mayankbajpai.rs.civ22@iitbhu.ac.in*

Groundwater (GW) level prediction is pivotal in sustainable water resource management, especially in regions vulnerable to water scarcity. Traditional prediction methods often struggle with the complexities of spatial and temporal variations in GW systems, necessitating advanced machine learning (ML) approaches. In this study, we investigated the effectiveness of the Kolmogorov-Arnold Network (KAN) in predicting GW levels and compare its performance with a standard Artificial Neural Network (ANN). KAN, known for its ability to explain, alongwith robust feature representation, integrates domain knowledge, potentially enhancing model interpretability and prediction accuracy. We assessed each model's accuracy and ability to capture GW fluctuations over time using a dataset generated from a hypothetical MODFLOW model simulating transient groundwater dynamics under varied conditions. Our findings suggest that KAN offers advantages over ANN regarding precision and insight, making it a promising tool for improved, transparent groundwater management.

In this study, we developed a GW model of the Ain River basin in France using MODFLOW 2005 to simulate spatial and temporal head variations. By repeatedly running the MODFLOW model under various boundary conditions and pumping scenarios, we generated a comprehensive dataset to capture the dynamic behavior of GW heads in the basin. Specifically, the inputs were the spatial coordinates (X, Y), time (t), external flow (W), hydraulic conductivity (Hk), and specific yield (S). We applied two machine learning models for GW-level prediction: the Kolmogorov-Arnold Network (KAN) and a conventional Artificial Neural Network (ANN). The ANN, widely used for GW modelling due to its flexibility and data-driven approach, captures patterns in the dataset but often lacks interpretability, making it challenging to connect predictions to underlying physical processes.

In contrast, KAN employs a structure that explicitly integrates knowledge of input-output relationships, enhancing interpretability by reflecting the underlying interactions within the data. This feature makes KAN particularly suitable for GW prediction, where understanding complex interactions is crucial for effective management decisions. The performance of both models was evaluated in terms of prediction accuracy and interpretability, highlighting KAN's potential as a more insightful tool for sustainable GW management. For spatial coordinates (X and Y), sharp changes in spline values might indicate areas with higher hydraulic gradients or more permeable subsurface structures, while smoother trends suggest uniform groundwater conditions. In time (t), periodic patterns in the spline could reflect seasonal groundwater fluctuations, while gradual trends may indicate long-term changes like groundwater depletion. Sudden changes might show responses to short-term events like rainfall or pumping variations.

The spline for K reflects the sensitivity of the groundwater head to aquifer permeability. Steep gradients suggest dynamic responses in areas of high permeability, while flatter trends indicate more stable conditions. Finally, the spline trend for S reveals the influence of specific yield on groundwater head. Steep changes suggest high sensitivity, often seen in unconfined aquifers, while flatter trends imply less impact, typical of unconfined or transmissive areas. These spline patterns in KAN provide interpretable insights into aquifer dynamics, enhancing understanding of groundwater behavior in the Ain River basin and aiding in data-driven resource management.

The results indicate distinct performance differences between the Kolmogorov-Arnold Network (KAN) and the Artificial Neural Network (ANN) predicting groundwater head. KAN achieved a training RMSE of 181.1 and a test RMSE of 212.3, showing a stable performance across training and testing. This suggests that KAN's in-built spline activation function effectively captures complex relationships between input variables and groundwater head, enhancing generalization to unseen data. The spline functions likely provide a better interpretive framework, allowing the model to capture non-linear trends in the dataset with reduced overfitting. In contrast, the ANN exhibited a much lower training RMSE of 109.7 but a significantly higher test RMSE of 652.3, indicating overfitting. The sharp discrepancy suggests that the ReLU activation function, while effective in minimizing the training error, struggles to generalize in this application, especially with only three neurons in the hidden layer. The ANN's poor test performance implies limited interpretability and an inability to accurately predict groundwater head outside the training data.

In conclusion, the KAN demonstrates superior generalization and interpretability over traditional ANNs for groundwater head prediction, particularly in the complex hydrogeological setting of the Ain River basin. While the ANN achieved a lower training RMSE, its high test RMSE indicates overfitting and limited robustness in predicting unseen data. The KAN, equipped with a spline-based activation function, provided more consistent results across training and testing, highlighting its capability to capture intricate non-linear relationships in groundwater dynamics. This stability and interpretability make KAN a promising tool for environmental modeling, where understanding variable interactions is crucial. Thus, KAN's performance in this study underscores its potential for enhancing groundwater management strategies through accurate and interpretable predictive modeling.

Keywords: *Groundwater modelling, artificial neural network, machine learning, MODFLOW, transient groundwater flow*