Groundwater Level Prediction Using a Hybrid.LSTM–Prophet Model

Sagar Shandilya[1], K. Venkata Rao[2]

[1,2.] *Department of Computer Science & Systems Engineering Andhra University College of Engineering (AUCE) Andhra University, Visakhapatnam Andhra Pradesh, India*

***Abstract*—Although groundwater is crucial for residential, agricultural and industrial use but the water security is now under tremendous risk owing to over-extraction, climate change and changes in land use. Traditional groundwater level prediction methods, such as manual monitoring and empirical models, struggle with automation, scaling and capturing complex non-linear interaction between the influencing factors. To solve these problems, this research proposes a prediction framework based on machine learning (ML) that increases efficiency, accuracy and adaptability. The system integrates a range of information, including historical groundwater levels, rainfall, temperature, soil characteristics, land use and also handles missing data using imputation techniques. Long Short-Term Memory (LSTM) networks and Prophet are the examples of advanced ML models that are used to predict outcomes and capture complex patterns far better than the conventional statistical models. The robustness of the model is further improved by incorporating feature engineering and cross-validation. This suggested ML-driven solution performs better as it is automated, scalable across contexts and accurate in prediction. These ML-based solutions address the shortcomings of the pre-existing methods and offers a pragmatic sustainable groundwater management. Thereby, aiding in better planning for water resources as well as the environmental preservation.**

***Index Terms*—Groundwater Storage (GWS), Machine Learning (ML), LSTM (Long-Short Term Memory), Prophet, Soil Moisture, Terrestrial Water Storage (TWS), GRACE/GRACE-OF**

I. INTRODUCTION

Groundwater is a vital resource that sustains household, industrial and agricultural demands. [1], [2]. Groundwater provides 85% of the drinking water supply in the rural areas and more than 60% of irrigation supply in places like India [3]. The depletion of the groundwater due to irresponsible extraction, the shifting of land-use patterns as well as the climatic changes has led to a serious challenge in the environment sustainability and the water security [4],[5].

The pre-existing conventional methods that help predict the groundwater levels mostly rely on hydrological models, empirical calculations and manual observations [6], [7]. These methods may provide basic understanding, but their accuracy and their effectiveness however are frequently constrained. When it comes to dynamic settings where there are non-linear interactions between the affecting variables like the temperature, rain and the land use [8], [9], these models are not feasible for large-scale or real-time applications as they constantly face issues like inadequate data, low scalability and are known for their lack of adaptability [10], [11]. This study suggests that our ML-based predictive framework uses data-driven methods to increase the precision, automation and flexibility of the groundwater level prediction [12]-[14].

This technology goes through intricate patterns and produces precise short-term as well as long-term forecasts taking groundwater levels as well as the environmental factors into account [15],[16].In order to address the drawbacks of the current methods, we present a hybrid prediction model in this work. This model combines Prophet as well as Long Short-Term Memory (LSTM) networks [12], [14]. Prophet was created for a scalable time-series analysis and is very good at detecting recurrent hydrological cycles like monsoon-driven variations in groundwater as it breaks data down into trends, seasonality as well as residual components [12], [16]. Prophet on its own, however, is not able to adjust to transient abnormalities or non-linear aberrations [12], [14]. So, in addition, Prophet's residual errors are modeled using LSTM networks, which correct the discrepancies through the sequential learning [12], [16]. The system can identify trends in the recent residuals and is able to enhance accuracy in the dynamic conditions due to the LTSM's ability to capture temporal dependencies [15], [16].

Prophet's baseline predictions have LTSM-predicted residuals added to it, this helps in the creation of a hybrid forecast, the forecast then yields output that are locally refined as well as structurally coherent [14], [16]. The methodology is further supported and is guaranteed a solid quantitative assessment by the evaluation measurements like Mean Absolute Error (MAE), Root mean Squared Error (RMSE) and R2 [15], [17].

This hybrid design has useful benefits, i.e., it embeds the flexibility of LSTM and maintains Prophet's interpretability as well as its computational efficiency simultaneously [12], [16]. This deems the model's dependability for a longer-term forecasting, making it adaptable for utilization with a variety of hydrological parameters, such as Soil Moisture, Groundwater Storage (GWS) and Terrestrial Water Storage (TWS) [14], [16]. This suggested methodology provides a reliable tool for sustainable water management under artificial and climatic stresses by filling the methodological voids as well as enhancing the prediction resilience [1]-[17].

II. LITERATURE REVIEW

Groundwater Level (GWL) prediction is critical for water resource management especially in regions that face depletion due to the change in climate as well as the overexploitation of the resources. MODFLOW, one of the traditional empirical and physically based models struggles with nonlinear hydrological dynamics and it also requires extensive data deeming it not feasible for such regions. The ML as well as hybrid approaches have advanced significantly in terms of forecasting GWL, offering robust alternatives to the conventional methods over the past two decades. Due to their capability of representing intricate and non-linear interactions, the Artificial Neural Networks (ANNs) have been considered as a key component of GWL Prediction. Raj et al. [1] used the ANNs to estimate GWL in the coastal aquifers. By combining historical GWL, rainfall, and tide levels, they were able to produce reliable predictions; however, for longer lead times, support vector machines (SVMs) shown better generalization. In semi-arid environments, Nayak et al. [2] showed that ANNs are superior, capturing recharge-discharge patterns with R2 values as high as 0.91. Kim et al. [3] compared ANNs with multiple linear regression (MLR) across 17 Japanese sites, reporting better agreement (e.g., lower RMSE) for ANN-predicted GWL.

By combining historical GWL, rainfall, and tide levels, they were able to produce reliable predictions; however, for longer lead times, support vector machines (SVMs) shown better generalization. Nayak et al. [2] showed that ANNs are superior as they captured the recharge-discharge patterns with R2 values as high as 0.91, in the semi-arid environments. Several ANN architectures were investigated by Coulibaly et al. [5, 13] for monthly forecasting, and their adaptability to hydrometeorological inputs such as temperature and rainfall was confirmed. Although they noted issues with seasonality, Yoon et al. [8] showed that ANNs are effective in shallow systems. Coppola and associates. [6] and Uddameri [7] confirmed ANNs in unconfined aquifers under various pumping and climate conditions. ANN applications were examined by Maier and Dandy [14] and Elsherbiny et al. [15], who emphasized their adaptability while warning against overfitting and limited interpretability.

In terms of prediction accuracy, the Hybrid models outperform the standalone ANNs. By optimizing ANN training with GA, Sharma et al. [4] proposed an ANN-Genetic Algorithm (GA) model in the Mahanadi River basin, achieving higher performance (e.g., RMSE decrease). Similarly, wavelet-based hybrid models, such as Wavelet-ANN and Wavelet-ANFIS, enhanced forecasting by breaking down time series into elements that captured trends and seasonality. Wavelet-ANFIS [4] outperformed the standalone ANN by 8.8% in R2, especially in the forecasts made one to two months in advance. The shortcomings of ANNs in managing non-stationary data, as highlighted by Adamowski and Chan [9], are addressed by these hybrid techniques.

Alternative ML methods also have showed promising results. Shiri and Kisi [12] examined Support Vector Machines (SVM), ANNs, Gene Expression Programming (GEP) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) for the GWL forecasting in South Korea. Due to its capacity to create symbolic expressions, GEP fared better than the others, obtaining lower RMSE for predictions up to seven days. Yoon et al. [8]

discovered that SVMs seemed to have better generalization ability than the ANNs in coastal aquifers, especially for longer lead times.

Unconventional approaches, such as dendrochronology, offer novel insights. Watson and Luckman [16] and Ferguson and St. George [17] used tree-ring chronologies to reconstruct hydroclimatic variables, K. Elsherbiny et al. [15] integrated dendrochronology with ANNs to simulate GWL fluctuations in Iran from 1912 to 2013. This approach achieved high accuracy by leveraging tree-ring diameter as a proxy for precipitation and GWL, demonstrating potential for historical reconstructions in data-scarce regions. Multisite cascading calibration techniques [16] further refined hydrological models like VIC, reducing negative Nash-Sutcliffe efficiency values from 69% to 3% in the Red River Basin.

GWL prediction has been revolutionized by the Satellite Remote Sensing as it provides large-scale data. Missions such as NASA's GRACE and GRACE-FO [10] monitor the Terrestrial Water Storage, enabling the global assessment as the Global Land Data Assimilation System (GLDAS) [11] provides soil as well as other parameters. These datasets help ML models capture enhanced spatiotemporal variability [3, 8], where GRACE data helped improve the prediction accuracy.

The challenges in predictions still exist, despite all these advancements. The ANN as well as other ML models often struggle when it comes to long-term seasonality and abrupt shifts and their black box nature limits the interpretability [12, 15]. These issues are mitigated by the Hybrid models as well as the ensemble methods, but they require computational resources. To address these drawbacks, this study proposes a hybrid framework which integrates Long Short-Term Memory (LSTM) networks as well as Prophet. Prophet excels at decomposing seasonal as well as trend components while LSTM captures non-linear temporal dependencies. Building on prior approaches, this model aims to deliver interpretable and accurate GWL forecasts, enhancing sustainable water management.

III. DATA AND METHODS

The subject of the study here is the coastal city of Visakhapatnam, located along the Bay of Bengal in Andhra Pradesh, India. The boundaries of the Region of Interest (ROI) were 83.0°–83.5° longitude and 17.6°–18.5° latitude. Groundwater Storage (GWS), Root Zone Soil Moisture (SoilMoist_RZ), Surface Soil Moisture (SoilMoist_S) and Terrestrial Water Storage (TWS) are among the environmental characteristics that are examined. The datasets were acquired in NetCDF4 format from the GRACE-DA1 V2.2, the Global Land Data Assimilation System (GLDAS) and Catchment Land Surface Model (CLSM), which are generated by NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC). The GRACE (NASA/DLR) and GRACE-FO (NASA/GFZ) satellite missions' observations of TWS are incorporated into these. NASA's Panoply visualization tool was used to identify and validate the variables before they were preprocessed into a consistent daily period. Python and the netCDF4 library were used for data parsing, quality checks and metadata alignment. Geospatial distributions were visualized using Matplotlib and Cartopy. After the preprocessing, Microsoft Excel was used to export the daily averages of every parameter into a CSV format for further analysis. The line plots were created for each parameter in order to investigate the temporal fluctuations. Then the ML techniques were utilized in order to predict the data 365 days into the future. A hybrid framework was created using LTSM and Prophet. The LTSM adjusted the non-linear residuals in order to increase the flexibility for the short-term variations, Prophet modeled the seasonality as well as the long-term patterns. The performance of the model was then evaluated with the use of the Coefficient of Determination (R2), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) which were then displayed in the console. This strategy guaranteed the framework for evaluating groundwater that used predicted modeling with satellite-driven datasets.

IV. RESULTS AND DISCUSSION

The findings are broken down into two sections: a five-year analysis of observed hydrological parameters and a hybrid Prophet–LSTM framework forecast of future values. In order to demonstrate the hybrid method's ability to capture seasonal swings, long-term patterns and forecast uncertainty across a variety of water-related indicators, this section compares historical variability with model forecasts.

© September 2025 | IJIRT | Volume 12 Issue 4 | ISSN: 2349-6002

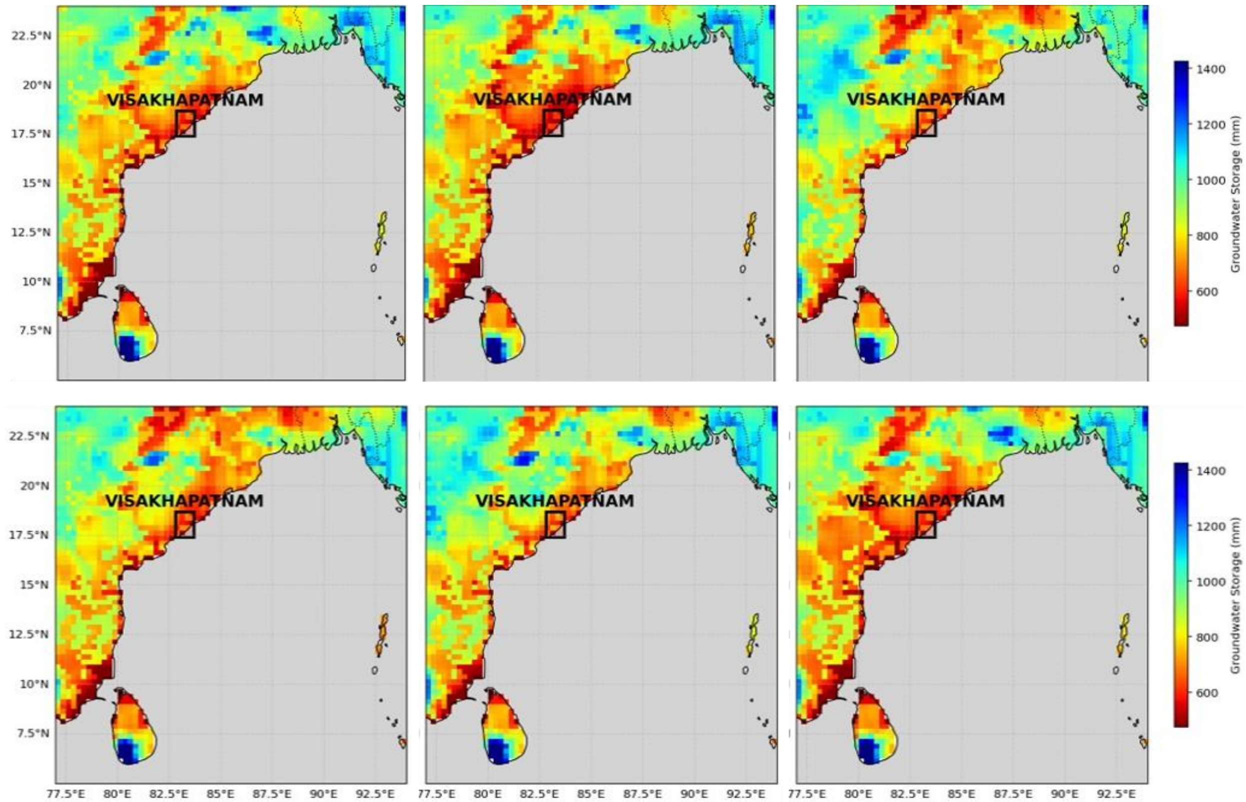

Fig. 1 Temporal Variation of Groundwater Storage in Visakhapatnam (17.6°–18.5°N, 83.0°–83.5°E), indicated by black rectangle on 12th August, 2020–2025.

Groundwater Storage (Gws)

Significant fluctuations were observed in Visakhapatnam's Groundwater Storage (GWS) during August 2020 to August 2025, which reflected the seasonal cycles of recharge as well as depletion, a common occurrence for semi-humid coastal areas. The mean being roughly at 594 mm, the observed values varied between 474 to 811 mm. Although the monsoon months recharged the GWS levels quickly due to precipitation and decreased extractions, but still extended decreases were noted during the dry months, frequently resulting in the GWS reaching the minimal values. This suggested a significant reliance on the groundwater extractions for household and agricultural purposes during the times of lower rainfall. The predictive time series also reveals multi-year oscillations, thereby indicating the interaction of climatic variables including land-use change and the variability during the rainy seasons.

This pattern is carried over into the following year by the hybrid Prophet–LSTM forecast, which offers information on anticipated storage dynamics under comparable meteorological conditions. GWS levels are predicted to climb first until late 2025, then gradually fall in early 2026 before leveling off in mid-2026. Crucially, the model displays recurrent fluctuations instead of the flat-line tendency of tree-based regressors, maintaining seasonal periodicity. The inherent uncertainty in long-term hydrological prediction is highlighted by the widening of the 95% CI in the other half of the forecast, especially when these are subjected to uncontrollable external factors like harsh weather or policy-driven changes in water demand. However, the hybrid model's ability to account for both short-term anomalies and long-term seasonal cycles clearly shows its usefulness in planning of water resources. These results confirm that in susceptible areas like Visakhapatnam, incorporating data-driven forecasting techniques can yield a more precise and accurate early warnings of possible groundwater stress.

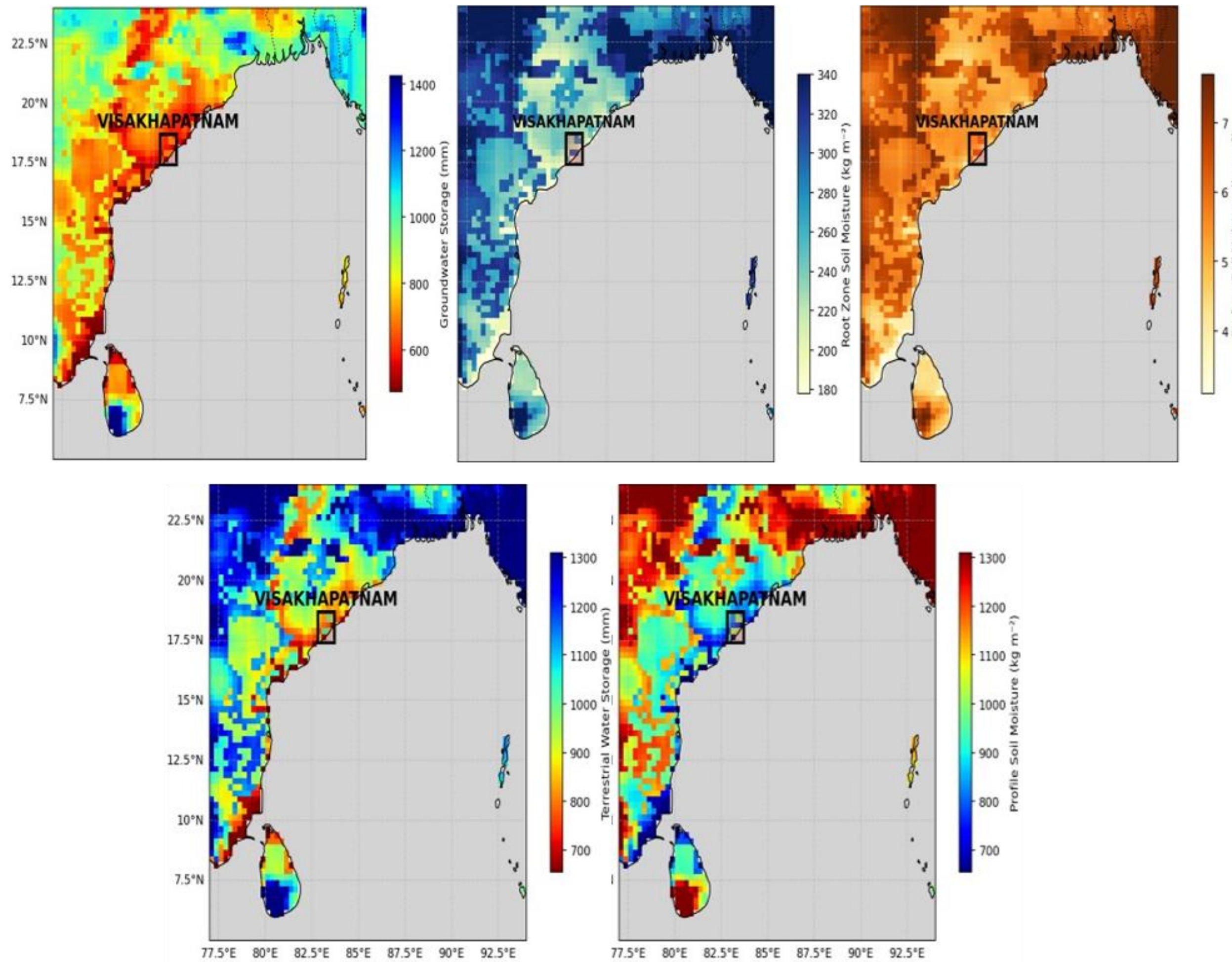

Fig. 2 Hydrological Parameters on 18 August 2025 – Groundwater Storage, Root Zone Soil Moisture, Surface Soil Moisture, Terrestrial Water Storage and Profile Soil Moisture

Root Zone Soil Moisture (Soilmoist_Rz)

A significant number of variations could be observed during the five-year observation period for the Root Zone Soil Moisture. It was evident that the measured values were significantly influenced by the rain, with greater soil moisture levels during periods of heavy precipitation and a slow decline during the dry months when the water seeped into the deeper layers or drawn out by the vegetative uptake. The Root Zone Soil Moisture is known for directly supporting the plant growth both during the periods of peak rainfall and after dry spells, thereby highlighting these variations as the significance of this factor as a crucial regulator of crop yield.

Across the subsequent years, the variability indicated that the region's soil water availability was shaped by the land-use changes and the interannual rainfall anomalies. The dynamics were then carried over into the annual cycle by the predicted values. The hybrid LSTM-Prophet model predicted and then produced the values. The Root Zone Soil Moisture, according to the prediction pattern, increased during the 2025 monsoons before steadily declining during the beginning of the year 2026. What's crucial here is that the model doesn't flatten to the mean value, it preserves the seasonal periodicity instead and incorporates the anticipated recharge-depletion sequences. The seasonal cycles show the framework's resilience in the expected persistence and the growing prediction uncertainty is represented by the widening CI over time. This enhances the model's ability to direct the drought preparedness and agricultural planning

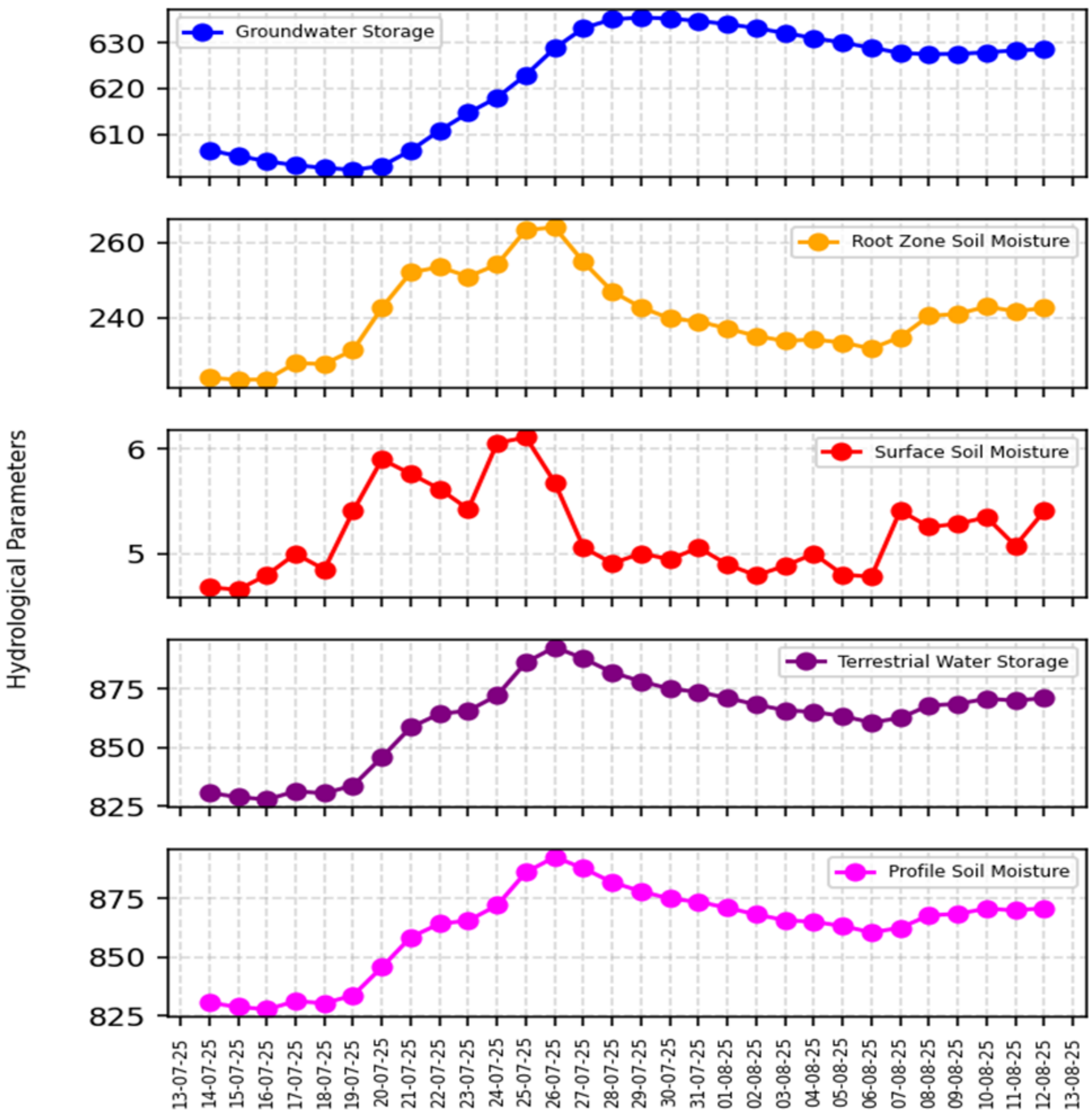

Fig. 3 Daily Trends of Hydrological Parameters from 13 July to 13 August 2025 – Groundwater Storage, Root Zone Soil Moisture, Surface Soil Moisture, Terrestrial Water Storage and Profile Soil Moisture

Surface Soil Moisture (Soilmoist_S)

Surface Soil Moisture reacts directly to precipitation, evaporation and land surface processes. It is a crucial metric for comprehending short-term hydrological dynamics. During the course of the five-year sample, these values fluctuated quickly i.e., peaking during the monsoon rainfall events and declining sharply during dry months. The surface moisture levels fluctuated rapidly, frequently within days, highlighting their ephemeral character in contrast to root zone moisture, which represents longer-term water retention. With recurrent wetting phases during subsequent monsoons and gradual drying phases as rainfall decreased, seasonal cycles were clearly discernible. This fluctuation emphasizes how vulnerable surface layers are to changes in the climate and land use.

According to the mean values for the observed period, the surface soil of Visakhapatnam underwent a fairly balanced alternation between saturation and depletion, which is in line with the semi-humid coastal climate. The hybrid Prophet–LSTM forecast predicts seasonal spikes during the 2025 monsoon and consistent reductions through the 2026 dry season, indicating that these dynamics will persist into the upcoming year. The predictions show how the framework captures the sudden spikes and sharp drops characteristic of surface processes while maintaining periodicity. Unlike the tree-based models that flatten the outcomes, this method provides CI that show uncertainty in long-range projections while reflecting natural variability.

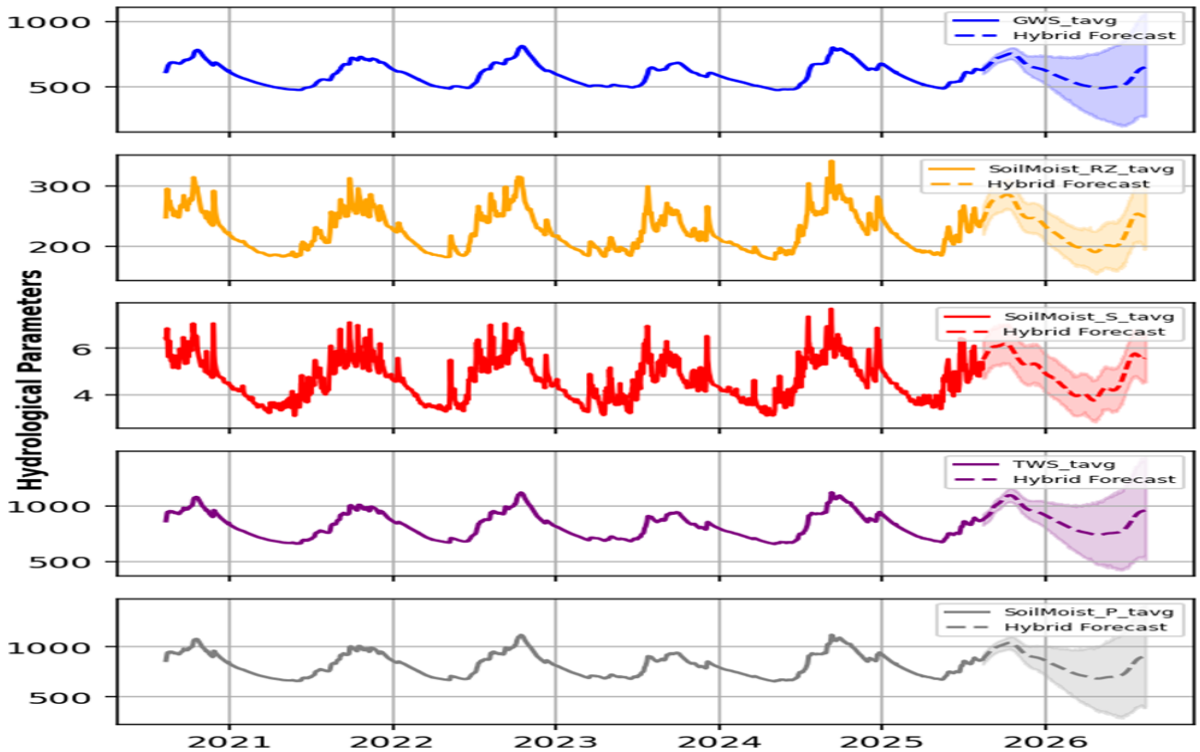

Fig. 4 Observed and Forecasted Hydrological Parameters from 13 August 2020 to 13 August 2025 – Groundwater Storage, Root Zone Soil Moisture, Surface Soil Moisture, Terrestrial Water Storage and Profile Soil Moisture

Terrestrial Water Storage (Tws)

The Terrestrial Water Storage (TWS) is all the water components in the land system, such as the surface water, groundwater, vegetation etc. Under the five-year period of observation, the combined effects of TWS were evident in Visakhapatnam. During the monsoon months, recurring peaks were found in the datasets due to predominant precipitation and rapid increase in terrestrial water stocks. However, during the dry seasons, notable decreases were noticed as the groundwater withdrawal increased, resulting in an increased depletion. These variations ranging from the lowest to the highest storage value, drew attention to this delicate balance between the human needs and the natural inputs. Thereby, indicating the broader climatic influences and the region's reliance on the monsoonal constancy for maintaining the water resources.

Insights into anticipated terrestrial water behavior under ongoing climate patterns were provided by the hybrid Prophet–LSTM projections, which carried these dynamics beyond 2026. In line with past recharge cycles, model projections showed that TWS would first increase during the 2025 monsoon before gradually declining throughout early 2026 as precipitation decreased and its use increased. Additionally, the model avoided the tendency of regression-based models to converge toward mean values in the long-range projections by maintaining the inherent periodicity of TWS. Predicting integrated hydrological systems is difficult because of the increasing confidence ranges, which increases uncertainty over time due to several interacting variables. The prediction's ability to capture both seasonal peaks and subsequent decreases, also shows how well the framework can replicate the behavior of

complex systems. From the standpoint of management, the anticipated trends indicate that rapid post-monsoon losses pose a threat to the long-term sustainability, even when seasonal recharge is still adequate to momentarily replenish water supplies.
The region may experience stressful times due to the uncontrolled extraction rates, especially during the successful years when the monsoons have very less rainfall. Policymakers may foresee the future shortages and rank remedies by adopting better groundwater regulation, artificial recharge structures and effective irrigation techniques due to the hybrid LSTM-Prophet model's capabilities. The combination of ML predictions and historical data shows how sophisticated modeling may fill in the knowledge voids in Terrestrial Water Storage and provide useful insights for climate and sustainable resource governance.

Profile Soil Moisture (SoilMoist_P)
The Profile Soil Moisture (SoilMoist_P) is basically an assessment of Soil Moisture Content across the soil column. It integrates the surface as well as the dynamics of the root zone. The Profile Soil Moisture is shown to have both long-term storage behavior as well as high relationships with the seasonal rainfall cycles throughout the five-year observation period. Periodic wetting periods are showed in the datasets where prolonged precipitation restored deeper layers during the rainy seasons. Profile Soil Moisture, unlike the Surface Moisture, adjusts more gradually, offering a stable reflection of subsurface water availability, rather than acting as a short-term buffer during the dry spells.
According to the predictions projected by the hybrid LSTM-Prophet model, the patterns of Profile Soil Moisture persist in 2026, predicting the seasonal increase during the 2025 monsoon and reductions during the dry months. Also, the persistence of water in the deeper soil layers were captured by the model. This showed that the moisture remained stable before gradually diminishing. Unlike the Surface and the Root Zone Soil Moisture that comparatively declined more rapidly. This demonstrated the capability of hybrid modeling over the traditional regressors which often tend to fail to retain the system-specific temporal dynamics.
Since the crop sustainability, ecosystem services as well as the groundwater recharge are directly influenced by the Profile Soil Moisture, it can be implied that these findings have practical implications. In order to maintain a long-term soil-water balance, adaptive strategies such as efficient irrigation scheduling, soil conservation measures and artificial recharge initiatives are required. The hybrid approach offers early warning signals of potential stress by aligning the historical patterns with anticipated dynamics. This enables the policy makers to implement the water conservation as well as the agricultural adaptations. The LTSM-Prophet model also provides a reliable method for monitoring the soil moisture by balancing short-term fluctuations with long-term stability. The importance of incorporating ML-based projections into regional governance are reinforced by these insights, ensuring the soil-water interactions are effectively accounted for in the sustainable planning for Visakhapatnam.

Parameter	MAE	RMSE	R2
Groundwater Storage	1.477913	3.670977	0.998306
Root Zone Soil Moisture	2.184491	4.019324	0.985396
Surface Soil Moisture	0.160983	0.241311	0.924173
Terrestrial Water Storage	3.728609	7.706267	0.995881
Profile Soil Moisture	3.979860	7.902674	0.995668

Table 1: Performance Metrics of the LSTM–Prophet Model for Groundwater and Soil Moisture Variables.

V. CONCLUSION

This study showed how a hybrid ML framework can be utilized for forecasting of hydrological parameters essential for Visakhapatnam's water resource management. The hybrid method predictions accuracy was pretty decent and structurally compatible with the observed variability by combining the LSTM and Prophet to identify sequential dependencies and rectifying non-linear residuals while simulating long-term seasonal cycles. The model provided a solid foundation for analyzing regional water dynamics by accurately reproducing seasonal recharge–depletion patterns and extending them into future projections across five parameters: groundwater storage, root zone soil moisture, surface soil moisture, terrestrial water storage and profile soil moisture.
After the comparison of the historical data as well as the expected trends, it is revealed that, while the seasonal monsoon recharge remains a vital source of

replenishment, the subsequent reductions offer long-term hazards of water stress. The increasing confidence intervals highlighted the difficulties of long-term hydrological prediction while simultaneously enhancing the model's transparency by quantifying forecast uncertainties. These results are for critical decision-making as the model provides early warning signs of any potential vulnerabilities and lead to adaptive solutions for sustainable water governance.

The need of data-driven forecasting methods in management of water resources under the climatic variability and the human pressure is highlighted by the findings. The methodological voids left by the conventional models are filled by the LSTM-Prophet framework while also improving the practical applicability by balancing the interpretability as well as the flexibility. The incorporation of these findings into policy and planning initiatives may boost resilience in semi-humid coastal regions, providing a long-term water security for agricultural as well as the urban needs in Visakhapatnam.

REFERENCES

[1] V. S. Raj, S. S. Mahajan and R. S. Patil, "Groundwater level prediction using artificial neural networks in coastal aquifers," J. Hydrol., vol. 587, 2020.

[2] M. K. Nayak, S. Das and P. S. Raju, "Artificial neural network-based groundwater level forecasting in semi-arid regions," Environ. Earth Sci., vol. 78, no. 5, 2019.

[3] H. Kim, Y. Park and S. Choi, "Comparative assessment of MLR and ANN for groundwater prediction in Japan," Hydrogeol. J., vol. 27, 2019.

[4] S. Sharma, A. Kumar and R. Tiwari, "Hybrid ANN-GA model for groundwater prediction," Water Resour. Manage., vol. 32, 2018.

[5] M. Coulibaly, H. Anctil and D. Bobée, "Groundwater level forecasting using ANN," Water Resour. Res., vol. 37, 2001.

[6] L. Coppola, F. Dimakopoulos and P. Tarantino, "Use of ANNs for groundwater modeling," J. Hydrol., vol. 302, 2005.

[7] S. Uddameri, "Forecasting groundwater levels using artificial neural networks," Hydrol. Process., vol. 21, 2007.

[8] Y. Yoon, S. Lee and K. Jung, "ANN-based prediction of groundwater in unconfined aquifers," Hydrogeol. J., vol. 19, 2011.

[9] M. Adamowski and H. Chan, "Neural networks for groundwater prediction," J. Hydrol. Eng., vol. 16, 2011.

[10] NASA Goddard Space Flight Center, "GRACE-DA1 V2.2 and GRACE-FO satellite data products," GES DISC, 2021.

[11] C. Reichle, "Global Land Data Assimilation System (GLDAS) Overview," NASA, 2017.

[12] S. Shiri and M. Kisi, "Predicting groundwater fluctuations using GEP, ANFIS, ANN and SVM techniques," Hydrol. Sci. J., vol. 56, 2011.

[13] V. Coulibaly, H. Anctil and D. Bobée, "ANN for monthly groundwater level modeling," Water Resour. Res., vol. 37, 2001.

[14] J. Maier and G. Dandy, "Neural networks for water resources modeling," Water Resour. Res., vol. 34., 1998

[15] K. Elsherbiny et al., "Recent advances in ANN for hydrological modeling," Hydrol. Sci. J., vol. 55, 2010.

[16] S. Watson and B. Luckman, "Dendrochronology and groundwater modeling," J. Hydrol., vol. 312, 2005.

[17] Ferguson and C. St. George, "Tree-ring chronologies for hydroclimatic reconstructions," Hydrogeol. J., vol. 11, 2003.