hydrology

MDPI

Article

An Enhanced Informer Deep Learning Model for Nationwide Groundwater Level Predictions: A Comparative Study Across 34 Monitoring Stations in China

Yi Zhang [1], Gan Luo [1] and Yanxia Liu [1,2,3,4,*]

[1] College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; yizhang@sdust.edu.cn (Y.Z.); 15170154765@163.com (G.L.)
[2] Laboratory of Marine Geology and Environment, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[3] Laboratory for Marine Geology, Qingdao Marine Science and Technology Center, Qingdao 266061, China
[4] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: siner_1234@163.com

Abstract

Groundwater resources are essential to global freshwater supply, and accurate groundwater level prediction is critical for sustainable water resource management. To overcome the limitations of traditional deep learning models in long-sequence groundwater forecasting, including weak generalization, reduced long-term prediction accuracy, and limited interpretability, this study proposes a dual-path Informer-p model integrated with residual theory. The main path captures nonlinear temporal dependencies and long-term hydrological patterns, while the residual path provides a stable linear prediction baseline to enhance local fluctuation representation and robustness to extreme events. The model was validated using long-term groundwater observations from 34 monitoring stations across five major ecosystems in China. Results from representative stations, including Ailao Mountain, showed that Informer-p achieved excellent predictive performance with RMSE = 0.05 m, MAPE = 1.2%, R^2 = 0.95, and KGE = 0.95, reducing RMSE and MAPE by 37.5% and 52%, respectively, compared with the original Informer. Across all stations, Informer-p outperformed the original Informer at 22 stations, with the greatest improvement observed in forest ecosystems. SHAP analysis identified window maximum, original groundwater level, and window minimum as the dominant predictive features. The proposed model provides an effective tool for national-scale groundwater level prediction and sustainable groundwater management.

Keywords: groundwater level prediction; time series forecasting; Informer; machine learning; hydrological modeling

Check for updates

Academic Editor: Emanuele Barca

Received: 25 April 2026
Revised: 30 May 2026
Accepted: 3 June 2026
Published: 8 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Against the background of global water scarcity, where 1 in 4 people around the world lack safely managed drinking water [1], groundwater is a critical global resource, meeting nearly one-third of global water demand [2] and providing clean water to more than two billion people [3]. However, groundwater resources are highly susceptible to the effects of climate change and anthropogenic pollution. Increased demand and unsustainable exploitation—particularly in developing regions—have strained aquifers [4–6]. For instance, regions such as northern India, Iran, and the North China Plain are experiencing

critical declines in groundwater availability [7–9]. The groundwater level (GWL), defined as the depth from the ground surface to the saturated zone, serves as a key indicator of aquifer health and is typically measured through monitoring wells. Analyzing GWL fluctuations is essential for scientifically managing groundwater resources amid growing demand and intensifying climate impacts [10]. Moreover, accurate predictions of groundwater levels support agricultural irrigation [11], food security [12], and sustainable water resource utilization [13] and provide a scientific basis for ecological conservation. These benefits are particularly critical in regions such as northern Ethiopia, which are highly vulnerable to climate variability because of their reliance on rain-fed agriculture and recurrent droughts [14]. With the increasing demand for groundwater resources in these areas, enhancing the understanding of groundwater level dynamics to guide sustainable extraction is vital for local livelihoods and socioeconomic activities.

Groundwater level predictions serve as a crucial tool for the allocation and management of groundwater extraction. Traditional groundwater level predictions employ physical statistical models, which are typically based on geology and numerical simulations and require substantial time and computational resources. With advancements in machine learning (ML), its potential in groundwater quality assessment and prediction has also begun to emerge [15]. ML models have begun to be applied to groundwater level forecasting. Conventional machine learning models include artificial neural networks [16–18], support vector machines (SVMs) [19–21], and random forests [22]. Studies have shown that these machine learning models can achieve comparable or even better accuracy in groundwater level predictions than traditional physics-based approaches can [23]. With further progress in machine learning, models specifically designed for time series forecasting have been proposed. For instance, the LSTM model, developed by S. Hochreiter and J. Schmidhuber, which is based on RNNs [24], effectively captures long-term dependencies in sequences through its gating mechanism [25]. Moreover, the Informer model, improved by Haoyi Zhou et al. and based on the Transformer [26] architecture, significantly enhances long-sequence prediction efficiency via the ProbSparse self-attention mechanism [27]. The Prophet model, proposed by Taylor S J and Letham B, decomposes time series into explicit components, such as trend, seasonality, and holiday effects, thereby strengthening its forecasting capability [28]. In recent studies, additional improved machine learning models have been applied in groundwater level predictions [29–31]. Li et al. further advanced this field [32] by using climate data, hydro-meteorological records, and topographic attributes processed through the AutoGluon automated ensemble framework to predict groundwater level categories.

However, despite these advances, the machine learning applications that are used in groundwater level modeling still face critical limitations. Overfitting, underfitting, and poor generalizability, particularly within deep learning frameworks, remain recurring issues that are often exacerbated by the inclusion of noisy or irrelevant input features [33]. Additionally, in groundwater time series forecasting, models often demonstrate acceptable accuracy in short-term predictions but experience significant performance degradation as the forecast horizon extends [34–36].

Groundwater resources are crucial for water security in many regions of the world, including China. However, managing these resources effectively is becoming increasingly difficult because of climate change and over-extraction. While traditional groundwater level prediction models have proven effective in certain areas, they face limitations when addressing the complex spatiotemporal dynamics of groundwater across diverse environmental gradients. These models often fail to capture long-term dependencies and complex hydrological patterns, especially in heterogeneous regions such as China, where the hydrogeological conditions vary significantly. To address this gap, we propose a novel dual-path Informer model (Informer-p),

which integrates residual networks to better model the fluctuations in groundwater levels. This research focuses on the development of more adaptive, accurate models that are capable of forecasting groundwater levels across various regions with diverse climates and ecological systems. The model is applied to groundwater monitoring stations across 34 regions in China for fitting and forecasting. Furthermore, traditional time series forecasting models, including Informer, LSTM, and Prophet, are implemented with data obtained from these stations for fitting and predictions, and their performance levels are compared against those of Informer-p. The research procedure is as follows:

(a) Data preprocessing was conducted;
(b) A dual-path Informer-p model integrating residual networks and Informer was developed;
(c) Using the Ailaoshan groundwater monitoring station as a representative site, Informer-p was compared with Prophet, LSTM, and Informer in terms of predictive accuracy. The dominant factors influencing groundwater level dynamics were analyzed, and groundwater level variations over the next 1000 days at the representative site were predicted. In addition, the performance improvement of Informer-p relative to Informer was evaluated across 34 monitoring stations within five different ecosystem types.

2. Data and Methods

2.1. Data

The dataset was obtained from the Chinese Academy of Science Discipline Data Center for Ecosystem [37]. This study encompasses 34 typical groundwater monitoring stations located across China, which are widely distributed and exhibit diverse geographical and ecological characteristics, demonstrating strong regional representativeness and scientific value. The locations of these stations are shown in Figure 1. On the basis of the dominant landform types, these stations can be categorized into five major ecosystems, namely, farmland, forestland, grassland, desert, and wetland, covering various environmental gradients from low-elevation plains to high-elevation mountains and from humid monsoon regions to arid inland zones. The stations show significant elevation variations, with the lowest station in Changshu (3.1 m) and the highest in Lhasa (3688 m), reflecting the notable topographic differences from coastal areas to the plateau margin. Geographically, the stations extend from Sanjiang Station in the east (E133°30′) to Cele Station in the west (E80.43°) and from Xishuangbanna Station in the south (N21.85°) to Haibei Station in the north (N37°36′), covering three major natural regions: the eastern monsoon zone, the northwestern arid zone, and the Qinghai–Tibet alpine region. In terms of their observation histories, most stations have conducted continuous monitoring since 2005, which has accumulated more than ten years of sequential data records. The number of data entries varies from three hundred to more than seven thousand, with stations such as Ailaoshan (7731 entries) and Qianyanzhou (6157 entries) having the most extensive datasets. The data used in this study are characterized by comprehensive ecosystem coverage, a wide geographical span, a long observation history, and high data quality. The original dataset includes 34 stations with elevations ranging from 3.1 m to 3100 m and vegetation types covering farmland (15 stations), forestland (10 stations), grassland (2 stations), desert (6 stations), and wetland (1 station). Table 1 displays 11 out of the 34 stations, and the remainder are provided in the appendix. For stations with multiple subsites, only the most representative subsite is selected for this study to enhance regional representativeness.

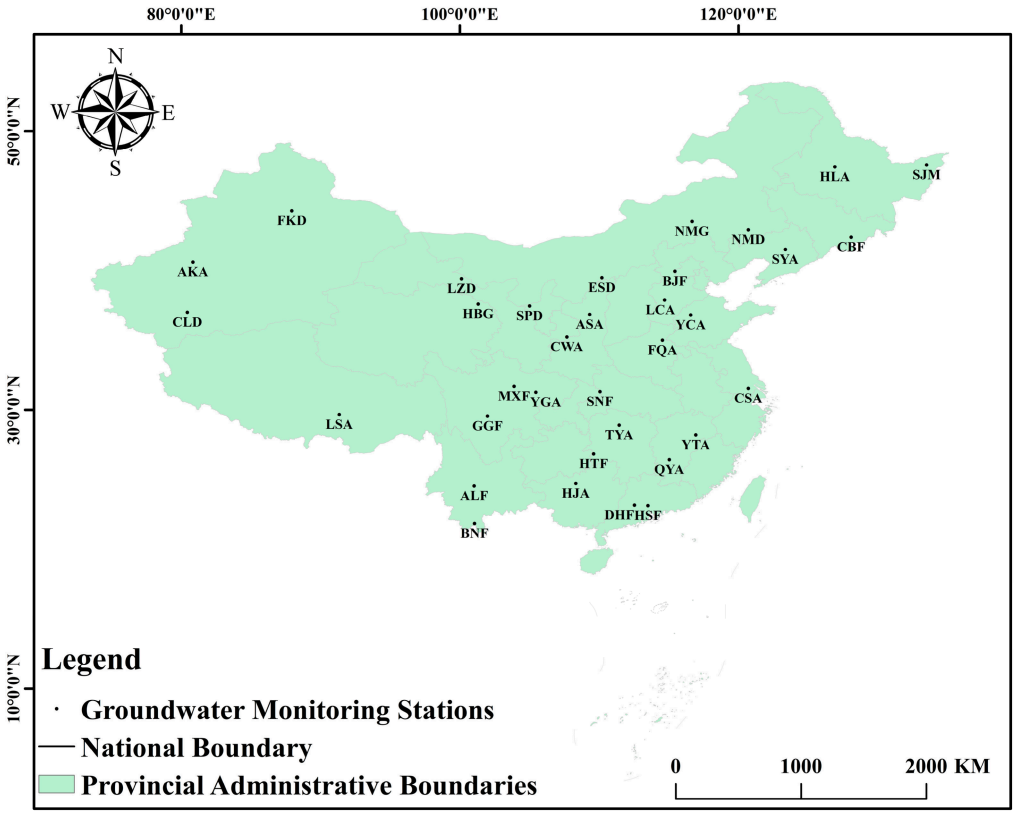


Figure 1. Distribution of 34 monitoring stations in China.

Table 1. Overview of selected monitoring stations.

No.	Station Code	Longitude	Latitude	Elevation (m)	Eco-Type	Start Year	Number of Records
1	AKA	80.83° E	40.62° N	1028	Farmland	2008	2509
2	ASA	109.32° E	36.85° N	1033.26	Farmland	2005	751
3	CSA	120.70° E	31.55°N	3.1	Farmland	2005	1417
4	CWA	107.68° E	35.23° N	1220	Farmland	2005	3136
5	FQA	114.54° E	35.01° N	67.5	Farmland	2005	2944
6	HJA	108.32° E	24.73° N	275.4865	Farmland	2008	1568
7	HLA	126.92° E	47.45° N	234.64	Farmland	2005	961
8	LCA	114.69° E	37.89° N	50.1	Farmland	2005	792
9	LSA	91.34° E	29.67° N	3688	Farmland	2005	1338
10	QYA	115.03° E	26.44° N	67	Farmland	2006	6157
				...			
34	SJM	133.50° E	47.58° N	55.6	Wetland	2005	2532

Ailaoshan (22°30′–24°30′ N, 100°30′–102°00′ E) is located in central Yunnan Province and trends northwest–southeast. The Ailaoshan station dataset contains 7731 records, of which 3286 have been collected from sites that have *Maojuecai* shrub-grassland vegetation, while the remaining data originate from sites that are covered by mid-mountain humid evergreen broad-leaved forest. The stations with *Maojuecai* shrub–grassland vegetation exhibit relatively few instances of missing, damaged, or anomalous data, ensuring a sufficient data volume for analysis. Before the experiment, 3286 valid groundwater observations were visualized and analyzed using the Rescaled Adjusted Partial Sums (RAPS) and Innovative Trend Analysis (ITA) methods, with the methodological details described in Section 2.2.5.

Figure 2a shows that the RAPS curve of the Ailao Mountain station exhibits clear stage-wise fluctuations in the groundwater-level series. The series contains 3286 valid time steps, with a long-term mean groundwater level of 2.4 m and a standard deviation of

0.44 m. The RAPS curve rises rapidly from time step 1 to 133 and reaches its maximum value of 134.0679 at time step 133, indicating that groundwater levels in the early part of the series remained consistently above the long-term mean. The curve then generally shifts downward and reaches its minimum value of −305.5599 at time step 1395, suggesting a prolonged negative deviation during which groundwater levels remained below the mean.

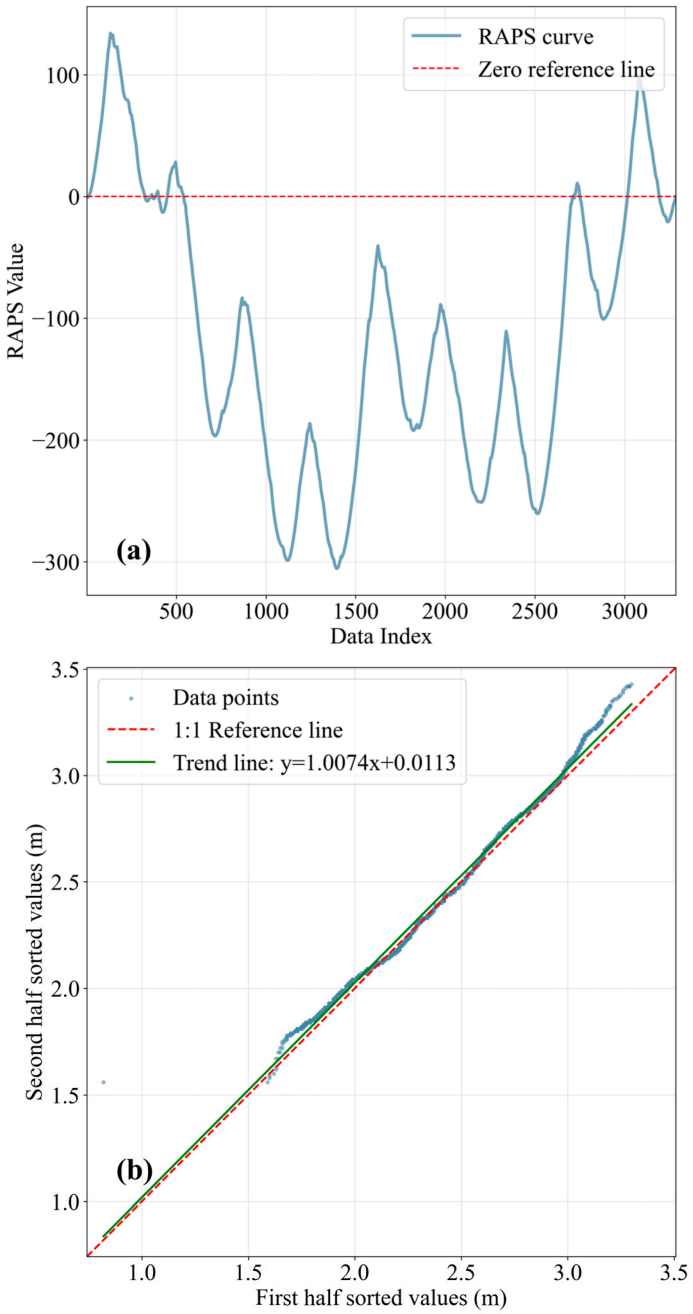


Figure 2. Groundwater level trend analysis of Ailaoshan. (**a**) RAPS curve, (**b**) ITA.

After time step 1395, the RAPS curve shows several recovery rises and staged declines, indicating that groundwater levels did not change monotonically but instead exhibited evident periodic or stage-wise fluctuation characteristics. In particular, in the latter half of the series, the RAPS curve approaches the zero reference line several times and becomes positive during some local stages, suggesting a recovery of groundwater levels relative to the earlier low-level state. Overall, the RAPS results reveal that groundwater levels at this

station experienced a process characterized by "initially high levels, sustained low levels in the early-to-middle period, and fluctuating recovery in the later period."

The ITA results in Figure 2b further demonstrate the distributional differences between the two sub-series. The first half corresponds to time steps 1 to 1643, while the second half corresponds to time steps 1644 to 3286. The mean value of the first half is 2.3917 m, whereas that of the second half is 2.4206 m, representing an increase of 0.0289 m, or a relative increase of 1.2089%. The ITA fitted line is y = 1.0074x + 0.0113, with a slope slightly greater than 1. In addition, 1197 points lie above the 1:1 reference line, accounting for 72.85%, indicating that groundwater levels in the second half are generally slightly higher than those in the first half.

However, 1347 points lie within the ±0.05 m range of the 1:1 line, accounting for 91.0%, which suggests that the overall increasing trend is weak. In terms of value ranges, the low-value range increases by an average of 0.0506 m, while the high-value range increases by an average of 0.0523 m; both show more pronounced increases than the middle-value range. The middle-value range increases by only 0.0139 m, and most points are close to the 1:1 line, indicating that the medium groundwater-level state is relatively stable. Therefore, the ITA plot shows that groundwater levels at this station generally exhibit a slight upward trend, but this trend is mainly reflected at the low- and high-water-level ends, while changes in the middle-water-level range are relatively small.

2.2. Methods

To comprehensively evaluate the performance of the Informer-p model in groundwater time series prediction tasks, three representative time series prediction models are selected as comparative benchmarks in this study: Prophet, the classic long short-term memory (LSTM) network, and the base Informer model upon which Informer-p is built. Prophet, developed by Facebook, is a classic time series prediction method that is based on an additive model. By explicitly decomposing trends, seasonality, and holiday effects, this model is particularly suitable for traditional time series data with strong periodicity and known external influencing factors; it has demonstrated interpretability and robustness in preliminary applications within the field of groundwater hydrology. LSTM, as a typical variant of recurrent neural networks (RNNs), can capture long-term dependencies in time series through its gating mechanism and is a widely used deep learning benchmark model in hydrological time series predictions. The Informer model, proposed by Zhou et al. in 2021 [27], is an improved time series prediction model that is based on the Transformer architecture.

2.2.1. Prophet

Prophet is an additive time series model whose basic formulation is given as (1)–(3), where $\boldsymbol{y}(\boldsymbol{t})$ is the observed value at time $\boldsymbol{t}$, $\boldsymbol{g}(\boldsymbol{t})$ is the trend component, $\boldsymbol{s}(\boldsymbol{t})$ is the seasonal component, $\boldsymbol{h}(\boldsymbol{t})$ is the holiday component, and $\boldsymbol{\epsilon_t}$ is the error term.

$$\boldsymbol{y}(\boldsymbol{t}) = \boldsymbol{g}(\boldsymbol{t}) + \boldsymbol{s}(\boldsymbol{t}) + \boldsymbol{h}(\boldsymbol{t}) + \boldsymbol{\epsilon_t} \tag{1}$$

The trend component offers two trend models, the piecewise linear trend model and the saturating growth trend model, whose mathematical expressions are as follows:

$$g(t) = \left(k + a(t)^T\delta\right)\cdot t + \left(m + a(t)^T\gamma\right) \tag{2}$$

and

$$g(t) = \frac{C}{1 + \exp(-k(t-m))} \tag{3}$$

To unify the data scale and accelerate model convergence, StandardScaler was employed in the Prophet model to standardize the input features and target values using Z-score normalization, and the calculation formula is given in Equation (4):

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

2.2.2. LSTM

LSTM is an enhanced RNN architecture that utilizes gating mechanisms for update operations.

$$f_t = \sigma\left(w_{x,f} + w_{h,f} h_{t-1}\right) \tag{5}$$

$$o_t = \sigma(W_{x,o}\, x_t + W_{h,0} h_{t-1}) \tag{6}$$

$$c_t = c_{t-1} \odot f_t + i_t \odot \tanh(W_{x,c} x_t + W_{h,c} h_{t-1}) \tag{7}$$

$$h_t = \tanh(c_t) \odot o_t \tag{8}$$

In Equations (5)–(8), f is the forget gate, i is the input gate, o is the output gate, t is the current time step, σ is the sigmoid activation function, c is the memory cell, h is the short-term hidden state, W is the weight matrix with the subscript indicating the connected layer, x is the input vector, and $\odot$ is the elementwise multiplication operation. LSTM employs three critical gating units—a forget gate, an input gate, and an output gate—to collaboratively regulate information flow and state updates. The forget gate determines the degree of historical information retention in the memory cell, simulating the memory attenuation process of groundwater systems toward prior states. The input gate controls the update intensity of the memory cell by current inputs, reflecting how newly observed data correct the system state. The output gate modulates the transmission ratio from the memory cell to the hidden state, embodying the mechanisms through which internal system states influence external outputs. The update process of the memory cell dynamically integrates historical memory and current inputs, while the hidden state builds the bridge between internal system memory and external outputs. The main structure of the LSTM is shown in Figure 3.

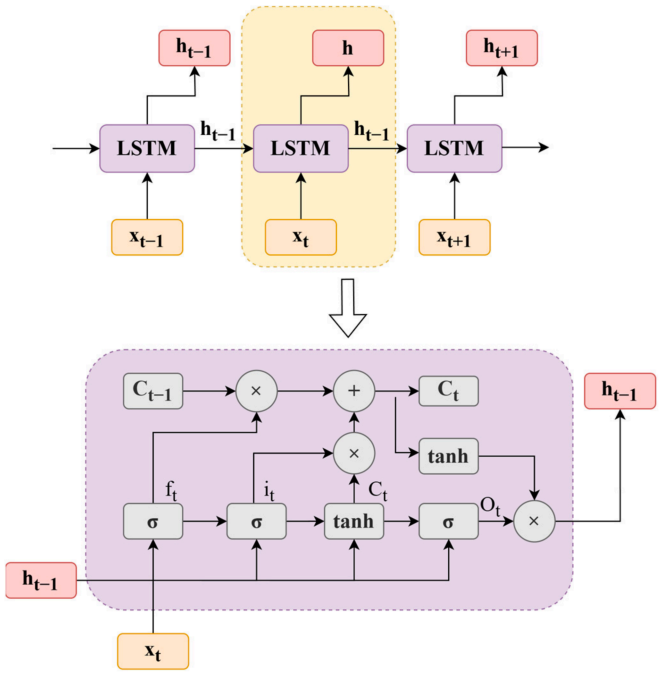


Figure 3. LSTM structure.

For the specific implementation, we employ the PyTorch (version: 1.12.0+cu113) deep learning framework to construct a two-layer LSTM network. Each layer contains 128 hidden units, with a dropout rate of 0.3 applied between layers and a recurrent dropout rate of 0.2 along the temporal dimension. This configuration ensures that the capacity of the model can capture the nonlinear characteristics of hydrological sequences while effectively preventing overfitting. In terms of sequence processing, the differenced data are restructured into supervised learning samples with 20 time steps. The selection of this time window is based on the memory characteristics of groundwater systems with respect to prior states, enabling the effective capture of medium- to long-term dependencies in hydrological processes.

To eliminate nonstationarity and enhance the stationary characteristics of the data, we introduce first-order differencing. For the target series y_t, the difference series is calculated as $\Delta y_t = y_t - y_{t-1}$. A sliding window approach is used to reshape the univariate time series into supervised learning samples. With the window length set to $L = 20$, for each time point t, the features from the preceding L time steps are used as the model input. The target value y_t at the current time step serves as the prediction output. StandardScaler is applied to standardize both the input features and target values via Z-score normalization, after which they are transformed to have zero means and unit variances. This preprocessing step is aimed at unifying data scales and accelerating model training convergence.

The training process involves the mean squared error (MSE) as the loss function and the Adam optimization algorithm, with an initial learning rate of 0.0005. These steps are combined with an adaptive learning rate adjustment strategy based on plateau detection: the learning rate is halved if the validation loss does not improve for five consecutive epochs. Gradient clipping (threshold of 1.0) is implemented to ensure training stability. The model is trained for a total of 200 epochs using a batch training approach with a batch size of 64.

2.2.3. Informer and Informer-p

Informer

The Informer framework is an improved variant of the Transformer architecture specifically designed for long-sequence time series forecasting (LSTF). While retaining the fundamental structure of the original Transformer architecture, Informer introduces key innovations to address efficiency challenges. The relationships between the Informer and Transformer frameworks can be summarized as Equation (9):

$$Informer = Transformer + ProbAttention + Distillation + generative\ decoding \quad (9)$$

The computational complexity of the traditional self-attention matrix $A \in R^{L \times L}$ is $O(L^2)$; however, the Informer reduces this complexity through sparsification.

$$A_{i,j} = \frac{\left(QK^T\right)_{i,j}}{\sqrt{d_k}} \times M_{i,j} \quad (10)$$

$$M_{i,j} = I\{i \in Top - k(q_i)\} \quad (11)$$

As shown in Equations (10) and (11), where Q and K represent the query and key matrices, respectively, and $Top - k(q_i)$ selects the $Top - k$ key vectors for query vector q_i, reducing the complexity to $O(L\ log\ L)$. Additionally, the Informer employs self-attention distillation, which progressively reduces the sequence length layer by layer to decrease computational cost.

$$X_l + 1 = MaxPool(ELU(Conv1D(X_l))) \quad (12)$$

As shown in Equation (12), where X_l represents the output of the l-th layer. The convolutional kernel employs a stride of 2, reducing the sequence length by half.

To capture short-term dynamic patterns in the sequence, a sliding window approach is adopted for feature construction. For the raw value s_t at each time point t, statistical features within its window (window size $w = 5$) are calculated, including the window mean μ_t, standard deviation σ_t, minimum min_t, and maximum max_t. Ultimately, the feature vector for each sample point is expanded as $s_t{}' = \{s_t, \mu_t, \sigma_t, min_t, max_t\}$. Detailed formulas and specifications are presented in Section Dual-Path Informer Model (Informer-p) Design.

Since both Informer and the Informer-p model employ the Transformer-based self-attention mechanism, which is highly susceptible to interference from abrupt hydrological processes, RobustScaler was applied to normalize both the input features and labels. This scaler removes the median and scales the data according to the interquartile range (in this study, the quantile range is set to (5, 95)), effectively reducing the impact of extreme values and demonstrating greater robustness than traditional Z-score normalization.

Dual-Path Informer Model (Informer-p) Design

The Informer-p model extends the original Informer architecture by integrating residual network paths. This dual-path mechanism allows the model to capture both high-order interactions between input features and low-order dependencies. The main path learns the complex nonlinear interactions and temporal patterns, whereas the residual path provides a stable baseline prediction. This architecture is designed to overcome the overfitting and underfitting issues that are commonly encountered in forecasting long-sequence time series by balancing the complexity and computational efficiency of the model.

In traditional residual networks (ResNets), the residual connection is formulated as Equation (13):

$$H(x) = F(x) + x \tag{13}$$

where x is the input feature, $F(x)$ is the residual mapping function to be learned, and $H(x)$ is the final output. Given an input $x \in \mathbb{R}^d$, the forward propagation of a traditional residual block can be formalized as Equation (14):

$$y_l = f\left(W_l \cdot y_{\{l-1\}}\right) + y_{\{l-1\}} \tag{14}$$

where y_l is the output of the l-th layer, W_l represents the layer parameters, and $f(\cdot)$ represents the activation function. In this design, the residual path performs only simple identity mapping without introducing any learnable parameters.

By building upon the foundations of the Informer and residual networks described above, we propose a dual-path residual mechanism capable of effectively capturing fluctuations in the data. This mechanism can be expressed as Equation (15):

$$y = f_{\{main\}}\left(x; \Theta_{\{main\}}\right) + f_{\{residual\}}\left(x; \Theta_{\{residual\}}\right) \tag{15}$$

where the main path $f_{\{main\}}(\cdot)$ and the residual path $f_{\{residual\}}(\cdot)$ are two independent learnable functions with separate parameter sets $\Theta_{\{main\}}$ and $\Theta_{\{residual\}}$. The $f_{\{main\}}(\cdot)$ path captures the high-order interactions and complex dependencies among the input features through multilayer nonlinear transformations, specifically learning nonlinear patterns and fluctuation characteristics in the data. The $f_{\{residual\}}(\cdot)$ path establishes a direct linear mapping from input to output, providing a stable prediction baseline for the model and ensuring basic predictive performance even when the main path is undertrained or struggles to capture temporal features. While the main path focuses on complex nonlinear transformations to capture high-order feature interactions, the residual path is responsible

for establishing an absolute numerical baseline. This characteristic enables the model to concentrate on learning incremental improvements, where the output becomes the linear baseline plus nonlinear refinements.

As shown in Figure 4a, in the main path, the preprocessed feature vector s_t' is projected into a high-dimensional space (some key hyperparameters are shown in Figure 4c, hidden dimension $d_{model} = 128$) via a linear projection layer. The GELU activation function subsequently introduces nonlinearity, and layer normalization is applied to stabilize the training process. This design enhances the model's representational capacity for input features. To equip the model with an awareness of sequence order, a positional encoding composed of superimposed sine and cosine functions is added to the embedded vectors, following the same computational method as in the Transformer model. The calculation formula is given as Equations (16) and (17), where pos is the position index and i is the dimension index. Figure 4b shows the structure of encoder.

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{16}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{17}$$

Compared with the original Informer, which was primarily designed for general long-sequence forecasting tasks, the Informer-p model proposed in this study places greater emphasis on optimizing data characteristics and engineering adaptability for groundwater dynamic prediction scenarios. The original Informer reduces the computational complexity of ultra-long sequence modeling through the ProbSparse Self-Attention mechanism and sequence distillation strategy; however, its ability to explicitly model local fluctuation statistics and abnormal disturbances remains relatively limited. Groundwater level variation processes typically exhibit pronounced nonstationarity, local fluctuations, and anomalous disturbance characteristics. Moreover, due to the limited scale of monitoring samples, the advantages of traditional sparse attention mechanisms for long-sequence modeling cannot be fully exploited in groundwater prediction tasks.

Based on these considerations, Informer-p retains the temporal dependency modeling framework and positional encoding mechanism of Informer while introducing several task-oriented improvements specifically for groundwater forecasting. First, statistical features constructed using sliding windows, including mean, standard deviation, and extreme values, were incorporated to enhance the model's capability to capture local groundwater fluctuation patterns. Notably, to ensure a fair comparison between Informer and Informer-p, the same statistical features were also used as inputs for the original Informer. Second, RobustScaler was adopted to improve the model's robustness against abnormal groundwater fluctuations and noisy observations. In addition, a residual linear prediction path was introduced, enabling the model to simultaneously preserve long-term linear trends and deep nonlinear temporal features in groundwater dynamics. Furthermore, Informer-p replaces the sparse Encoder–Decoder structure with a TransformerEncoder architecture, thereby maintaining global temporal dependency modeling capability while improving training stability and generalization performance on small- and medium-scale groundwater monitoring datasets.

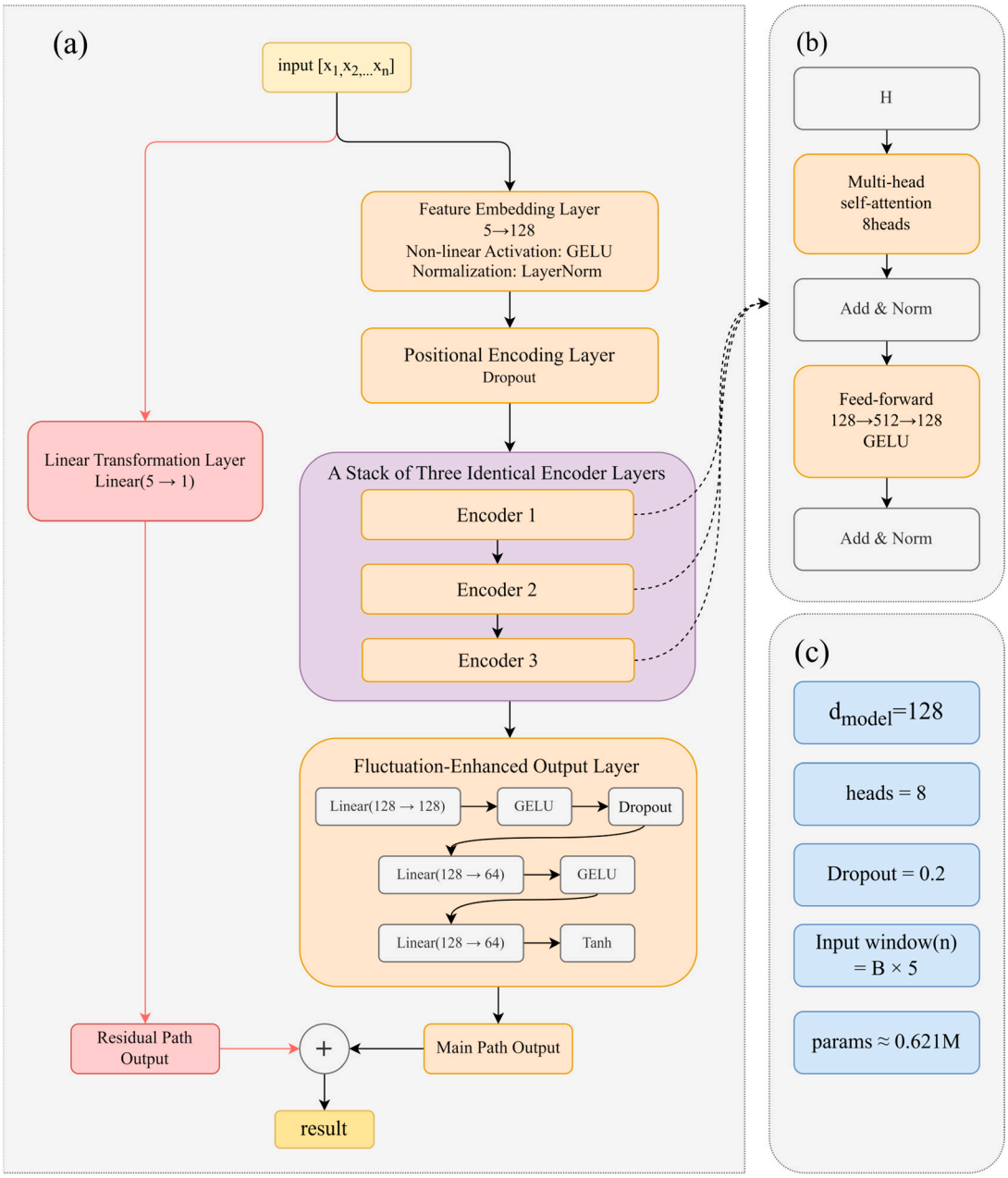


Figure 4. Informer-p. (**a**) Main structure. (**b**) Encoder structure. (**c**) Key hyperparameters and params.

Informer and Informer-p take five customized time-series features constructed by sliding window operation as the raw input, including *Original Value*, *Window Max*, *Window Min*, *Window Mean*, and *Window Std*, the standard deviation characterizing the fluctuation intensity within the sliding time window. Equations (18)–(22) present the calculation procedures for the respective performance features. The original groundwater level sequence is expressed as $x = (x_1, x_2, \ldots, x_t)$, where x_t denotes the observed groundwater level at time t, L represents the length of the sliding window, which is set to $L = 10$ in this study, and W_t stands for the sliding window sequence. The encoder consists of a stack of $N = 3$ identical Transformer encoder layers. Each encoder layer contains a multihead self-attention mechanism (with $n_{head} = 8$ heads) and a feedforward neural network (FFN; with a hidden dimension of $d_{model} \times 4$), both of which utilize the GELU activation function. Residual connections and dropout ($p = 0.2$) are applied between layers to mitigate gradient vanishing and overfitting. The output of the encoder passes through a fluctuation-enhanced output layer, which is implemented as a multilayer perceptron that progressively reduces

the dimensionality to 1. Finally, a tan h activation function is applied to constrain the output within a specific range. The final output is the sum of the two paths.

$$Original\ Value(t) = x_t \tag{18}$$

$$Window\ Max(t) = max\{x_i | x_i \in W_t\} \tag{19}$$

$$Window\ Min(t) = min\{x_i | x_i \in W_t\} \tag{20}$$

$$Window\ Mean(t) = \frac{1}{L}\sum x_t \tag{21}$$

$$Window\ Std(t) = \sqrt{\frac{1}{L}\sum (x_t - Window\ Mean(t))^2} \tag{22}$$

During model training, the Huber loss function is selected because it combines the advantages of the mean squared error and the mean absolute error and is less sensitive to outliers than the MSE while maintaining differentiability for large errors. The AdamW optimizer is chosen for its decoupled weight decay, which helps improve generalizability. The initial learning rate is set to 0.001. Gradient clipping (with a threshold of 1.0) is applied during training to prevent gradient explosion, and an early stopping mechanism (with a patience level of 15) is implemented to halt training when the validation loss ceases to improve, thereby avoiding overfitting.

To further illustrate the improvements of the proposed Informer-p model over the original Informer architecture, a systematic comparison of the key structures of the two models was conducted, as presented in Table 2. Compared with the original Informer, Informer-p incorporates several optimization strategies specifically designed for groundwater level forecasting tasks and hydrological time-series characteristics. First, statistical temporal features, including sliding-window mean, standard deviation, minimum value, and maximum value, were introduced at the input stage to enhance the model's ability to characterize groundwater fluctuation patterns. Second, the input embedding module adopts a "Linear–GELU–LayerNorm" structure, improving nonlinear feature representation capability and training stability. Informer-p also employs a deeper Transformer Encoder architecture with an increased number of attention heads and a larger feed-forward network dimension, thereby strengthening the model's ability to learn long-term temporal dependencies. In addition, a residual prediction branch was introduced to preserve shallow trend information and alleviate potential information loss during deep feature extraction.

Table 2. Key component comparison between Informer and Informer-p.

Key Component	Informer	Informer-p
Input Features	Enhanced temporal statistical features	Enhanced temporal statistical features
Input Dimension	5	5
Embedding Strategy	Single linear projection	Linear + GELU + LayerNorm
Positional Encoding	Sinusoidal positional encoding	Sinusoidal positional encoding
Attention Mechanism	ProbSparse self-attention	Transformer encoder attention
Encoder Layers	3	3
Attention Heads	8	8
Feed-Forward Dimension	512	512
Activation Function	GELU	GELU
Residual Path	Not used	Added residual regression branch
Loss Function	Huber Loss	Huber Loss
Optimizer	AdamW	AdamW
Params	≈0.53 M	≈0.62 M

2.2.4. Introduction of Metrics

The evaluation metrics for the experimental data are the RMSE, MAPE, R^2 and KGE, calculated as (23)–(27) where n is the total number of samples, y_i is the true value of the i-th sample, $\hat{y}_i$ is the predicted value from the model, and $\overline{y}$ is the arithmetic mean of the true values. In the KGE formula, r is the Pearson correlation coefficient between the predicted and true values, which measures the strength of their linear relationship; α is the ratio of the standard deviation of the predicted values to the true values; and β is the ratio of the mean of the predicted values to the true values.

It should be noted that among the 34 monitoring stations nationwide, several special stations exist, such as CSA, which is located in a farmland ecosystem suitable for rice cultivation. During specific periods of the year, the measured groundwater level at the CSA station reaches 0, indicating that the groundwater in the entire farmland area becomes integrated with surface water. Therefore, MAPE was only adopted as one of the evaluation metrics for model comparison at the representative Ailao Mountain station. In the nationwide comparative analysis across all stations, MAPE was not used as an evaluation metric.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{23}$$

$$RMSE = \sqrt{MSE} \tag{24}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{26}$$

$$KGE = 1 - \sqrt{\left[(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2\right]} \tag{27}$$

2.2.5. Data Trend Analysis Methods

We introduced the Rescaled Adjusted Partial Sums (RAPS) and Innovative Trend Analysis (ITA) methods to characterize fluctuation trends in groundwater-level data.

RAPS is a statistical diagnostic method used to identify long-term fluctuations, stage-wise changes, and potential abrupt-change characteristics in time series. By mean-centering and standardizing the original series using the standard deviation, this method transforms the deviation of each time step from the overall mean state into a cumulative anomaly series. This reduces the interference of short-term random fluctuations and highlights persistent increases, decreases, or stage transitions in the series. For a groundwater-level time series x_t, the cumulative standardized deviation $RAPS_k$ up to the k-th time step is calculated as shown in Equation (28). Here, x_t is the groundwater-level value at the t-th time step, $\overline{x}$ is the mean value of the series, and s is the standard deviation. If the RAPS curve shows an upward trend, it indicates that groundwater levels during that stage are generally higher than the series mean; if the curve shows a downward trend, it indicates that groundwater levels during that stage are generally lower than the mean.

$$RAPS_k = \sum_{t=1}^{k}\frac{x_t - \overline{x}}{s} \tag{28}$$

ITA is a non-parametric trend analysis method based on graphical identification. ITA does not require the time series to satisfy the assumption of normal distribution and is highly adaptable to serial autocorrelation and outliers. In addition, ITA can not

only determine the overall trend direction but also reveal different trend characteristics that may exist in the low-, medium-, and high-value ranges. Therefore, it is suitable for analyzing hydrological time series with nonlinear, non-stationary, and multi-stage variation characteristics. The basic idea of ITA is to divide the complete time series into two sub-series of equal length and then sort the two sub-series separately in ascending order. The sorted values of the first sub-series are then used as the horizontal coordinates, while those of the second sub-series are used as the vertical coordinates to generate a scatter plot, with the 1:1 line serving as the no-trend reference line. If the scatter points are mainly distributed above the 1:1 line, the values in the later period are generally higher than those in the earlier period, indicating an upward trend. If the scatter points are mainly distributed below the 1:1 line, the series shows a downward trend. If the scatter points are approximately distributed near the 1:1 line, the series has no significant trend or only a weak trend.

2.2.6. Data Preprocessing

Taking Ailao Mountain as an example, after the raw data were processed, a total of 3286 groundwater depth records spanning 8 years were obtained from the Ailaoshan Maojuecai shrub-grassland station, with a daily measurement interval. To ensure data quality, a preprocessing procedure is applied. Missing values are first filled using interpolation to maintain data completeness. The time series data are then partitioned as follows. The groundwater burial depth sequence is denoted as $S = \{s_0, s_1 \dots s_n\}$ and is divided into $S_{train} = \{s_0, s_1 \dots s_{k-1}\}$, $S_{test} = \{s_k, s_{k+1} \dots s_{t-1}\}$, and $S_{val} = \{s_t, s_{t+1} \dots s_n\}$. Here, k, $t-k$, $n-t$, and n represent the numbers of training samples, test samples, validation samples, and total samples, respectively. Given the annual periodicity of groundwater, data from the same year should be grouped together in the training set, test set, or validation set. A total of 8 years of observations are available, with one full year reserved for predictive validation (the validation set is excluded from model training), resulting in 7 years of data being allocated to the training and test sets. We have tested three training:test:validation split ratios: 6:1:1, 5:2:1, and 4:3:1.

3. Results

3.1. Timing Fitting and Prediction Results Using Ailao Mountain as an Example

3.1.1. Timing Fitting Results and Analysis

Table 3 presents the model performances under different dataset splitting strategies. The 6:1:1 training–test–validation split yields favorable fitting performances for all four models. In particular, both Informer and Informer-p achieve the best results across all the evaluation metrics under this splitting scheme; LSTM also attains optimal values for the RMSE, MAPE, and R^2. On the basis of the 6:1:1 dataset split, Figure 5 displays comparisons between the observed and predicted values of different models on the validation set, and Figure 6 analyzes the prediction errors of each model.

Table 3. Time series fitting results.

Models	Data Split Ratio (Training Set:Test Set:Validation Set)	Metrics of Validation Set			
		RMSE (mm)	MAPE	R^2	KGE
Informer-p	6:1:1	0.05	1.2%	0.95	0.95
Informer		0.08	2.5%	0.95	0.94
LSTM		0.42	23.8%	0.72	0.60
Prophet		0.24	8.3%	0.66	0.78

Table 3. *Cont.*

Models	Data Split Ratio (Training Set:Test Set:Validation Set)	Metrics of Validation Set			
		RMSE (mm)	MAPE	R^2	KGE
Informer-p	5:2:1	0.27	13.9%	0.84	0.87
Informer		0.36	17.5%	0.87	0.82
LSTM		0.56	33.8%	0.65	0.73
Prophet		0.79	40.1%	0.51	0.67
Informer-p	4:3:1	0.39	14.5%	0.75	0.65
Informer		0.39	17.6%	0.77	0.71
LSTM		0.67	24.9%	0.71	0.61
Prophet		0.65	29.7%	0.81	0.71

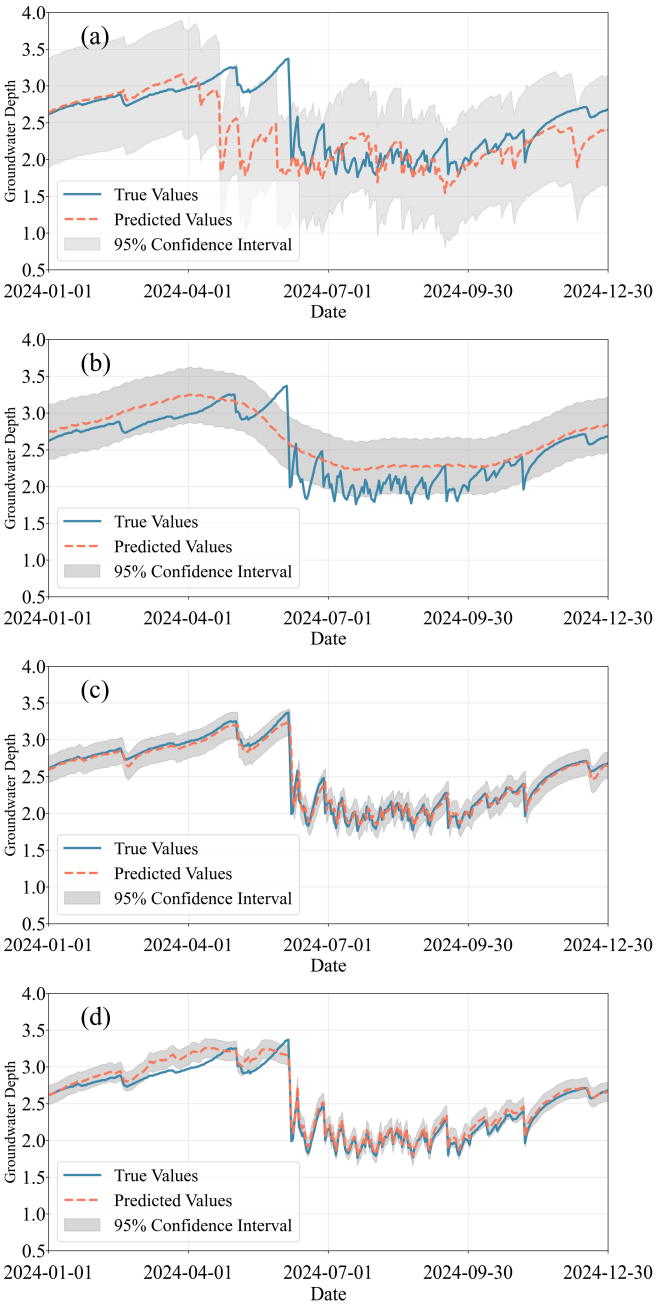


Figure 5. Fitting results. (**a**) LSTM fitting results; (**b**) Prophet fitting result; (**c**) Informer fitting results; (**d**) Informer-p fitting results.

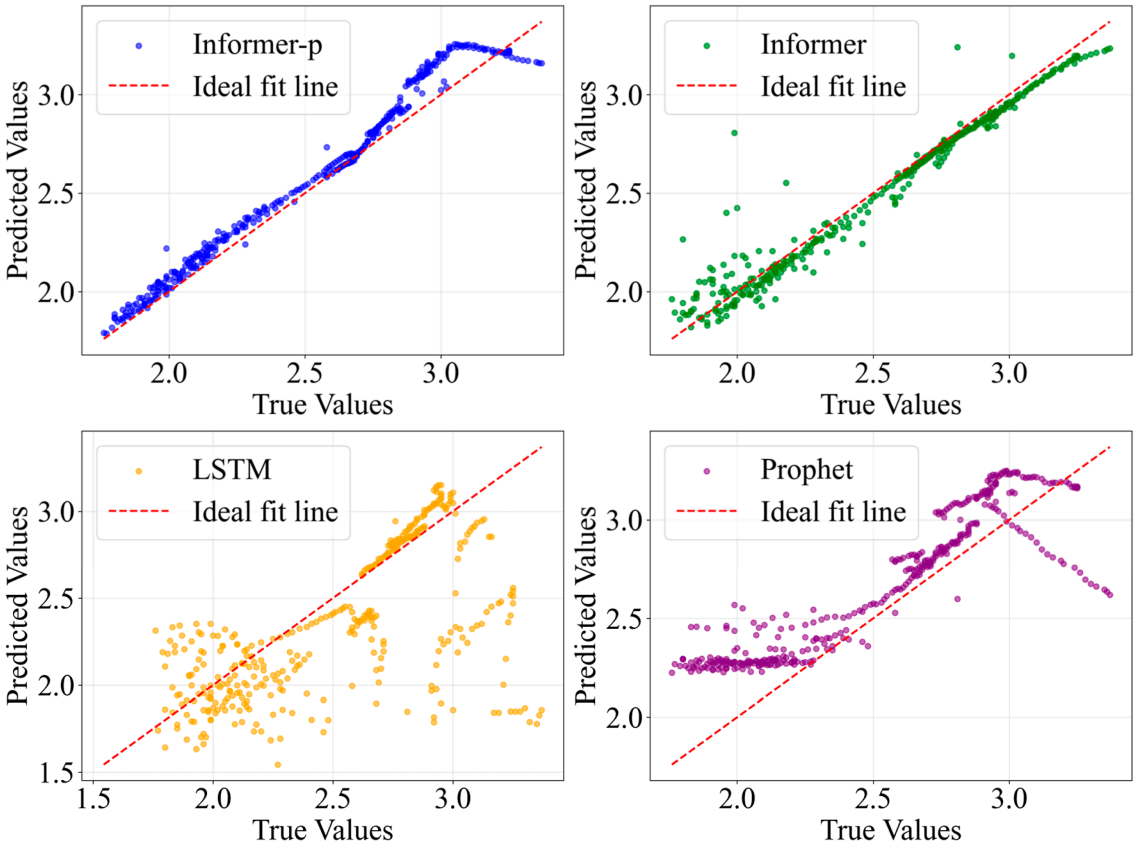


Figure 6. Comparison of the results of the 4 models.

According to the experimental results, our proposed Informer-p model demonstrates the best performance in the groundwater time series forecasting task, achieving an RMSE of 0.05 and a MAPE of 1.2%, which are the lowest among all the compared models. Compared with the original Informer model (RMSE: 0.08; MAPE: 2.57%), Informer-p reduces the RMSE by 37.5% and the MAPE by more than 53%, indicating significant improvement through model optimization.

In the model performance comparison, Informer-p shows noticeable improvement over the original Informer model. Although the original Informer model outperforms traditional models, its prediction accuracy remains limited. In contrast, the LSTM model performs the worst, with an RMSE of 0.42 and a MAPE of 23.81%, indicating the clear limitations of traditional recurrent neural networks in handling long-sequence groundwater data. The Prophet model performs moderately well, with an RMSE of 0.24 and a MAPE of 8.29%, as its additive model structure has a relatively limited ability to capture complex spatiotemporal dependencies.

Figures 7 and 8 present the RAPS and ITA analyses of the groundwater-level prediction results on the test set. The results show that the observed groundwater-level series exhibits clear stage-wise cumulative deviation characteristics. For the Ground Truth series, the maximum RAPS value is 152.2070, occurring at the 158th sample point, while the minimum value is −10.5215, occurring at the 327th sample point. This indicates that groundwater levels during the test period underwent a transition from a high-water-level accumulation stage to a recession stage. All four models reproduce the cumulative variation process of the observed series to some extent. Among them, the maximum RAPS value of Informer-p occurs at the 159th sample point, which is closest to the observed peak position, and its mean value is 2.4987 m, close to the observed mean of 2.5149 m. The ITA in Figure 8 further shows that the trend-line slope of the observed series is 0.7523, indicating that the ranked groundwater levels

in the second half of the period are generally lower than those in the first half, suggesting a decreasing trend. The ITA slopes of Informer-p and Informer are 0.7289 and 0.7245, respectively, both of which are close to the observed trend. In contrast, the slopes of Prophet and LSTM are 0.5662 and 0.4711, respectively, indicating a clear underestimation of their ability to maintain the groundwater-level distribution in the second half of the period. Considering the RAPS peak positions and mean differences in Figure 7 together with the ITA slope results in Figure 8, Informer-p provides a characterization of the stage-wise cumulative changes and distributional trend of groundwater levels that is closer to the observed process.

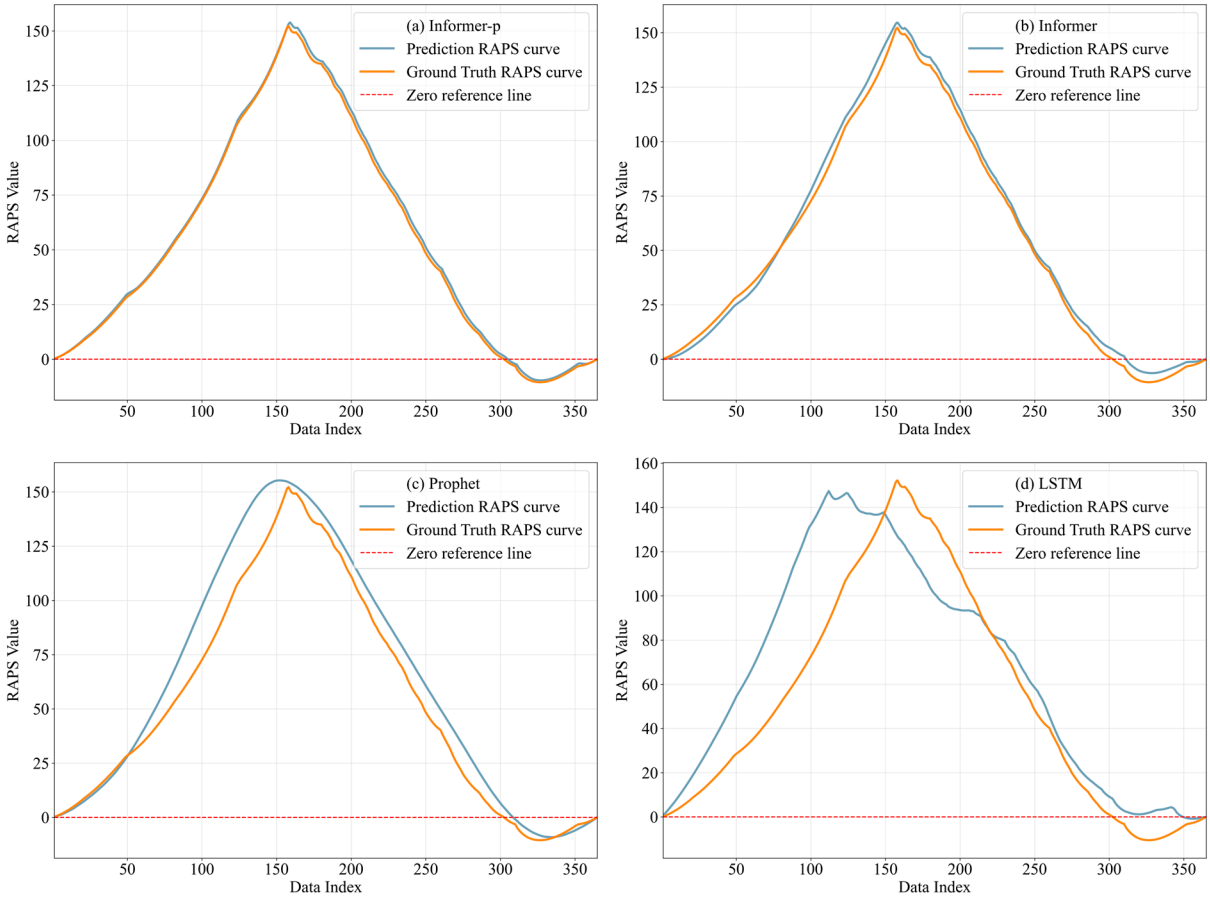


Figure 7. RAPS analysis of groundwater level predictions on the test set. (**a**) RAPS analysis of Informer-p; (**b**) RAPS analysis of Informer; (**c**) RAPS analysis of Prophet; (**d**) RAPS analysis of LSTM.

Figures 9 and 10 show the RAPS and ITA analyses of the absolute error series, further revealing the temporal cumulative characteristics of errors among different models. Informer-p has the lowest mean absolute error, at 0.0584 m, which is lower than those of Informer, Prophet, and LSTM, at 0.0743 m, 0.2053 m, and 0.2711 m, respectively. This indicates that Informer-p has the smallest overall prediction error magnitude. According to the error RAPS results in Figure 9, the maximum and minimum RAPS values of Informer-p are 24.2462 and −40.2301, respectively, indicating relatively limited cumulative error fluctuations. The maximum error RAPS value of Informer reaches 92.1074, suggesting a persistent cumulative deviation above the mean error level during certain stages. Although LSTM has a maximum RAPS value of 36.0863, its mean absolute error and standard deviation reach 0.2711 m and 0.3163 m, respectively, indicating larger error magnitude and greater instability. The ITA results in Figure 10 show that the trend-line slope and intercept of the absolute error series for Informer-p are 0.7035 and 0.0152, respectively, suggesting relatively smooth changes in the error distribution between the first and second halves of the period. Overall, the absolute-

error RAPS and ITA results in Figures 9 and 10 indicate that Informer-p performs better in controlling error magnitude and maintaining cumulative error stability.

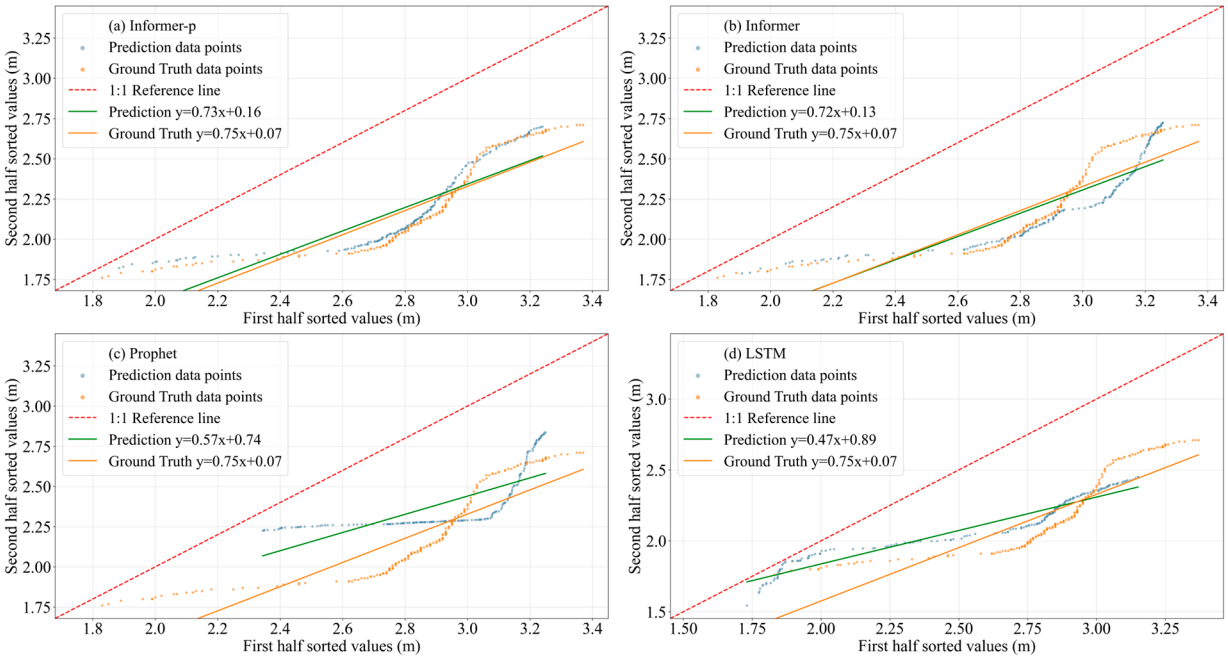


Figure 8. ITA of groundwater level predictions on the test set. (**a**) ITA of Informer-p; (**b**) ITA of Informer; (**c**) ITA of Prophet; (**d**) ITA of LSTM.

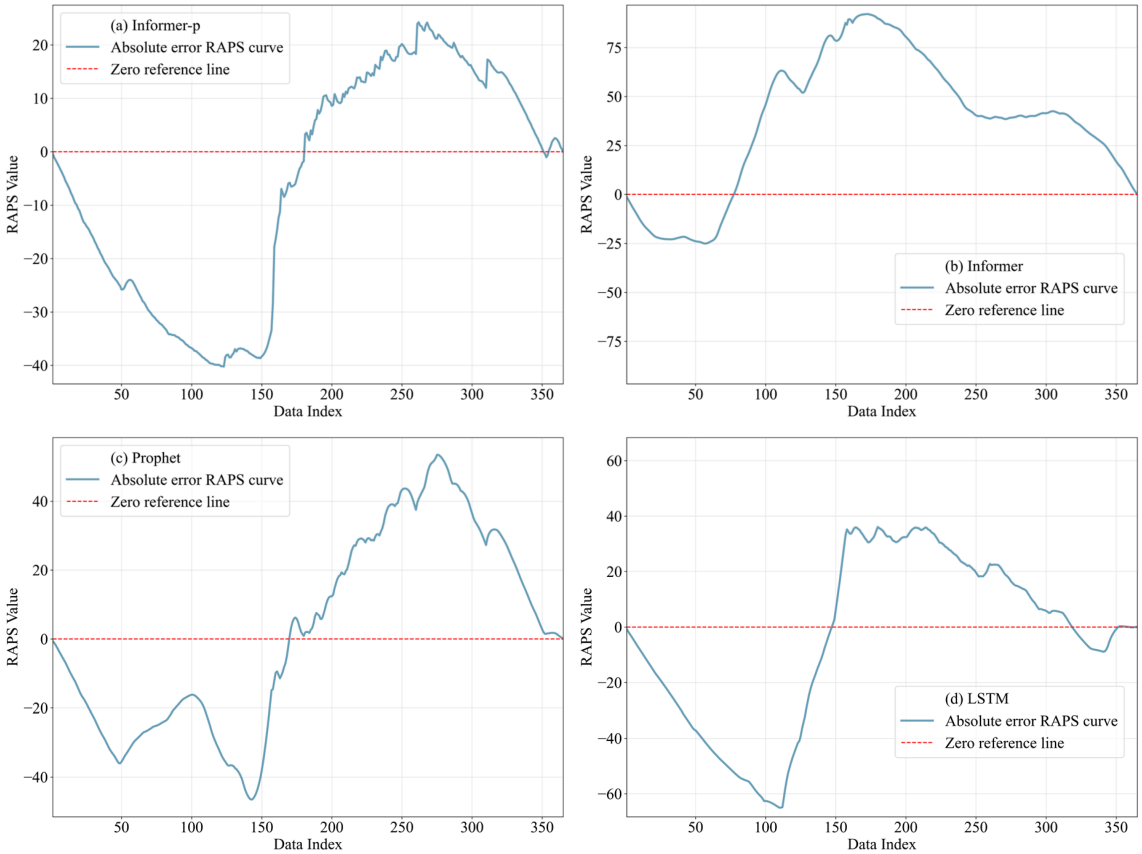


Figure 9. RAPS analysis of absolute prediction errors on the test set. (**a**) RAPS analysis of the absolute errors of Informer-p; (**b**) RAPS analysis of the absolute errors of Informer; (**c**) RAPS analysis of the absolute errors of Prophet; (**d**) RAPS analysis of the absolute errors of LSTM.

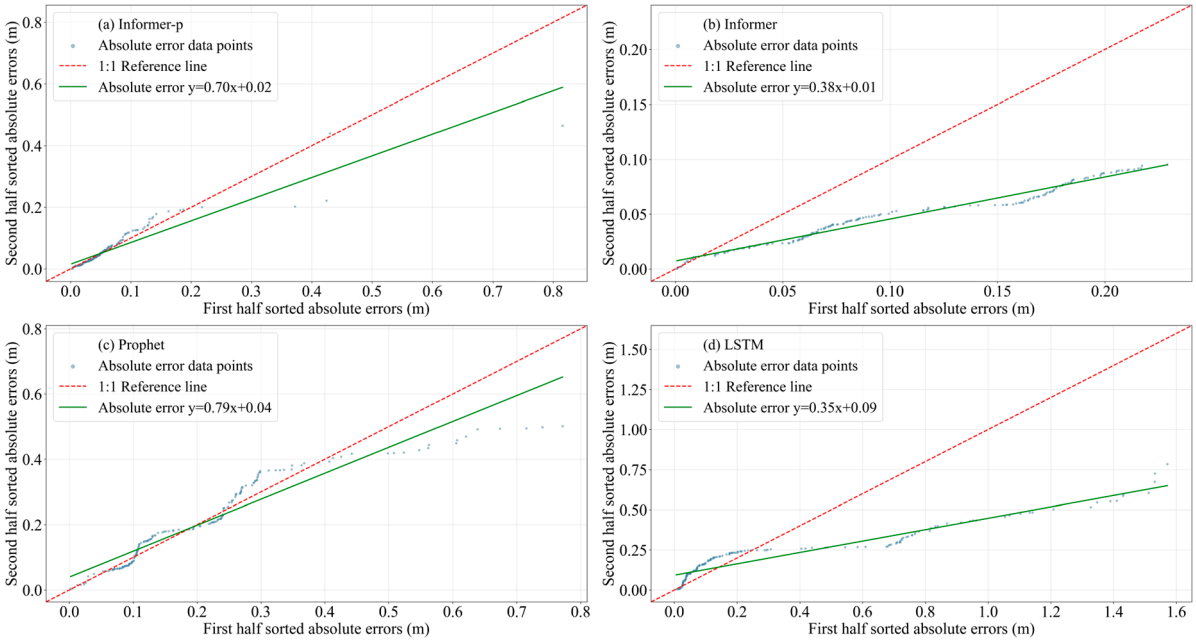


Figure 10. ITA of absolute prediction errors on the test set. (**a**) ITA of the absolute errors of Informer-p; (**b**) ITA of the absolute errors of Informer; (**c**) ITA of the absolute errors of Prophet; (**d**) ITA of the absolute errors of LSTM.

Figure 11 illustrates the SHAP analysis results of the Informer-p model, which quantifies the contribution of each time-series feature to groundwater level prediction. The upper scatter plot indicates that Window Max, Original Value and Window Min present generally positive effects. High feature values (red dots) correspond to positive SHAP values, whereas low feature values (blue dots) exert relatively weak impacts. In contrast, Window Mean shows a prominent negative effect, with its SHAP values concentrated on the left side of the zero axis. The SHAP values of Window Std are generally close to zero, suggesting its limited influence on model outputs. The mean heatmap at the bottom further quantifies the overall contribution of each feature. Window Max has an average SHAP value of 0.045, ranking as the most critical positive influencing factor of the model. Original Value (0.041) and Window Min (0.027) provide secondary positive contributions. The average SHAP value of Window Mean is −0.012, which acts as the major negative factor, while Window Std delivers the weakest contribution at only 0.006. Overall, for the groundwater level prediction model in the Ailaoshan region, time-series extreme values and original values serve as the core driving features for model prediction, and the time-series mean imposes a certain inhibitory effect.

The feature importance sequence revealed by SHAP analysis is highly consistent with the fundamental hydrological behaviors of the groundwater system in the Ailaoshan area, which verifies that the Informer-p model can effectively capture the physical control mechanisms of groundwater dynamics. The dominant positive effects of Window Max and Window Min reflect the system's sensitivity to extreme hydrological events. Specifically, the maximum water level within the window directly reflects the intensity of rainfall infiltration and recharge in the rainy season, and the minimum water level represents the base flow condition and discharge boundary constraints in the dry season. Jointly, these two features determine the overall pattern of groundwater dynamics. The strong positive contribution of Original Value reflects the hysteresis and memory effect inherent to groundwater systems, and the current water level state remains a core basis for predicting future trends. The negative effect of Window Mean can be explained by physical implications: a higher average water level within the window compresses the aquifer regulation and storage

capacity and stabilizes the hydraulic gradient between recharge and discharge, thereby restraining the further rise or decline of groundwater levels.

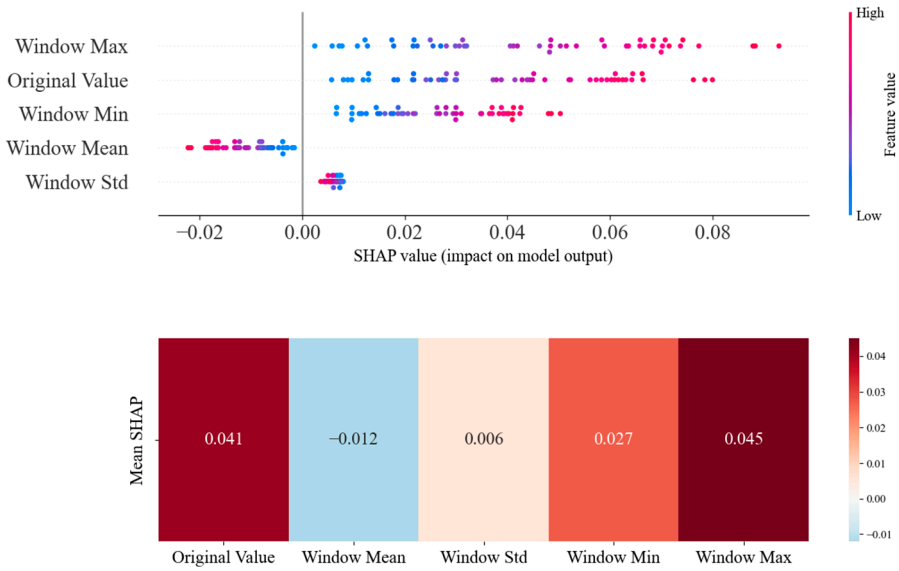


Figure 11. SHAP analysis of Informer-p.

3.1.2. Time Series Forecasting Results

On the basis of the test set predictions, we further employ relatively high-performance Informer and Informer-p models to conduct mid- to short-term forecasts of future groundwater level changes in the Ailaoshan area, with a forecast horizon of 1000 days. The results are shown in Figure 12. The results indicate that both models predict clear periodic patterns in the future groundwater levels for the region. The prediction trajectory of the Informer model is relatively smooth, demonstrating its good ability to capture overall trends. In contrast, the Informer-p model exhibits more pronounced short-term fluctuations, reflecting its sensitive response to local features. Seasonal fluctuations consistent with historical periodic characteristics are still observable in the forecasted sequences, indicating that the models effectively retain cyclical information from the time series. As the forecast horizon extends, some divergence in predictions between the models emerges in the later stages, highlighting the inherent uncertainty in long-term forecasting.

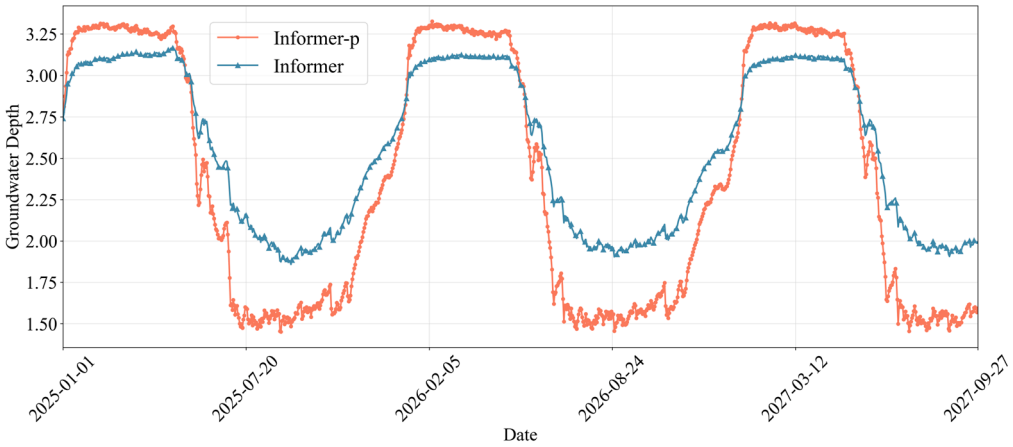


Figure 12. Time series forecasting results with the Informer and Informer-p models.

Figures 13 and 14 present the RAPS and ITA analyses of the partial future prediction series generated by Informer-p and Informer. In this analysis, a two-year period comprising

730 daily time steps was used rather than the complete 1000-day prediction series. This is because groundwater levels exhibit pronounced annual periodicity. If the full 1000-day series were used directly, the sequence would contain two complete annual cycles plus an additional incomplete cycle of 270 days, causing the ITA comparison between the first and second halves to be affected by inconsistent seasonal phases and the incomplete annual cycle. To ensure that the trend comparison has clear hydrological meaning, this study extracted the first 365 × 2 = 730 days of prediction results, so that the first and second halves of the ITA each correspond to one complete annual cycle.

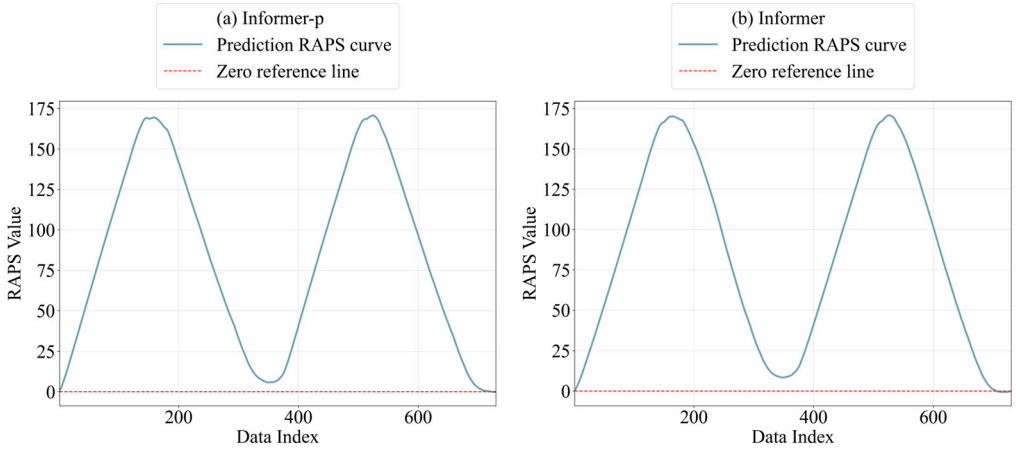


Figure 13. RAPS analysis of two-year future groundwater level predictions at the Ailaoshan station. (**a**) RAPS analysis of Informer-p future predictions; (**b**) RAPS analysis of Informer future predictions.

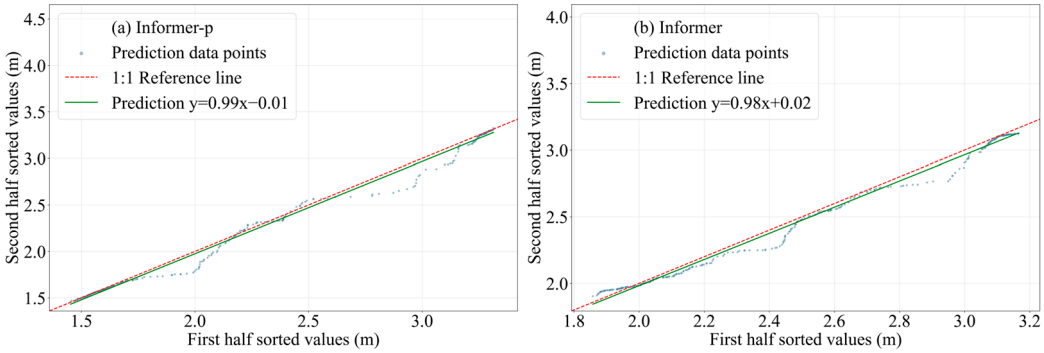


Figure 14. ITA of two-year future groundwater level predictions at the Ailaoshan station. (**a**) ITA of Informer-p future predictions; (**b**) ITA of Informer future predictions.

The two-year prediction results show that the maximum RAPS values of Informer-p and Informer are 170.8444 and 170.8042, occurring near the 524th and 526th prediction points, respectively. This indicates that both models predict a significant high-water-level accumulation stage around the middle of the second year. Their minimum RAPS values are −0.1619 and −0.5530, respectively, both occurring near the end of the prediction period, suggesting that the cumulative deviation tends to decline toward the end of the two-year scale. The ITA results show that the trend-line slopes of Informer-p and Informer are 0.9919 and 0.9821, respectively, both close to 1. This indicates strong consistency in the ranked groundwater-level distributions between the two complete annual cycles, with only a slight decreasing trend. The predicted mean and standard deviation of Informer-p are 2.3611 m and 0.7328 m, respectively, while those of Informer are 2.5418 m and 0.4724 m, respectively, indicating that Informer-p produces stronger interannual fluctuation amplitudes. Since this part is a future prediction analysis without observational constraints, the results are

mainly used to characterize the internal trend structure and annual-cycle stability of the model-predicted series, rather than to directly evaluate prediction accuracy.

3.2. Groundwater Prediction Results at 34 Stations in China

Since LSTM and Prophet demonstrate far inferior performance to Informer and Informer-p in the national-scale historical fitting task, we only adopted the Informer and Informer-p models to conduct historical fitting for 34 monitoring stations across the country, as presented in Table 4. The fitting results indicated that, among the 34 stations, Informer-p outperformed the original Informer across all three evaluation metrics at 17 stations, and achieved superior performance in two metrics at additional 5 stations. In total, 22 stations exhibited more prominent fitting performance with the Informer-p model.

Table 4. Model Performance across 34 Stations in China.

No.	Station Code	Informer-p RMSE	Informer RMSE	Informer-p R^2	Informer R^2	Informer-p KGE	Informer KGE
1	AKA	0.04	0.21	0.96	0.93	0.95	0.94
2	ASA	0.12	0.12	0.96	0.96	0.89	0.84
3	CSA	0.04	0.26	0.97	0.92	0.91	0.89
4	CWA	0.97	0.97	0.91	0.92	0.81	0.93
5	FQA	0.40	0.30	0.93	0.93	0.85	0.86
6	HJA	1.13	0.25	0.81	0.93	0.75	0.86
7	HLA	0.40	1.28	0.97	0.94	0.92	0.87
8	LCA	1.95	1.29	0.85	0.79	0.87	0.75
9	LSA	0.25	1.76	0.89	0.83	0.87	0.82
10	QYA	0.79	0.83	0.73	0.77	0.72	0.69
11	SYA	0.44	0.49	0.82	0.79	0.81	0.75
12	TYA	0.73	2.32	0.93	0.84	0.91	0.87
13	YCA	0.12	1.72	0.95	0.94	0.93	0.91
14	YGA	0.09	0.13	0.97	0.97	0.97	0.97
15	YTA	0.10	0.09	0.96	0.94	0.92	0.93
16	ALF	0.05	0.08	0.95	0.95	0.95	0.94
17	BJF	0.05	0.54	0.95	0.81	0.87	0.79
18	BNF	0.14	0.13	0.93	0.85	0.82	0.79
19	CBF	0.07	0.58	0.95	0.82	0.91	0.89
20	DHF	0.08	0.94	0.93	0.82	0.89	0.78
21	GGF	0.48	0.02	0.82	0.89	0.81	0.9
22	HSF	0.04	0.27	0.98	0.94	0.97	0.92
23	HTF	0.14	1.93	0.94	0.91	0.89	0.88
24	MXF	0.03	0.12	0.96	0.94	0.95	0.95
25	SNF	0.08	0.08	0.93	0.93	0.89	0.9
26	HBG	0.10	0.02	0.91	0.89	0.83	0.9
27	NMG	0.02	0.12	0.96	0.91	0.93	0.89
28	CLD	0.14	0.30	0.89	0.87	0.84	0.78
29	ESD	2.14	1.83	0.82	0.89	0.74	0.89
30	FKD	1.08	1.98	0.81	0.73	0.82	0.81
31	LZD	0.42	0.28	0.71	0.74	0.71	0.69
32	NMD	0.29	0.28	0.91	0.87	0.92	0.91
33	SPD	0.15	0.21	0.93	0.82	0.95	0.8
34	SJM	0.37	1.78	0.87	0.74	0.84	0.65

However, it should also be noted that several stations exhibited clear counterexamples. For instance, at the GGF station, the original Informer achieved a substantially lower RMSE (0.02), whereas the RMSE of Informer-p reached 0.48. Meanwhile, Informer also outperformed Informer-p in terms of both R^2 and KGE metrics. These results indicate that the improvements introduced by Informer-p do not provide universal advantages under all hydrogeological conditions. One possible explanation is that groundwater dynamics at the GGF station are relatively stable and exhibit weaker nonlinear temporal characteristics, allowing the original Informer to effectively capture the dominant temporal dependencies. In contrast, the enhanced feature extraction structure of Informer-p may introduce unnecessary complexity for such relatively simple sequences, thereby leading to local performance degradation.

In addition, at stations such as NMD, the performance differences between the two models were relatively small, suggesting that the performance gains provided by Informer-p exhibit a certain degree of site dependency and are closely related to regional groundwater dynamic characteristics and temporal fluctuation patterns. Therefore, Informer-p should be regarded as an improved model with enhanced overall robustness and broader applicability, rather than as a unified replacement that is consistently superior to the original Informer at all stations. Figure 15 visualizes the quantitative fitting metrics of the Informer-p model for each individual station.

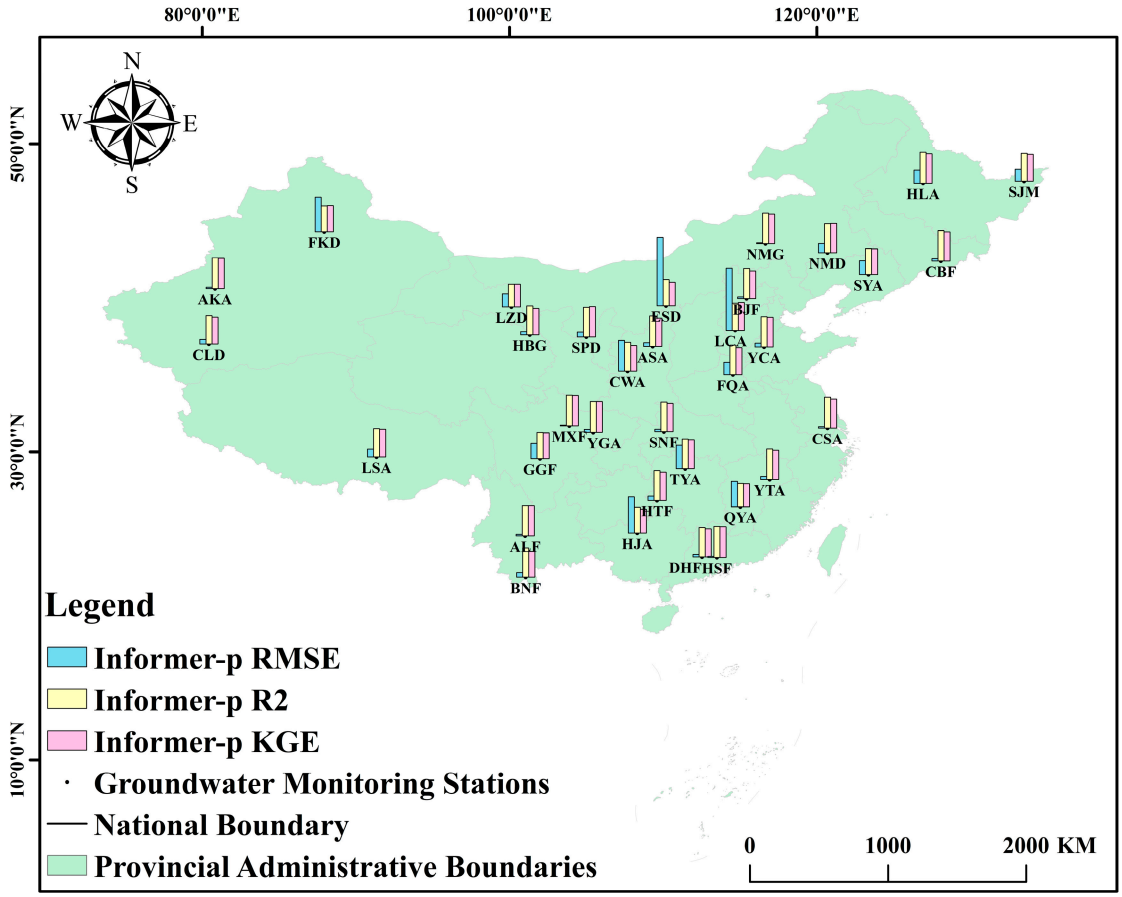


Figure 15. Informer-p Performance across 34 Stations in China.

To further evaluate the adaptability of the model across different ecosystem types, the 34 monitoring stations were grouped according to ecosystem category for statistical analysis. Since the Wetland and Grassland groups contained only one and two samples, respectively, the sample sizes were insufficient for meaningful statistical analysis; therefore, these two ecosystem types were excluded from the analysis. Table 5 presents the mean values and standard deviations of Informer-p and the original Informer in terms of RMSE, R^2, and KGE metrics, while Table 6 presents the results of paired *t*-tests ($\alpha = 0.05$).

Table 5. Statistical Summary of Performance Metrics by Ecosystem Type.

Eco-Type	Number of Stations	Informer-p RMSE	Informer RMSE	Informer-p R^2	Informer R^2	Informer-p KGE	Informer KGE
Farmland	15	0.50 ± 0.53	0.80 ± 0.72	0.91 ± 0.07	0.89 ± 0.07	0.87 ± 0.07	0.86 ± 0.08
Forest	10	0.12 ± 0.13	0.47 ± 0.59	0.93 ± 0.04	0.89 ± 0.06	0.89 ± 0.05	0.87 ± 0.06
Desert	6	0.7 ± 0.78	0.81 ± 0.85	0.84 ± 0.08	0.82 ± 0.07	0.83 ± 0.1	0.81 ± 0.08

Table 6. Paired *t*-test Results of Performance Metrics by Ecosystem Type ($\alpha = 0.05$).

Eco-Type	Number of Stations	Metrics	Value of *t*	Value of *p*
Farmland	15	RMSE	−1.51	0.15
		R^2	1.1	0.29
		KGE	0.84	0.42
Forest	10	RMSE	−1.8	0.11
		R^2	2.28	0.05
		KGE	1.23	0.25
Desert	6	RMSE	−0.64	0.55
		R^2	0.91	0.4
		KGE	0.42	0.69

The results indicate that model performance varies across different ecosystem types. Overall, Informer-p exhibited lower average RMSE values and higher average R^2 and KGE values in most ecosystem categories. This trend can also be observed from the *t*-values reported in Tables 5 and 6. The significance test results show that only the R^2 metric in the forest ecosystem demonstrated a statistically significant difference ($t = 2.17$, $p = 0.05$), indicating that Informer-p achieved significantly better goodness of fit than the original Informer. In farmland ecosystems, Informer-p showed clear performance improvement trends in RMSE, R^2, and KGE metrics, although these improvements did not reach statistical significance. In desert ecosystems, no significant differences were observed between the two models for any metric (all $p > 0.4$). Moreover, strong regional variability in hydrological characteristics, such as the reversed prediction performance observed at the ESD station, further prevented Informer-p from establishing a statistically stable advantage in desert regions.

4. Discussion

From the perspectives of both a representative single-station case and the national spatial scale, the Informer-p model demonstrated overall performance advantages over the original Informer while also exhibiting clear limitations. This result is consistent with previous groundwater-level forecasting studies showing that data-driven and deep learning models can effectively capture nonlinear water-table dynamics, but their performance is strongly affected by station-specific hydrogeological conditions, data quality, and input-feature design [16–18,29,30]. At the representative Ailao Mountain station, Informer-p reduced RMSE by 37.5% and MAPE by 52% during the historical fitting stage while maintaining a high coefficient of determination ($R^2 = 0.95$) and slightly improving the KGE value. Across the 34 monitoring stations covering different ecosystem types and hydrogeological conditions nationwide, Informer-p outperformed the original Informer at 22 stations, among which 17 stations achieved simultaneous superiority in RMSE, R^2, and KGE. These results indicate that the proposed structure is effective in many, but not all, groundwater systems.

However, the performance improvement of Informer-p relative to Informer exhibited strong site dependency. Similar conclusions have been reported in comparative groundwater-modeling studies, where no single machine-learning model consistently performs best across all aquifer settings, especially when hydrological responses differ in linearity, noise level, and external forcing complexity [17,18,23,30]. At stations where groundwater dynamics are relatively stable and nonlinear characteristics are weak, the enhanced nonlinear feature-learning capability of Informer-p may introduce unnecessary model complexity and increase the risk of overfitting. Therefore, the proposed model should not be regarded as a universal solution applicable to all hydrogeological conditions,

but rather as a model more suitable for stations with large groundwater-depth fluctuations and strong nonlinear characteristics.

The superior performance of Informer-p at stations characterized by strong nonlinear groundwater dynamics mainly stems from the synergistic effect between the dual-path residual architecture and the fluctuation-enhanced feature construction strategy. Informer-p inherits the long-term dependency modeling ability of Transformer-based architectures [26] and the efficient long-sequence forecasting design of Informer [27]. Compared with statistical forecasting frameworks such as Prophet [28] and recurrent models such as LSTM [29], Informer-p combines a residual prediction path with a nonlinear attention-based main path. The residual path provides a relatively stable linear baseline, whereas the main path learns nonlinear temporal relationships and abrupt fluctuation characteristics. This "linear baseline + nonlinear correction" structure helps reduce training instability under limited groundwater-monitoring samples.

The SHAP results further verified the physical interpretability of Informer-p to a certain extent. SHAP provides an additive feature-attribution framework for interpreting complex prediction models [38], and its use is particularly valuable for data-driven hydrological models whose internal nonlinear mappings are difficult to directly explain. In this study, Window Max and Window Min were the most influential positive features, suggesting that groundwater systems are sensitive to extreme hydrological states. The high contribution of Original Value also indicates a pronounced memory effect in groundwater dynamics, which is consistent with previous groundwater forecasting studies emphasizing the importance of antecedent groundwater levels and lagged hydrological information [16,17,29]. Nevertheless, SHAP analysis quantifies statistical feature contributions rather than proving physical causality; therefore, the interpretation should be regarded as supportive evidence rather than a complete mechanism explanation.

In addition, the structural design of Informer-p exhibits transferability and extensibility. Although this study focused on groundwater-depth prediction, related studies have shown that deep-learning models can be applied to groundwater forecasting in agricultural, coastal, and water-scarcity regions [29,35,36]. These tasks commonly involve long-term dependency, periodic fluctuation, and abrupt extreme responses, which are also the main targets of the Informer-p design. Therefore, Informer-p can be regarded not only as a groundwater forecasting model but also as a potentially extensible hydrological time-series forecasting framework.

Nevertheless, Informer-p remains a data-driven temporal prediction model and has not yet explicitly coupled external physical variables such as precipitation, evapotranspiration, lithological conditions, land-use change, and anthropogenic groundwater extraction. Previous comparisons between machine learning and numerical groundwater models suggest that data-driven models are efficient and accurate under data-rich conditions, whereas physical or numerical models remain important for mechanism interpretation and scenario extrapolation [23,30]. Future studies should incorporate multi-source hydrometeorological variables and combine them with physical constraints or groundwater numerical simulation models to improve interpretability, robustness, and generalization under nonstationary climate and human-disturbance conditions.

5. Conclusions

This study proposed a dual-path Informer-p deep learning model integrated with residual theory to improve groundwater level prediction accuracy and generalization ability under complex hydrological conditions. Validation using data from 34 representative groundwater monitoring stations in China demonstrated that Informer-p outperformed the original Informer at most stations, particularly in forest and farmland ecosystems. At

the representative Ailaoshan station, the model significantly reduced RMSE and MAPE while more accurately capturing seasonal fluctuations and local variations in groundwater levels. SHAP analysis further revealed that window extreme values and the original groundwater level were the dominant predictive factors, confirming the model's capability to effectively learn groundwater dynamic patterns. The results indicate that the dual-path architecture, combining a linear baseline with nonlinear temporal modeling, enhances the model's adaptability to complex groundwater dynamics and extreme hydrological events. However, the model showed limited advantages at stations with relatively stable and weakly nonlinear groundwater dynamics, and external driving factors such as precipitation, evapotranspiration, and anthropogenic activities were not explicitly incorporated. Overall, Informer-p provides an efficient and reliable approach for national-scale groundwater dynamic prediction and water resource management.

Author Contributions: Conceptualization, Y.Z. and Y.L.; Methodology, Y.Z. and Y.L.; Software, Y.Z.; Validation, Y.Z.; Formal analysis, G.L.; Investigation, Y.Z.; Resources, Y.Z.; Data curation, Y.Z.; Writing—original draft, G.L.; Writing—review & editing, G.L.; Visualization, G.L.; Supervision, G.L.; Project administration, G.L.; Funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Shandong Province (grant number: ZR2023MD105) and Hebei Key Laboratory of Geological Resources and Environment Monitoring and Protection (grant number: JCYKT202501).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data was cited from https://www.escience.org.cn/ (accessed on 2 June 2026).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO/UNICEF Joint Monitoring Programme (JMP). *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2024*; UNICEF: New York, NY, USA; WHO: New York, NY, USA, 2025.
2. Stevanović, Z. Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.* **2019**, *78*, 662. [CrossRef]
3. Famiglietti, J.S. The global groundwater crisis. *Nat. Clim. Change* **2014**, *4*, 945–948. [CrossRef]
4. Evans, R.G.; Sadler, E.J. Methods and technologies to improve efficiency of water use. *Water Resour. Res.* **2008**, *44*, W00E04. [CrossRef]
5. White, I.; Falkland, T. Management of freshwater lenses on small Pacific Islands. *Hydrogeol. J.* **2010**, *18*, 227–246. [CrossRef]
6. Karandish, F.; Salari, S.; Darzi-Naftchali, A. Application of virtual water trade to evaluate cropping pattern in arid regions. *Water Resour. Manag.* **2015**, *29*, 4061–4074. [CrossRef]
7. Mehmood, K.; Tischbein, B.; Flörke, M.; Usman, M. Spatiotemporal analysis of groundwater storage changes, controlling factors, and management options over the transboundary Indus basin. *Water* **2022**, *14*, 3254. [CrossRef]
8. Jadav, K.; Yadav, B. Identifying the suitable managed aquifer recharge (MAR) strategy in an overexploited and contaminated river basin. *Environ. Monit. Assess.* **2023**, *195*, 1014. [CrossRef]
9. Kunwar, G.; Saharia, M.; Getirana, A.; Pandey, A. Detection and socio economic attribution of groundwater depletion in India. *Hydrogeol. J.* **2024**, *32*, 1801–1815. [CrossRef]
10. Butler, J., Jr.; Stotler, R.; Whittemore, D.; Reboulet, E. Interpretation of water level changes in the High Plains aquifer in western Kansas. *Groundwater* **2013**, *51*, 180–190. [CrossRef]
11. Shekhar, S.; Kumar, S.; Densmore, A.; Van Dijk, W.; Sinha, R.; Kumar, M.; Joshi, S.K.; Rai, S.P.; Kumar, D. Modelling water levels of northwestern India in response to improved irrigation use efficiency. *Sci. Rep.* **2020**, *10*, 13452. [CrossRef]
12. Chang, K.-H.; Chiu, Y.-T.; Su, W.-R.; Yu, Y.-C.; Chang, C.-H. A spatial–temporal deep learning-based warning system against flooding hazards with an empirical study in Taiwan. *Int. J. Disaster Risk Reduct.* **2024**, *102*, 104263. [CrossRef]
13. Adnan, R.M.; Mostafa, R.R.; Dai, H.-L.; Heddam, S.; Kuriqi, A.; Kisi, O. Pan evaporation estimation by relevance vector machine tuned with new metaheuristic algorithms using limited climatic data. *Eng. Appl. Comput. Fluid Mech.* **2023**, *17*, 2192258. [CrossRef]

14. Tesfahunegn, G.B.; Gebru, T.A. Smallholder farmers' level of understanding on the impacts of climate change on water resources in northern Ethiopia catchment. *Geojournal* **2022**, *87*, 565–583. [CrossRef]
15. Rajeev, A.; Shah, R.; Shah, P.; Shah, M.; Nanavaty, R. The Potential of Big Data and Machine Learning for Ground Water Quality Assessment and Prediction. *Arch. Comput. Methods Eng.* **2024**, *32*, 927–941. [CrossRef]
16. Taormina, R.; Chau, K.W.; Sethi, R. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* **2012**, *25*, 1670–1676. [CrossRef]
17. Jeong, J.; Park, E. Comparative applications of data-driven models representing water table fluctuations. *J. Hydrol.* **2019**, *572*, 261–273. [CrossRef]
18. Jeong, J.; Park, E.; Chen, H.; Kim, K.Y.; Han, W.S.; Suk, H. Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. *J. Hydrol.* **2020**, *582*, 124512. [CrossRef]
19. Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [CrossRef]
20. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **2015**, *520*, 224–243. [CrossRef]
21. 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*, 965–976. [CrossRef]
22. Sun, K.; Hu, L.; Guo, J.; Yang, Z.; Zhai, Y.; Zhang, S. Enhancing the understanding of hydrological responses induced by ecological water replenishment using improved machine learning models: A case study in Yongding River. *Sci. Total Environ.* **2021**, *768*, 145489. [CrossRef]
23. Chen, C.; He, W.; Zhou, H.; Xue, Y.; Zhu, M. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, Northwestern China. *Sci. Rep.* **2020**, *10*, 3904. [CrossRef]
24. Rumelhart, D.E.; McClelland, J.L. Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*; MIT Press: Cambridge, MA, USA, 1987; pp. 318–362.
25. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.
27. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *arXiv* **2020**, arXiv:2012.07436. [CrossRef]
28. Taylor, S.J.; Letham, B. Forecasting at Scale. *Am. Stat.* **2018**, *72*, 37–45. [CrossRef]
29. 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*, 918–929. [CrossRef]
30. Boo, K.B.W.; El-Shafie, A.; Othman, F.; Khan Md, M.H.; Birima, A.H.; Ahmed, A.N. Groundwater level forecasting with machine learning models: A review. *Water Res.* **2024**, *252*, 121249. [CrossRef] [PubMed]
31. Wu, Z.; Lu, C.; Sun, Q.; Lu, W.; He, X.; Qin, T.; Yan, L.; Wu, C. Predicting Groundwater Level Based on Machine Learning: A Case Study of the Hebei Plain. *Water* **2023**, *15*, 823. [CrossRef]
32. Li, C.; Yang, R. Machine Learning for Proactive Groundwater Management: Early Warning and Resource Allocation (Version 1). *arXiv* **2025**, arXiv:2506.22461.
33. Sharghi, E.; Nourani, V.; Zhang, Y.; Ghaneei, P. Conjunction of cluster ensemble-model ensemble techniques for spatiotemporal assessment of groundwater depletion in semi-arid plains. *J. Hydrol.* **2022**, *610*, 127984. [CrossRef]
34. Peng, Z.; Mo, S.; Sun, A.Y.; Wu, J.; Zeng, X.; Lu, M.; Shi, X. An explainable Bayesian TimesNet for probabilistic groundwater level prediction. *Water Resour. Res.* **2025**, *61*, e2025WR040191. [CrossRef]
35. Zhang, X.; Dong, F.; Chen, G.; Dai, Z. Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 83–96. [CrossRef]
36. Li, W.; Finsa, M.M.; Laskey, K.B.; Houser, P.; Douglas-Bate, R. Groundwater level prediction with machine learning to support sustainable irrigation in water scarcity regions. *Water* **2023**, *15*, 3473. [CrossRef]
37. Zhu, Z.; Tang, X.; Yuan, G.; Zhang, X.; Sun, X.; Chang, X.; Cheng, Y.; Chu, G.; Dai, G.; Dou, S.; et al. CERN groundwater level dataset from 2005 to 2014. *China Sci. Data* **2017**, *2*, 45–53. [CrossRef]
38. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.