Forecasting-Surface-water-and-Groundwater-Level-in-Florida-using-Advanced-ML-Approaches

Rohith Reddy Nedhunuri[1], Golmar Golmohammadi[2], Saman javadi[3]

1. Department of Soil, Water and Ecosystem Sciences, University of Florida, IFAS/RCREC, Ona, Florida, USA
2. Department of Computer and Information Science, University of Florida, Gainesville, Florida, USA
3. Department of Soil, Water and Ecosystem Sciences, University of Florida, IFAS/SWFREC, Immokalee, Florida, USA

Abstract

Effective water resource management and planning rely heavily on accurate forecasting of surface water discharge and groundwater levels, particularly in hydrologically intricate regions like Florida. This study conducts a large-scale analysis across the entire state of Florida, systematically evaluating the efficacy of five advanced deep learning models—Long Short-Term Memory (LSTM), Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS), Neural Hierarchical Interpolation for Time Series (N-HiTS), Temporal Fusion Transformers (TFT), and Informer—using extensive datasets from 45 surface water stations and 45 groundwater monitoring wells, spanning 23 years (2001-2023). Unlike traditional methods that use different models for predicting surface water and groundwater, we demonstrate that the same models can effectively forecast both surface water discharge and groundwater levels. After rigorous data preprocessing, including cleaning and normalization, these models were trained and tested to predict daily water dynamics. Our analysis reveals that the N-HiTS model consistently delivers superior performance for both surface water discharge and groundwater levels. Specifically, N-HiTS was the best-performing model in 23 out of 45 groundwater wells based on Root Mean Square Error (RMSE), and in some wells, it achieved an RMSE that was 70% less than that of the next best model, TFT. N-HiTS led 34 out of 45 stations using Normalized RMSE (NRMSE) for surface water discharge, consistently achieving at least a 20% lower NRMSE than the next-best model in all leading stations. These findings demonstrate N-HiTS's exceptional ability to capture temporal dynamics in hydrological data, offering a powerful tool for enhancing predictive accuracy in water resource management. This research provides crucial insights into the application of deep learning models in hydrological forecasting, with significant implications for resource planning in Florida and similar regions.

1. Introduction

The accurate forecasting of surface water discharge and groundwater levels is crucial for effective water resource management and planning, particularly in regions like Florida, where water resources play a pivotal role in supporting the local ecosystem and economy. The state of Florida, with its unique hydrological characteristics, presents a challenging yet essential area for hydrological studies. The region's water dynamics are influenced by the interaction between the extensive surface water network, including rivers, lakes, and wetlands, and the underlying Floridan Aquifer, one of the world's most productive aquifers (Vu, M. T. et al., 2023).

Using large-scale data from the entire state of Florida is important for improving the accuracy of hydrological models, especially because of the complex interactions between surface water and groundwater in the area. Florida’s water system is highly connected, with significant exchanges between surface water bodies and the Floridan Aquifer, one of the most active and productive aquifers in the world (Sutton, J. E. et al., 2015). These exchanges vary across different locations and times, making it essential to use data from all over the state to accurately capture these changes. A large-scale approach helps models learn from diverse conditions, including differences in rainfall, evaporation, and human activities like groundwater pumping, which are key to making reliable predictions (Gordu, F., & Nachabe, M. H., 2023). This approach reduces regional biases and improves the models' ability to work well under different weather and water conditions, supporting more effective water management and planning across Florida.

Recent advancements in deep learning have shown promising potential in improving the accuracy of hydrological forecasting. Deep learning models, with their ability to capture complex nonlinear relationships and temporal dependencies in data, are well-suited for predicting hydrological variables. In this study, we explored the application of various deep learning models, including LSTM, N-BEATS, N-HiTS, TFT, and Informer, for forecasting daily surface water discharge and groundwater levels across multiple stations in Florida.

This study utilizes a diverse suite of deep-learning models, each engineered to address different aspects of time-series forecasting. The LSTM networks fall under the category of Recurrent Neural Networks (RNNs), which are particularly adept at capturing long-term dependencies in sequential data. This makes them suitable for applications that require memory of past events, such as natural language processing and sequential prediction. In contrast, both N-BEATS and N-HiTS belong to the category of Feedforward Neural Networks. N-BEATS employs a unique architecture designed to forecast time series directly through stacks of fully connected layers, focusing on a data-driven approach without reliance on traditional time-series forecasting methods. N-HiTS, similarly, introduces a hierarchical architecture that enhances the model's ability to interpolate and forecast at various time scales effectively.

Additionally, the research incorporates models that leverage advanced attention mechanisms within the transformer framework, namely the TFT and the Informer. TFT combines high-capacity transformers with a recurrent-style gating mechanism to adeptly handle dynamic relationships between variables in multivariate, multi-horizon time-series data. The Informer extends the transformer architecture to efficiently manage long-sequence forecasting, implementing a novel self-attention mechanism that reduces computational demands while retaining effectiveness over extended data sequences. These models represent the forefront of machine learning techniques, offering sophisticated solutions to the complex challenges presented by modern time-series analysis.

Our research is motivated by the need to enhance the predictive capabilities of hydrological models for both surface water and groundwater, which are essential for understanding water dynamics and making informed decisions in water resource management, flood prevention, and drought mitigation.By comparing the performance of these deep learning models, we aim to identify the most effective approach for capturing the temporal dynamics of hydrological variables in the entire state of Florida.

Various machine learning (ML) models have been effectively utilized to improve groundwater and surface water forecasting accuracy. For instance, SVM and NARX have been applied to forecast groundwater levels in South Africa (Aderemi et al., 2023). Random Forest algorithms have been used to predict groundwater levels from passive seismic wavefields (Abi Nader et al., 2023; Muhumuza, 2020; Onyebuchi, 2020). Additionally, ML has been employed to predict groundwater potentiality in semi-arid mountainous regions (Jadoud et al., 2023). Despite their successes, ML models also exhibit several disadvantages. They often struggle with long-term dependencies and complex nonlinear relationships in time series data, which are essential for accurate hydrological predictions. Additionally, they require extensive feature engineering and preprocessing, which can be resource-intensive and may not scale well with large datasets.

These limitations can be potentially overcome by deep learning models. Deep learning models, particularly those using architectures like LSTM, are well-suited to address these challenges. LSTMs are capable of learning long-term dependencies without the need for manual feature extraction, making them ideal for modeling the temporal dynamics of groundwater and surface water levels. For example, LSTMs have been used to predict groundwater recharge (Huang et al., 2023; Patra et al., 2023) and multi-step-ahead groundwater levels (Koch et al., 2023). They have also been applied to sustainable groundwater management (Alabdulkreem et al., 2023) and groundwater quality prediction (Valadkhan et al., 2022; Jaffar et al., 2022). While LSTM models are powerful for handling sequential data, they come with certain limitations. They are computationally inefficient with long sequences, making training and inference slow. LSTMs can overfit small datasets due to their complexity, and careful tuning of hyperparameters like layer count and unit size is needed to avoid underfitting or overfitting.

The aforementioned challenges can be addressed by using more recent advancements in neural network architectures, specifically, those falling into the categories of feedforward neural networks like N-BEATS and N-HiTS, as well as attention mechanisms like TFT and Informer. These models offer several improvements over LSTMs .

In terms of practical applications, N-BEATS has been implemented in various studies related to groundwater and surface water predictions. For instance, it has provided robust long-term forecasts from intermittent observation data (Vu et al., 2023; Li et al., 2023), groundwater level predictions across multiple locations (Mbouopda et al., 2022), and integrating multiple time series data sources to enhance water level prediction accuracy (Kenda et al., 2020). These examples highlight the versatility and efficiency of N-BEATS in environmental data analysis, offering a promising alternative to LSTM for hydrological modeling tasks. Building upon the successful application of the N-BEATS model, the N-HiTS model advances forecasting with hierarchical interpolation and multi-rate data sampling techniques. It has shown improvements in accuracy and efficiency in large-scale forecasting (Challu et al., 2023), has been used to predict traffic flow (Lu et al., 2024), forecast product demand in retail (Kollipara et al., 2022), and predict solar and wind energy outputs in renewable energy forecasting (Mazen et al., 2023; Di Grande, 2024). These implementations underline the N-HiTS model's versatility and robust performance across different forecasting scenarios.

Following the discussion on the N-HiTS model, another advanced approach in the field of time series forecasting is the TFT. TFT has shown remarkable success in hydrology-related forecasting tasks by effectively integrating multiple data inputs and handling complex temporal dynamics. TFT enhances prediction accuracy and provides superior alternatives to traditional models like LSTM and standalone Transformer models. For example, TFT improves streamflow prediction accuracy (Koya et al., 2024; Wei, 2023; Vu, 2023), river level prediction with improved probabilistic forecasting and uncertainty quantification (Wang et al., 2023), and flood forecasting by modeling spring flood formation for long-term maximum water level predictions (Castangia et al., 2023). These applications underline the TFT model's versatility and robust performance in hydrology.

Building on the insights gained from N-HiTS and TFT models, the Informer model emerges as another powerful tool in the forecasting landscape, particularly noted for its efficiency in handling long sequences and reducing computational overhead, which is critical in hydrology. It has been applied in drought forecasting, demonstrating superior performance over traditional models like ARIMA and LSTM, especially at longer timescales, using the Standardized Precipitation Evapotranspiration Index (SPEI) (Shang et al., 2023).

Despite significant advancements in hydrological modeling, most existing studies have limitations that our research aims to address. Previous research often focuses on smaller, localized datasets, limiting the models’ ability to generalize across broader regions with varying hydrological conditions. By conducting our study on a large scale across the entire state of Florida, we address this gap, capturing the complex and diverse interactions between surface water and groundwater that smaller-scale studies may miss. This comprehensive approach allows for a more robust understanding of hydrological dynamics, enhancing the predictive capabilities of the models under various climatic and hydrological scenarios (Vu, M. T. et al., 2023). Additionally, many prior studies typically use separate models for predicting surface water discharge and groundwater levels, which increases computational costs and resource requirements. Our research breaks new ground by demonstrating that a single model can effectively forecast both surface and groundwater dynamics, streamlining the forecasting process and reducing the need for multiple models. This unified approach not only improves efficiency but also ensures consistent and comparable predictions across different water systems, setting our study apart from traditional methods.

In summary, this study presents a novel approach by 1) developing and comparing the performance of five deep learning models—LSTM, N-BEATS, N-HiTS, TFT, and Informer—on extensive hydrological data from Florida to identify the most accurate model for this region; 2) training these models on both surface water and groundwater data, unlike many studies that focus solely on one type of water data, allowing us to use these models for predicting both surface and groundwater dynamics; and 3) utilizing a large-scale dataset that includes daily records from 45 stations each for surface water discharge and groundwater levels, spanning 23 years, with these stations strategically distributed across Florida to ensure comprehensive regional coverage. This integrated and extensive dataset ensures robust model training and testing, thereby enhancing the reliability of our predictions. The findings from this study provide valuable insights for water resource management and planning in Florida and can be adapted for use in other regions with similar hydrological characteristics.

2. Materials and Methodology

2.1 Study Area

In this study, the focus is on the hydrological dynamics of the state of Florida, which serves as the study area. The region is characterized by a complex interaction between its abundant surface water and groundwater resources. The state is underlain by the Floridan Aquifer, one of the most productive aquifers in the world, providing a significant portion of the freshwater supply for both residential and agricultural purposes. The aquifer spans an area of about 100,000 square miles, extending into neighboring states, and is composed of porous limestone and dolomite formations. Groundwater levels in Florida are influenced by factors such as rainfall patterns, evapotranspiration rates, and water withdrawal for human use. On average, the Floridan Aquifer discharges about 10 billion gallons of water per day, with recharge primarily occurring through precipitation and infiltration.

Surface water in the study area is equally vital and diverse, encompassing over 1,700 rivers and streams, numerous lakes, and extensive wetland areas, including the famous Everglades. The state's surface water network has a total length of approximately 50,000 miles, with the St. Johns River being the longest at about 310 miles. These surface water bodies play crucial roles in supporting biodiversity, providing recreational opportunities, and sustaining agricultural activities. The interaction between groundwater and surface water is particularly evident in Florida's numerous springs, where groundwater from the aquifer emerges at the surface, contributing to the flow of rivers and streams. Effective management of both groundwater and surface water resources is essential for maintaining the ecological balance and supporting the water needs of Florida's growing population and economy.

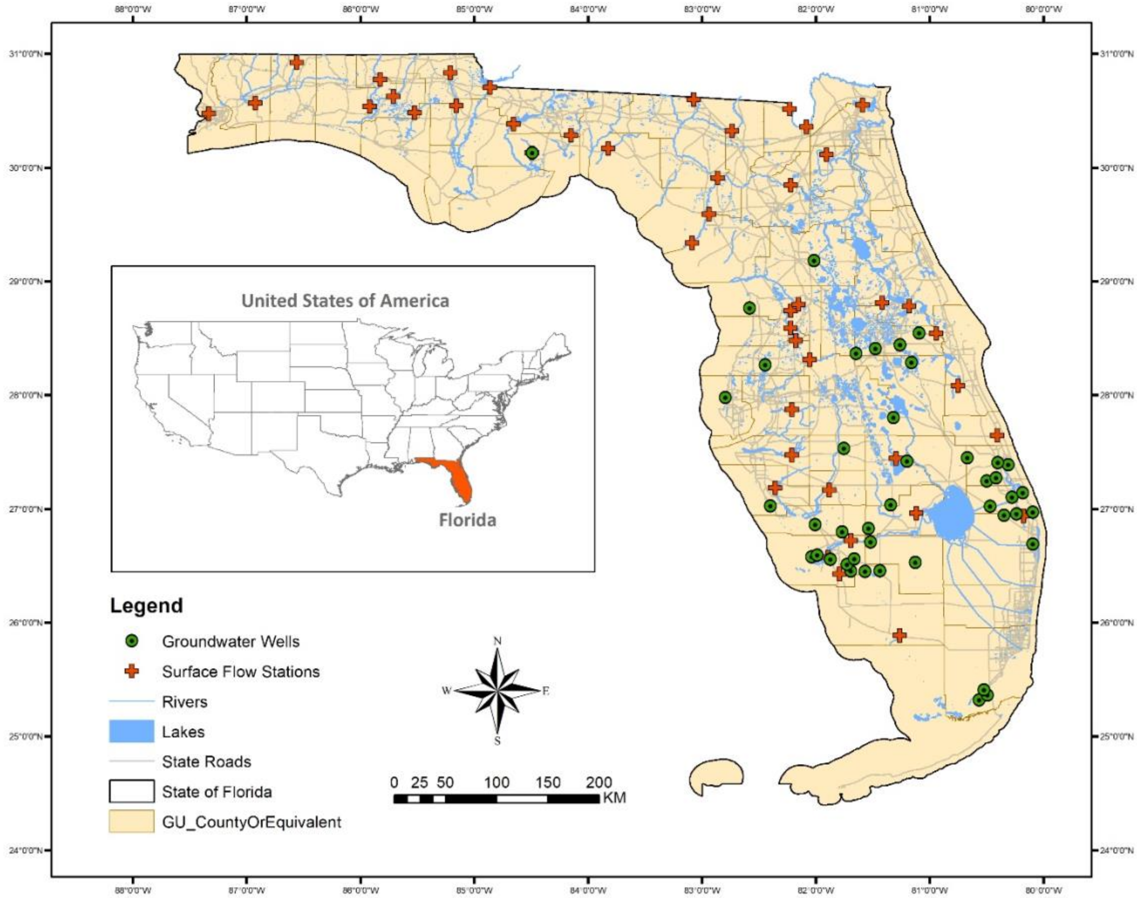


Figure 1: Study Area Image

2.2 Input Data

The data used in this study were obtained from the United States Geological Survey (USGS) website, comprising daily records of surface water discharge and groundwater levels from 45 stations each, respectively, spanning 23 years from 2001 to 2023. The datasets underwent rigorous preprocessing, including data cleaning and data normalization, to ensure the quality and consistency of the input data for model training and testing.

The data for this study was sourced from the USGS website, encompassing hydrological measurements from various stations across Florida. For groundwater analysis, data from 45 stations were considered, each providing daily measurements spanning 23 years from 2001 to 2023. The groundwater dataset comprises two columns: 'Daily Date' and 'Groundwater level above NGVD 1929, meter (Maximum)'. The predictive model for groundwater levels utilized the previous day's groundwater level as an input to forecast the current day's level. Similarly, for surface water analysis, data from 45 stations were included, with each station offering daily records over the same 23-year period. The surface water dataset consists of two columns: 'Daily Date' and 'Discharge, cubic feet per second (Mean).' The model for predicting surface water discharge employed the previous day's discharge values as a predictor for the current day's discharge.

Prior to analysis, the datasets underwent several preprocessing steps to ensure data quality and consistency. These steps included data cleaning to remove any anomalies, normalization to standardize the data range, and splitting the data into training and testing sets for model validation. During the preprocessing phase, gaps were filled, and negative measurements were substituted with the average values of the respective datasets. The data was divided chronologically, ensuring that the predictive models were trained on historical data and tested on recent data to simulate real-world prediction scenarios.

2.3. Deep Learning Algorithms

After determining the effective input combinations and optimum values for each operator, five different models, including LSTM, N-BEATS, N-HiTS, TFT, and Informer, were used to forecast surface water discharge and groundwater levels. In this section, a general description of the behavior of the methods, as well as the differences between them, is described. The details of the LSTM, N-BEATS, N-HiTS, TFT, and Informer algorithms are presented in the Supplementary material document. For the LSTM model, it was chosen for its ability to capture long-term dependencies within time-series data. LSTM is a variant of recurrent neural networks that incorporates memory cells and gates to regulate long-term information storage. These memory cells allow the network to retain information over extended periods while the gates control the flow of information, deciding which information to keep and which to discard. This makes LSTM particularly effective in modeling sequential data such as hydrological measurements. The N-BEATS model is an innovative deep-learning architecture designed specifically for the time series competition. Developed for its simplicity and high accuracy in predicting univariate data points, N-BEATS utilizes a hierarchical architecture composed of stack blocks, each subdivided into multiple sub-blocks. The input to each sub-block is refined by subtracting the backcast output from the previous sub-block, and each sub-block generates a forecast output that contributes to the overall prediction. This structure allows N-BEATS to capture diverse patterns efficiently, making it suitable for time-series forecasting. N-HiTS, an extension of N-BEATS, was employed for its enhanced accuracy and computational efficiency in long-horizon forecasting. N-HiTS uses multi-rate sampling and multiscale synthesis to construct hierarchical forecasts. The model consists of multiple blocks, each containing a multilayer perceptron (MLP) that projects inputs onto basis functions to produce backcast and forecast outputs. These outputs are aggregated to form the final prediction, with each stack specializing in capturing different data characteristics. This hierarchical approach makes N-HiTS particularly effective for capturing complex temporal patterns. The TFT is an advanced deep neural network designed for multi-horizon forecasting. TFT leverages attention mechanisms to enhance performance and introduces interpretability in forecasting models. Key features of TFT include static covariate encoders for context vectors, gating mechanisms, and sample-dependent variable selection to reduce irrelevant inputs, a sequence-to-sequence layer for processing observed inputs, and a temporal self-attention decoder to capture long-term dependencies. These innovations allow TFT to deliver high forecasting accuracy and provide insights into the importance of various inputs and temporal patterns. The Informer model was chosen for its ability to handle long sequence time-series forecasting (LSTF) problems effectively. Informer introduces a ProbSparse self-attention mechanism, which replaces traditional self-attention to efficiently handle long sequence inputs with reduced time complexity and memory usage. The model's architecture includes an encoder that processes long sequence inputs and a decoder that generates output predictions in a generative manner. This structure enables the Informer to predict long sequence outputs in a single forward step, minimizing cumulative errors during inference. Each of these models has its unique strengths. LSTM excels at capturing long-term dependencies, N-BEATS offers simplicity and accuracy for univariate data, N-HiTS provides enhanced accuracy and efficiency for long-horizon forecasts, TFT delivers high performance and interpretability for multi-horizon tasks, and Informer is optimized for handling long sequence inputs efficiently. These diverse capabilities make them well-suited for the complex task of hydrological forecasting.

2.4 Error Criteria

The Nash-Sutcliffe efficiency (NSE), introduced by Nash and Sutcliffe in 1970, is a prominent criterion for the calibration and evaluation of hydrological models using observed data. Unlike measures that depend on the units of the predicted variable, NSE is dimensionless, ranging from negative infinity to 1.0. It is calculated by subtracting the ratio of the mean squared error (MSE) to the variance of the observations from 1.0. Due to its dimensionless nature, NSE is often preferred for reporting and comparing model performance. Additionally, NSE can be regarded as a classic skill score, as described by Murphy in 1988, indicating the relative performance of a model compared to a baseline model, which in this case is the mean of the observations. A model is considered no better than using the observed mean as a predictor if its NSE is less than or equal to zero. In the context of optimization, while MSE is minimized, NSE is maximized to improve model performance.

$$NSE = 1 - \frac{\sum_{t=1}^{n}\left(x_{s,t}-x_{o,t}\right)^2}{\sum_{t=1}^{n}\left(x_{o,t}-\mu_o\right)^2} = 1 - \frac{MSE}{\sigma_o^2} \quad (1)$$

In the given equations, 'n' represents the total number of time steps, 'xs,t' denotes the simulated value at a specific time step 't,' and 'xo,t' indicates the observed value at that same time step. The symbols' μo' and 'σo' represent the mean and standard deviation of the observed values, respectively.

The Root Mean Square Error (RMSE) is a widely used metric for quantifying the accuracy of a predictive model. It measures the average magnitude of the errors between predicted and observed values, providing an indication of the model's overall performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2} \quad (2)$$

The Normalized Root Mean Square Error (NRMSE) is a variation of the Root Mean Square Error (RMSE) that provides a relative measure of model accuracy by scaling the RMSE with a normalization factor. Typically, this factor is the range or standard deviation of the observed values, allowing for a comparison of model performance across different datasets or variables with varying scales.

$$nRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2} \times \frac{100}{\bar{O}} \quad (3)$$

3. Results

3.1 Evaluation of models performance and comparison

In this study, we employed three days of preceding data to predict the current day's value for both surface water discharge and groundwater levels. Our analysis involved evaluating five different models across 45 stations for surface water and 45 stations for groundwater, all situated in Florida. The dataset spanned 23 years, from January 1, 2001, to May 20, 2023. We divided the data into training and testing sets, with the first 80% of the time series data used for training and the remaining 20% for testing. To assess and compare the performance of these models, we utilized quantitative efficiency criteria, including Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), and Nash-Sutcliffe Efficiency (NSE), along with graphical analyses such as scatter plots, time variation graphs, and Taylor diagrams. The performance metrics were derived from the testing dataset.

Overall, N-HiTS emerged as the best-performing model across most stations for both surface water and groundwater. The next best model was TFT. For RMSE, N-HiTS delivered the best performance in 23 out of 45 groundwater wells, often outperforming the second-best model by at least 20%. In certain stations, N-HiTS was up to 70% better than the next model. For instance, at stations 252-3701, 272-1601, 281-4601, 282-4601, and 300-3001, N-HiTS achieved RMSE values of 0.007, 0.091, 0.036, 0.03, and 0.068, respectively, surpassing the nearest competitor by at least 75%. TFT was the top performer in 12 stations, occasionally outperforming N-HiTS by 30%, notably at stations 270-5301, 272-2701, 274-0301, and 284-5301. Informer led in 4 stations, LSTM in 3, and N-BEATS in 2.

For surface water discharge predictions, N-HiTS excelled in more stations in terms of NRMSE due to the infrequent and anomalous data in these stations, which rendered RMSE less effective. The NRMSE metric was advantageous because it normalizes the error, providing a more accurate comparison when data variability is high. N-HiTS outperformed other models in 34 stations, often by at least 20%. At stations such as 2288800, 2310947, and 2312000, N-HiTS's performance was nearly 50% better than the second-ranked model in terms of NRMSE. TFT was the best performer in 6 stations, surpassing N-HiTS by 50% at stations 2292900, 2293230, 2297100, and 2370000. LSTM led in 4 stations and N-BEATS in 1. The robustness of the results was confirmed by consistency across RMSE, NRMSE, and NSE metrics. NRMSE offers a normalized measure of prediction error, making it suitable for comparing models across datasets with different scales, while NSE assesses how well the plot of observed versus predicted data fits the 1:1 line.

N-HiTS emerged as the most effective model among the five evaluated, demonstrating superior performance in a significant majority of both surface and groundwater stations. To support our findings that the N-HiTS model is the best-performing model, we refer to studies where N-HiTS has demonstrated superior performance compared to the TFT model. For instance, in the study by (Gobato, 2023), N-HiTS outperformed established benchmarks, including the TFT model, in forecasting stock realized volatility, showcasing its robustness and accuracy. Additionally, a study on time series forecasting compared N-HiTS with TFT and found that N-HiTS consistently outperformed TFT across various forecasting metrics, highlighting its advanced capability in handling complex datasets (Olivares, 2023).

The Temporal Fusion Transformer (TFT) model ranked second, with it being the best model in 13 surface water stations and 6 groundwater stations, followed by Informer, LSTM, and N-BEATS in varying capacities across the dataset. This extensive comparative analysis confirms the robustness of the N-HiTS model in handling both surface and groundwater data under varied conditions, marking it as a preferred tool for hydrological modeling and prediction. Similarly, to support the effectiveness of the TFT model, we refer to studies where TFT has outperformed models like N-BEATS and LSTM. Google's research on TFT demonstrated its superior performance in multi-horizon forecasting tasks, outperforming various deep learning models, including N-BEATS and LSTM (Lim et al., 2021). Additionally, a study published in ScienceDirect highlighted TFT's ability to provide accurate forecasts and interpretability, making it a valuable tool in domains requiring dynamic relationship management between variables (Maldonado et al., 2024).

In addition to the quantitative metrics, graphical analyses such as time variation graphs, scatter plots, and taylor diagrams were employed to visualize the performance of the models. The graphical analysis included observed versus predicted graphs to demonstrate model performance on the time series data generated by predicting both the training data (first 80% from January 1, 2001, to June 1, 2017) and the testing data (last 20% from June 2, 2017, to May 30, 2023). We selected one station each from surface water and groundwater stations to display five graphs per station, one for each model. These graphs illustrated the accuracy of the models in tracking the fluctuations of surface water discharge and groundwater levels. For both surface water and groundwater stations, the time variation graphs revealed that N-HiTS closely followed the observed data trends, indicating its superior predictive ability. The TFT model also performed well, although it did not match the accuracy of N-HiTS in most cases. The remaining models, including Informer, LSTM, and N-BEATS, showed more significant deviations from the observed values, confirming their comparatively lower performance.

Additionally, scatter plots were used to visualize the model predictions against actual observations for selected stations in both surface water and groundwater datasets. In these plots, the closer the data points were to the 1:1 line, the better the model's performance. For both surface water and groundwater stations, scatter plots showed that N-HiTS had the highest concentration of points near the 1:1 line, reflecting its high accuracy. The TFT model, while also showing a good fit, had a slightly wider dispersion of points, indicating more variability in its predictions. The other models exhibited even greater dispersion, further validating the superior performance of N-HiTS and TFT.

The Taylor diagram displayed above is crucial for evaluating the comparative performance of the five models—LSTM, N-BEATS, N-HiTS, TFT, and Informer—at a single depth for each station. This diagram uses reference points to illustrate centered Root Mean Square Differences (RMSD), and the distance from these points indicates the accuracy of each model. Models positioned closer to the reference point demonstrate a higher accuracy, exhibiting a correlation coefficient close to 1 and a variation range similar to the observations, which are indicative of superior model performance.

From the Taylor diagrams, it is evident that N-HiTS consistently shows exceptional performance. For instance, in the Taylor diagrams for groundwater stations, N-HiTS is depicted as the best model in 23 out of 45 stations, aligning closely to the reference point, which indicates its higher accuracy and reliability in simulating groundwater dynamics. Similarly, for surface water stations, N-HiTS continues to dominate in model accuracy and consistency, being the best performer in 34 stations, as evidenced in the diagrams. The diagrams also highlight that the TFT ranks as the second most consistent model, particularly in stations where it surpasses other models like LSTM and N-BEATS.

Table 14: Outcomes from the optimal model for every groundwater station.

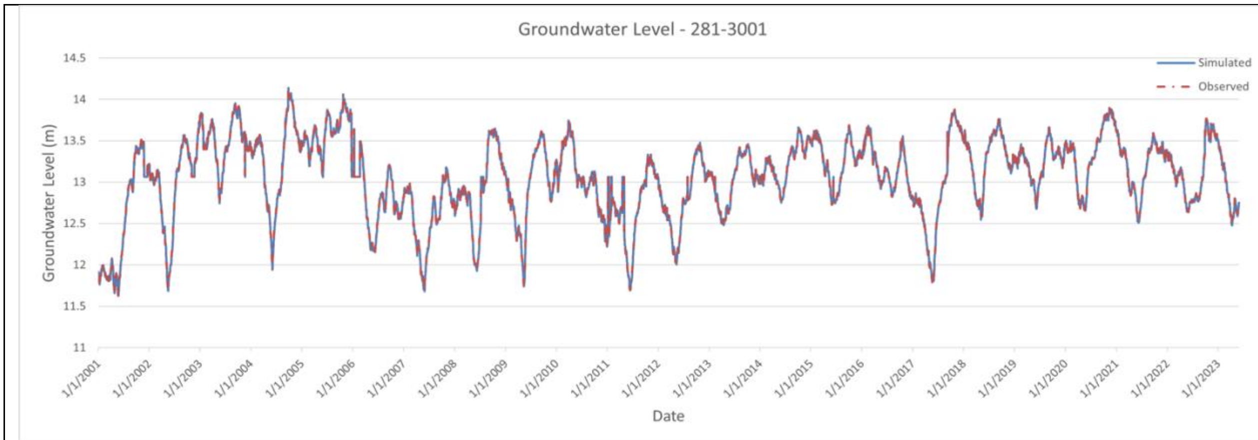
Stations	Models	Test		
		RMSE	NRMSE	NSE
252-3701	N-HiTS	0.002	0.011	0.999
251-0701	N-HiTS	0.006	0.021	0.996
252-0001	N-HiTS	0.010	0.094	0.913
262-0201	LSTM	0.119	0.027	0.99
262-3701	LSTM	0.095	0.02	0.982
262-0701	N-HiTS	0.071	0.036	0.991
263-3501	LSTM	0.038	0.005	0.987
263-3102	N-HiTS	0.245	0.034	0.992
263-3103	TFT	0.051	0.031	0.980
263-2401	N-HiTS	0.071	0.013	0.999
263-4301	N-BEATS	0.228	0.066	0.984
263-2001	TFT	0.082	0.009	1
263-2201	N-HiTS	0.123	0.009	1
264-3801	TFT	0.021	0.018	0.991
264-0601	N-HiTS	0.035	0.032	0.994
264-0602	N-HiTS	0.013	0.029	0.993
264-0802	N-HiTS	0.073	0.023	0.996
264-1301	N-BEATS	0.317	0.034	0.962
265-2201	N-HiTS	0.026	0.065	0.968
265-3001	TFT	0.017	0.031	0.979
265-1801	N-HiTS	0.018	0.063	0.967
265-3901	TFT	0.020	0.014	0.997
270-0202	N-HiTS	0.013	0.015	0.999
270-5301	TFT	0.025	0.031	0.985
270-3101	TFT	0.018	0.025	0.99
270-3401	N-HiTS	0.013	0.013	0.999
270-5801	N-HiTS	0.019	0.04	0.985
271-1201	TFT	0.025	0.016	0.994
271-5801	TFT	0.021	0.014	0.998
272-2701	TFT	0.013	0.01	0.999
272-0201	N-HiTS	0.010	0.014	0.999
272-0101	TFT	0.029	0.031	0.988
272-1601	N-HiTS	0.027	0.106	0.904
273-1401	N-HiTS	0.187	0.044	0.984
274-0301	TFT	0.024	0.027	0.989
275-4201	LSTM	0.041	0.024	0.899
281-4601	N-HiTS	0.010	0.016	0.999
281-3001	N-BEATS	0.022	0.002	0.996

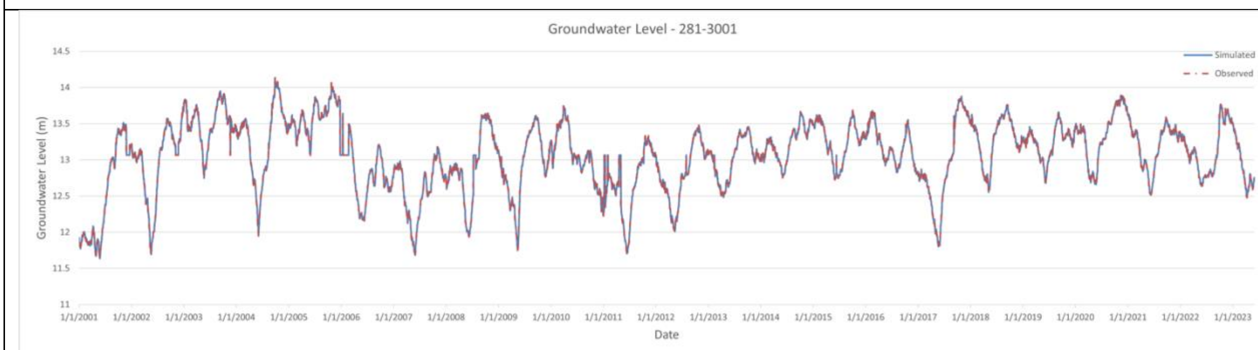
282-4601	N-HiTS	0.009	0.009	1
282-3102	N-BEATS	0.076	0.005	0.986
282-3801	N-HiTS	0.053	0.042	0.988
283-3201	N-HiTS	0.026	0.029	0.994
284-5301	TFT	0.020	0.044	0.98
291-0003	N-HiTS	0.008	0.008	1
300-3001	N-HiTS	0.020	0.03	0.993

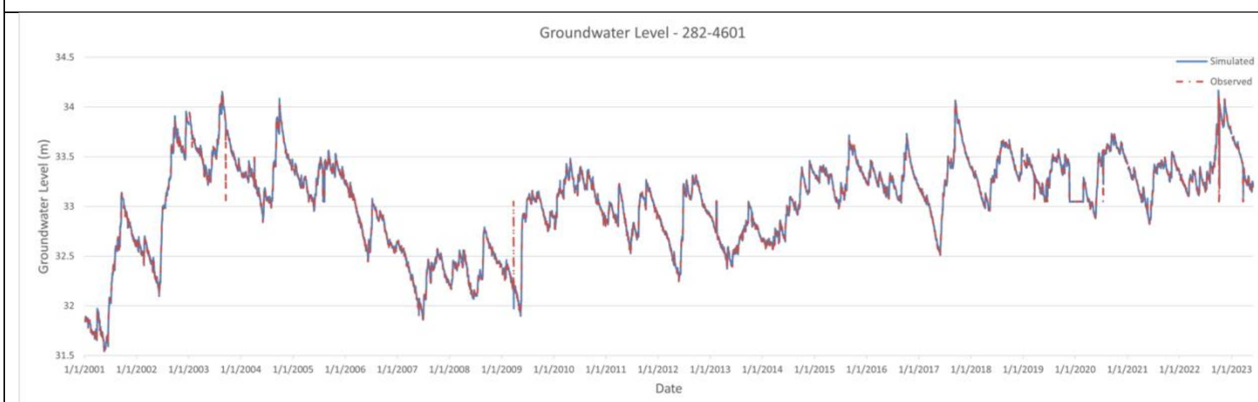
Table 13: Outcomes from the optimal model for every surface water station.

Stations	Best Model	Test		
		RMSE	NRMSE	NSE
2228500	N-HiTS	0.053	0.016	0.999
2231000	N-HiTS	0.728	0.026	0.995
2231291	N-HiTS	10.746	0.304	0.588
2232000	N-HiTS	0.729	0.058	0.977
2232500	N-HiTS	0.800	0.019	0.997
2234435	TFT	4.118	0.176	0.533
2235000	TFT	0.237	0.026	0.988
2246000	N-HiTS	0.355	0.072	0.953
2253000	N-HiTS	0.080	0.063	0.972
2257000	N-HiTS	0.403	0.087	0.946
2270500	N-HiTS	0.343	0.061	0.971
2277600	N-HiTS	0.216	0.109	0.906
2288800	N-HiTS	0.180	0.012	1
2291597	N-HiTS	0.003	0.038	0.989
2292900	TFT	9.531	0.134	0.743
2293230	TFT	0.022	0.138	0.706
2297100	TFT	0.123	0.03	0.98
2298880	N-HiTS	0.050	0.008	1
2299950	N-HiTS	0.435	0.291	0.626
2301500	N-HiTS	0.615	0.074	0.96
2310947	N-HiTS	0.036	0.026	0.996
2312000	N-HiTS	0.134	0.033	0.993
2312500	N-HiTS	0.125	0.014	0.999
2312600	N-HiTS	0.112	0.012	0.999
2312700	N-HiTS	0.157	0.032	0.994
2315500	N-HiTS	0.610	0.015	0.999
2317620	N-HiTS	1.491	0.032	0.99
2320700	N-HiTS	0.040	0.042	0.992
2322800	N-HiTS	1.096	0.039	0.984
2323500	N-BEATS	11.722	0.043	0.994
2323592	LSTM	17.049	0.062	0.975
2326000	N-HiTS	0.116	0.019	0.998
2326900	N-HiTS	0.424	0.043	0.989
2328522	N-HiTS	4.377	0.062	0.971
2330100	N-HiTS	1.354	0.333	0.745
2358000	LSTM	122.650	0.176	0.946
2358789	N-HiTS	1.894	0.082	0.931
2359000	N-HiTS	1.552	0.035	0.988

2359500	N-HiTS	0.640	0.124	0.878
2365769	N-HiTS	4.183	0.799	0.453
2366000	N-HiTS	0.982	0.079	0.945
2366500	LSTM	35.083	0.188	0.945
2368000	LSTM	14.909	0.485	0.803
2370000	TFT	8.722	0.115	0.72
2376293	N-HiTS	1.879	2.274	0.318







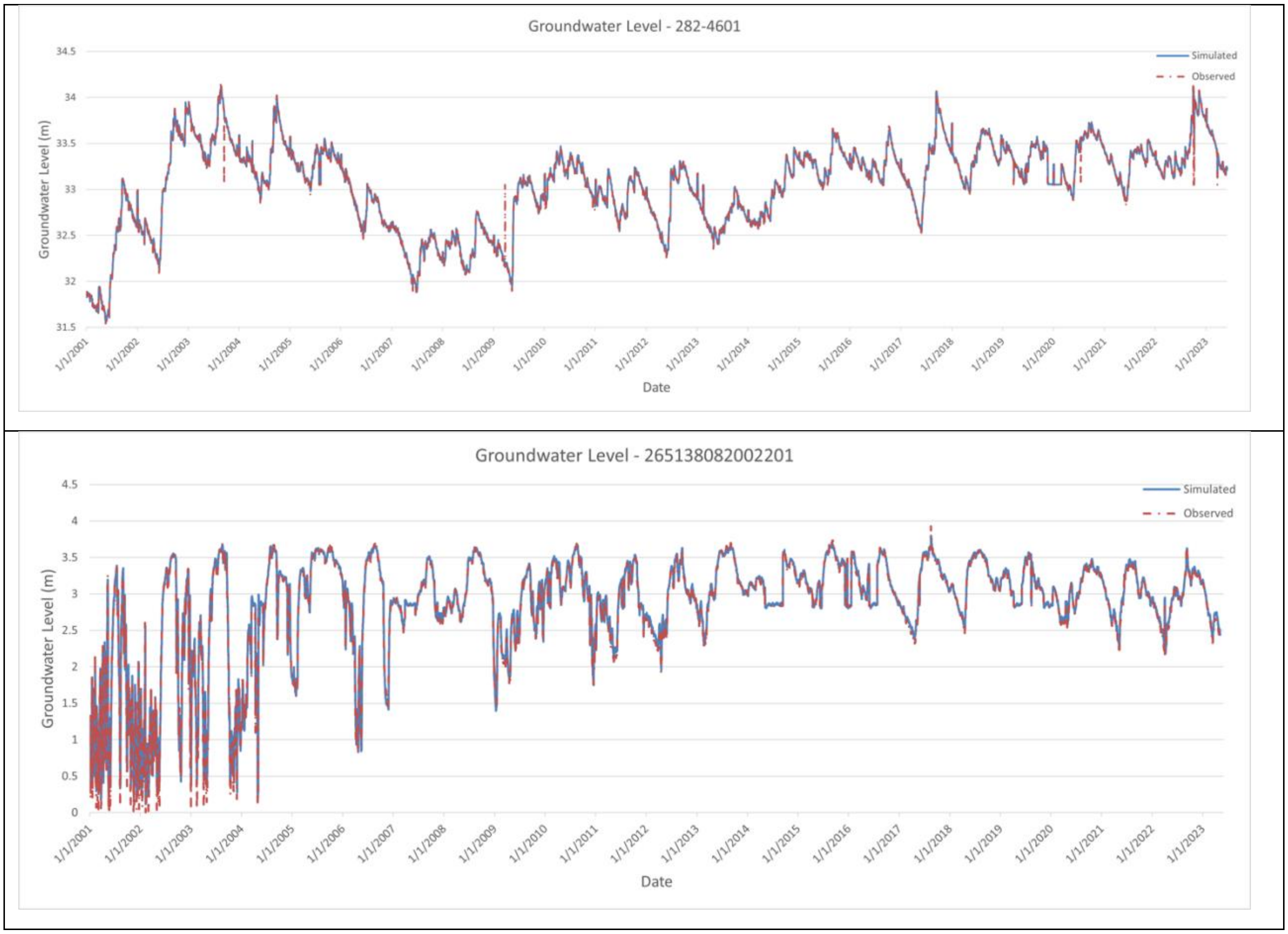
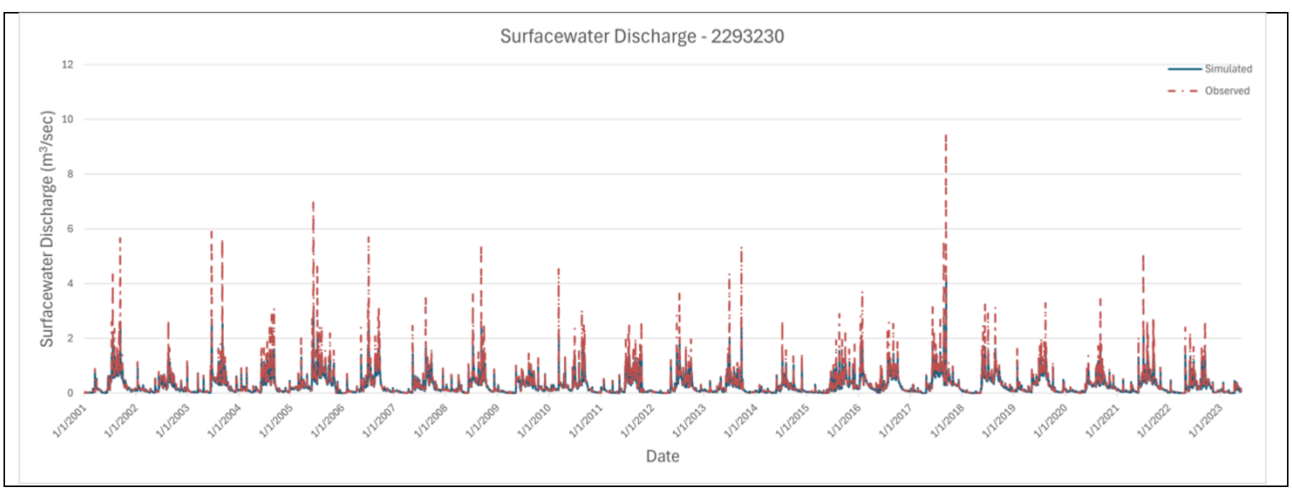
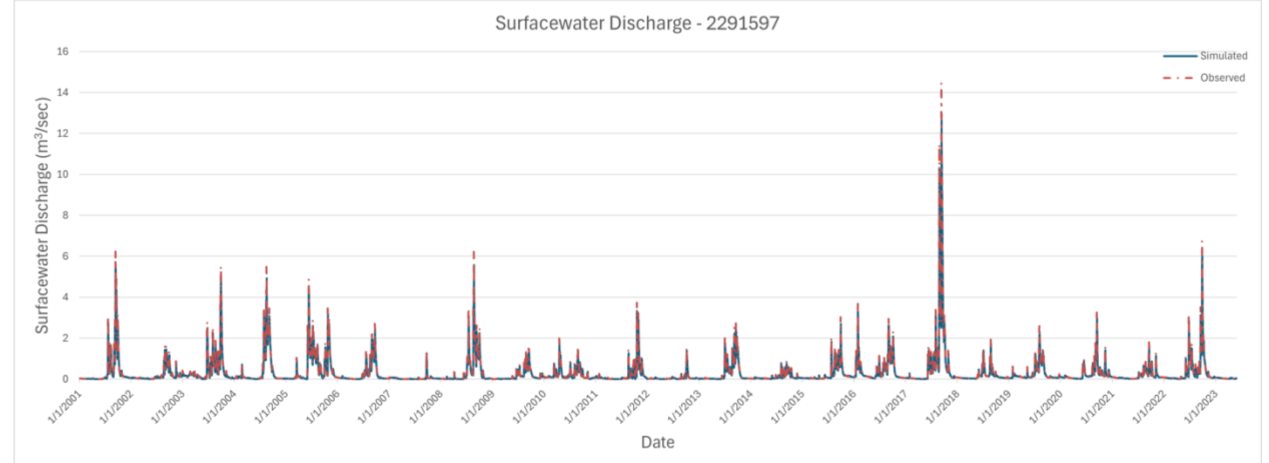
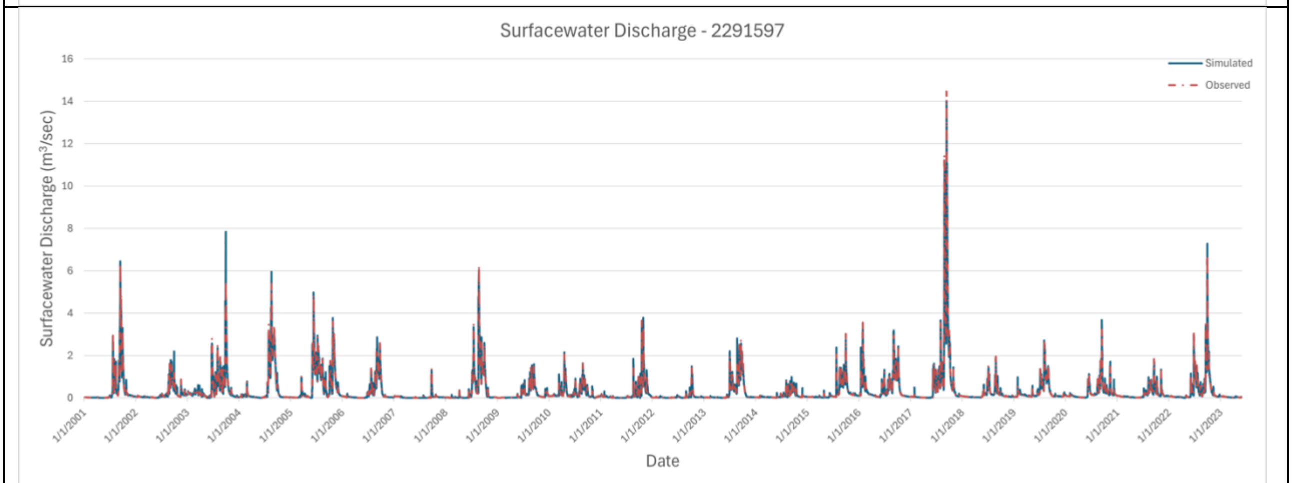
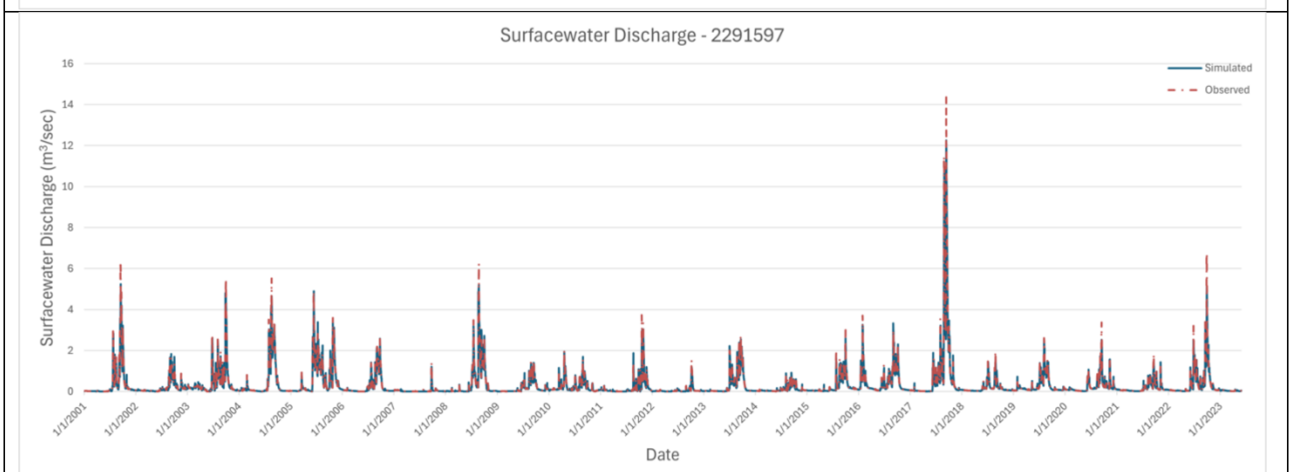


Fig 2: Observed and Predicted Ground Water Level on the whole dataset: a) LSTM, b) N-BEATS, c) N-HiTS, d) TFT, e) Informer.




Surfacewater Discharge - 2291597
Simulated
Observed
Surfacewater Discharge (m³/sec)
Date


Surfacewater Discharge - 2291597
Simulated
Observed
Surfacewater Discharge (m³/sec)
Date


Surfacewater Discharge - 2291597
Simulated
Observed
Surfacewater Discharge (m³/sec)
Date

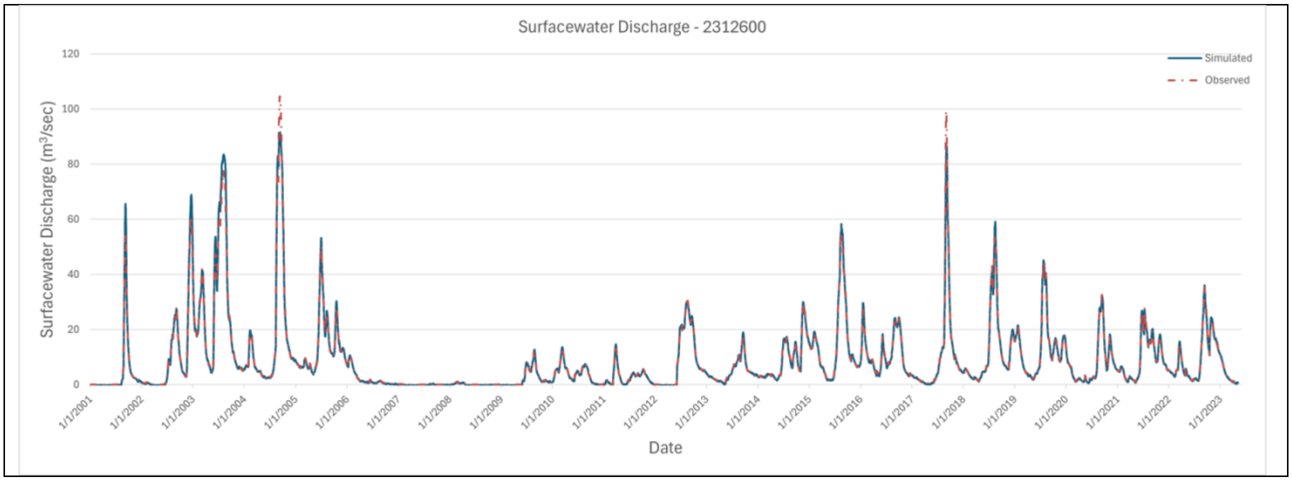
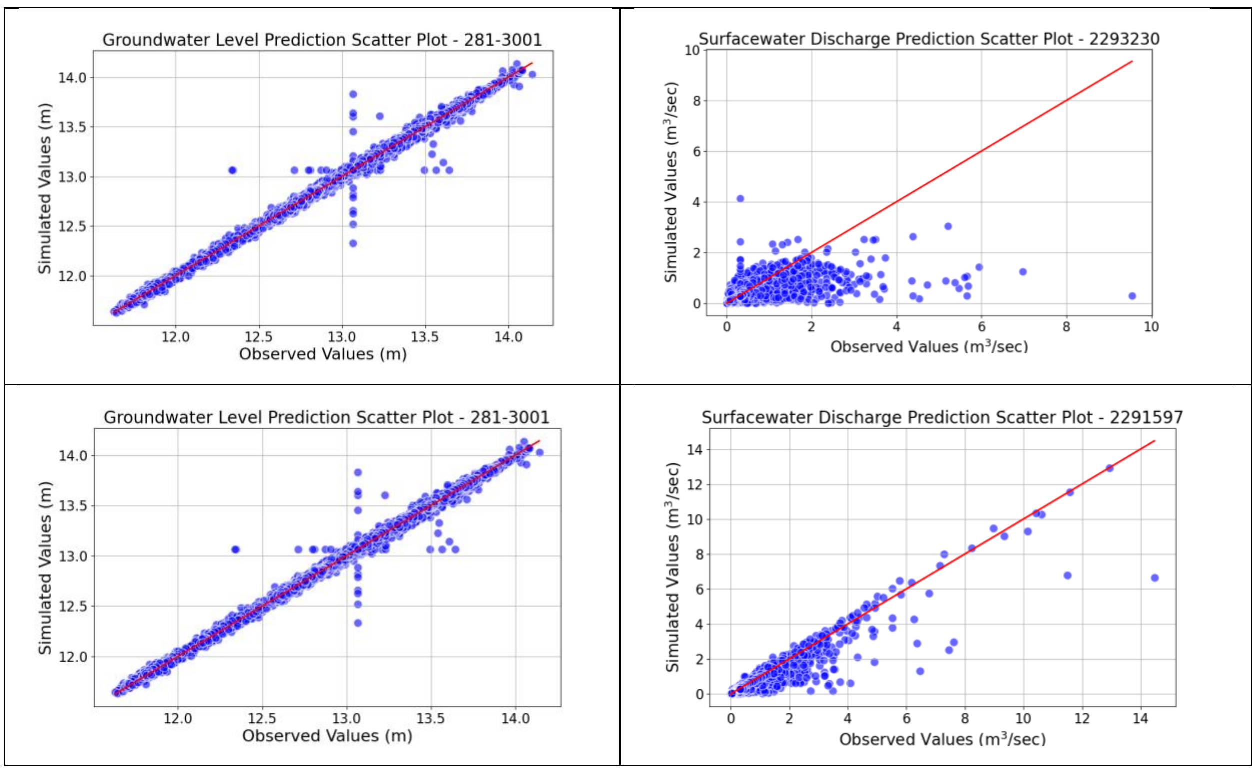


Fig 3: Observed and Predicted Surface Water Discharge on the whole dataset: a) LSTM, b) N-BEATS, c) N-HiTS, d) TFT, e) Informer.



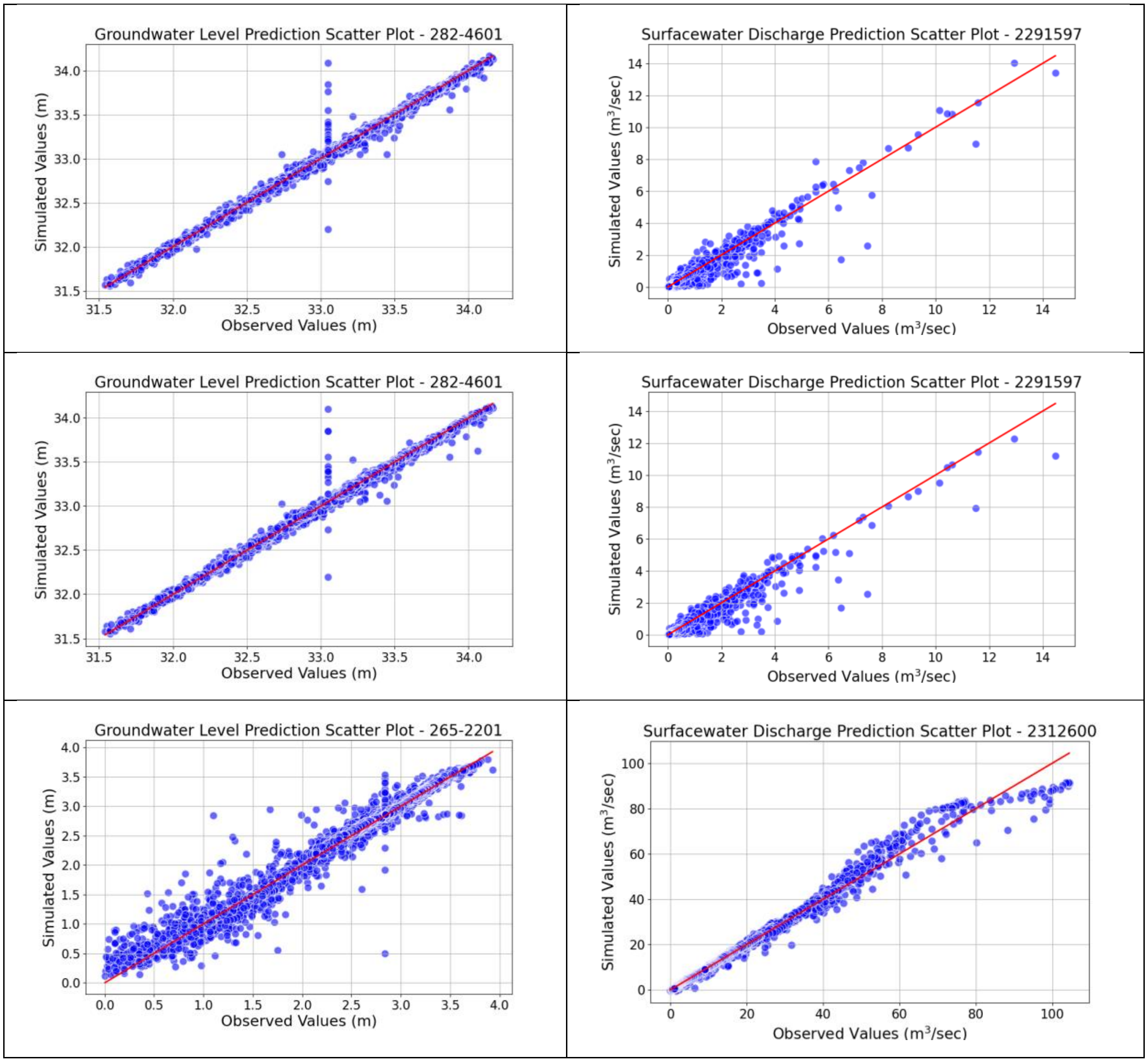


Fig 4: Observed and Predicted Ground water Level and Surface water Discharge Scatter Plots on the whole dataset: a) LSTM, b) N-BEATS, c) N-HiTS, d) TFT, e) Informer.

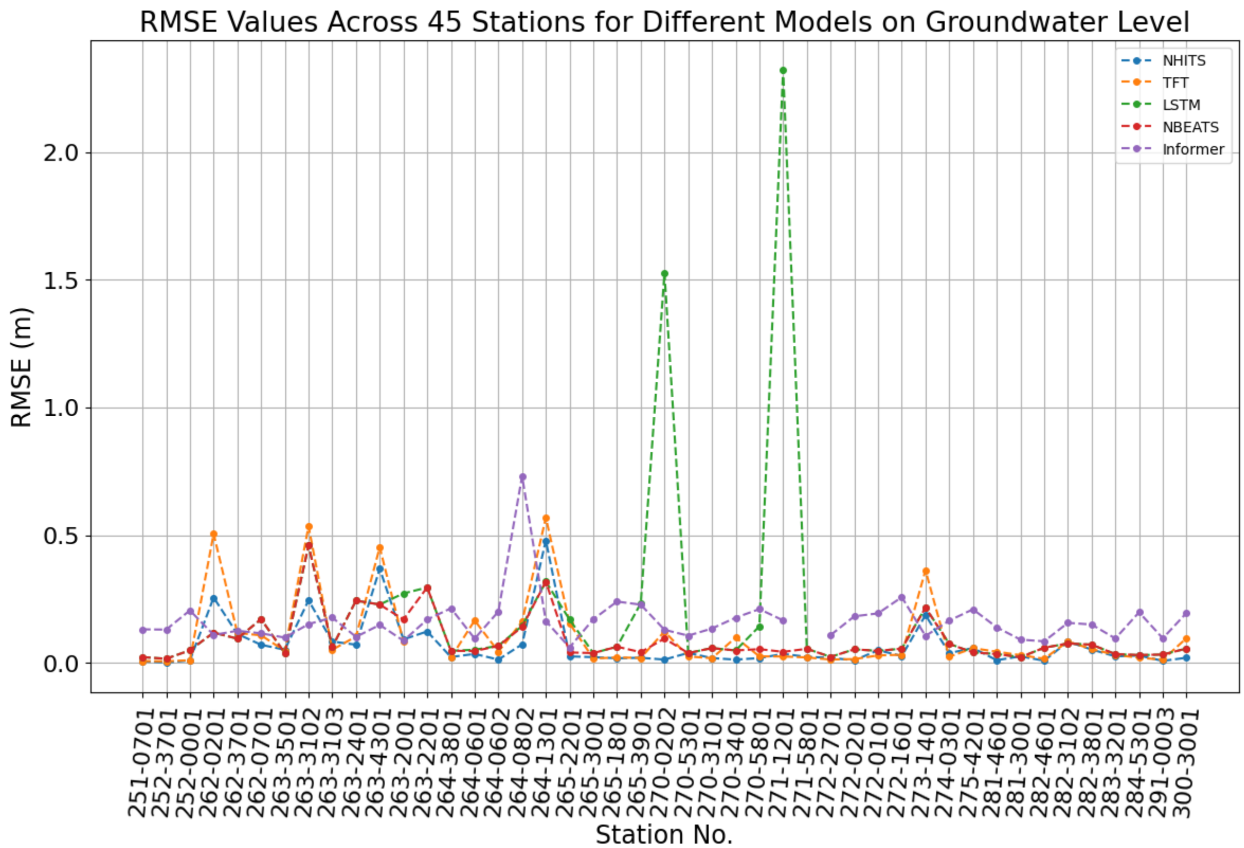


Fig 4(a): The above graph shows the RMSE values of five predictive models (N-HiTS, TFT, LSTM, N-BEATS, and Informer) across 45 ground water stations, illustrating the comparative accuracy of each model at different locations.

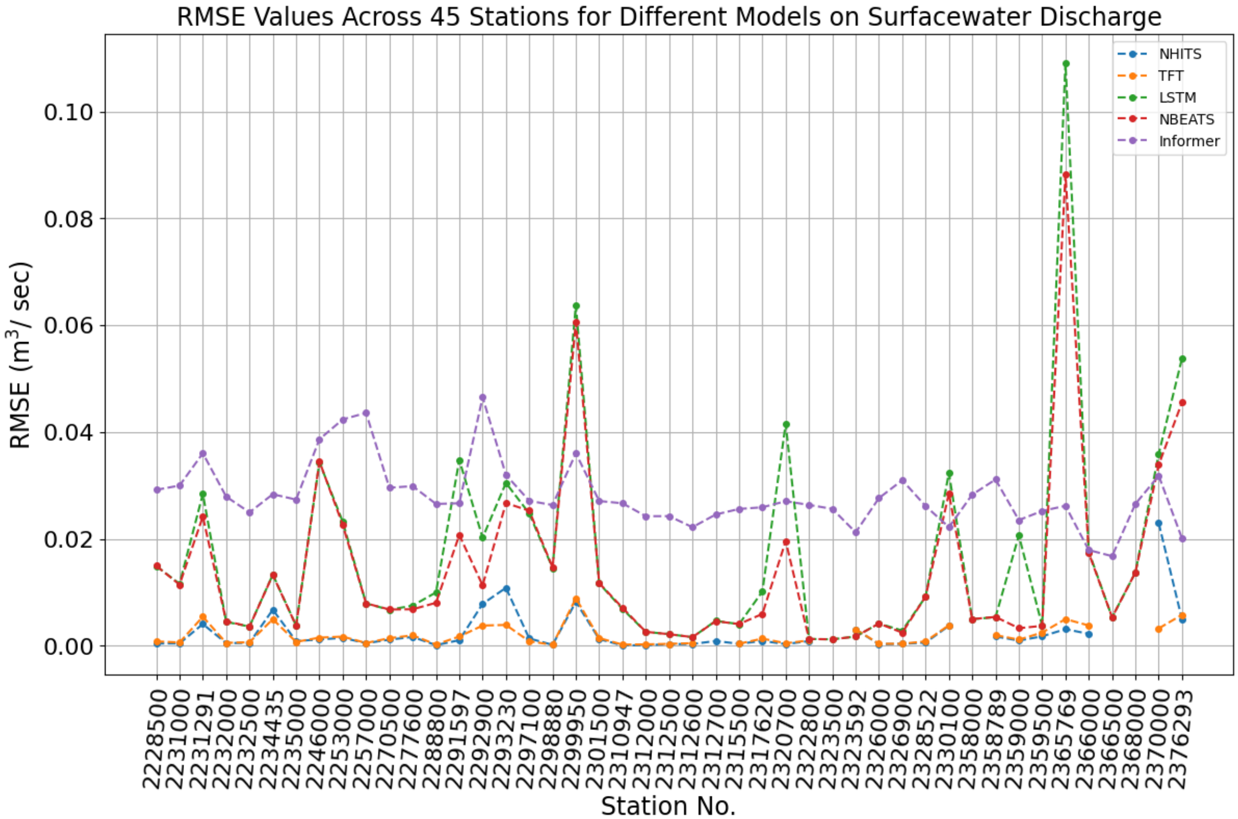


Fig 4(b): The above graph shows the RMSE values of five predictive models (N-HiTS, TFT, LSTM, N-BEATS, and Informer) across 45 surface water stations, illustrating the comparative accuracy of each model at different locations.

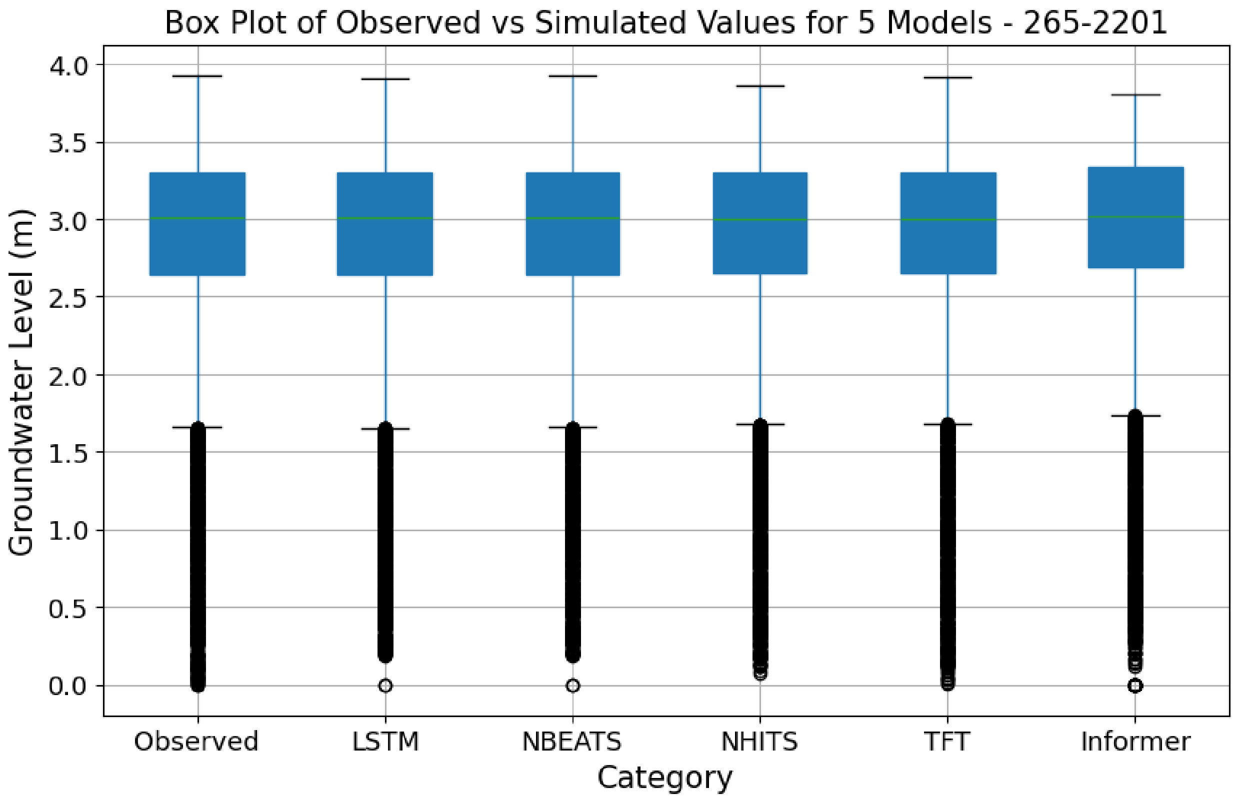


Fig 5(a): The above box plot illustrates the observed values against the predicted values from five different models across 265-2201 ground water stations.

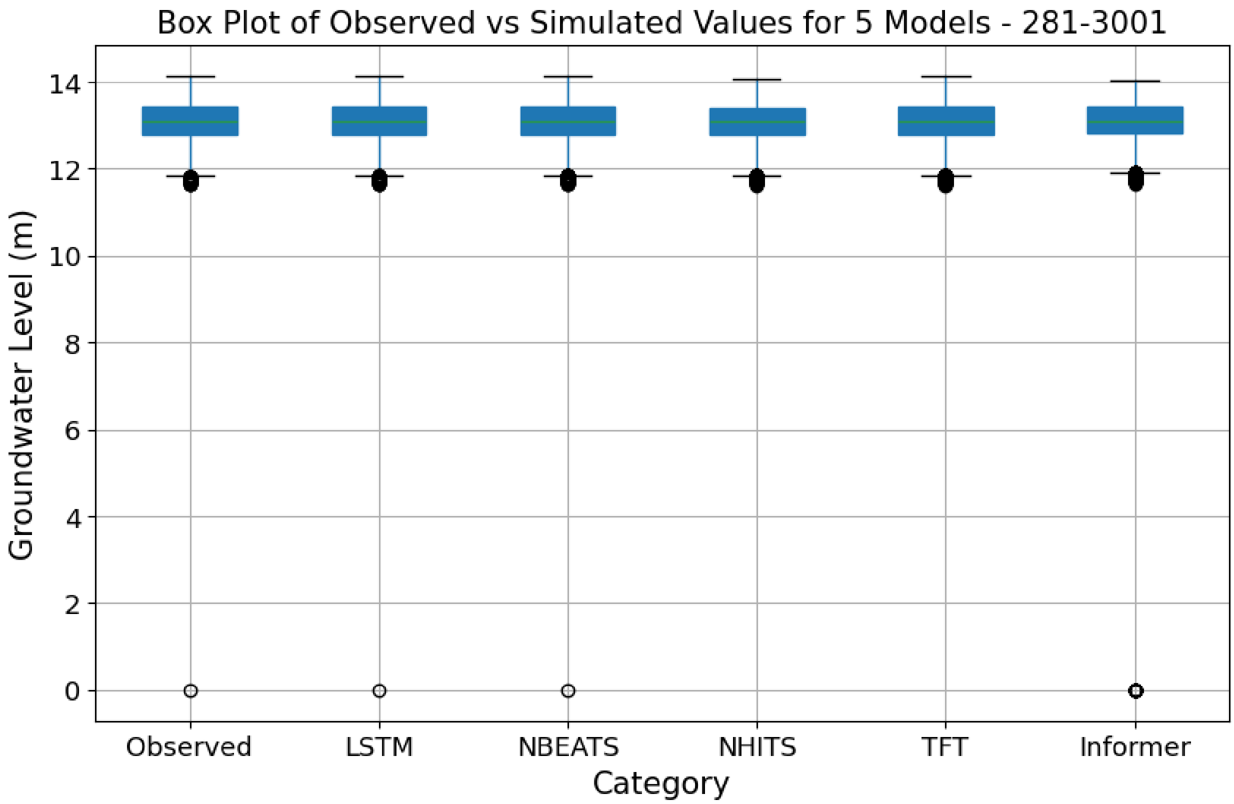


Fig 5(b): The above box plot illustrates the observed values against the predicted values from five different models across 281-3001 ground water stations.

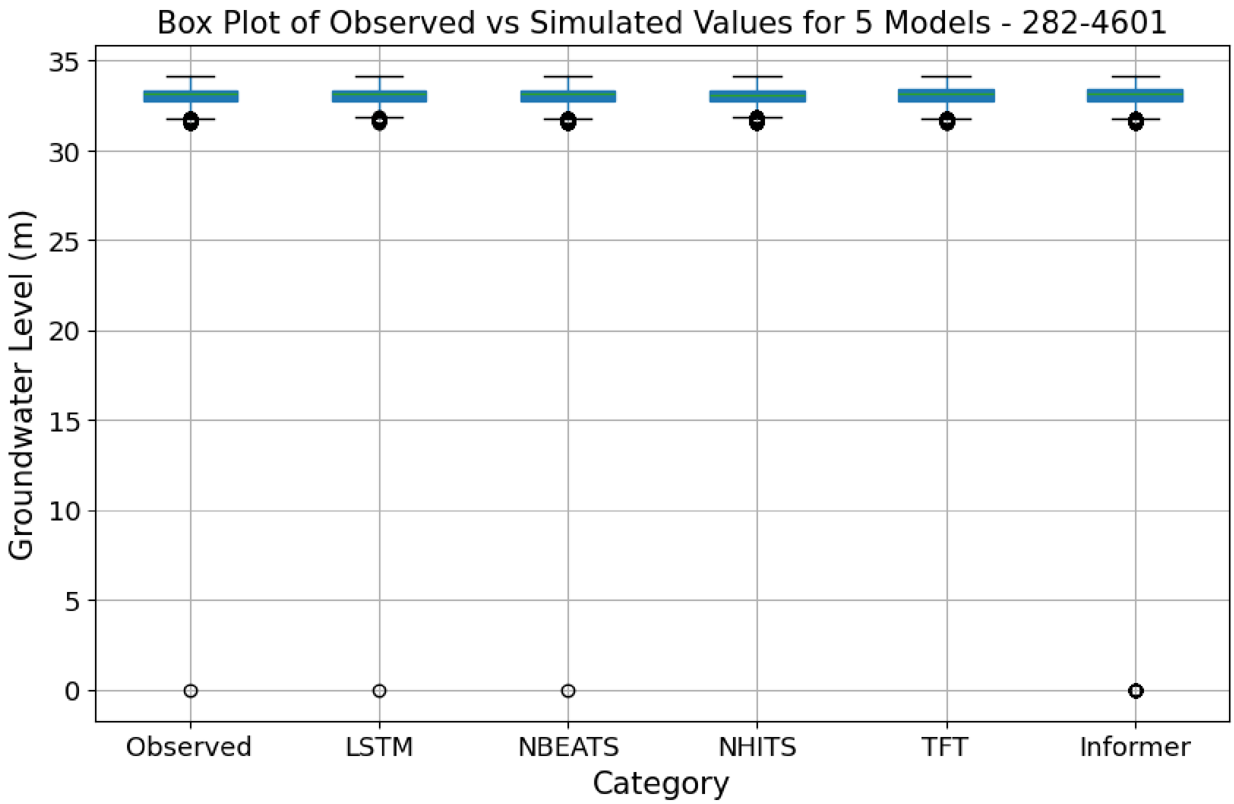


Fig 5(c): The above box plot illustrates the observed values against the predicted values from five different models across 282-4601 ground water stations.

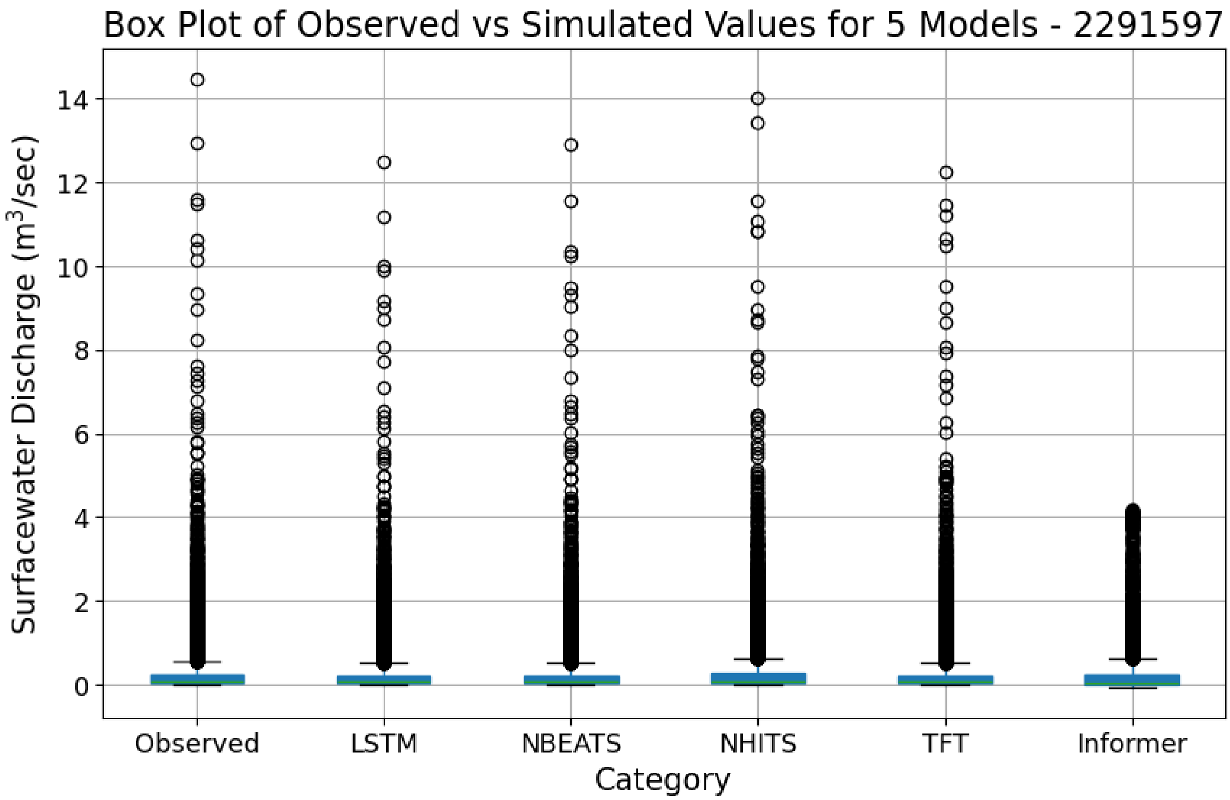


Fig 5(d): The above box plot illustrates the observed values against the predicted values from five different models across 2291597 surface water stations.

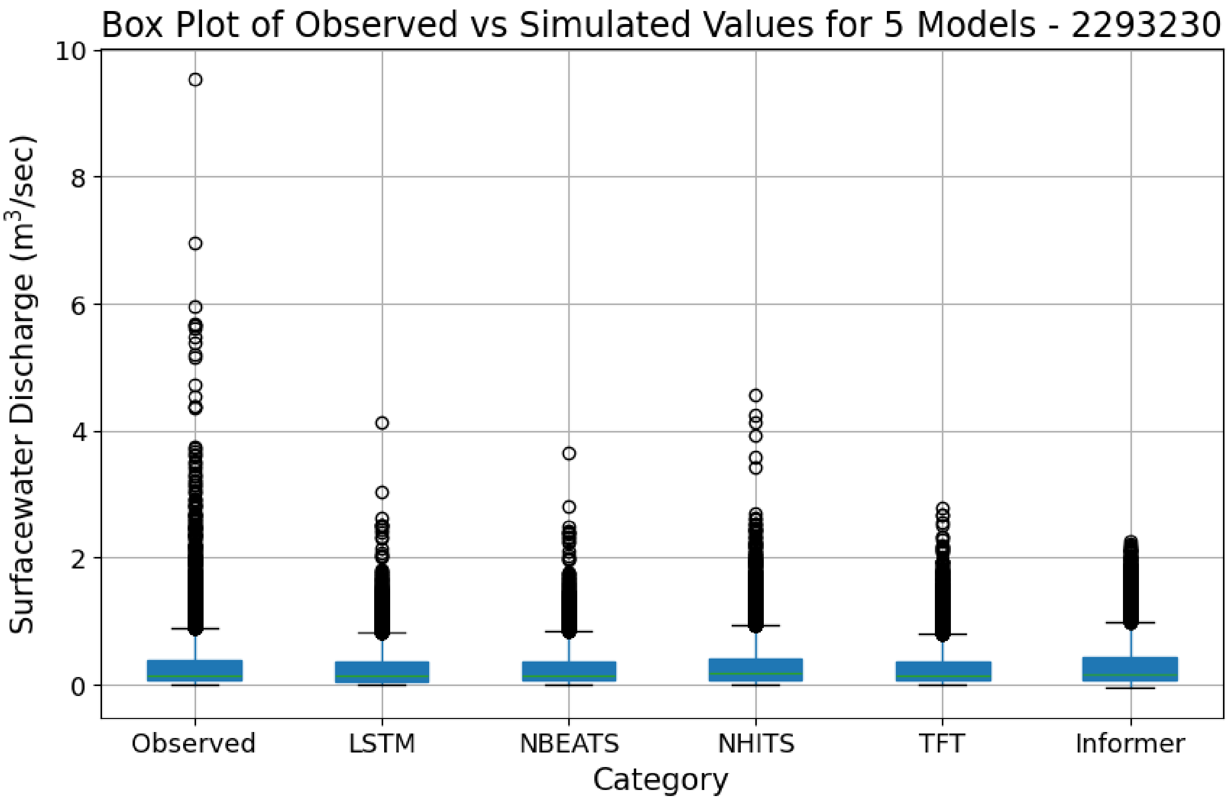


Fig 5(e): The above box plot illustrates the observed values against the predicted values from five different models across 2293230 surface water stations.

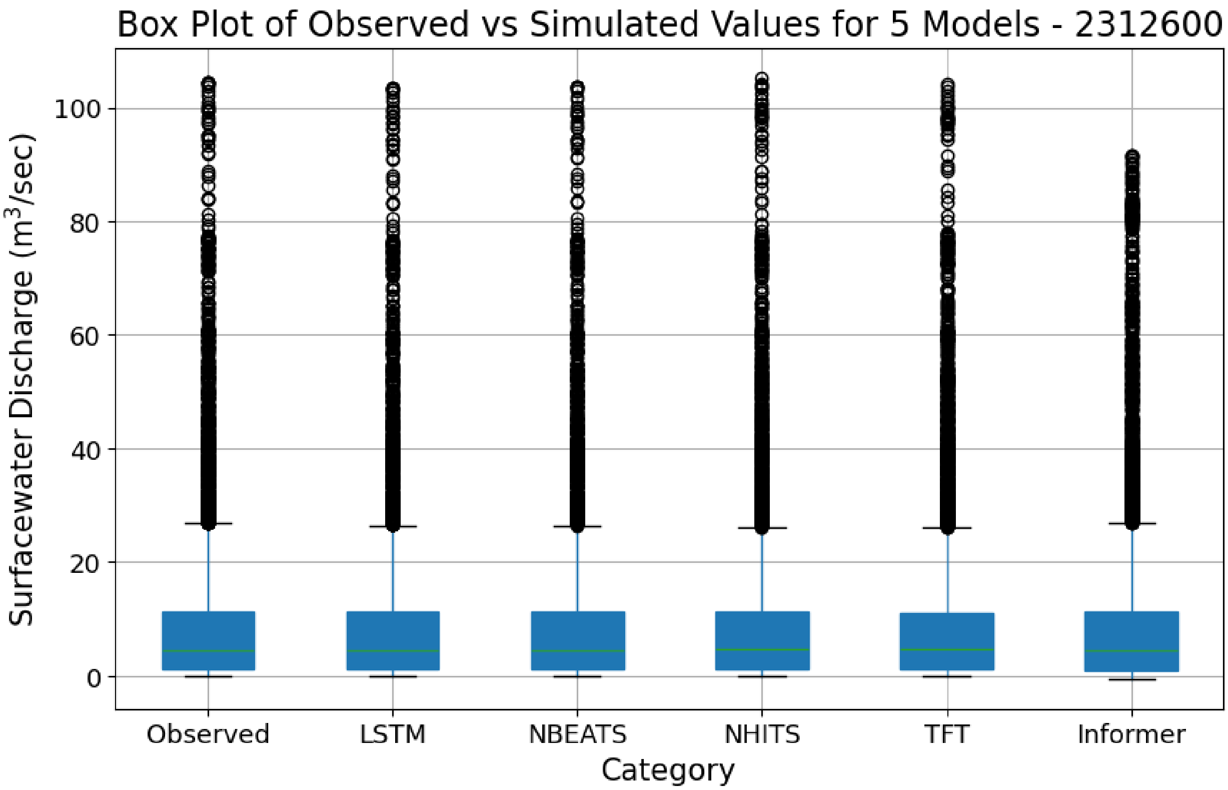
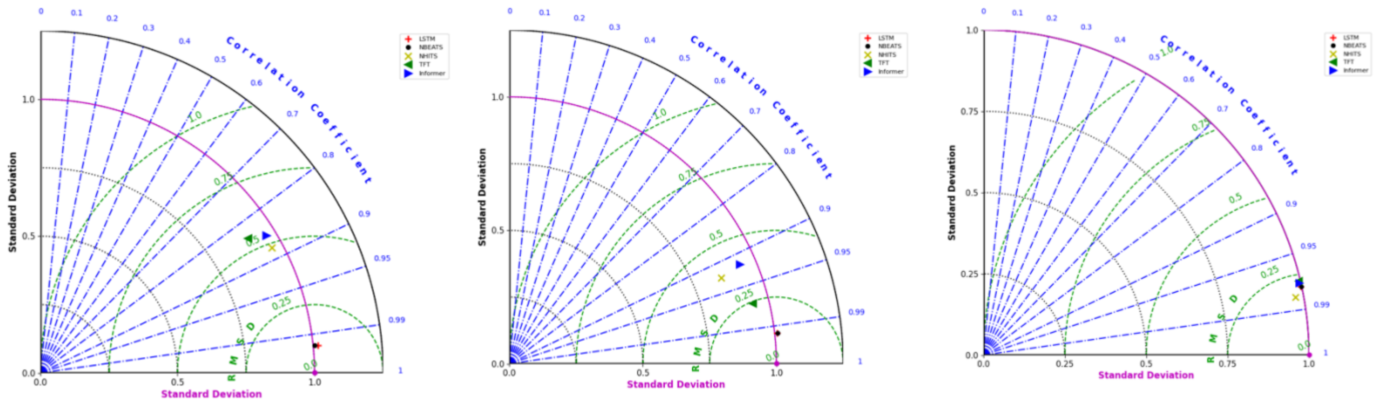
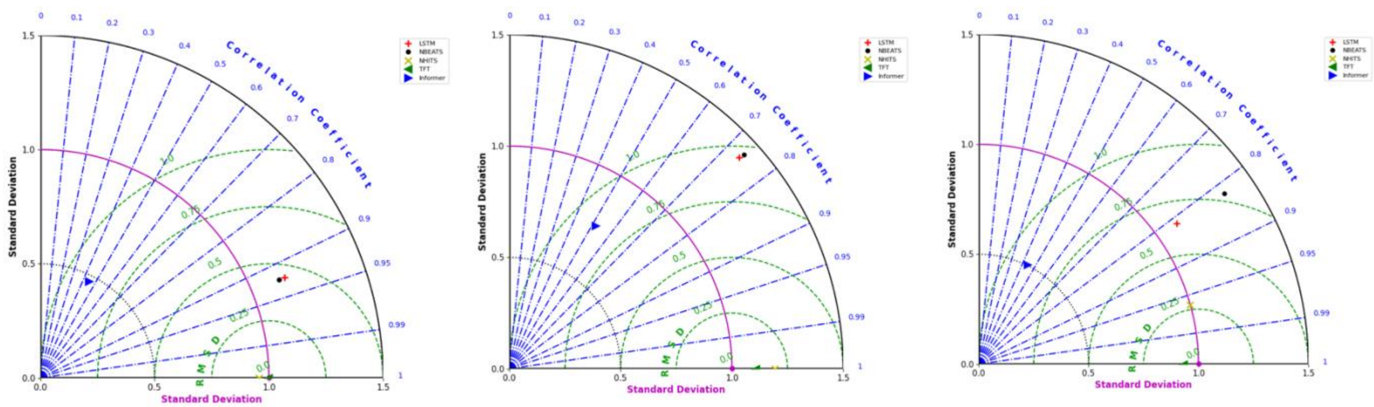


Fig 5(f): The above box plot illustrates the observed values against the predicted values from five different models across 2312600 surface water stations.



265-2201	281-3001	282-4601

Fig 6(a): Taylor Diagrams for selected 3 Ground Water stations in the testing phase



2291597	2293230	2312600

Fig 6(b): Taylor Diagrams for selected 3 Surface Water stations in the testing phase

These diagrams are an effective tool not only for visualizing model performance in terms of accuracy and correlation with observed data but also for quickly identifying which models are best suited for specific types of hydrological data, as demonstrated by the comparative analysis of surface and groundwater stations.

3.2 Forecasting

In this section, we extend our analysis by using the N-HiTS model, our best-performing model, to forecast ground water levels and surface water discharge for two selected stations — one groundwater station, 282202081384601, and one surface water station, 2291597. We performed forecasting for the next 3, 6, 9, and 12 months based on the available 23 years of observed data. The forecasting results are presented visually, where a dashed line represents the predicted values following the observed data.

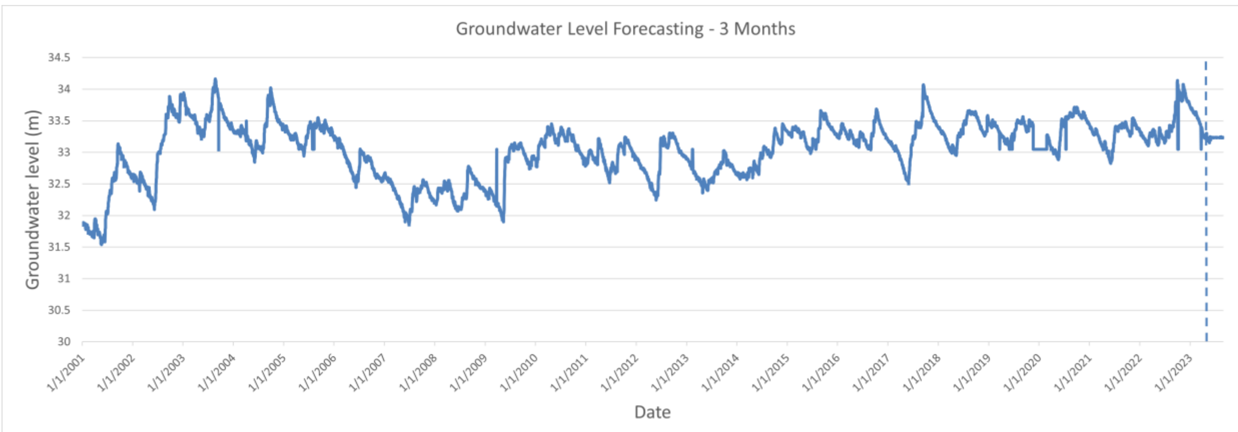


Fig 7(a): The graph above showcases the observed groundwater data over time, followed by a forecast for the next three months, providing a visual comparison between historical trends and predicted future values.

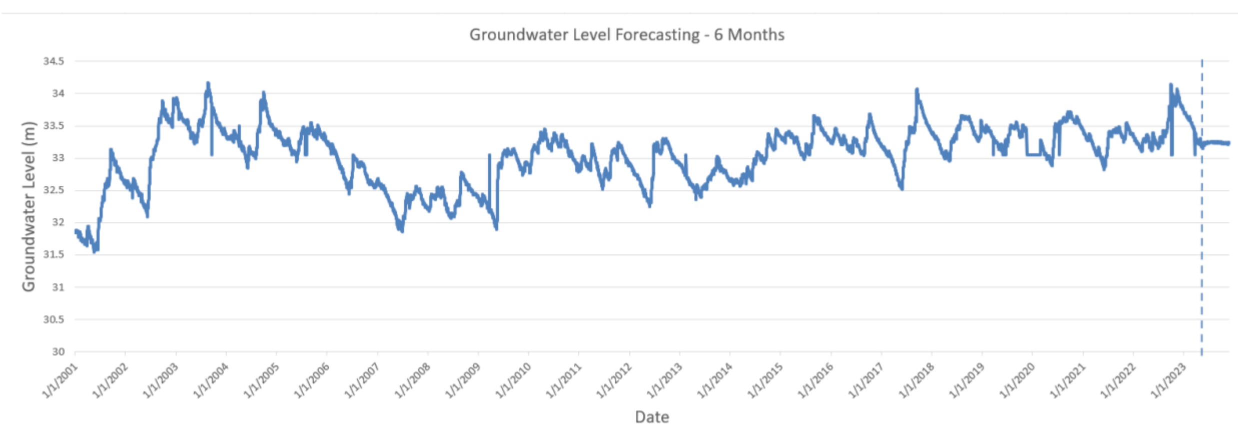


Fig 7(b): The graph above showcases the observed groundwater data over time, followed by a forecast for the next six months, providing a visual comparison between historical trends and predicted future values.

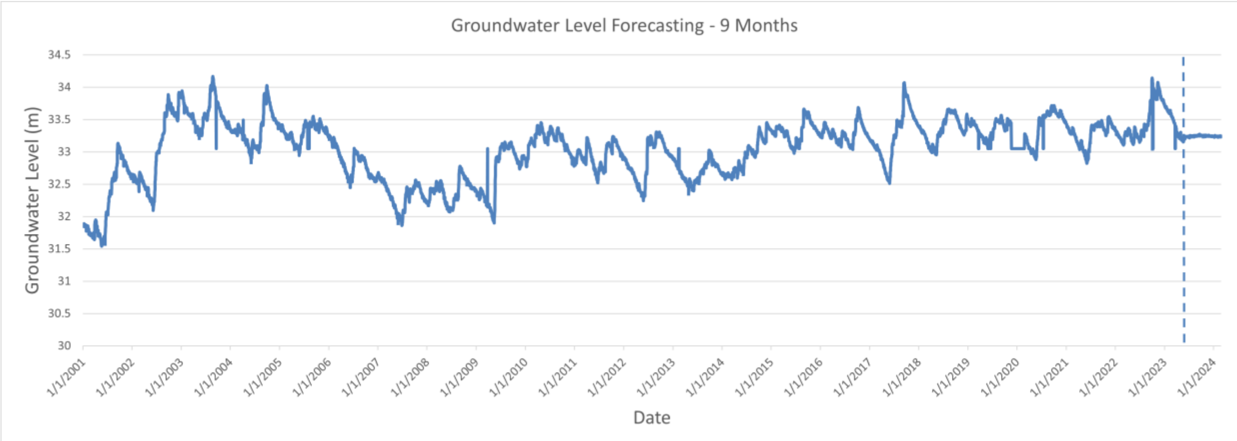


Fig 7(c): The graph above showcases the observed groundwater data over time, followed by a forecast for the next nine months, providing a visual comparison between historical trends and predicted future values.

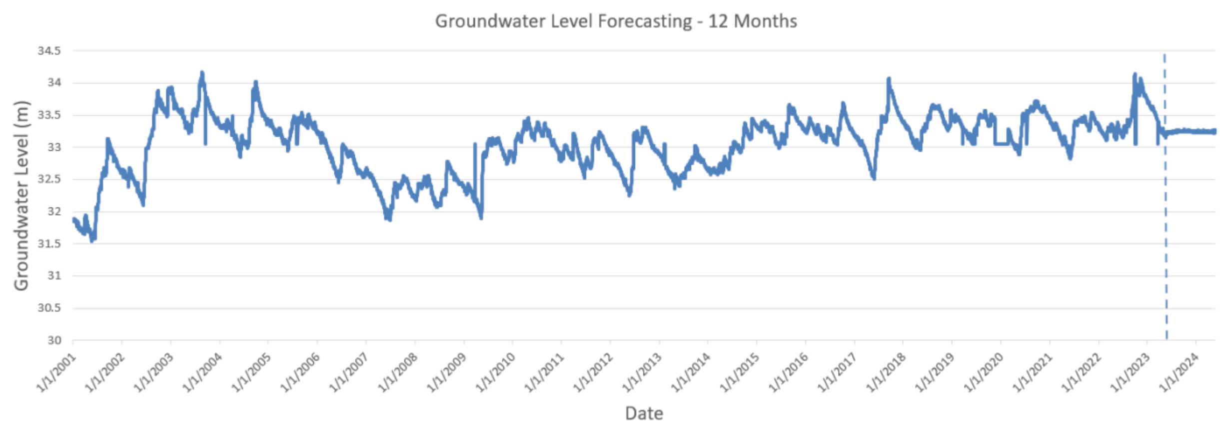


Fig 7(d): The graph above showcases the observed groundwater data over time, followed by a forecast for the next twelve months, providing a visual comparison between historical trends and predicted future values.

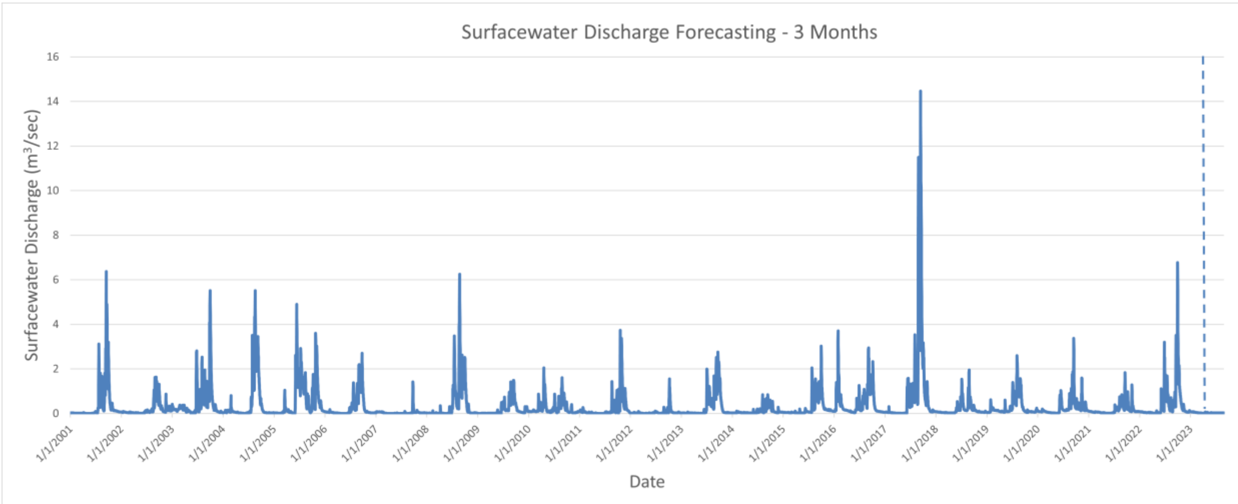


Fig 8(a): The above graph showcases the observed surface water data over time, followed by a forecast for the next three months, providing a visual comparison between historical trends and predicted future values.

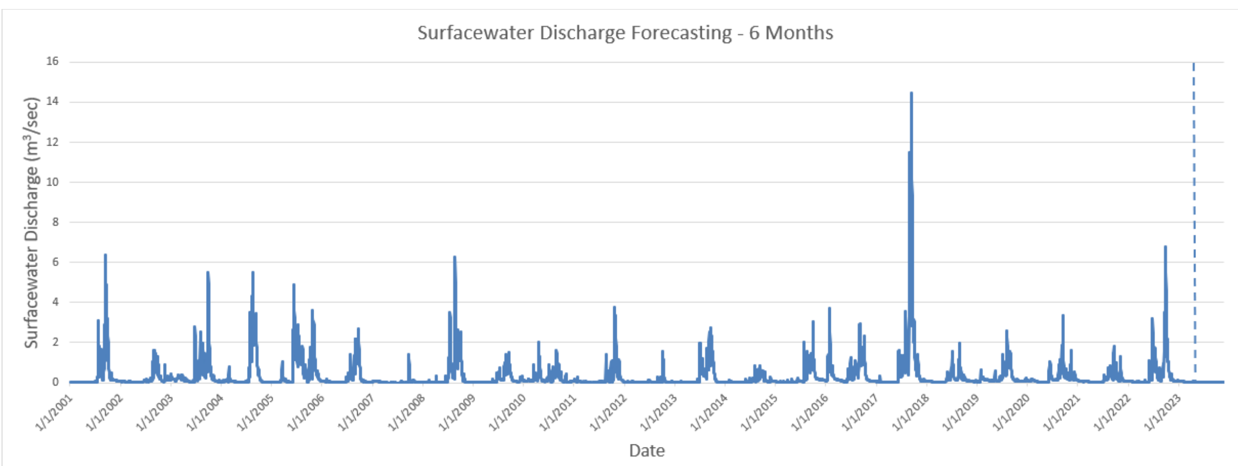


Fig 8(b): The above graph showcases the observed surface water data over time, followed by a forecast for the next six months, providing a visual comparison between historical trends and predicted future values.

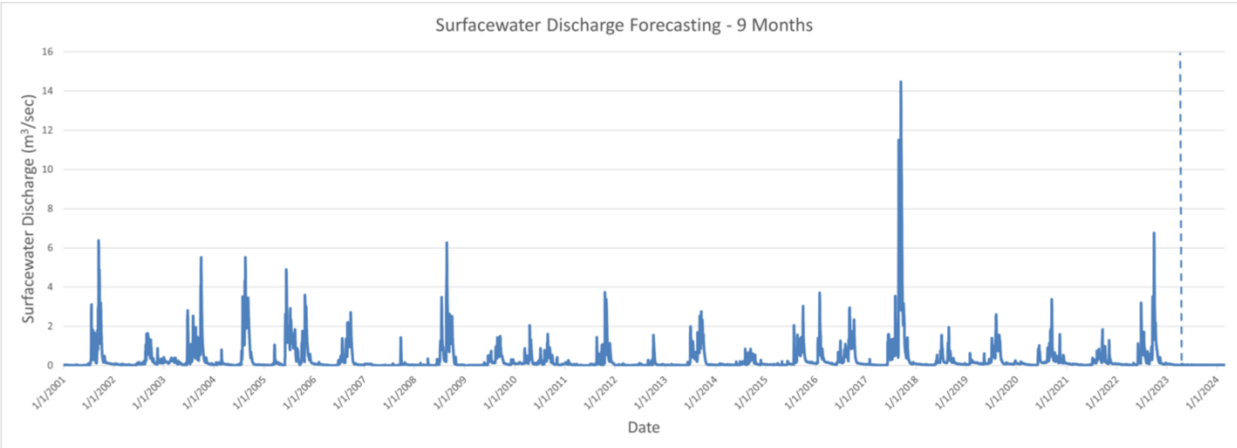


Fig 8(c): The above graph showcases the observed surface water data over time, followed by a forecast for the next nine months, providing a visual comparison between historical trends and predicted future values.

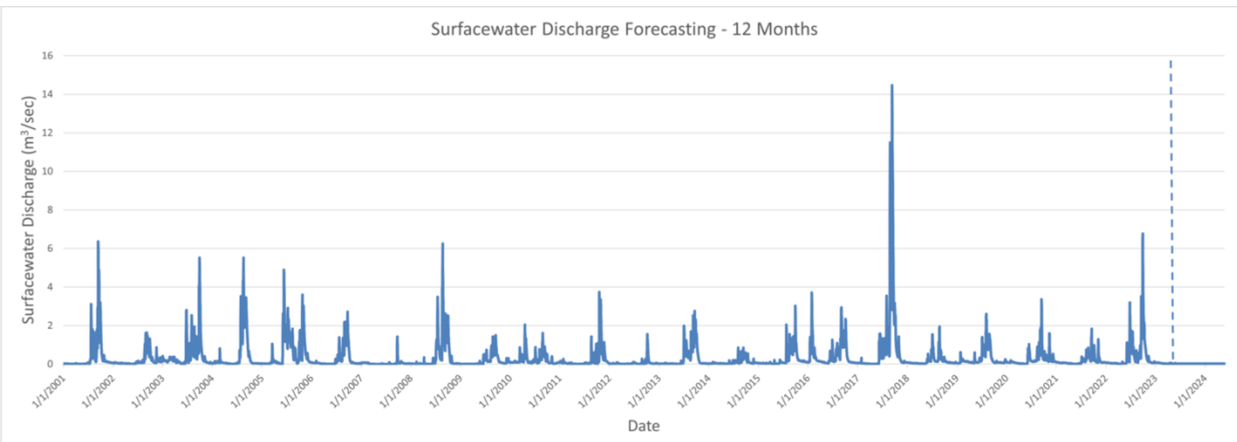


Fig 8(d): The graph above showcases the observed groundwater data over time, followed by a forecast for the next twelve months, providing a visual comparison between historical trends and predicted future values.

However, the forecasting results did not match the expected accuracy. The observed data exhibits significant fluctuations and variability, but the predictions from N-HiTS failed to capture this complexity, leading to less accurate forecasts. One possible reason for this underperformance is the limited input variables used in the model. Specifically, we relied solely on the previous day's water levels or surface water discharge to predict future values, which may not provide enough context to effectively predict long-term trends in such dynamic systems.

4. Discussion

The results of the present study determined that all employed models—LSTM, N-BEATS, N-HiTS, TFT, and Informer—can accurately predict water levels and discharge, showcasing their potential in hydrological forecasting. Several studies have compared these models in various contexts, providing a broader understanding of their capabilities and performance.

For instance, in a study by Nixtla (2022), various neural forecasting methods, including LSTM, N-BEATS, and N-HiTS, were evaluated for time series forecasting. The study found that N-HiTS consistently outperformed LSTM and N-BEATS across multiple datasets, aligning with our findings, where N-HiTS demonstrated superior performance in predicting water discharge and levels in Florida. Similarly, (Challu et al., 2023) highlighted N-HiTS's advanced architecture and superior performance compared to TFT in time series forecasting, further supporting our results.
In contrast, some studies have reported different outcomes. For example, a study by Towards Data Science (2022) found that TFT outperformed N-HiTS in specific time series forecasting tasks, suggesting that TFT may be more suitable for certain scenarios. Additionally, a study on multi-horizon building energy forecasting highlighted TFT's robustness and effectiveness, often surpassing models like N-BEATS and N-HiTS under specific conditions, such as when handling multivariate time series data (Zhang et al., 2024).

These contrasting results highlight the importance of context and application when selecting a forecasting model. While N-HiTS and TFT showed strong performance in our study, other models like LSTM and N-BEATS may be more effective in different settings or when specific types of data patterns are present. This variability underscores the need for continued research and model evaluation across diverse datasets and conditions to fully understand the strengths and limitations of each approach. The diversity in model performance suggests that no single model can be deemed universally superior. Each model has unique strengths tailored to specific types of data and forecasting requirements. Therefore, it's crucial to select the most appropriate model based on the specific characteristics of the dataset and the forecasting goals.

In light of the findings, N-HiTS appears particularly promising for several reasons. First, N-HiTS focuses on constructing hierarchical time series forecasts, which enhances both accuracy and computational efficiency. Second, it has demonstrated superior performance across various datasets, including finance and energy, indicating its versatility. Third, N-HiTS is designed to handle large-scale time series data efficiently, making it suitable for extensive datasets like the one used in this study. Fourth, the model's architecture allows it to be adapted to different types of forecasting problems, enhancing its adaptability. Lastly, N-HiTS has shown robustness in capturing nonlinear relationships and temporal dependencies, which are critical for accurate hydrological forecasting. Moreover, N-HiTS has been noted for its computational efficiency, requiring less training time compared to transformer-based and fully connected models, as demonstrated in a study by Towards AI (2023).

The computational efficiency of each model varied significantly, reflecting their differing demands on system resources. Notably, the Informer model required the most substantial amount of time to execute, consuming nearly 30 min on a 32GB Windows system and approximately 25 min on a MacBook Air equipped with an M2 chip. This extensive runtime can be attributed to the model's complex architecture, which processes large-scale data sequences, making it computationally intensive.

In stark contrast, the LSTM model demonstrated remarkable speed, completing its run in just 4 minutes for 100 epochs on both tested systems. This efficiency underscores LSTM's streamlined data processing capabilities, which are less resource-intensive while still delivering robust results. Following closely, the N-BEATS model required roughly 8 minutes to complete the same number of epochs, showcasing its ability to efficiently handle time series data without the computational overhead associated with more complex models. The N-HiTS model displayed intermediate efficiency, taking about 15 minutes to execute 100 epochs. Its performance strikes a balance between computational demand and forecasting precision, making it a practical choice for moderate-sized datasets. Lastly, the TFT required approximately one hour to complete 100 epochs, reflecting its intricate internal mechanisms designed to capture and integrate multiple time-dependent variables effectively.

These runtime variations highlight the importance of considering both computational efficiency and predictive performance when selecting models for hydrological forecasting, especially in scenarios where resource constraints are a significant concern. The choice of model can thus greatly influence the overall feasibility and speed of research projects, particularly in data-intensive environments like hydrological forecasting. This analysis not only aids in model selection but also informs infrastructure decisions, ensuring that computational resources are aligned with the demands of specific models to optimize both performance and efficiency in practical applications.

Table 15: Models with their Best and worst run times on two different systems.

Model	System 1 (Windows 32GB RAM)		System 2 (M2 AIR 8GB RAM)	
	Max Execution Time (100 epochs)	Min Execution Time (100 epochs)	Max Execution Time (100 epochs)	Min Execution Time (100 epochs)
LSTM	5 Min	3 Min	4 Min	2 Min
N-BEATS	15 Min	12 Min	10 Min	6 Min
TFT	20 Min	15 Min	15 Min	12 Min
N-HiTS	20 Min	15 Min	15 Min	12 Min
Informer	30 Min	25 Min	25 Min	22 Min

Limitations

Despite the promising results obtained in this study, several limitations should be acknowledged. Firstly, the datasets used span 23 years from 2001 to 2023 and cover daily records from 45 stations each for surface water discharge and groundwater levels in Florida. While extensive, this dataset is geographically limited. The unique hydrological characteristics of Florida may differ significantly from other regions, potentially limiting the generalizability of our findings. The deep learning models employed, including LSTM, N-BEATS, N-HiTS, TFT, and Informer, were trained and tested on this specific dataset. Although these models performed well, their effectiveness might vary with different datasets or hydrological conditions.

Another limitation is the computational resources required for training and fine-tuning these deep-learning models. Training these models is computationally intensive and time-consuming, which may pose challenges for practical implementation, particularly in real-time forecasting scenarios or in regions with limited computational infrastructure. Our approach focused on predicting variables using only their historical values, which may not fully capture long-term dependencies or trends crucial for certain hydrological applications. Future research should integrate additional temporal features and external factors to enhance model accuracy and robustness.

Future Enhancement

For future enhancements, we recommend integrating additional temporal features and external factors such as climatic conditions, land use changes, and human activities to improve model accuracy and robustness. Exploring newer and more advanced deep learning models could further enhance predictive performance. Expanding the geographical scope to include diverse hydrological regions would improve the generalizability of the findings. Additionally, incorporating multivariate analysis to capture interactions between different hydrological variables could provide deeper insights. Real-time data assimilation and the development of more efficient computational techniques would support practical implementation in water resource management. Lastly, implementing ensemble modeling approaches could leverage the strengths of multiple models, potentially leading to more reliable and accurate forecasts.

5. Conclusion

This study explored the application of deep learning models for forecasting surface water discharge and groundwater levels across Florida. By comparing the performance of LSTM, N-BEATS, N-HiTS, TFT, and Informer models, we demonstrated the potential of deep learning approaches in capturing the complex temporal dynamics of hydrological variables. Our findings reveal that the N-HiTS model consistently outperformed the other models regarding RMSE, NRMSE, and NSE across all stations, establishing its superiority in forecasting accuracy.

The use of deep learning models, particularly N-HiTS, offers a promising avenue for enhancing the predictive capabilities of hydrological models. This advancement is crucial for informed decision-making in water resource management, enabling better anticipation of water availability, flood risks, and drought conditions. Furthermore, the insights gained from this study contribute to the ongoing development of more accurate and reliable hydrological forecasting models, which are essential for sustainable water resource management in the face of climate change and increasing water demands.

Future research should focus on further refining deep learning models, exploring their integration with traditional hydrological models, and expanding their application to other regions with varying hydrological characteristics. Additionally, investigating the potential of deep learning models in predicting other hydrological variables, such as water quality parameters, could provide a more comprehensive understanding of water systems and their management.

References

[1] Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023, June). N-HiTS: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 6, pp. 6989-6997).

[2] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764.

[3] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 12, pp. 11106-11115).

[4] Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.

[5] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235-1270.

[6] Rahman, A. S., Hosono, T., Quilty, J. M., Das, J., & Basak, A. (2020). Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms—Advances *in Water Resources*, *141*, 103595.

[7] Liu, W., Yu, H., Yang, L., Yin, Z., Zhu, M., & Wen, X. (2021). Deep learning-based predictive framework for groundwater level forecast in arid irrigated areas. *Water*, *13*(18), 2558.

[8] Dogan, A., Demirpence, H., & Cobaner, M. (2008). Prediction of groundwater levels from lake levels and climate data using the ANN approach. *Water Sa*, *34*(2), 199-208.

[9] Lee, S., Lee, K. K., & Yoon, H. (2019). Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. *Hydrogeology Journal*, *27*(2).

[10] Soebroto, A. A., Cholissodin, I., Pratiwi, D. E., & Putra, G. P. W. (2022). Comparative Study of SVR, Regression and ANN Water Surface Forecasting for Smart Agriculture. *HABITAT*, *33*(1), 86-92.

[11] Aderemi, B. A., Olwal, T. O., Ndambuki, J. M., & Rwanga, S. S. (2023). Groundwater levels forecasting using machine learning models: A case study of the groundwater region 10 at Karst Belt, South Africa. Systems and Soft Computing, 5, 200049.

[12] Abi Nader, A., Albaric, J., Steinmann, M., Hibert, C., Malet, J. P., Sue, C., ... & Boetsch, A. (2023). Machine learning prediction of groundwater heights from passive seismic wavefield. Geophysical Journal International, 234(3), 1807-1818.
[13] Muhumuza, K. (2020). Computational methods for seismic imaging and monitoring (Doctoral dissertation, Itä-Suomen yliopisto).
[14] Onyebuchi, O. E. (2020). Integrated geophysical methods for near-surface site characterization in South Africa (Doctoral dissertation, Faculty of Science, University of the Witwatersrand, Johannesburg).
[15] Abi Nader, A., Albaric, J., Steinmann, M., Hibert, C., Malet, J. P., Pohl, B., & Sue, C. (2023, May). Groundwater Heights Prediction from Seismic Waves with Machine Learning. In EGU General Assembly Conference Abstracts (pp. EGU-1601).
[16] Jadoud, M., El Achheb, A., Laftouhi, N., Namous, M., Khouz, A., Trindade, J., ... & Rachidi, S. (2023). Machine Learning Models for Spatial Prediction of Groundwater Potentiality in a Large Semi-Arid Mountainous Region: Application to the Rherhaya Watershed, High Atlas, Morocco (No. EGU23-16375). Copernicus Meetings.
[17] Huang, X., Gao, L., Zhang, N., Crosbie, R. S., Ye, L., Liu, J., ... & Bryan, B. A. (2023). A top-down deep learning model for predicting spatiotemporal dynamics of groundwater recharge. Environmental Modelling & Software, 167, 105778.
[18] Patra, S. R., & Chu, H. J. (2023). Regional groundwater sequential forecasting using global and local LSTM models. Journal of Hydrology: Regional Studies, 47, 101442.
[19] Koch, J., Kidmose, J., Liu, J., Schneider, R., Stisen, S., & Troldborg, L. (2023, May). Multi-step-ahead forecasting of groundwater levels using a Long Short-Term Memory based Encoder-Decoder (LSTM-ED) model. In EGU General Assembly Conference Abstracts (pp. EGU-3056).
[20] Alabdulkreem, E., Alruwais, N., Mahgoub, H., Dutta, A. K., Khalid, M., Marzouk, R., ... & Drar, S. (2023). Sustainable groundwater management using stacked LSTM with a deep neural network. Urban Climate, 49, 101469.
[21] Valadkhan, D., Moghaddasi, R., & Mohammadinejad, A. (2022). Groundwater quality prediction based on LSTM RNN: An Iranian experience. International Journal of Environmental Science and Technology, 19(11), 11397-11408.

[22] Jaffar, A. Z. H. A. R., Thamrin, N. M., Amin, M. S., Ali, M., Misnan, M. F., & Yassin, A. I. M. (2022). WATER QUALITY PREDICTION USING LSTM-RNN: A REVIEW. Penerbit UMT Journal of Sustainability Science and Management, 17, 205-226.
[23] Vu, M. T., Jardani, A., Massei, N., Deloffre, J., Fournier, M., & Laignel, B. (2023). Long-run forecasting surface and groundwater dynamics from intermittent observation data: An evaluation for 50 years. Science of The Total Environment, 880, 163338.
[24] Li, D., Zhao, W., Hu, J., Zhao, S., & Liu, S. (2023). A long-term water quality prediction model for marine ranch based on a time-graph convolutional neural network: Ecological Indicators, 154, 110782.
[25] Mbouopda, M. F., Guyet, T., Labroche, N., & Henriot, A. (2022, September). Experimental study of time series forecasting methods for groundwater level prediction. In International Workshop on Advanced Analytics and Learning on Temporal Data (pp. 34-49). Cham: Springer International Publishing.
[26] Kenda, K., Peternelj, J., Mellios, N., Kofinas, D., Čerin, M., & Rožanec, J. (2020). Usage of statistical modeling techniques in surface and groundwater level prediction. Journal of Water Supply: Research and Technology—AQUA, 69(3), 248-265.
[27] Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023, June). N-HiTS: Neural hierarchical interpolation for time series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 6, pp. 6989-6997).
[28] Lu, J., & Wu, M. (2024, February). Long-term traffic flow prediction based on multiscale perceptual change. In Seventh International Conference on Traffic Engineering and Transportation System (ICTETS 2023) (Vol. 13064, pp. 126-130). SPIE.
[29] Kollipara, S. C., Krishna, S., Golem, S. V., Fursule, P., & Depuru, B. K. Harnessing Deep Learning for Enhanced Demand Forecasting in Wooden Pallet Manufacturing.
[30] Mazen, F. M. A., Shaker, Y., & Abul Seoud, R. A. (2023). Forecasting of Solar Power Using GRU–Temporal Fusion Transformer Model and DILATE Loss Function. Energies, 16(24), 8105.
[31] Di Grande, S., Berlotti, M., Cavalieri, S., & Gueli, R. (2024). Optimizing Planning Strategies: A Machine Learning Forecasting Model for Energy Aggregators and Hydropower Producers.

[32] Koya, S. R., & Roy, T. (2024). Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence. Journal of Hydrology, 131301.

[33] Wei, X., Wang, G., Schmalz, B., Hagan, D. F. T., & Duan, Z. (2023). Evaluate the Transformer model and Self-Attention mechanism in the Yangtze River basin runoff prediction—Journal of Hydrology: Regional Studies, 47, 101438.

[34] Vu, D. Q., Mai, S. T., & Dang, T. D. (2023). Streamflow prediction in the Mekong River Basin using deep neural networks. IEEE Access.

[35] Wang, C., & Tang, W. (2023). Temporal Fusion Transformer-Gaussian Process for Multi-horizon River Level Prediction and Uncertainty Quantification. Journal of Circuits, Systems and Computers, 32(18), 2350309.

[36] Castangia, M., Grajales, L. M. M., Aliberti, A., Rossi, C., Macii, A., Macii, E., & Patti, E. (2023). Transformer neural networks for interpretable flood forecasting. Environmental Modelling & Software, 160, 105581.

[37] Shang, J., Zhao, B., Hua, H., Wei, J., Qin, G., & Chen, G. (2023). Application of Informer Model Based on SPEI for Drought Forecasting. Atmosphere, 14(6), 951.

[38] Gobato Souto, H. (2023). N-HiTS for Forecasting Stock Realized Volatility. *Available at SSRN 4650761*.

[39] Olivares, K. G. (2023). *Applied Mathematics of the Future* (Doctoral dissertation, NASA).

[40] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764.

[41] Maldonado-Cruz, E., & Pyrcz, M. J. (2024). Multi-horizon well performance forecasting with temporal fusion transformers. *Results in Engineering*, *21*, 101776.

[42] Nixtlaverse, weblink : "https://nixtlaverse.nixtla.io/statsforecast/docs/tutorials/statisticalneuralmethods.html"

[43] Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023, June). N-HiTS: Neural hierarchical interpolation for time series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 6, pp. 6989-6997).

[44] Ni, Z., Zhang, C., Karlsson, M., & Gong, S. (2024). A study of deep learning-based multi-horizon building energy forecasting. Energy and Buildings, 303, 113810.

[45] TowardsDataScience, weblink: https://towardsdatascience.com/temporal-fusion-transformer-time-series-forecasting-with-deep-learning-complete-tutorial-d32c1e51cd91
[46] Sutton, J. E., Screaton, E. J., & Martin, J. B. (2015). Insights on surface-water/groundwater exchange in the upper Floridan aquifer, north-central Florida (USA), from streamflow data and numerical modeling. *Hydrogeology Journal*, *23*(2), 305.
[47] Gordu, F., & Nachabe, M. H. (2023). Inferences of Groundwater Response to Projected Hydroclimatic Changes in North Florida. Journal of Hydrologic Engineering, 28(4), 04023001.
[48] Vu, M. T., Jardani, A., Massei, N., Deloffre, J., Fournier, M., & Laignel, B. (2023). Long-run forecasting surface and groundwater dynamics from intermittent observation data: An evaluation for 50 years. Science of The Total Environment, 880, 163338.
[49] Vu, M. T., Jardani, A., Krimissa, M., Zaoui, F., & Massei, N. (2023). Large-scale seasonal forecasts of river discharge by coupling local and global datasets with a stacked neural network: Case for the Loire River system. Science of The Total Environment, 897, 165494.