Research

A maximal overlap discrete wavelet packet transform coupled with an LSTM deep learning model for improving multilevel groundwater level forecasts

Dilip Kumar Roy[1,2,3] · Ahmed A. Hashem[1,4,5] · Michele L. Reba[2] · Deborah L. Leslie[6] · John Nowlin[1,4]

Abstract

Developing precise groundwater level (GWL) forecast models is essential for the optimal usage of limited groundwater resources and sustainable planning and management of water resources. In this study, an improved forecasting accuracy for up to 3 weeks ahead of GWLs in Bangladesh was achieved by employing a coupled Long Short Term Memory (LSTM) network-based deep learning algorithm and Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) data preprocessing. The coupled LSTM-MODWPT model's performance was compared with that of the LSTM model. For both standalone LSTM and LSTM-MODWPT models, the Random Forest feature selection approach was employed to select the ideal inputs from the candidate GWL lags. In the LSTM-MODWPT model, input GWL time series were decomposed using MODWPT. The 'Fejér-Korovkin' mother wavelet with a filter length of 18 was used to obtain a collection of scaling coefficients and wavelets for every single input time series. Model performance was assessed using five performance indices: Root Mean Squared Error; Scatter Index; Maximum Absolute Error; Median Absolute Deviation; and an a-20 index. The LSTM-MODWPT model outperformed standalone LSTM models for all time horizons in GWL forecasting. The percentage improvements in the forecasting accuracies were 36.28%, 32.97%, and 30.77%, respectively, for 1-, 2-, and 3-weeks ahead forecasts at the observation well GT3330001. Accordingly, the coupled LSTM-MODWPT model could potentially be used to enhance multiscale GWL forecasts. This research demonstrates that the coupled LSTM-MODWPT model could generate more precise GWL forecasts at the Bangladesh study site, with potential applications in other geographic locations globally.

Highlights

- An improved forecasting accuracy for up to 3 weeks ahead of groundwater levels is proposed
- A coupled LSTM and Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) data preprocessing are utilized
- The Random Forest feature selection approach is employed to select the most influential input variables
- The coupled LSTM-MODWPT model outperformed the standalone LSTM model.

✉ Dilip Kumar Roy, dilip.roy@my.jcu.edu.au | [1]College of Agriculture, Arkansas State University, Jonesboro, AR 72467, USA. [2]United States Department of Agriculture, Agricultural Research Service Delta Water Management Research Unit, Jonesboro, AR 72467, USA. [3]Irrigation and Water Management Division, Bangladesh Agricultural Research Institute, Gazipur 1701, Bangladesh. [4]The University of Arkansas System Division of Agriculture, Little Rock, AR 72467, USA. [5]Center for No-Boundary Thinking, Arkansas State University, Jonesboro, AR 72467, USA. [6]Department of Earth Sciences, University of Memphis, Memphis, TN 38152, USA.

Discover

1 Introduction

Groundwater aquifers are regarded as the primary sources of the world's clean water supplies and play a crucial role in the stability of irrigated agriculture, domestic water requirements, and industrial water supplies in locations where high-quality surface water is insufficient [1–3]. Groundwater supplies are under increased pressure due to population growth, rising water demand, and the inevitable influences of climate change [4]. Consequently, groundwater systems are undergoing an accelerated decline. Although human action, such as excessive pumping, is thought to be the significant determinant of groundwater level (GWL) decline, recent forecasts suggest that the situation will worsen even sooner than predicted due to climate change [4]. Due to the over-extraction of limited groundwater reserves, groundwater resources will continue to be depleted, causing various environmental, operational, and economic problems [5]. Groundwater is a significant source of water supply in Bangladesh, where approximately 80% of the population relies on groundwater supplies primarily for basic water requirements [6]. Consequently, prudent management and sustainable use of the aquifer's limited groundwater reserves are imperative to secure steady groundwater supplies for future generations. Precise prediction and projection of impending GWL fluctuations may generate a practical groundwater management approach in Bangladesh and globally [3, 7].

Numerical simulation models of groundwater flow processes have historically been used in groundwater hydrology to forecast GWLs and better understand the system's underlying mechanisms [8–10]. However, accurate predictions of GWLs using simulation models require a thorough knowledge of the aquifer's characteristics and skilled modelers with a thorough understanding of aquifer geometry and modeling strategies. As such, data-driven modeling techniques have been developed and deployed to reduce modeling complexities in the hydrological research domains [11–16]. The attributes of the underlying physical processes do not need to be explicitly defined to develop data-driven models. In Machine Learning (ML) based data-driven modeling techniques, a model's inputs and outputs are directly mapped or correlated using an iterative learning process [17]. In the realm of the prediction of GWL time series data, it has been discovered that a data-driven model developed using an Artificial Neural Network (ANN) performed as well as or had improved predictions compared to a numerical simulation model [18, 19]. As such, there has been increasing interest in data-driven modeling techniques as alternatives to complex numerical simulation models. Sufficient and accurate prediction of GWL over the short- to medium-term helps develop a groundwater management plan in regions where droughts initiated by climate change or over-pumping are the primary driving forces [20–22]. Various data-driven modeling techniques are gaining popularity due to their ability to achieve comparable results with conventional hydrogeological modeling while requiring fewer data points and being easier to execute [23]. Numerous techniques have recently been employed in the research area related to predictions of GWL fluctuations. These approaches ranged from standalone data-driven modeling to hybrid modeling [24–28].

The standalone modeling approaches comprise ML-based modeling [21, 22, 29–31], ANNs [32–34], NARX neural networks [21], ANFIS [3, 35–39], Gaussian Process Regression [37], Prophet modeling technique [40], Support Vector Machine (SVM) [35, 41], and Discrete Space-State model [7]. On the other hand, the hybrid approaches used to predict GWL fluctuations are hybrid Wavelet Transform–ML approaches [24, 37, 42, 43], hybrid ML and Ensemble Empirical Mode Decomposition [2], Wavelet—ANFIS [44], and Nonlinear System Identification models coupled with Linear Polynomials [45]. Another domain of hybrid ML-based GWL prediction includes the usage of evolutionary algorithms to tune the model parameters. These evolutionary algorithm tuned ML models include the application of Whale Algorithm—ANN [46], Particle Swarm Optimization (PSO)—ARIMA [47], hybrid Self Organizing Map- and Multi-objective Genetic Algorithm-SVM [48], and hybrid SVM-PSO [49, 50]. A thorough analysis of ML-based methods for modeling GWLs is provided by Rajaee et al. [51] and by Sarma and Singh [52]. It is evident that a variety of modeling techniques have been used to forecast changes in GWL with differing extents of prediction accuracy. It is also clear that recommending a specific prediction model for projecting GWL changes is technically challenging, and possibly unachievable. Consequently, improving the prediction accuracy of GWL variations still necessitates more sophisticated approaches.

Deep Learning (DL) has recently been used as an emerged and well-developed sub-area of ML-based approaches. A growing number of scientific fields have successfully used DL-based modeling [53–56]. The recent application of DL has also been noted in the prediction of time series data [57, 58], forecasting of GWLs [59, 60], and in estimating future water quality variables in the short term [61]. Consequently, several recent groundwater modeling studies [21, 25, 62]

have focused on the effective usage of DL-based Recurrent Neural Networks (RNNs). However, standard RNN designs struggle to capture long-term dependence between variables due to the presence of two issues: vanishing and exploding gradients [63]. These challenges can be addressed by Long Short-Term Memory (LSTM) networks, an advanced version of traditional RNN topologies. Despite their widespread applicability in numerous study fields [23, 64–66], LSTMs have only recently been employed to predict hydrological time series. A recent study investigated the potential of two stand-alone DL models for daily water demand forecasting, employing a Convolutional Neural Network (CNN) and LSTM, alongside their hybrid CNN–LSTM model [67]. Additionally, a deep learning-based bi-directional LSTM model, integrated with CNN, was introduced to predict daily water consumption in London, referred to as the CNN–BiLSTM hybrid model. The authors concluded that the CNN–BiLSTM approach shows promise for accurate forecasting of urban water demand globally. In another study, Swagatika et al. [68] proposed a hybrid DL model, Fourier Transform LSTM (FT-LSTM), aiming to enhance the prediction accuracy of monthly discharge time series in the Brahmani river basin at the Jenapur station. The findings demonstrate that the FT-LSTM model effectively improves the accuracy of monthly runoff forecasts, offering a promising solution for water resource management and river basin decision-making processes. Dehghani et al. [69] explored the efficacy of three DL-based methodologies, namely LSTM, CNN, and Convolutional LSTM (ConvLSTM), in short-term streamflow forecasting within the Kelantan and Muda River basins in Malaysia. Their findings revealed that all three DL-based techniques exhibited high accuracy in predicting streamflow. Additionally, the study identified that LSTM performed exceptionally well in smaller basins characterized by well-distributed rainfall stations, whereas CNN and ConvLSTM showed greater effectiveness in regions experiencing moderate to high streamflow and in larger river basins. Jeong et al. [70] utilized LSTM-based models to predict GWLs utilizing real-world 'faulty' data containing outliers and noise. They found that the predictive power of an LSTM network outperformed that of an RNN when predicting hourly GWL values for a seaside city in the USA prone to episodic inundations [59]. In order to achieve accuracy and resilience in irrigation flow forecasting, Mouatadid et al. [71] combined a Maximal Overlap Discrete Wavelet Transform (MODWT) and an LSTM model. Zhang et al. [23] proposed an LSTM-based model for forecasting water table elevation in agricultural areas and achieved reliable prediction outcomes by employing a straightforward data preprocessing method for data standardization. Considering recent literature, LSTMs have the potential to be utilized in the research domain of hydrological time series projections. Therefore, this study represents the first attempt to employ LSTM-based models coupled with a data preprocessing tool for forecasting multi-step forward GWLs at designated observation wells in the Gazipur Sadar Upazilla, Bangladesh.

Wavelet decomposition and wavelet packet transform have been successfully utilized in numerous research domains to enhance the forecasting capability of ML-based models. However, recent studies focusing on wavelet packet-based prediction modeling techniques have encountered issues related to leveraging "future data," making them unsuitable for real-life forecasting scenarios [72, 73]. In contrast, the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) approach is better suited for real-world applications as it can overcome these "future data" issues. This study utilizes MODWPT as a preprocessing tool to enhance the forecasting capabilities of an LSTM model for multiple-step forward GWL forecasting. Therefore, the key motivation and focus of this study are to assess the potential use of MODWPT as a preprocessing tool to improve the forecasting capability of LSTM in predicting multiple-step forward GWLs at designated observation well locations.

2 Methods

2.1 Study area and input data

The study area encompasses the Gazipur Sadar Upazilla, with an approximate surface area of 446.38 km^2. It is situated between the longitudes of 90.33° and 92.50° E and the latitudes of 23.88° and 24.18° N. This area falls within the physiographic unit known as the Madhupur jungle tract, locally referred to as *Bhawal Garh*. The terrace topography in many parts of the Madhupur jungle tract ranges from flat areas to lower-rounded mountains and ridges separated by closely spaced shallow *'bides'* (valleys) [74]. Additionally, a few *'beels'* (lakes) are situated near the periphery of the study area. The soil conditions in *Bhawal Garh* are diverse and often complex. A wide range of soil variability exists, from red laterite soils at the extreme to almost undeveloped soil of raw Pleistocene clay [74], with numerous intermediate soil layers. Four major constituents of land formation were reported [74]: clay (70.02%), sandy loam (13.84%), sand (6.92%), and pebble (4.61%). The wet season rainfall plays a crucial role as a source of water for groundwater replenishment and recharge,

albeit constrained by the thick clay soil surface and low rainfall periods, leading to decreased natural recharge. Pumped groundwater serves as the primary water source for residential and agricultural water requirements.

In this study, secondary GWL data were collected from the Processing and Flood Forecasting Circle (PFFC) of the Bangladesh Water Development Board (BWDB) [75]. GWL data were collected from various observation well locations. Among these locations, two observation wells, GT3330001 and GT3330002, were selected based on the criteria of having the fewest missing GWL entries. A data quality assurance procedure is often deployed to warrant the qualities of the obtained GWL datasets, which boosts the trustworthiness of GWL forecasts using ML tools [76]. While a thorough quality assurance procedure was not conducted for the current dataset, the quality of the acquired GWL data was systematically assessed for accuracy and comprehensiveness using range/limit tests. Range testing involves straightforward verification that any observation falls within a given range [76]. Measurements outside of this range are flagged as invalid, and only measurements within this threshold are accepted [77, 78]. Data within the acceptable range were utilized to simulate future GWL changes in the selected observation wells, focusing on providing multiple-step ahead GWL projections. The study domain and locations of observation wells are depicted in Fig. 1.

The weekly GWL data collected from BWDB for GT3330001 spanned from January 7, 1980, to September 17, 2018, comprising a total of 2019 data points. GWL data are often subject to missing values due to factors such as sensor malfunction, data transmission errors, or gaps in monitoring. At GT3330001, there were 29 missing values, accounting for 1.4% of the data, leaving 1983 data points available for analysis. On the other hand, the weekly GWL data at GT3330002 spanned from January 7, 1980, to December 26, 2016, comprising a total of 1929 data points. There were 18 missing values, accounting for 0.93% of the data, resulting in 1919 data points for analysis. Failure to address missing data can introduce biases and inaccuracies in the model, potentially leading to erroneous predictions. By implementing strategies to handle missing data, such as interpolation techniques or data imputation methods, the model can effectively
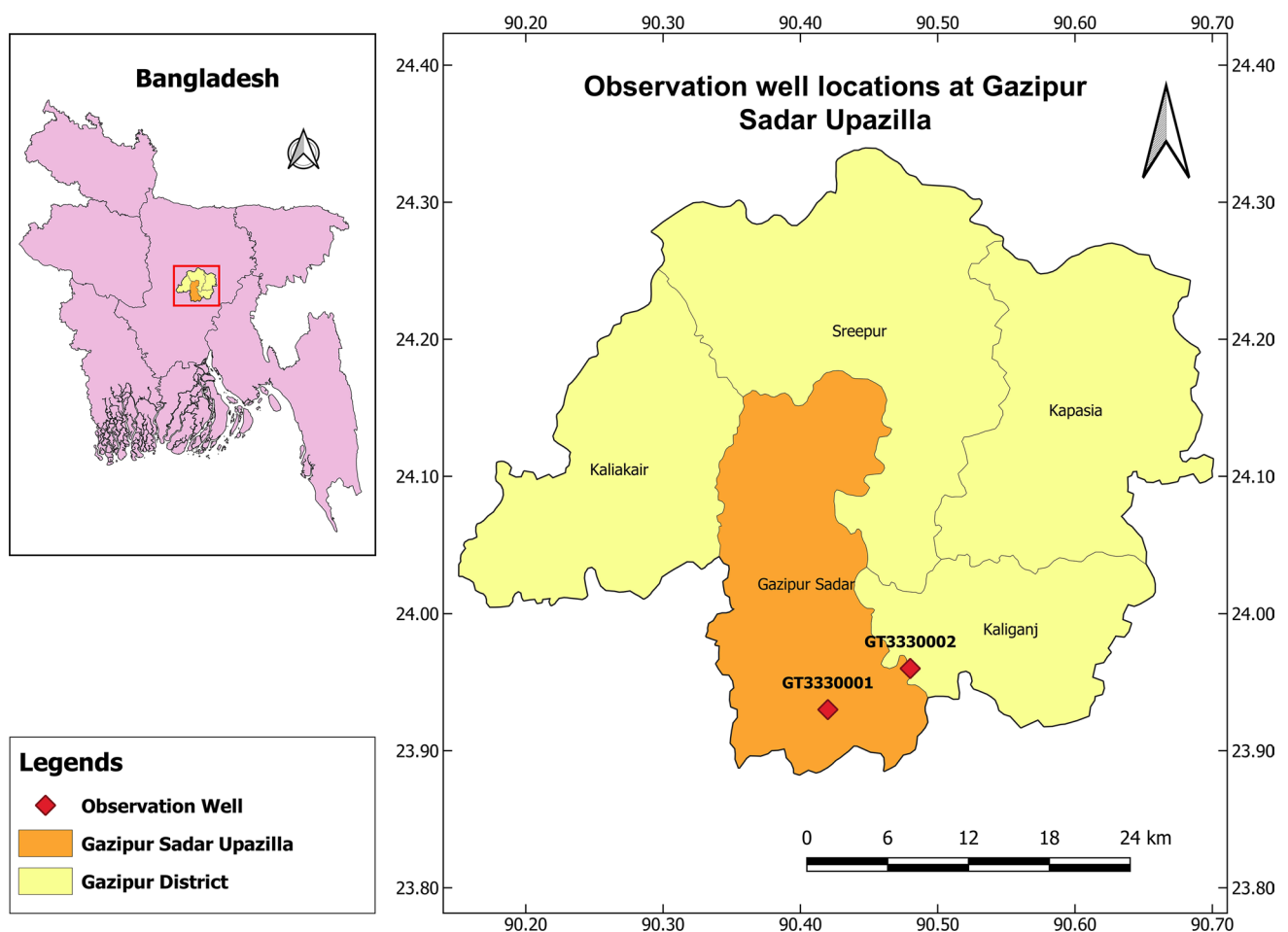

Fig. 1 Study area with two observation well locations in the Gazipur Sadar Upazilla

utilize available information while mitigating the impact of missing values on the forecasting process. By addressing missing data, the model can maintain continuity and consistency in the input data, reducing the likelihood of errors and distortions in the forecasting results. The omitted GWL data were replaced using the 'moving median' method of data imputation, in which a moving median with a specified window length was employed. Finally, the observation wells GT3330001 and GT3330002 had 2012 readings (from January 7, 1980, to September 17, 2018) and 1937 readings (from January 7, 1980, to December 26, 2016) of weekly GWL entries after the imputation of missing entries.

The weekly GWL dataset's descriptive statistics are presented in Table 1. The mean of the GWL data ranges from 12.96 m (at GT3330001) to 13.26 m (at GT3330002), while the standard deviation values range from 3.39 m (at GT3330002) to 7.92 m (at GT3330001) (Table 1). The positively skewed values indicate that the statistical distribution of the data at both observation wells has a larger right tail than a left tail (Table 1). The GWL data at GT3330001 had negative kurtosis values and light-tailed distributions. On the other hand, the datasets at GT3330002 showed heavy-tailed distributions due to the positive kurtosis value.

2.2 Modelling approaches

Brief descriptions of various approaches, including the Long Short Term Memory Network (LSTM), Maximal Overlap Discrete Wavelet Packet Transform (MODWPT), integrated LSTM-MODWPT, input variable selection using Random Forest (RF), and data partitioning, are provided in the following sections.

2.2.1 Long-Short Term Memory (LSTM) Networks

An LSTM model is a subtype of advanced Recurrent Neural Networks (RNNs) able to acquire longer-term reliance amongst sequence data time steps. Because LSTMs incorporate state dynamics and gating functions, they overcome the issues of exploding and vanishing gradients found in conventional RNNs, making them particularly effective for predicting sequence data [79]. The LSTM network's architecture is made up of multiple memory blocks connected by layers, each of which contains a substantial number of memory cells with recurrent connections. Examples of the three multiplicative parts (called gates) that make up an LSTM memory cell are the forget, input, and output gates [80]. A sequential input layer is used to feed time series data into an LSTM model network, comprising the majority of a primary LSTM network. There are four layers in an LSTM network designed to solve a simple regression problem: the LSTM network commences with a sequential input layer preceded by an LSTM layer, and the network concludes with an entirely interconnected layer and a regression output layer. The straightforward and deeper LSTM architectures, as well as the mechanism by which LSTMs perform forecasting tasks are illustrated in Figs. S1, S2, S3 of the Supplemental Information (SI).

An LSTM network architecture with three hidden layers was utilized. Model overfitting was prevented by assigning a dropout layer to each hidden layer. The number of hidden neurons and the associated dropout rates in the 1st, 2nd, and 3rd hidden layers were selected upon several iterations. The optimal numbers of hidden neurons were 100, 50, and 20 for the 1st, 2nd, and 3rd hidden layers, respectively, while the corresponding optimal dropout rates were 0.4, 0.3, and 0.4, respectively.

The LSTM architecture with multiple hidden units was utilized. The numbers of 'hidden neurons' were decided via several trials by varying the number of 'hidden neurons' in each iteration. The LSTM architecture parameters were obtained through numerous trials, and the optimal sets of parameter values are presented in Table 2. These optimum parameter values were used to develop the LSTM models for predicting 1, 2, and 3-week(s) ahead of GWL forecasting at the two observation wells.

Table 1 Values of the statistical parameters computed on the GWL data (m) at the designated observation wells	Obs. wells	Min	Max	Mean	Median	STD	Skewness	Kurtosis
	GT3330001	0.10	25.57	12.96	9.83	7.92	0.50	− 1.39
	GT3330002	6.30	23.50	13.26	13.26	3.39	0.74	0.44

Table 2 The best pairings of various training alternatives

Training options	Relevant option values
Solver used to optimize	'adam'
Maximum number of epochs	1000
Gradient threshold value	1
Initial rate of learning	0.001
Minimum size of the batch	150
Length of the sequence	1000

2.2.2 Maximal Overlap Discrete Wavelet Packet Transform (MODWPT)

The formulation of wavelet decomposition using MODWPT may be expressed as [81]:

$$\tilde{W}^P_{j,n,t} = \sum_{l=0}^{L-1} \tilde{f}_{n,l} \tilde{W}^P_{j-1,[n/2],(t-2^{j-1}l) \bmod N} \tag{1}$$

$$\tilde{f}_{n,l} = \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3 \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2 \end{cases} \tag{2}$$

where $\tilde{W}^P_{j,n,t}$ represents the coefficients of MODWPT at time t for the level of decomposition j and band n (where $n = 0, 1, \ldots, 2^j - 1$), which is technically corresponding to frequencies for an interval of $\left[\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}} \right]$.

The MODWPT is a wavelet decomposition algorithm that utilizes energy, and its output consists of time-delayed signals relative to the input signals. It partitions the energy over the entire wavelet packets at each level, with the total energy of any input signal equaling the sum of the energies over all the wavelet packets. Moreover, the features in the MODWPT closely align with those present in the input signal. These details, such as energies, can produce the original signal when summing the details in each sample for a given level. The MODWPT performs a discrete wavelet packet transform and returns a 'sequency-ordered' wavelet packet tree. The time series decomposition process via MODWPT using a sequency-ordered wavelet packet tree can be found in Fig. S4 of the SI.

2.2.3 Proposed model development (integrated LSTM-MODWPT)

The essential phases for developing the proposed LSTM-MODWPT model are presented in the following sub-sections and depicted in Fig. 3. The first step in developing DL-based forecasting models is the selection of potential input variables. Given that only the weekly GWLs at two observation wells (GT3330001 and GT3330002) were utilized in this effort, the possible inputs consisted of the time-lagged forms of the measured GWLs at every station, along with their wavelet packet transformed components. The determination of candidate GWL lags was based on a thorough analysis of temporal dependencies within the dataset. This included utilizing Partial Autocorrelation Function (PACF) plots to extract time-lagged information from the GWL time series, alongside considering the typical temporal scales of hydrological processes. The aim of this selection was to capture significant lagged effects that enhance the predictive accuracy of GWL forecasting. A PACF plot was utilized to extract time-lagged information from the GWL time series and determine which lags to incorporate as potential inputs. The PACF plots for the GWL data obtained from the two selected observation wells are illustrated in Fig. 2. From Fig. 2, it is possible to deduce that the present and previous five time lags are essential for 3-weeks ahead GWL forecasting (GWL_{t+1}, GWL_{t+2}, and GWL_{t+3}). For observation well GT3330001, the present and previous time lags were GWL_t, GWL_{t-1}, GWL_{t-3}, GWL_{t-4}, GWL_{t-5}, and GWL_{t-6}. Conversely, for observation well GT3330002, the time lags were GWL_t, GWL_{t-1}, GWL_{t-2}, GWL_{t-3}, GWL_{t-4}, and GWL_{t-5}. Consequently, the potential input variables for the non-wavelet decomposition-based LSTM model comprised six time-lagged variables as inputs. For the MODWPT-based LSTM model, the candidate input variables included the same inputs along with their MODWPT decomposed counterparts. Each of the time-lagged GWL time series was wavelet decomposed independently. The MODWPT produced many
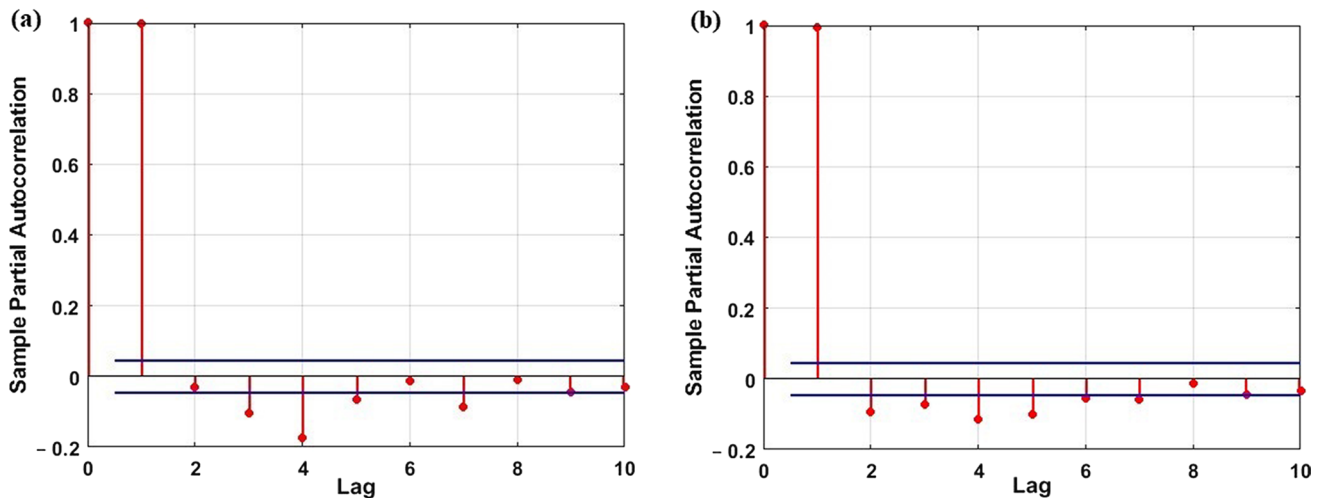
Fig. 2 PACF plots of the weekly GWL timeseries for the observation wells **a** GT3330001, **b** GT3330002

candidate input variables for the wavelet-based LSTM models at both observation wells. Therefore, the most significant input variables were selected using an RF-based modeling approach.

The choice of wavelet filters and decomposition levels, as well as the elimination of boundary-impacted wavelet coefficients, is primary crucial consideration during wavelet decomposition [72]. This study utilized the Fejér-Korovkin scaling filter with a filter length of 18 and a decomposition level of 3. The selection of the 'Fejér-Korovkin' mother wavelet was driven by its desirable characteristics such as orthogonality, compact support, vanishing moments, and smoothness. These properties enable effective decomposition and reconstruction of signals, which are essential for wavelet-based analysis [82]. Moreover, Fejér-Korovkin wavelets demonstrated strong performance in capturing the essential features of input groundwater level signals. This determination was made through a comparison of various mother wavelet approaches, considering different filter lengths (up to 22) and decomposition levels (up to seven). The choice of filter length in the MODWPT is crucial as it directly impacts the level of decomposition and resolution at each level. While longer filter lengths typically offer improved frequency resolution, they may also introduce higher computational complexity. Therefore, a filter length of 18 was chosen to strike a balance between frequency resolution and computational efficiency. This decision was made by experimentation and deemed adequate for capturing relevant frequency components while ensuring manageable computational overhead. As the wavelet-decomposed data is influenced by "boundary conditions" [73], it was essential to eliminate the initial few wavelet and scaling coefficients affected by these conditions. This essential elimination of 188 boundary-impacted wavelet coefficients from the start of the sets of possible input and target variables was grounded in the boundary correction approach established by Quilty and Adamowski [72]. Similar principles were applied in the LSTM models that were not wavelet-based. A flowchart illustrating the model development processes can be seen in Fig. 3.

MATLAB [83] commands and functions were employed to develop the proposed models using the observed GWL data from the two observation wells.

MODWPT plays a crucial role in decomposing the input GWL time series into different frequency bands, thereby capturing multiscale variations and temporal dynamics inherent in groundwater systems. This enables the model to extract valuable information from the data across various time scales, which might not be effectively captured by traditional forecasting methods. MODWPT allows the model to capture complex temporal patterns and fluctuations in GWLs across multiple scales, enhancing its ability to make accurate forecasts. These techniques enable the model to extract valuable insights from the data, mitigate potential sources of error, and enhance its predictive capabilities, ultimately supporting informed decision-making and sustainable management of groundwater resources. The combination of MODWPT and missing data handling techniques provides the model with greater flexibility and adaptability to varying data conditions and quality, thereby increasing its robustness and reliability in real-world applications.

2.2.4 Input variable selection

The RF approach was employed to select the scaling coefficients and wavelet family that are most effective in producing precise predictions of the output variable. The decision to utilize the RF approach over other variable selection methods

Fig. 3 Flowchart of the main model development steps
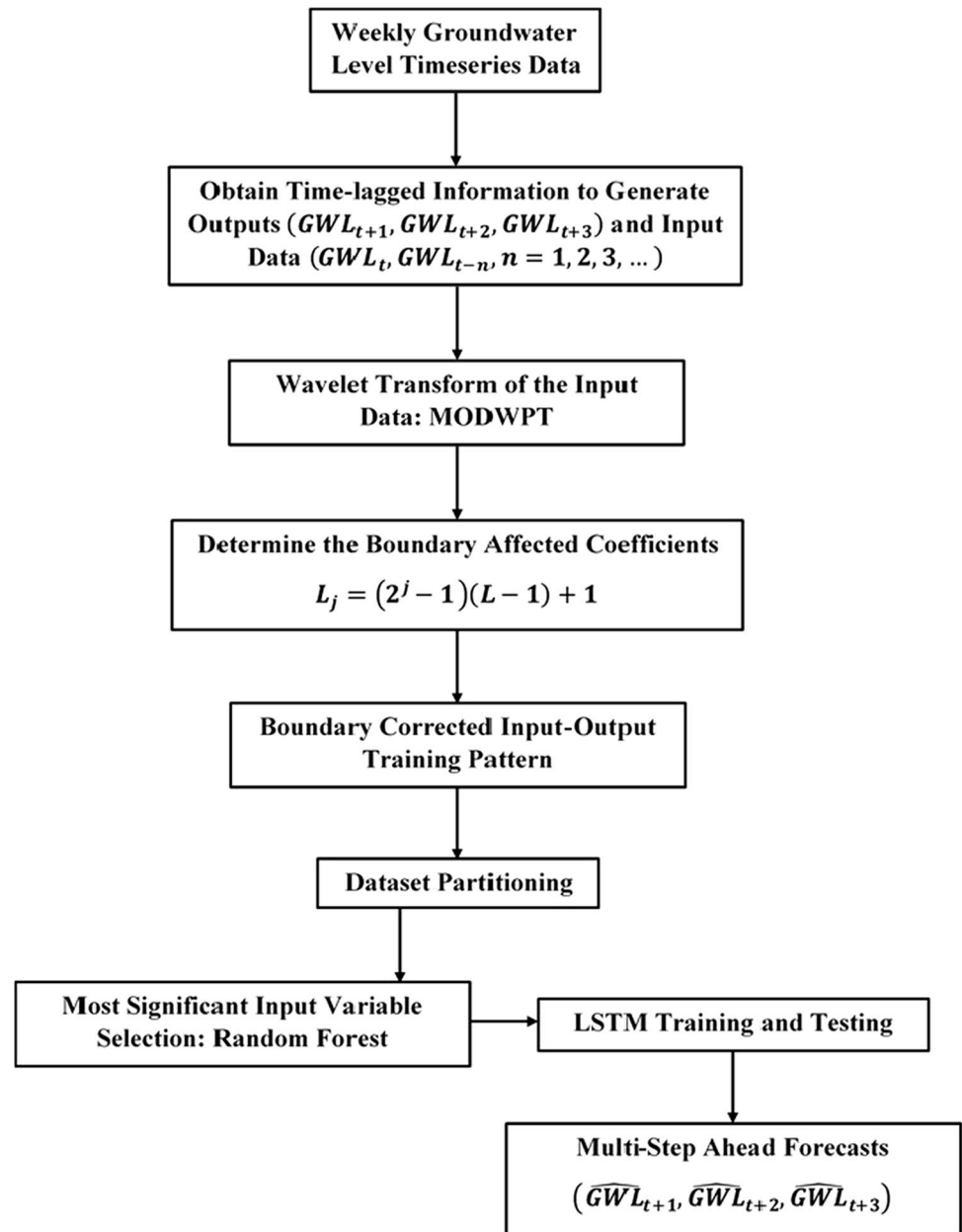
was driven by several factors. Firstly, RF is known for its robustness to overfitting and handling of high-dimensional datasets, which is particularly advantageous when dealing with complex hydrological systems. Its ensemble nature and built-in mechanism for bootstrapping and aggregation help mitigate overfitting issues, especially in high-dimensional datasets. Additionally, RF inherently provides feature importance scores, allowing for transparent and interpretable variable selection. This helps identify the most influential variables in the model, aiding in understanding the underlying processes driving the system. Furthermore, its ability to handle both categorical and continuous variables without the need for preprocessing simplifies the modeling process. Moreover, RF is relatively robust to multi-collinearity, a common issue in datasets where predictor variables are highly correlated. It can still provide accurate predictions even when multi-collinearity is present, making it suitable for diverse datasets. Lastly, RF has demonstrated effectiveness in capturing nonlinear relationships and interactions among variables, which are common in hydrological data, thereby enhancing the model's predictive capability. Overall, RF is relatively easy to implement and requires minimal parameter tuning compared to other complex algorithms. Its versatility and ease of use make it a popular choice for variable selection in various fields, including hydrology.

The selection of inputs using the RF feature selection approach for both standalone LSTM and LSTM-MODWPT models was influenced by several factors. Firstly, the importance of capturing relevant temporal patterns and relationships between predictors and GWL fluctuations guided the inclusion of features with significant predictive power. Additionally, considerations such as the physical significance of input variables in hydrological processes and their potential impact on GWL dynamics played a crucial role. Furthermore, the ability of selected features to provide robustness against noise and irrelevant information was essential for model generalization and performance. Lastly, the aim to strike a balance between model complexity and predictive accuracy led to the inclusion of a subset of inputs that maximized information gain while minimizing redundancy, thus enhancing the efficiency and interpretability of the models. The RF approach's main task was to ensure the selection of only necessary variables for projecting the output variable while omitting unnecessary or inappropriate variables. The goal was to create accurate models that are not overly complex or limited in their predictive ability. This study used a 'bagged' ensemble of 500 regression trees to develop the RF model. All input variables were assigned at each tree node to ensure that each regression tree utilized all input variables to attain better accuracy. The number of levels among the inputs varied using a standard CART algorithm for selecting split inputs at each node of the trees in a RF model, which could result in less accurate estimates. To address this issue, a curvature or interaction test was performed to select split inputs [84]. Variable importance was estimated by performing permutation of the out-of-bag observations among the trees.

Time-lagged input variables, along with their wavelet packet coefficients, resulted in 102 input variables [GWL at the present time and five times lagged GWLs (6) + 16 wavelet packet coefficients for Fejér-Korovkin scaling filter with a specified filter length (16) × 6]. Using all the input variables is associated with a computational burden. Therefore, only the most significant input variables were selected for model development to reduce computational load and enhance computational efficiency. This study used the 20 most influential input variables determined by the RF modeling technique to develop LSTM-MODWPT models at the observation wells for one-, two-, and three-step ahead GWL forecasts. Figure 4 depicts the plots of variable importance.

2.3 Data partitioning

Due to time lagging and removal of the boundary-affected coefficients, the observed GWL data were reduced at each observation well. At the observation well GT3330001, a total of 1816 records remained (from 09 October 1983 to 17 September 2018) after removing 188 boundary-affected coefficients and eight records due to time lagging (three time lags forward + five time lags backward) from the entire GWL time series of 2012 readings (from 07 January 1980 to 17 September 2018). At GT3330002, a total of 1741 records remained (from 09 October 1983 to 26 December 2016) after removing 188 boundary-affected coefficients and eight records due to time lagging (three time lags forward + five time lags backward) from the entire GWL time series of 1937 readings (07 January 1980 to 26 December 2016). The remaining dataset was separated into two distinct sets of training and testing samples where 80% of the data records were allocated for training, and the remaining 20% was allotted for testing.

For GT3330001, the remaining 1816 readings were split into 1453 records (from 09 October 1983 to 03 October 2011) and 363 records (from 04 October 2011 to 17 September 2018), respectively, for training and testing purposes. For GT3330002, the remaining 1741 readings at GT3330002 were divided into 1393 records (from 09 October 1983 to 26 April 2010) and 348 records (from 27 April 2010 to 26 December 2016), respectively, for training and testing purposes.

3 Statistical indices for performance evaluation

Five statistical parameters were utilized to evaluate the model's performance (Eqs. 3–7). Generally, the Root Mean Squared Error (RMSE) criterion measures the error of the model. A lower RMSE value indicates higher prediction power of the model. However, the value of RMSE largely depends on the magnitude of the data; therefore, a lower RMSE value does not necessarily signify better prediction performance. To address this issue, the Scatter Index (SI) criterion was used to eliminate the dimensionality effect of the data. Model performance assessment criteria based on the SI indeAx values were: Excellent when SI is less than 0.1, good when SI is between 0.1 and 0.2, fair when SI is between 0.2 and 0.3, and poor when SI is greater than 0.3 [85]. The $a^{20} - index$ value ranges between 0 and 1, and for an ideal model, the $a^{20} - index$ value is 1 [58].
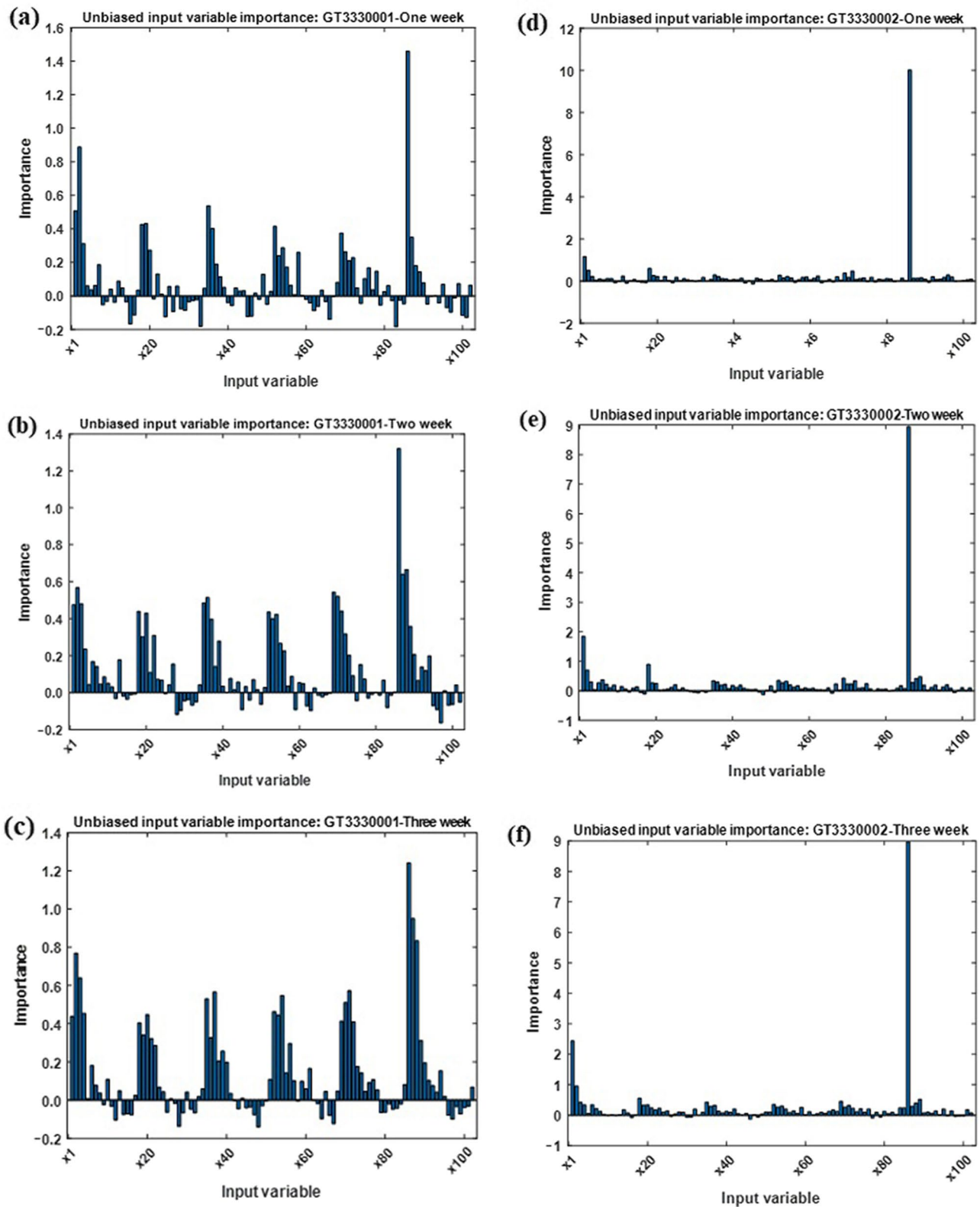
Root Mean Squared Error (RMSE):

Fig. 4 Most significant inputs variables determined by RF approach: left panes (**a**, **b**, and **c**) represent one-, two-, and three-step weekly lead times, respectively, for GT3330001, whereas the second panes (**d**, **e**, and **f**) represent one-, two-, and three-step weekly lead times, respectively, for GT3330002

$$RMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(GWL_i^A - GWL_i^P\right)^2}}{\overline{GWL^A}} \tag{3}$$

Scatter Index, SI [85]:

$$SI = \frac{RMSE}{\overline{GWL^A}} \tag{4}$$

Maximum Absolute Error (MAE):

$$MAE = max\left[\left|GWL_i^A - GWL_i^P\right|\right] \tag{5}$$

Median Absolute Deviation, MAD [86]:

$$MAD\left(GWL^A, GWL^P\right) = median\left(\left|GWL_{i=1}^A - GWL_{i=1}^P\right|, \left|GWL_{i=2}^A - GWL_{i=2}^P\right|, \ldots, \left|GWL_{i=n}^A - GWL_{i=n}^P\right|\right) \tag{6}$$

a^{20} – index [58]:

$$a^{20} - index = \frac{k^{20}}{n} \tag{7}$$

where GWL_i^A and GWL_i^P are the actual and predicted GWL values for the ith data points of the dataset, respectively; $\overline{GWL^A}$ and $\overline{GWL^P}$ are the mean values of the actual and predicted GWL, respectively; n is the number of entries in the GWL time series data; k^{20} is the amount of data that is associated with a GWL_i^A / GWL_i^P value varying from 0.80 to 1.20 [58].

While the performance indices used in the study, including RMSE, SI, MAE, MAD, and the a-20 index, offer valuable insights into the model's forecasting accuracy, there are certain limitations and uncertainties associated with them. For instance, RMSE gives equal weight to all deviations, which might not adequately capture the model's performance for extreme values. Similarly, while the SI provides a measure of forecast dispersion, it does not account for systematic bias. Moreover, the a-20 index focuses on error magnitudes exceeding a specific threshold, potentially overlooking smaller but still significant deviations.

Alternative metrics that could complement these indices include measures of forecast skill such as the Nash–Sutcliffe Efficiency (NSE) or the Kling–Gupta Efficiency (KGE), which assess the ability of the model to replicate observed variability and patterns. Another metric may be the mean absolute percentage error (MAPE), which provides a relative measure of forecasting accuracy and is useful for comparing performance across different datasets or time periods. Additionally, quantile regression loss functions can assess the model's performance at different quantiles of the forecast distribution, offering insights into its reliability under varying levels of uncertainty. Furthermore, probabilistic scoring rules such as the Brier score or logarithmic score can evaluate the model's calibration and probabilistic forecasts, enhancing its utility in decision-making contexts. Integrating a combination of these metrics can provide a more comprehensive assessment of the coupled LSTM-MODWPT model's performance, capturing both the central tendency and dispersion of forecast errors while accounting for uncertainties inherent in GWL forecasting. Therefore, while the performance indices used in the study offer valuable insights, future studies may consider incorporating alternative metrics and approaches to provide additional perspectives on model performance and enhance the robustness of GWL forecasting assessments.

4 Results and discussion

The findings of the non-wavelet and wavelet-based LSTM models for multi-step [i.e., 1-, 2-, and 3-week(s)] ahead GWL forecasting were evaluated using several performance evaluation indices. Additionally, graphical approaches were employed to evaluate the performances of the developed models. Generally, the performances of all models during both the training and testing phases exhibited a good tradeoff, indicating a reasonably fair generalization capability of the developed models in multi-step ahead GWL forecasting. However, the comparison of model performances between

the non-wavelet and wavelet-based LSTM models was primarily based on their testing phase performances. Both training and testing phase model performances are presented in the following sub-sections.

4.1 Performance of the standalone LSTM models

Determining the optimal LSTM model architecture is crucial in DL-based forecasting approaches. In this study, different combinations of various numbers of hidden layers and neurons were evaluated to determine the optimal LSTM model structure in multi-step ahead GWL forecasting. The RMSE criterion was utilized to assess how well the developed models performed during the training and testing phases across various scenarios of hidden layer and hidden neuron combinations. The RMSE values on the training and test datasets for different numbers of neurons are depicted in Fig. 5.

For GT3330001, the minimum values of the absolute difference between the training and test RMSE were 0.09 m (hidden neurons: 180-150-80), 0.03 m (hidden neurons: 150-100-50), and 0.06 m (hidden neurons: 150-100-50) for 1-, 2-, and 3-week(s) ahead forecasting, respectively (Fig. 5). Conversely, at GT3330002, the RMSE values were 3.13 m (hidden neurons: 150-120-80-50), 3.22 m (hidden neurons: 140-120-60), and 3.14 m (hidden neurons: 100-80-50-20) for 1-, 2-, and 3-week(s) ahead forecasting, respectively. Therefore, the LSTM models with these hidden neurons were selected as the best-performing models.

The findings (RMSE, Scatter index, MAE, MAD, and a-20 index) of the standalone LSTM models for forecasting GWLs at 1-, 2-, and 3-week(s) ahead are presented in Table 3. The overall accuracy across the two observation wells for the
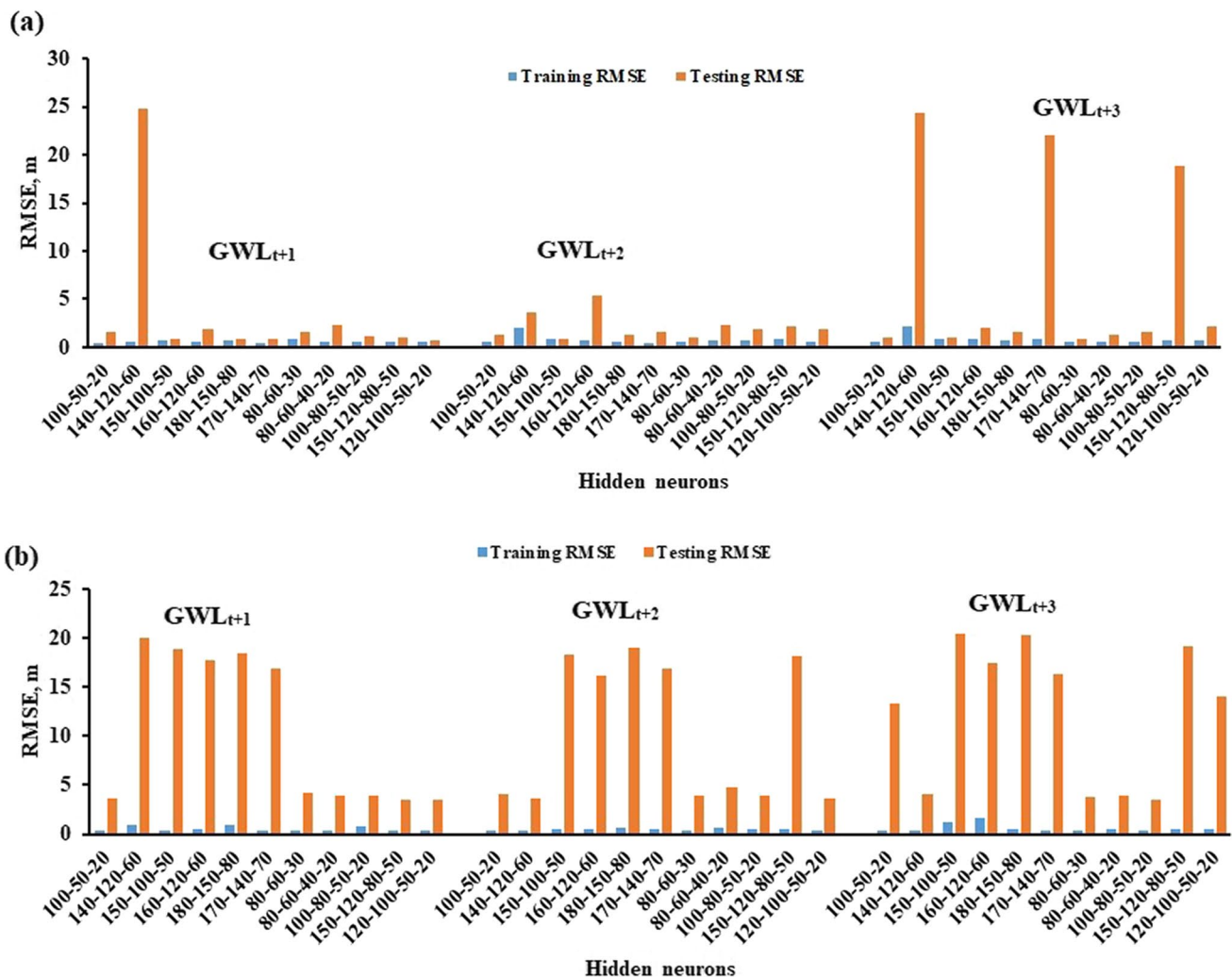

Fig. 5 Training and testing RMSE values of the standalone LSTM models for different combinations of hidden layers and hidden neurons for the observation wells **a** GT3330001 and **b** GT3330002

Table 3 One-, two-, and three-week(s) ahead forecasting performance of the developed standalone LSTM models on test dataset

	Performance evaluation indices				
	RMSE, m	Scatter index	MAE, m	MAD, m	a-20 index
GT3330001					
One week ahead	0.827	0.034	12.480	0.179	0.997
Two weeks ahead	0.910	0.037	10.237	0.319	0.997
Three weeks ahead	0.975	0.040	10.754	0.405	0.997
GT3330002					
One week ahead	3.466	0.188	7.561	1.200	0.589
Two weeks ahead	3.623	0.196	7.727	1.326	0.531
Three weeks ahead	3.493	0.189	7.393	1.327	0.573

standalone LSTM models demonstrated very good performance in terms of scatter index [85], MAD, and a-20 index [58] criteria whereas the performances were reasonably good with reference to the RMSE and MAE criteria. It is noted that the performances of the standalone LSTM models at the observation well GT3330001 were generally superior to those at observation well GT3330002. The plausible reason for this discrepancy in modeling performances may be attributed to the quality and quantity of the observed data as well as the number of missing values that were imputed. Nevertheless, the developed LSTM models produced acceptable results at both observation wells (Table 3).

Although the accuracy of LSTM forecasts typically decreases with increasing lead times [87], the LSTM models developed at the observation well GT3330001 exhibited very good performance (RMSE = 0.975 m, Scatter index = 0.040, MAD = 0.405, m and a-20 index = 0.997) for 3-weeks-ahead GWL forecasting. For monitoring well GT3330002, differences in forecasting performances were also not substantial with respect to increased lead times. It is worth mentioning that several previous studies have compared the forecasting accuracies of various ML algorithms such as SVR, MLR, ANN, and RF [51], with SVR often being identified as the best performing model [87]. Additionally, it has been reported that the LSTM model demonstrated satisfactory performance in forecasting one-step ahead reference evapotranspiration within a subtropical climatic zone [88]. Therefore, direct comparison of these findings with previous studies in GWL forecasting may be challenging, especially considering the differences in study locations. Moreover, different ML approaches have been found to provide superior performance over others for different observation well locations within the same study area [87]. However, based on the findings of this study, it can be argued that LSTM models can effectively be utilized to provide reasonable GWL-level forecasts.

4.2 Performance of the MODWPT coupled LSTM models

The findings of the MODWPT coupled LSTM models for the two observation wells are presented in this section. The obtained best numbers of hidden layers and hidden neurons for the standalone LSTM models were utilized to develop MODWTP-based LSTM models (LSTM-MODWTP). The training performance of the generated LSTM-MODWTP models at the two observation well locations is illustrated in Fig. 6. These results support that MODWPT, as a preprocessing tool, significantly improved the training and testing performance of the LSTM models (Fig. 6). The minimal differences

Fig. 6 Model training and testing phase errors for the developed LSTM-MODWTP for the two observation wells **a** GT3330001, **b** GT3330002
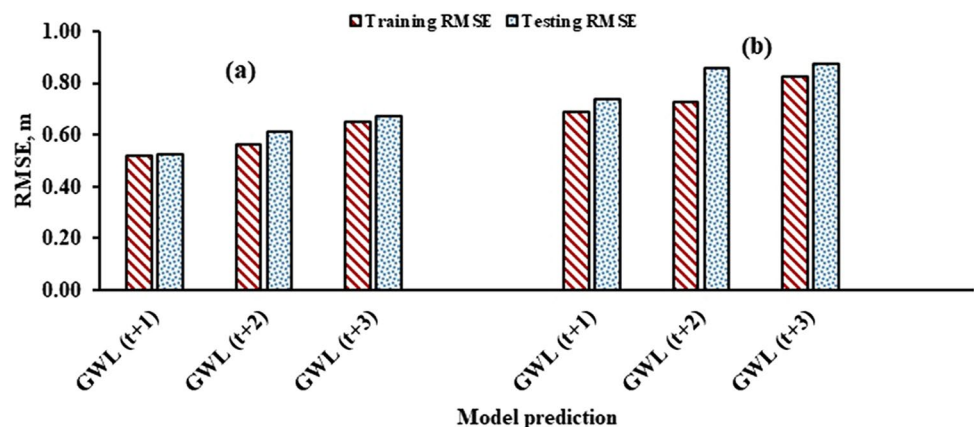
between the training and test RMSE values suggest that the coupled LSTM-MODWPT models were trained adequately and no model overfitting was observed during the training process. In addition, this study adopted the best practices proposed in Quilty and Adamowski [73] to correct the data against boundary affected datasets. This approach ensured the correct utilization of wavelet transforms as a data preprocessing tool for GWL forecasting and minimized the likelihood of MODWPT universally leading to enhanced performances of the LSTM-MODWPT models over the standalone LSTM models. After confirming the absence of model overfitting using the RMSE criterion, the trained models were employed to compute several other statistical performance evaluation indices on the test dataset.

The statistical performance outcomes for the 1-, 2-, and 3-week(s)-ahead forecasting performance of the developed LSTM- MODWTP model on the test dataset are summarized in Table 4. Similar to the standalone LSTM models (results are presented in Table 3), the performances of the LSTM-MODWTP for multi-step forward forecasts at GT3330001 were generally better than the forecasting abilities of the LSTM-MODWTP developed at GT3330002. This performance deviation may have resulted from differences in the length and quality of data, as well as the number of missing data that were imputed. However, the developed LSTM-MODWPT models provided acceptable results at both observation wells. Similar metrics for evaluating the performance were calculated to assess the robustness of the proposed LSTM-MODWTP models (presented in Table 4). It is evident from Table 4 that MODWPT improved the LSTM performance at both observation wells and for all lead times, as evidenced by the higher values of a-20 index and lower values of RMSE, Scatter index, MAE, and MAD criteria. Although the forecasting performance slightly decreased with increased forecasting timeframes, the forecasting performance for the last forecasting horizon (3 weeks ahead) is reasonably good and within acceptable limits for the particular statistical indices.

The improvements in forecasting performance of the LSTM-MODWPT over the standalone LSTM models are quite satisfactory at both observation wells for all forecasting horizons. The percentage improvements in the forecasting accuracy based on RMSE criterion were 36.28% for 1-week-ahead forecasting at GT3330001, 32.97% for 2-weeks-ahead forecasting at GT3330001, 30.77% for 3-weeks-ahead forecasting at GT3330001, 78.68% for 1-week-ahead forecasting at GT3330002, 76.37% for 2-weeks-ahead forecasting at GT3330002, and 74.92% for 3-weeks-ahead forecasting at GT3330001. The percentage improvements in the forecasting accuracy based on Scatter index criterion were 29.41% (1-week-ahead forecasting at GT3330001), 27.03% (2-weeks-ahead forecasting at GT3330001), 25% (3-weeks-ahead forecasting at GT3330001), 47.87% (1-week-ahead forecasting at GT3330002), 58.67% (2-weeks-ahead forecasting at GT3330002), and 54.50% (3-weeks-ahead forecasting at GT3330001). The percentage improvements in the forecasting accuracy based on MAE criterion were 56.49% (1-week ahead forecasting at GT3330001), 51.96% (2-weeks-ahead forecasting at GT3330001), 56.52% (3-weeks-ahead forecasting at GT3330001), 35.29% (1-week-ahead forecasting at GT3330002), 46.62% (2-weeks-ahead forecasting at GT3330002), and 46.04% (3-weeks-ahead forecasting at GT3330001). The percentage improvements in the forecasting accuracy based on MAD criterion were 14.53% (1-week-ahead forecasting at GT3330001), 31.97% (2-weeks-ahead forecasting at GT3330001), 11.85% (3-weeks-ahead forecasting at GT3330001), 27.75% (1-week-ahead forecasting at GT3330002), 32.58% (2-weeks-ahead forecasting at GT3330002), and 31.80% (3-weeks-ahead forecasting at GT3330001). The percentage improvements in the forecasting accuracy based on a-20 index criterion were 0.20% (1-week-ahead forecasting at GT3330001), 0% (2-weeks-ahead forecasting at GT3330001), 0.10% (3-weeks-ahead forecasting at GT3330001), 48.73% (1-week-ahead forecasting at GT3330002), 66.29% (2-weeks-ahead forecasting at GT3330002), and 59.69% (3-weeks-ahead forecasting at GT3330001). The most remarkable finding in Table 4 is that MODWPT especially improved the forecasting performance of the standalone LSTM models at the observation well GT3330002, where the standalone LSTM models performed poorly in all forecasting horizons.

Table 4 One-, two-, and three-week(s)-ahead forecasting performance of the developed LSTM- MODWPT model on test dataset	Performance evaluation indices				
	RMSE, m	Scatter index	MAE, m	MAD, m	a-20 index
GT3330001					
One week ahead	0.527	0.024	5.430	0.153	0.999
Two weeks ahead	0.610	0.027	4.918	0.217	0.997
Three weeks ahead	0.675	0.030	4.676	0.357	0.998
GT3330002					
One week ahead	0.739	0.098	4.893	0.867	0.876
Two weeks ahead	0.856	0.081	4.125	0.894	0.883
Three weeks ahead	0.876	0.086	3.989	0.905	0.915

The superior performance of LSTM-MODWPT models (in terms of percentage improvements in forecasting accuracies) compared to standalone LSTM models can be attributed to several factors. Firstly, the incorporation of the MODWPT as a preprocessing step enhances the model's ability to capture and represent the temporal patterns and multiscale dynamics present in GWL data. By decomposing the signal into different frequency components, the MODWPT facilitates a more comprehensive representation of the underlying variability, thereby enabling the LSTM model to better learn and exploit the intricate dependencies within the data. Additionally, the MODWPT helps in filtering out noise and irrelevant information, thereby improving the signal-to-noise ratio and enhancing the model's robustness to noisy data. Furthermore, the combination of LSTM with MODWPT allows for a more efficient extraction of relevant features from the data, enabling the model to better discern the important temporal patterns and make more accurate forecasts. Overall, the synergy between LSTM and MODWPT leverages the strengths of both approaches, resulting in improved forecasting performance for 1-, 2-, and 3-weeks ahead GWL forecasts. These enhanced forecasting accuracies imply a greater predictive capability of the LSTM-MODWPT models in capturing the underlying dynamics of groundwater level fluctuations, which is crucial for informed decision-making in water resource management. Moreover, the observed improvements underscore the potential of utilizing advanced signal processing techniques in conjunction with deep learning algorithms to achieve more accurate and reliable GWL forecasts, thereby facilitating better resource planning and allocation strategies. These findings highlight the importance of methodological advancements in improving the performance of hydrological forecasting models, ultimately contributing to the sustainable management of water resources in both local and global contexts.

In summary, Table 4 demonstrates the superiority of the proposed LSTM-MODWPT over the standalone LSTM models (Table 3) for the selected observation wells across the three forecast periods. This superior performance is evidenced through the five statistical performance evaluation indices considered for evaluating the models' performances in this study. The proposed LSTM-MODWPT model is expected to handle changes in data quality due to its architecture. Specifically, the proposed LSTM-MODWPT model is capable of managing missing or noisy data through the LSTM architecture and the multi-resolution analysis provided by MODWPT. However, the current study did not explore the model's sensitivity to variations in data sources. As a result, investigations into how the LSTM-MODWPT model would perform with diverse datasets were not conducted, suggesting avenues for future research to examine its generalizability. Moreover, future studies may focus on fine-tuning model parameters, exploring alternative data pre-processing techniques, and incorporating additional variables such as climatic factors or land use changes to enhance forecasting accuracy. Furthermore, aspects such as long-term prediction beyond the 3-week horizon and the assessment of model uncertainty may be incorporated in future research. Moreover, even though this study demonstrated the promise of LSTM-MODWPT models under the conditions studied, further assessments for such modelling aspects would be required for other geographic locations. The outcomes of this research have the potential to enhance overall performance accuracy, reduce modelling complexity, and simplify parameter selections for groundwater modelling. This finding is particularly important in the water resources management, as early forecasting of GWLs is crucial for decision-making in the fields such as irrigation scheduling, land development, and in many other research domains including environmental sciences.

The insights gleaned from this study hold profound implications for policymakers and water resource managers. Firstly, the coupled LSTM-MODWPT model demonstrates significantly improved forecasting accuracy for GWL fluctuations up to 3 weeks ahead in Bangladesh. This enhanced accuracy is crucial for the optimal utilization of limited groundwater resources and sustainable water resource planning and management practices. Secondly, its ability to handle missing or noisy data makes it adaptable to various environmental conditions. Thirdly, the model provides insights into the temporal dynamics of GWL fluctuations, aiding in the identification of trends and anomalies. Beyond Bangladesh, the coupled LSTM-MODWPT model holds promise for sustainable groundwater planning and management globally. Its applicability lies in its adaptability to diverse geographical regions and hydrological settings, enabling tailored solutions to specific challenges. Moreover, by providing reliable forecasts, it empowers policymakers and resource managers to implement proactive measures, such as optimized extraction strategies and conservation efforts, fostering long-term sustainability in groundwater management practices worldwide.

To implement the coupled LSTM-MODWPT model for GWL forecasting in a specific region, several considerations should be addressed. Firstly, it's crucial to thoroughly understand the hydrogeological characteristics and data availability of the target region to tailor the model architecture and data pre-processing techniques accordingly. Secondly, conducting comprehensive validation studies using local datasets and evaluating model performance against traditional forecasting methods will provide valuable insights into the model's effectiveness and reliability. Additionally, fostering collaboration between hydrologists, data scientists, and stakeholders can facilitate knowledge exchange and ensure

the successful implementation and adoption of coupled LSTM-MODWPT-based deep-learning approaches for GWL forecasting in diverse regions.

5 Conclusions

An efficient and sustainable groundwater management plan can be developed using accurate and reliable predictions of GWLs. Such planning will aid in recommending optimal abstraction rates and groundwater usage for agricultural, domestic, and industrial purposes. However, due to the nonlinear nature of GWLs, as well as their multiscale and time-varying behavior, it is frequently difficult to provide accurate GWL forecasts. One of the most influential pre-requisites of developing ML-based GWL forecast models is the appropriate choice of ML algorithm. LSTM models have demonstrated promising performance in hydrological and other time series predictions. Equally important is the incorporation of a suitable data preprocessing approach, believed to enhance the forecasting performance of ML-based algorithms.

To address these challenges, this research developed a reliable forecasting tool using coupled LSTM-MODWPT models for predicting 1-, 2-, and 3-week(s)-ahead GWL fluctuations. MODWPT was employed to capture multiscale information from the GWL time series, which was then integrated into LSTM models to enhance their forecasting capability. The suitable weekly lag times of GWLs and their wavelet packet transformed counterparts were utilized as input variables for the forecast models, while the outputs were the GWLs projected for 1, 2, and 3 weeks ahead. The selection of an ideal blend of input variables for the proposed models was achieved through a RF-based modeling approach. The proposed models' performance assessment was executed using various statistical performance assessment metrics by which LSTM-MODWPT models were benchmarked against their non-wavelet-based counterparts, e.g., the standalone LSTM models. The results of this study indicated that the LSTM-MODWPT models outperformed the standalone LSTM models across all three future time horizons and at each observation well. Therefore, it can be concluded that LSTM-MODWPT models have the capability to accurately predict multi-step-ahead fluctuations in GWLs quite accurately for the study area.

The coupled LSTM-MODWPT undoubtedly enhances the forecasting performance of the standalone LSTM models. However, a few challenges were encountered during the implementation of the coupled LSTM-MODWPT model for forecasting multi-step ahead GWL fluctuations. One challenge was the selection of optimal hyperparameters for both the LSTM and MODWPT components, considering their interplay and impact on model performance. This was addressed through rigorous experimentation and validation techniques to identify the most suitable parameter configurations. Additionally, integrating the MODWPT feature extraction with the LSTM architecture required careful alignment and synchronization to ensure compatibility and effectiveness. Extensive testing and validation procedures were conducted to fine-tune the integration process and optimize model performance. Lastly, the computational complexity associated with the MODWPT transformation presented resource constraints, necessitating optimization strategies and efficient utilization of computational resources to maintain scalability and practical feasibility. Computational efficiency in developing the proposed LSTM-MODWPT model was attained through the implementation of parallel computing in the MATLAB environment. This system automatically divided tasks and assigned them to a pool of MATLAB workers (equivalent to the physical cores of a multicore desktop computer), enabling computations to be executed in parallel. By addressing these specific challenges through appropriate techniques and methodologies, the implementation of the coupled LSTM-MODWPT model was optimized to achieve improved forecasting performance and scalability.

The proposed modeling approach was promising for short-term GWL forecasts at specified observation wells in a water scarce region of Bangladesh and could potentially be extended to other geographical areas facing similar challenges. Moreover, this promising modeling framework has broader applicability and can be adapted to research areas in hydrology and water resources for medium-term and long-term forecasting. However, there are specific considerations or adaptations that would be necessary when applying this model to different regions with varying hydrogeological characteristics. Specific considerations include the need to adapt input features to reflect regional hydrological processes, adjust model hyperparameters to account for different groundwater dynamics, and validate the model's performance against local data to ensure its effectiveness in capturing region-specific patterns and behaviors. Additionally, incorporating domain knowledge and local expertise can help refine the model's performance in capturing region-specific patterns and behaviors, enhancing its reliability across different hydrogeological contexts.

Data availability Datasets and other materials are available with the authors, and may be accessible at any time upon request.

Code availability MATLAB codes are available with the first author and may be accessible upon request.

Declarations

Ethics approval and consent to participate There are no potential conflicts of interest. The research does not include human participants and/or animals.

The research does not include human participants and/or animals.

Consent for publication The authors give their consent to publish in the Water Resources Management journal if accepted for publication.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Gong Y, Zhang Y, Lan S. A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida. Water Resour Manag. 2016;30:375–91. https://doi.org/10.1007/s11269-015-1167-8.
2. Gong Y, Wang Z, Xu G, Zhang Z. A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. Water. 2018;10:1–20. https://doi.org/10.3390/w10060730.
3. Roy DK, Biswas SK, Mattar MA, El-Shafei AA, Murad KFI, Saha KK, Datta B, Dewidar AZ. Groundwater level prediction using a multiple objective genetic algorithm-grey relational analysis based weighted ensemble of ANFIS models. Water. 2021;13(21):3130. https://doi.org/10.3390/w13213130.
4. Wada Y, Bierkens MFP. Sustainability of global water use: past reconstruction and future projections. Environ Res Lett. 2014;9: 104003. https://doi.org/10.1088/1748-9326/9/10/104003.
5. Banerjee P, Prasad RK, Singh VS. Forecasting of groundwater level in hard rock region using artificial neural network. Environ Geol. 2009;58:1239–46. https://doi.org/10.1007/s00254-008-1619-z.
6. Hoque MA, Adhikary SK. Prediction of groundwater level using artificial neural network and multivariate timeseries models. In: Proceedings of the 5th international conference on civil engineering for sustainable development (ICCESD 2020). KUET, Khulna, Bangladesh; 2020. p. 1–8.
7. Roy DK, Biswas SK, Saha KK, Murad KFI. Groundwater level forecast via a discrete space-state modelling approach as a surrogate to complex groundwater simulation modelling. Water Resour Manag. 2021;35(6):1653–72. https://doi.org/10.1007/s11269-021-02787-6.
8. Doble RC, Pickett T, Crosbie RS, Morgan LK, Turnadge C, Davies PJ. Emulation of recharge and evapotranspiration processes in shallow groundwater systems. J Hydrol. 2017;555:894–908. https://doi.org/10.1016/j.jhydrol.2017.10.065.
9. Masterson JP, Garabedian SP. Effects of sea-level rise on ground water flow in a coastal aquifer system. Ground Water. 2007;45:209–17.
10. Park E, Parker JC. A simple model for water table fluctuations in response to precipitation. J Hydrol. 2008;356:344–9. https://doi.org/10.1016/j.jhydrol.2008.04.022.
11. Fahimi F, Yaseen ZM, El-shafie A. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor Appl Climatol. 2017;128:875–903. https://doi.org/10.1007/s00704-016-1735-8.
12. Govindaraju RS. Artificial neural networks in hydrology. I: Preliminary concepts. J Hydrol Eng. 2000a; 5: 115–123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115).
13. Govindaraju RS. Artificial neural networks in hydrology. II: Hydrologic applications. J Hydrol Eng. 2000;5:124–37. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124).
14. Maier HR, Jain A, Dandy GC, Sudheer KP. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ Model Softw. 2010;25:891–909. https://doi.org/10.1016/j.envsoft.2010.02.003.
15. Sadler JM, Goodall JL, Morsy MM, Spencer K. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. J Hydrol. 2018;559:43–55. https://doi.org/10.1016/j.jhydrol.2018.01.044.

16. Yang T, Asanjan AA, Welles E, Gao X, Sorooshian S, Liu X. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour Res. 2017;53:2786–812. https://doi.org/10.1002/2017WR020482.

17. Solomatine DP, Ostfeld A. Data-driven modelling: some past experiences and new approaches. J Hydroinformatics. 2008;10:3–22. https://doi.org/10.2166/hydro.2008.015.

18. Karandish F, Šimůnek J. A comparison of numerical and machine-learning modeling of soil water content with limited input data. J Hydrol. 2016;543:892–909. https://doi.org/10.1016/j.jhydrol.2016.11.007.

19. Mohanty S, Jha MK, Kumar A, Panda DK. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India. J Hydrol. 2013;495:38–51. https://doi.org/10.1016/j.jhydrol.2013.04.041.

20. Feng S, Kang S, Huo Z, Chen S, Mao X. Neural networks to simulate regional ground water levels affected by human activities. Ground Water. 2008;46:80–90. https://doi.org/10.1111/j.1745-6584.2007.00366.x.

21. Guzman SM, Paz JO, Tagert MLM. The use of NARX neural networks to forecast daily groundwater levels. Water Resour Manag. 2017;31:1591–603. https://doi.org/10.1007/s11269-017-1598-5.

22. Sahoo S, Russo TA, Elliott J, Foster I. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. Water Resour Res. 2017;53:3878–95. https://doi.org/10.1002/2016WR019933.

23. Zhang J, Zhu Y, Zhang X, Ye M, Yang J. Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol. 2018;561:918–29. https://doi.org/10.1016/j.jhydrol.2018.04.065.

24. Adamowski J, Chan HF. A wavelet neural network conjunction model for groundwater level forecasting. J Hydrol. 2011;407:28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013.

25. Daliakopoulos IN, Coulibaly P, Tsanis IK. Groundwater level forecasting using artificial neural networks. J Hydrol. 2005;309:229–40. https://doi.org/10.1016/j.jhydrol.2004.12.001.

26. Obergfell C, Bakker M, Maas K. Identification and explanation of a change in the groundwater regime using time series analysis. Groundwater. 2019;57:886–94. https://doi.org/10.1111/gwat.12891.

27. Roshni T, Jha MK, Deo RC, Vandana A. Development and evaluation of hybrid artificial neural network architectures for modeling spatio-temporal groundwater fluctuations in a complex aquifer system. Water Resour Manag. 2019;33:2381–97. https://doi.org/10.1007/s11269-019-02253-4.

28. Sakizadeh M, Mohamed MMA, Klammler H. Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method. Water Resour Manag. 2019;33:1425–37. https://doi.org/10.1007/s11269-019-02208-9.

29. Samani S, Vadiati M, Azizi F, Zamani E, Kisi O. Groundwater level simulation using soft computing methods with emphasis on major meteorological components. Water Resour Manag. 2022;36:3627–47. https://doi.org/10.1007/s11269-022-03217-x.

30. Dong L, Guangxuan L, Qiang F, Mo L, Chunlei L, Abrar FM, Imran KM, Tianxiao L, Song C. Application of particle swarm 0ptimization and extreme learning machine forecasting models for regional groundwater depth using nonlinear prediction models as preprocessor. J Hydrol Eng. 2018;23:4018052. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001711.

31. Mohanty S, Jha MK, Raul SK, Panda RK, Sudheer KP. Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. Water Resour Manag. 2015;29:5521–32. https://doi.org/10.1007/s11269-015-1132-6.

32. Ghorbani MA, Deo RC, Karimi V, Yaseen ZM, Terzi O. Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. Stoch Environ Res Risk Assess. 2018;32:1683–97. https://doi.org/10.1007/s00477-017-1474-0.

33. Lee S, Lee K-K, Yoon H. Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeol J. 2019;27:567–79. https://doi.org/10.1007/s10040-018-1866-3.

34. Dadhich AP, Goyal R, Dadhich PN. Assessment and prediction of groundwater using geospatial and ANN modeling. Water Resour Manag. 2021;35:2879–93. https://doi.org/10.1007/s11269-021-02874-8.

35. Nadiri AA, Naderi K, Khatibi R, Gharekhani M. Modelling groundwater level variations by learning from multiple models using fuzzy logic. Hydrol Sci J. 2019;64:210–26. https://doi.org/10.1080/02626667.2018.1554940.

36. Nourani V, Mousavi S. Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method. J Hydrol. 2016;536:10–25. https://doi.org/10.1016/j.jhydrol.2016.02.030.

37. Raghavendra SN, Deka PC. Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. Cogent Eng. 2015;2: 999414. https://doi.org/10.1080/23311916.2014.999414.

38. Wen X, Feng Q, Yu H, Wu J, Si J, Chang Z, Xi H. Wavelet and adaptive neuro-fuzzy inference system conjunction model for groundwater level predicting in a coastal aquifer. Neural Comput Appl. 2015;26:1203–15. https://doi.org/10.1007/s00521-014-1794-7.

39. Zare M, Koch M. Groundwater level fluctuations simulation and prediction by ANFIS- and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: application to the Miandarband plain. J Hydro-environment Res. 2018;18:63–76. https://doi.org/10.1016/j.jher.2017.11.004.

40. Aguilera H, Guardiola-Albert C, Naranjo-Fernández N, Kohfahl C. Towards flexible groundwater-level prediction for adaptive water management: using Facebook's Prophet forecasting approach. Hydrol Sci J. 2019;64:1504–18. https://doi.org/10.1080/02626667.2019.1651933.

41. Tang Y, Zang C, Wei Y, Jiang M. Data-driven modeling of groundwater level with least-square support vector machine and spatial–temporal analysis. Geotech Geol Eng. 2019;37:1661–70. https://doi.org/10.1007/s10706-018-0713-6.

42. Barzegar R, Fijani E, Asghari Moghaddam A, Tziritis E. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci Total Environ. 2017;599–600:20–31. https://doi.org/10.1016/j.scitotenv.2017.04.189.

43. Peng T, Zhou J, Zhang C, Fu W. Streamflow forecasting using empirical wavelet transform and artificial neural networks. Water. 2017;9:406 (1-20). https://doi.org/10.3390/w9060406.

44. Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N. A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour Manag. 2013;27:1301–21. https://doi.org/10.1007/s11269-012-0239-2.

45. Makungo R, Odiyo JO. Estimating groundwater levels using system identification models in Nzhelele and Luvuvhu areas, Limpopo Province, South Africa. Phys Chem Earth Parts A/B/C. 2017;100:44–50. https://doi.org/10.1016/j.pce.2017.01.019.

46. Banadkooki FB, Ehteram M, Ahmed AN, Teo FY, Fai CM, Afan HA, Sapitang M, El-Shafie A. Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. Nat Resour Res. 2020;29:3233–52. https://doi.org/10.1007/s11053-020-09634-2.

47. Boubaker S. Identification of monthly municipal water demand system based on autoregressive integrated moving average model tuned by particle swarm optimization. J Hydroinformatics. 2017;19:261–81. https://doi.org/10.2166/hydro.2017.035.

48. Fang H-T, Jhong B-C, Tan Y-C, Ke K-Y, Chuang M-H. A two-stage approach integrating SOM- and MOGA-SVM-based algorithms to forecast spatial-temporal groundwater level with meteorological factors. Water Resour Manag. 2019;33:797–818. https://doi.org/10.1007/s11269-018-2143-x.

49. Wei Z-L, Wang D-F, Sun H-Y, Yan X. Comparison of a physical model and phenomenological model to forecast groundwater levels in a rainfall-induced deep-seated landslide. J Hydrol. 2020;586:124894. https://doi.org/10.1016/j.jhydrol.2020.124894.

50. Mozaffari S, Javadi S, Moghaddam HK, Randhir TO. Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. Water Resour Manag. 2022;36:1955–72. https://doi.org/10.1007/s11269-022-03118-z.

51. Rajaee T, Ebrahimi H, Nourani V. A review of the artificial intelligence methods in groundwater level modeling. J Hydrol. 2019;572:336–51. https://doi.org/10.1016/j.jhydrol.2018.12.037.

52. Sarma R, Singh SK. A comparative study of data-driven models for groundwater level forecasting. Water Resour Manag. 2022;36:2741–56. https://doi.org/10.1007/s11269-022-03173-6.

53. Plappert M, Mandery C, Asfour T. Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. Rob Auton Syst. 2018;109:13–26. https://doi.org/10.1016/j.robot.2018.07.006.

54. Fang W, Zhong B, Zhao N, Love PED, Luo H, Xue J, Xu S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. Adv Eng Inform. 2019;39:170–7. https://doi.org/10.1016/j.aei.2018.12.005.

55. Fan L, Zhang T, Zhao X, Wang H, Zheng M. Deep topology network: A framework based on feedback adjustment learning rate for image classification. Adv Eng Inform. 2019;42:100935. https://doi.org/10.1016/j.aei.2019.100935.

56. Cummins N, Baird A, Schuller BW. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. Methods. 2018;151:41–54. https://doi.org/10.1016/j.ymeth.2018.07.007.

57. Tien Bui D, Hoang N-D, Martínez-Álvarez F, Ngo PTT, Hoa PV, Pham TD, Samui P, Costache R. A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. Sci Total Environ. 2020;701:134413. https://doi.org/10.1016/j.scitotenv.2019.134413.

58. Xu H, Zhou J, Asteris PG, Jahed Armaghani D, Tahir MM. Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. Appl Sci. 2019;9(18):3715. https://doi.org/10.3390/app9183715.

59. Bowes BD, Sadler JM, Morsy MM, Behl M, Goodall JL. Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks. Water. 2019;11(5):1–38. https://doi.org/10.3390/w11051098.

60. Supreetha BS, Shenoy N, Nayak P. Lion algorithm-optimized long short-term memory network for groundwater level lorecasting in Udupi District, India. Appl Comput Intell Soft Comput. 2020. https://doi.org/10.1155/2020/8685724.

61. Barzegar R, Aalami MT, Adamowski J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stoch Environ Res Risk Assess. 2020;34:415–33. https://doi.org/10.1007/s00477-020-01776-2.

62. Chang F-J, Chang L-C, Huang C-W, Kao I-F. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. J Hydrol. 2016;541:965–76. https://doi.org/10.1016/j.jhydrol.2016.08.006.

63. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Networks Learn Syst. 1994;5(2):157–66. https://doi.org/10.1109/72.279181.

64. Hu C, Wu Q, Li H, Jian S, Li N, Lou Z. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water. 2018;10(11):1543. https://doi.org/10.3390/w10111543.

65. Liang C, Li H, Lei M, Du Q. Dongting lake water level forecast and its relationship with the three Gorges dam based on a long short-term memory network. Water. 2018;10(10):1389. https://doi.org/10.3390/w10101389.

66. Tian Y, Xu Y-P, Yang Z, Wang G, Zhu Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. Water. 2018;10(11):1655. https://doi.org/10.3390/w10111655.

67. Sahoo BB, Panigrahi B, Nanda T, Tiwari MK, Sankalp S. Multi-step ahead urban water demand forecasting using deep learning models. SN Comput Sci. 2023;4(6):752. https://doi.org/10.1007/s42979-023-02246-6.

68. Swagatika S, Paul JC, Sahoo BB, Gupta SK, Singh PK. Improving the forecasting accuracy of monthly runoff time series of the Brahmani River in India using a hybrid deep learning model. J Water Clim Change. 2024;15(1):139–56. https://doi.org/10.2166/wcc.2023.487.

69. Dehghani A, Moazam HMZH, Mortazavizadeh F, Ranjbar V, Mirzaei M, Mortezavi S, Ng JL, Dehghani A. Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. Ecol Inform. 2023;75: 102119. https://doi.org/10.1016/j.ecoinf.2023.102119.

70. Jeong J, Park E, Chen H, Kim K-Y, Shik Han W, Suk H. Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. J Hydrol. 2020;582: 124512. https://doi.org/10.1016/j.jhydrol.2019.124512.

71. Mouatadid S, Adamowski J, Tiwari MK, Quilty JM. Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. Agric Water Manag. 2019;219:72–85. https://doi.org/10.1016/j.agwat.2019.03.045.

72. Quilty J, Adamowski J. A maximal overlap discrete wavelet packet transform integrated approach for rainfall forecasting–a case study in the Awash River Basin (Ethiopia). Environ Model Softw. 2021;144: 105119. https://doi.org/10.1016/j.envsoft.2021.105119.

73. Quilty J, Adamowski J. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J Hydrol. 2018;563:336–53. https://doi.org/10.1016/j.jhydrol.2018.05.003.

74. Bangladesh Bureau of Statistics (BBS). District Statistics 2011: Gazipur District-Bangladesh Bureau of Statistics. Statistics and Informatics Division. Ministry of planning. Government of the People's Republic of Bangladesh Retrieved from; 2013. www.bbs.gov.bd.

75. Zahid A, Hossain A. Bangladesh Water Development Board: A bank of hydrological data essential for planning and design in water sector. In: 2nd International Conference on Advances in Civil Engineering 2014 (ICACE-2014). 2014; 26–28 December, 2014, CUET, Chittagong, Bangladesh.

76. Estévez J, García-Marín AP, Morábito JA, Cavagnaro M. Quality assurance procedures for validating meteorological input variables of reference evapotranspiration in mendoza province (Argentina). Water Manag. 2016;172:96–109. https://doi.org/10.1016/j.agwat.2016.04.019Agric.

77. Feng S, Hu Q, Qian Q. Quality control of daily meteorological data in China: 1951–2000: a new dataset. Int J Climatol. 2004;24:853–70. https://doi.org/10.1002/joc.1047.

78. Shafer MA, Fiebrich CA, Arndt DS, Fredrickson SE, Hughes TW. Quality assurance procedures in the Oklahoma Mesonet. J Atmos Oceanic Technol. 2000;17:474–94.

79. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

80. Yuan X, Chen C, Lei X, Yuan Y, Muhammad AR. Monthly runoff forecasting based on LSTM–ALO model. Stoch Environ Res Risk Assess. 2018;32:2199–212. https://doi.org/10.1007/s00477-018-1560-y.

81. Walden AT, Contreras CA. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. Proc R Soc A Math Phys Eng Sci. 1998;454:2243–66. https://doi.org/10.1098/rspa.1998.0257.

82. Nielsen M. On the construction and frequency localization of finite orthogonal quadrature filters. J Approx Theory. 2001;108(1):36–52. https://doi.org/10.1006/jath.2000.3514.

83. Mathworks (2020a) MATLAB Version R2020a. Mathworks, Natick

84. MathWorks, 2020b. Technical documentation [WWW Document]. Select predictors for random forests. URL https://au.mathworks.com/help/stats/select-predictors-for-random-forests.html. Accessed 23 Apr 2020.

85. Li M-F, Tang X-P, Wu W, Liu H-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers Manag. 2013;70:139–48. https://doi.org/10.1016/j.enconman.2013.03.004.

86. Pham-Gia T, Hung TL. The mean and median absolute deviations. Math Comput Model. 2001;34:921–36. https://doi.org/10.1016/S0895-7177(01)00109-1.

87. Rahman ATMS, Hosono T, Quilty JM, Da J, Basak A. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. Adv Water Resour. 2020;141: 103595. https://doi.org/10.1016/j.advwatres.2020.103595.

88. Roy DK. Long short-term memory networks to predict one-step ahead reference evapotranspiration in a subtropical climatic Zone. Environ Process. 2021;8:911–41. https://doi.org/10.1007/s40710-021-00512-4.