



OPEN

Enhancing groundwater level prediction with a hybrid deep learning model in Jinan City, China

Can Zhuang¹✉, Liangliang Cui² & Yi Cui³

Accurate prediction of groundwater levels (GWL) is critical for sustainable utilization and scientific management of groundwater resources. However, precise forecasting of GWL fluctuations faces significant challenges due to the complex nonlinear coupling effects of hydrogeological conditions and hydro-meteorological factors. In recent years, research on GWL prediction based on deep learning models has become a cutting-edge topic in the field of hydrogeology. This study focused on Jinan City, China, and constructed a novel hybrid deep learning model that integrates graph neural networks to capture spatial relationships and recurrent neural networks to model temporal dynamics, effectively learning the complex spatio-temporal patterns in the data, namely the Spatio-Temporal Graph Prediction Model (STGPM). Our approach uniquely captures both hydrological connectivity between monitoring wells and multi-scale temporal dependencies, overcoming key limitations of conventional time-series models. Comparative experiments demonstrate that STGPM outperforms the benchmark models on the test set, achieving the lowest prediction errors (MAE = 0.039, RMSE = 0.052) and the highest coefficient of determination ($R^2=0.988$). Notably, for the monitoring well data not involved in model training, the STGPM still maintains excellent predictive accuracy (MAE = 0.062, RMSE = 0.087, $R^2=0.980$), demonstrating the model's strong generalization ability to unmonitored locations. This study provides water resource managers with a reliable decision-support tool for sustainable groundwater management and spring conservation strategies. The proposed methodological framework also offers a transferable solution for addressing various environmental forecasting challenges characterized by spatial heterogeneity.

Keywords GWL prediction, Deep learning, Spatio-temporal features, Jinan city

Groundwater resources, as the most abundant and valuable freshwater resources globally, play a crucial role in various key areas vital to human activities, such as agricultural irrigation, industrial production, and potable water supply^{1–3}. However, global groundwater systems are currently facing multiple pressures, including overexploitation, environmental pollution, and climate change, which have led to a marked deterioration in both the quantity and quality of groundwater⁴. A representative case is Jinan City, China, renowned as the “Spring City” for its iconic karst spring system, with four major spring groups—Baotu Spring, Heihu Spring, Pearl Spring, and Wulongtan Spring—distributed within its territory, and it has a rich variety of groundwater types⁵. With the rapid development of socio-economy and the continuous advancement of urbanization, the area of spring recharge zones is sharply decreasing. The significant increase in surface imperviousness has severely impaired the infiltration and recharge capacity of karst water, disrupting the natural balance of the regional groundwater system. In recent years, the decline in groundwater levels in the spring distribution areas has posed a severe threat to sustainable spring outflow. Groundwater level (GWL) is a key indicator for measuring the availability and accessibility of groundwater and is closely related to various hydrological and ecological processes^{6–8}. Consequently, accurate GWL prediction constitutes not only an essential foundation for groundwater conservation and ecosystem protection but also a prerequisite for formulating sustainable water management strategies and realizing sustainable utilization^{9,10}.

However, GWL prediction constitutes a complex systemic process, where dynamic variations are a comprehensive response to coupled interactions of climatic, topographic, and hydrogeological factors^{11,12}. This inherent complexity poses significant challenges for precise GWL modeling. Recent advances have witnessed global efforts in developing quantitative and qualitative prediction approaches to establish high-accuracy,

¹School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo, China. ²Jinan Zhongan Digital Technology Co., Ltd, Jinan, China. ³Jinan Qilu Shutong Technology Co., Ltd, Jinan, China. ✉email: zhuangcan@sut.edu.cn

robust GWL forecasting models. Current methodologies for groundwater simulation primarily follow two paradigms: physically-based numerical models and data-driven artificial intelligence models. Physically-based numerical models, such as MODFLOW¹³ and FEFLOW¹⁴, simulate groundwater flow by solving governing partial differential equations derived from physical laws using numerical discretization techniques (e.g., finite difference¹⁵, finite element¹⁶, and finite volume methods¹⁷). These models offer a notable advantage by explicitly elucidating the physical processes driving GWL fluctuations¹⁸. Nevertheless, the prediction accuracy of such methods is inherently constrained by two critical limitations: the difficulty in accurately parameterizing complex surface water potential fields and the frequent unavailability of precise hydrogeological parameters^{10,19}. Furthermore, their high demands for computational resources and data volume often hinder the precise scenario simulation and real-time forecasting^{20,21}. These limitations have sparked growing interest among researchers in data-driven artificial intelligence approaches. Benefiting from their strong capability for nonlinear pattern recognition, artificial intelligence methods have demonstrated remarkable advantages and application potential in groundwater forecasting, effectively overcoming the constraints of traditional statistical techniques²².

Machine learning (ML), a vital research domain within artificial intelligence, uncovers complex mappings between predictors and response variables from historical data by eliminating the need for explicit representation of physical characteristics or underlying mechanisms, thereby providing a viable alternative to computationally intensive physical models²³. Numerous studies have successfully integrated meteorological data with GWL datasets to train ML models^{24,25}, including support vector machines^{26,27}, random forests²⁸, and artificial neural networks²⁹. However, these individual ML models often struggle to address prediction uncertainties arising from model parameterization and structural limitations²³. To address these challenges, hybrid ML models have emerged as valuable tools in groundwater simulation. By combining the predictive capabilities of multiple ML algorithms, the hypothesis space for groundwater dynamics prediction can be effectively expanded, thereby enabling more comprehensive analysis of complex factor interactions⁴. For instance, Pham, et al.³⁰ conducted an in-depth investigation into the performance of seven ML models for GWL prediction, demonstrating that the ensemble learning methods Bagging-RT and Bagging-RF outperformed the other five ML models. Despite the advancements represented by these ML and hybrid models, they primarily remain limited to point-based forecasting, failing to incorporate the spatial interdependencies between monitoring locations, a critical factor in aquifer systems. Furthermore, their performance is often hampered by sensitivity to hyperparameter selection and feature engineering³¹.

To enhance the robustness and accuracy of ML models, researchers integrated them with meta-heuristic optimization algorithms (e.g., Particle Swarm Optimization, Genetic Algorithm) for automated hyperparameter tuning^{9,32–34}. For instance, Saroughi, et al.³⁵ employed the Honey Badger Algorithm (HBA) to optimize parameters of ANN and SVR models, with systematic evaluations confirming that the optimized HBA-ANN and HBA-SVR models significantly outperformed their standalone counterparts. In further research, the team integrated ANN with both Coot and Honey Badger optimization algorithms for GWL prediction in the Tabriz plain of Iran³⁶. Statistical metric selection based on the Shannon entropy criterion verified the superior predictive performance of the Honey Badger optimization algorithm. This hybridization, as evidenced by studies like Thakur and Karmakar³⁷, led to noticeable performance improvements. Nevertheless, while these optimized hybrids addressed parameterization issues, their ability to learn and generalize from the complex, coupled spatio-temporal dynamics inherent in groundwater systems remained inadequate. In addition, although the intelligent optimization algorithms mentioned above demonstrate advantages in efficiency, the present study—considering the small number and discrete nature of the model's hyperparameters—employs the more comprehensive and stable Grid Search method to ensure the optimality and reproducibility of the results.

Deep learning, a significant branch of machine learning, leverages deep neural architectures with high-parameter capacity to effectively capture high-order nonlinear features and complex correlation patterns in data. Substantial empirical research has demonstrated the superior predictive performance of deep learning approaches over both standalone and hybrid ML methods in water resources management tasks^{12,20,38}. Models such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) have become the benchmark for time-series forecasting in hydrology^{39,40}. More recently, architectures like the Transformer have been explored for their superior ability to capture long-range dependencies through self-attention mechanisms⁴¹. Concurrently, Graph Neural Networks (GNNs), such as GraphSAGE⁴², have emerged as powerful tools for modeling relational data and spatial correlations, showing great potential in applications like water quality prediction. A nascent body of research has begun to explore the integration of these temporal and spatial architectures for spatio-temporal forecasting. Studies like Chen, et al.⁴³ have proposed hybrid models (e.g., STGCN) combining graph convolutional networks with temporal modules, demonstrating promising results for regional-scale GWL prediction. Despite these advancements, a significant research gap persists. Inherently, GWL prediction is a quintessential complex spatio-temporal problem⁴³. From a temporal perspective, GWL dynamics exhibit pronounced periodic fluctuations driven by meteorological conditions and seasonal cycles. Spatially, the inherent circulation mechanisms of groundwater cause water level changes in geographically adjacent areas to display strong spatial correlation. Existing research primarily leverages the nonlinear fitting capabilities of machine learning or deep learning to construct predictive models based on the autocorrelation (which measures the correlation of a time series with its own lagged values) and periodic characteristics of time-series data. However, these methods, relying solely on the autocorrelation features of time series, struggle to effectively characterize the spatial heterogeneity among monitoring wells, thereby limiting the prediction accuracy of the models.

To bridge this gap, this study developed a hybrid deep learning model that integrates spatiotemporal features, aiming to provide a scientific decision-support tool for groundwater management in Jinan City. The primary objectives and innovations of this work were threefold:

1. We developed a new deep learning framework that synergistically integrates GraphSAGE and a multi-branch GRU network, which successfully captured both the hydrological connectivity between monitoring wells and multi-scale temporal dynamics of groundwater systems. This design allowed the model to jointly capture immediate responses to rainfall events, seasonal fluctuations, and inter-annual trends;
2. We introduced a trainable cross-attention mechanism to dynamically fuse the multi-scale temporal features with the spatially-aware graph embeddings, replacing simple concatenation or averaging. This enabled more effective and context-aware integration of spatio-temporal information.
3. We designed a dedicated unseen-well test set to rigorously evaluate the model's spatial extrapolation ability. The superior performance on this test set demonstrates that our model learns universal hydrogeological patterns rather than merely memorizing site-specific sequences.

The remainder of this paper is organized as follows: Section “Materials and methods” describes the study area, data sources, preprocessing procedures, and the detailed architecture of the proposed STGPM model. Section “Results and discussion” presents the experimental results, including model performance comparisons, ablation studies, and interpretability analysis. Finally, section “Conclusion” concludes the study.

Materials and methods

The complete workflow of the methodology in this study was divided into three critical stages: data preprocessing, model construction, and model optimization and evaluation (Fig. 1). During the data preprocessing stage, a variety of techniques, including data cleaning and feature engineering, were employed to reconstruct raw data to meet the requirements for model training. The data were also reasonably partitioned to ensure their quality and usability. The modeling phase developed an architecture specifically designed to capture both the temporal

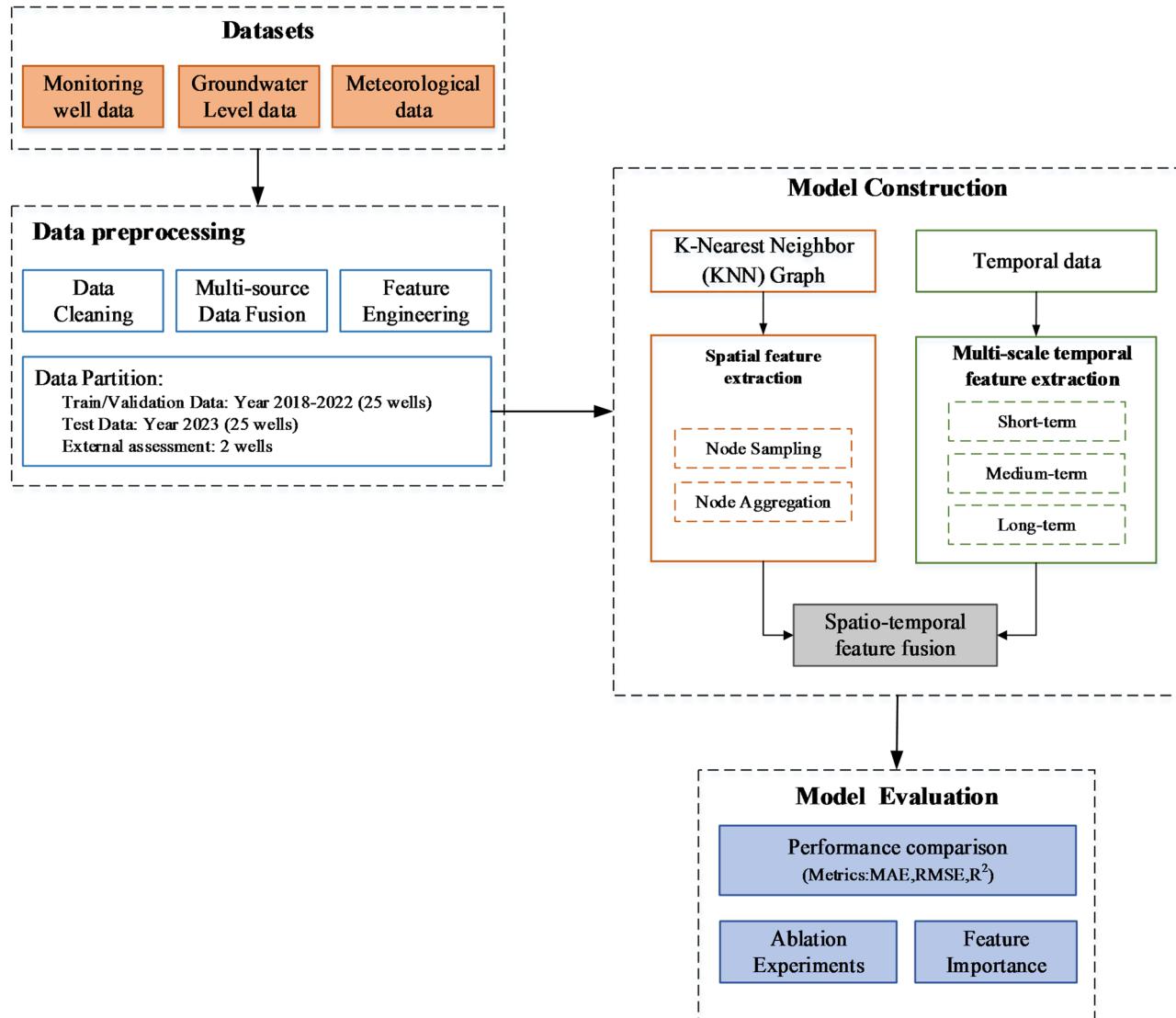


Fig. 1. Workflow diagram of the methodology.

dependencies and spatial correlations inherent in GWL dynamics, establishing a robust foundation for prediction tasks. During the model optimization and validation stage, hyperparameters were iteratively optimized using a loss function and an optimizer until convergence, resulting in the optimal model performance. A comprehensive performance validation was conducted through a multi-indicator, multi-dimensional evaluation system. The experimental design included performance comparison, ablation experiments, and model interpretability analysis to ensure the accuracy and reliability of the model.

Study area and datasets

Study area

This study focused on the administrative region of Jinan City, which is located in the eastern part of North China Plain, in the middle and western part of Shandong Province, East China (Fig. 2). The geographical location ranges from 36°02' N to 37°54' N and from 116°21' E to 117°59' E, with a total area of 8,154 km². The study area featured a warm-temperate continental monsoon climate situated in the mid-latitude inland area. This climatic characteristic results in significant seasonal differences in precipitation, with the majority of rainfall concentrated in the summer months (June to August), accounting for about 70% of the annual precipitation. Additionally, there is considerable inter-annual variability in precipitation. It is worth noting that atmospheric precipitation serves as the dominant recharge source for the karst aquifer system.

Jinan is situated in a transition zone between the low-mountain hills of central-southern Shandong and the alluvial plain of northwestern Shandong. The topography is higher in the south and lower in the north. The southern consists of Ordovician limestone karst aquifers, while the northern is characterized by igneous aquitards. This geological structure creates a natural hydrogeological unit with “southern recharge-northern barrier” that drives regional groundwater flow along a predominant south-to-north gradient⁴⁴. The karst water in the central piedmont plain serves as the main water supply for Jinan, with an average GWL of 45.68 m. In contrast, the southern hilly areas are primarily composed of fracture water, with an average GWL of 227.36 m. Both karst and fracture water levels exhibit significant fluctuations.

Datasets

In this study, we utilized GWL observation data as the target variable and integrated multiple input variables to construct the dataset for predictive modeling. We collected static attribute data from 27 monitoring wells located within the study area (Fig. 2c), including geographical coordinates (longitude and latitude), wellhead elevation, and aquifer type. Concurrently, we obtained GWL time series recorded at 7-day intervals from January 2018 to October 2023, providing absolute elevation values (meters) relative to the national vertical datum.

To capture the complex dynamics of groundwater fluctuations, this study comprehensively considered the lag effect and driving mechanism, and incorporated three key driving factors: (1) Historical GWL data (τ time-lagged terms): characterizing temporal autocorrelation in groundwater systems; (2) Meteorological variables (precipitation, temperature, evapotranspiration): representing external climatic forcing; (3) Spatial factors (water levels of adjacent monitoring wells): quantifying the spatial correlation. Among them, the lag effect was captured through historical GWL time-lags, while the external driving mechanisms were represented by meteorological and spatial factors. These meteorological data were sourced from the National Earth System Science Data Center (<http://www.geodata.cn/main/>), which provides a 1 km resolution monthly dataset for the Chinese region (1901–2023), with each product containing 12 monthly bands. Data from 2018 to 2023 were selected to ensure temporal consistency with the GWL observation data. Through data preprocessing methods, the gridded meteorological data were precisely matched with the locations of each monitoring well, constructing a spatiotemporally consistent multivariate analysis dataset.

Data preprocessing

Data preprocessing constitutes a critical and indispensable step in the construction of deep learning models, playing a decisive role in enhancing model performance³⁴. Our systematic preprocessing pipeline comprised four key phases: (1) data cleaning; (2) multi-source data fusion; (3) feature engineering; (4) dataset partitioning. The workflow of data processing is illustrated in Fig. 3.

Data cleaning

Given the susceptibility of groundwater level (GWL) observations to sensor errors and environmental disturbances, this study implemented rigorous noise reduction protocols to enhance signal-to-noise ratios. To address heterogeneity of multi-source data, we systematically unified the sampling frequency and measurement units across all monitoring wells, thereby eliminating potential biases from data inconsistencies²³. Specifically: (1) depth-to-water measurements were converted to elevation head values using wellhead benchmarks, (2) high-frequency daily data were resampled to 7-day resolution using arithmetic averaging to maintain temporal consistency, and (3) a stringent quality control filter was applied to select wells with a missing rate of less than 30% and no gaps exceeding one consecutive month during 2018–2023. Missing values were then imputed using seasonal-trend decomposition (STL) to preserve the statistical properties of hydrological time series⁴⁵.

Multi-source data fusion

This study employed a systematic data fusion approach to achieve spatio-temporal synchronization between meteorological variables and GWL observations. Utilizing the ArcGIS 10.8 platform, we first extracted monthly bands (2018–2023) from raster datasets for each meteorological element (precipitation, temperature, and potential evapotranspiration) through raster processing. The “Extract Values to Points” spatial analyst tool was then applied to derive precise time-series of meteorological elements at all 27 monitoring well locations. Temporal alignment was rigorously enforced by establishing unified timestamp indices that synchronize GWL

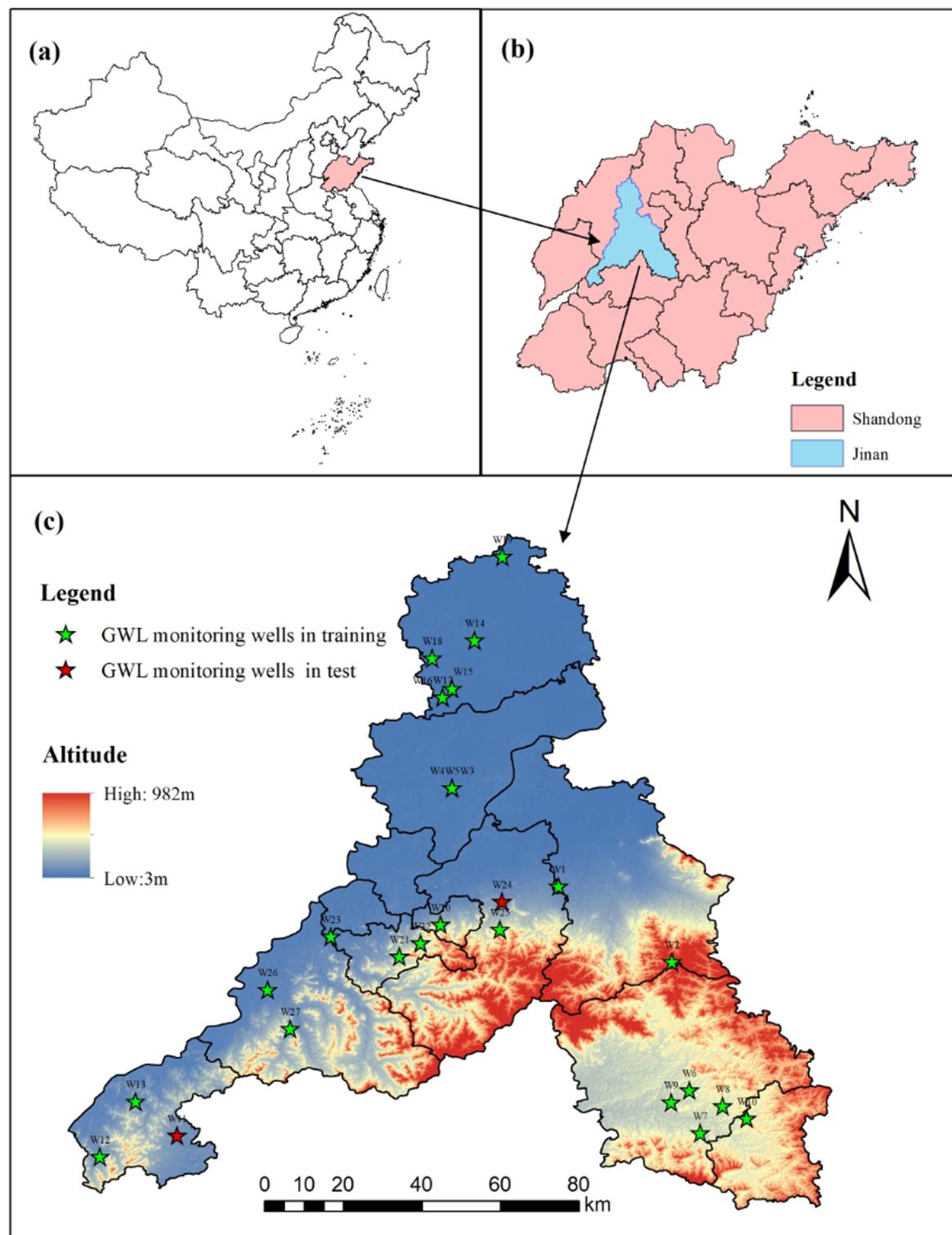


Fig. 2. Geographical location of the study monitoring well in Jinan City, Shandong Province. **(a)** Displays the location of Shandong Province within China; **(b)** provides a zoomed-in view of Shandong Province, with a focus on Jinan, shown in the highlighted area; **(c)** presents a detailed topographic map of the study area and GWL monitoring wells, with the elevation ranging from 3 to 982 m. This figure was created using ArcGIS 10.8.1. The provincial-level administrative boundary map of China and the administrative boundary map of Shandong Province were obtained from the Resource and Environmental Science Data Platform (<https://www.resdc.cn/>) under a free download policy. Note: Some monitoring wells (e.g., W3, W4, W5 and W16, W17) are in very close proximity, and their markers overlap visually.

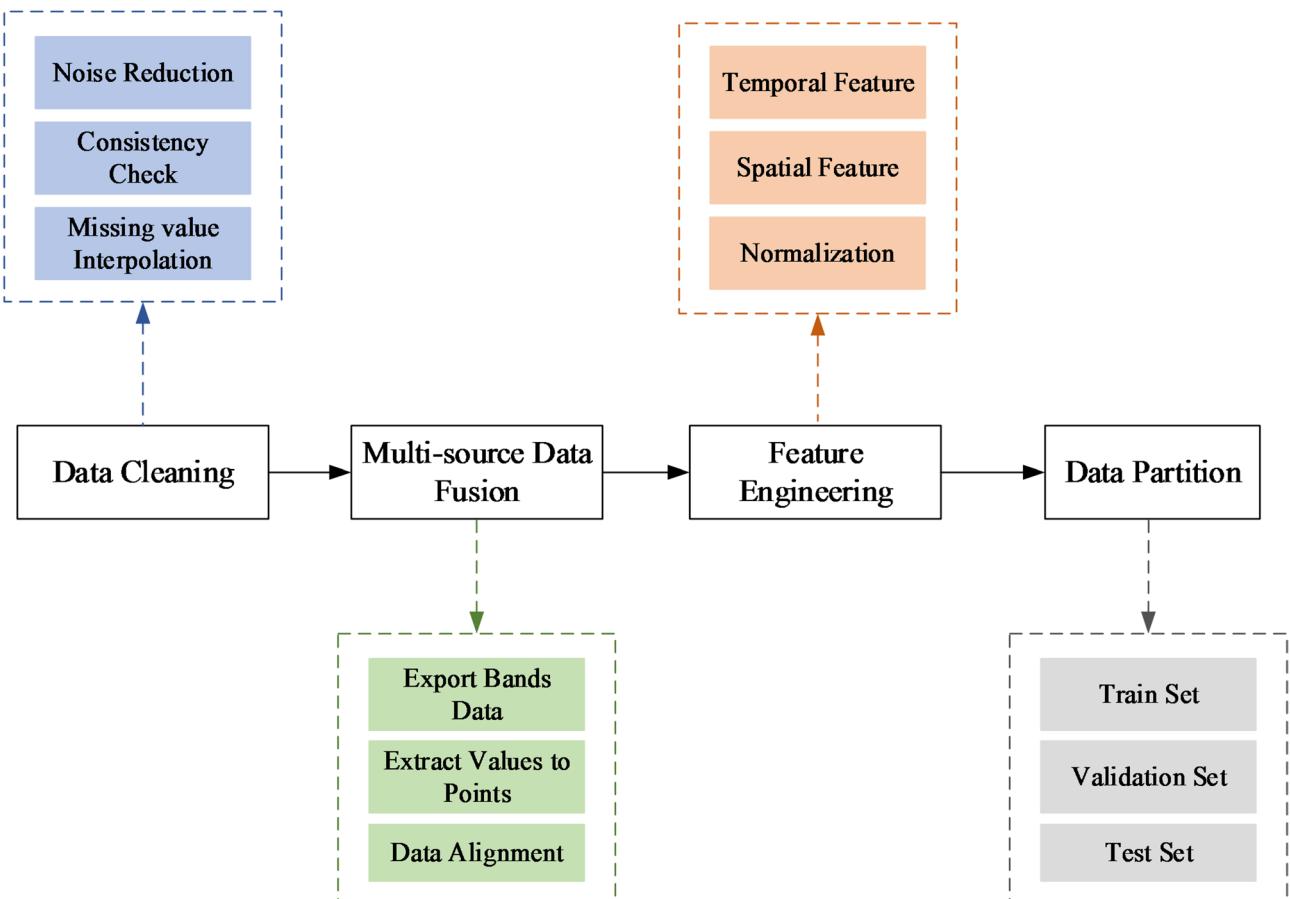


Fig. 3. The workflow diagram of data preprocessing for GWL prediction modeling.

	Input variables					Output variable
	Precipitation (mm)	Temperature (°C)	Evaporation (mm)	GWL_lag1 (m)	GWL_lag2 (m)	
Max	384	29	199.5	434.51	434.51	434.51
Min	0	-4	23.1	-42.89	-42.89	-42.89
Mean	63	14	98.62	76.58	76.27	76.85
Median	30	15	94.6	45.26	44.59	45.44
STD	80.44	10.23	55.36	93.02	93.00	93.03
Skew	1.87	-0.20	0.14	1.73	1.74	1.71

Table 1. Statistical description of the input and output variable.

records with corresponding meteorological measurements, ultimately generating a spatiotemporally coherent multivariate dataset. This fusion process ensured rigorous spatiotemporal alignment of multi-source datasets, establishing a robust foundation for subsequent spatiotemporal modeling.

Feature engineering

Temporal feature engineering: Beyond fundamental meteorological variables (precipitation, temperature, and evapotranspiration), we leveraged the time lag of GWLs to generate lagged GWL features. Autocorrelation function (ACF) analysis of GWL time series across monitoring wells revealed that 20 wells had a significant lag step of 2, while the remaining 7 wells demonstrated a significant lag step of 3. To optimize the trade-off between model complexity and feature representation capacity, this study adopted the maximal consensus lag order (lag 2) across all monitoring wells. Consequently, we constructed the GWL lag features for 7-day lagged values (GWL_lag1) and 14-day lagged values (GWL_lag2) as model inputs. Table 1 presents comprehensive statistics (Mean, Min, Max, STDEV, etc.) for both input and target variables (2018–2023), providing quantitative characterization of aquifer system dynamics and data basis for prediction model training.

Spatial feature engineering: The spatial dataset delineated the geographical coordinates (latitude and longitude) and static attributes (elevation referenced to national geodetic datum, aquifer type classification) for

all monitoring wells. Subsequently, we can precisely compute the hydraulic connectivity metrics between each monitoring well based on Euclidean distances. Table 2 provides representative examples of spatial dataset for some monitoring wells.

Normalization: Given the well-documented sensitivity of deep neural networks to input feature scales, this study implemented rigorous normalization using Scikit-learn machine learning library in the Python environment. This preprocessing step effectively eliminated the dimensional differences between features, ensuring the stability and convergence efficiency of model training and laying the data foundation for subsequent modeling.

Dataset partitioning

To ensure systematic and reliable evaluation of the model, this study adopted the following data partitioning strategy: Initially, the time-series data of two monitoring points randomly selected from the 27 monitoring points were reserved as an independent unseen-well test set to evaluate model performance on unseen monitoring wells. The data of the remaining 25 wells were divided according to the time series, with the data of 2023 year serving as the conventional test set for final validation of prediction accuracy. Data from 2018 to 2022 were used as the model development set, which was strictly divided into a training set (80%) and a validation set (20%) in chronological order. Here, the training set facilitated the learning and optimization of model parameters, while the validation set was used to monitor the generalization ability in real time during the training process and to prevent over-fitting. After partitioning, the training, validation, conventional test, and unseen-well test sets contained approximately 5,220, 1,305, 1,075, and 608 samples, respectively.

Ultimately, model performance was assessed through dual evaluation levels: The conventional test set was used to evaluate temporal extrapolation capability, that is, the predictive accuracy for future time points on known monitoring wells; The unseen-well test set was used to assess spatial extrapolation performance, that is, the predictive adaptability to new monitoring wells. This dual testing strategy evaluated the model performance across both temporal and spatial dimensions, ensuring the comprehensiveness and reliability of the model evaluation.

Model construction

Inherently, GWL prediction is a complex systems problem with significant spatiotemporal coupling characteristics, where dynamic variations are simultaneously influenced by temporal evolution and spatial interactions. In the temporal dimension, GWL exhibits sequential dependence through continuous evolution, with new observations dynamically correlated to their historical states. Spatially, fluctuations in GWL at adjacent monitoring wells show significant hydraulic interdependencies. In response to these characteristics, this study designed a hybrid GWL prediction model integrating spatio-temporal features (STGPM), whose core architecture was organically composed of three key modules: spatial feature extraction, multi-scale temporal feature extraction, and spatio-temporal feature fusion. The structure of the overall model was shown as Fig. 4, which fully incorporated the spatio-temporal coupling mechanisms of the groundwater system, providing a scientifically rigorous modeling paradigm for accurate GWL prediction.

Construction of the K-nearest neighbor graph

To effectively capture hydraulic connectivity between monitoring wells, this study constructed a K-nearest neighbor (KNN) graph based on the geographical coordinates of monitoring wells, where each monitoring well was regarded as a node of the undirected graph. This graph structure could effectively capture the local spatial correlation among monitoring wells, providing neighborhood information for subsequent spatial feature aggregation. The specific steps were as follows:

Coordinate extraction: Extracted latitude and longitude coordinates of each monitoring well from their spatial information to form an $N \times 2$ coordinate matrix (where N is the number of monitoring points, $N=25$).

K-Nearest neighbors calculation: Utilized the nearest neighbor algorithm to calculate the K nearest neighbors for each monitoring well and obtained the Euclidean distances to these nearest neighbors.

Edge construction: Traversed each node and established undirected edges between it and its nearest neighbors, with edge weights set as the inverse of Euclidean distance. This weighting scheme ensured stronger connections between geographically closer nodes, thereby more accurately reflecting the spatial relationships between wells.

The final undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ completely characterized the spatial topology structure of the monitoring well network, where $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$ represented the monitoring wells and $\mathcal{E} = \{e_{1,2}, e_{i,j}, \dots, e_{m,n}\}$ described the strength of spatial connections between them. This undirected graph served as input to the GraphSAGE model, providing accurate neighborhood information for subsequent spatial feature extraction.

	ID	Lon	Lat	Elev	AquiferType
W1	370,114,220,003	117.345	36.731	55.64	1
W2	370,114,220,004	117.606	36.558	481.13	3
...
W27	370,113,220,002	116.730	36.404	92.67	3

Table 2. Representative examples of Geospatial characteristics and static attributes of monitoring wells.

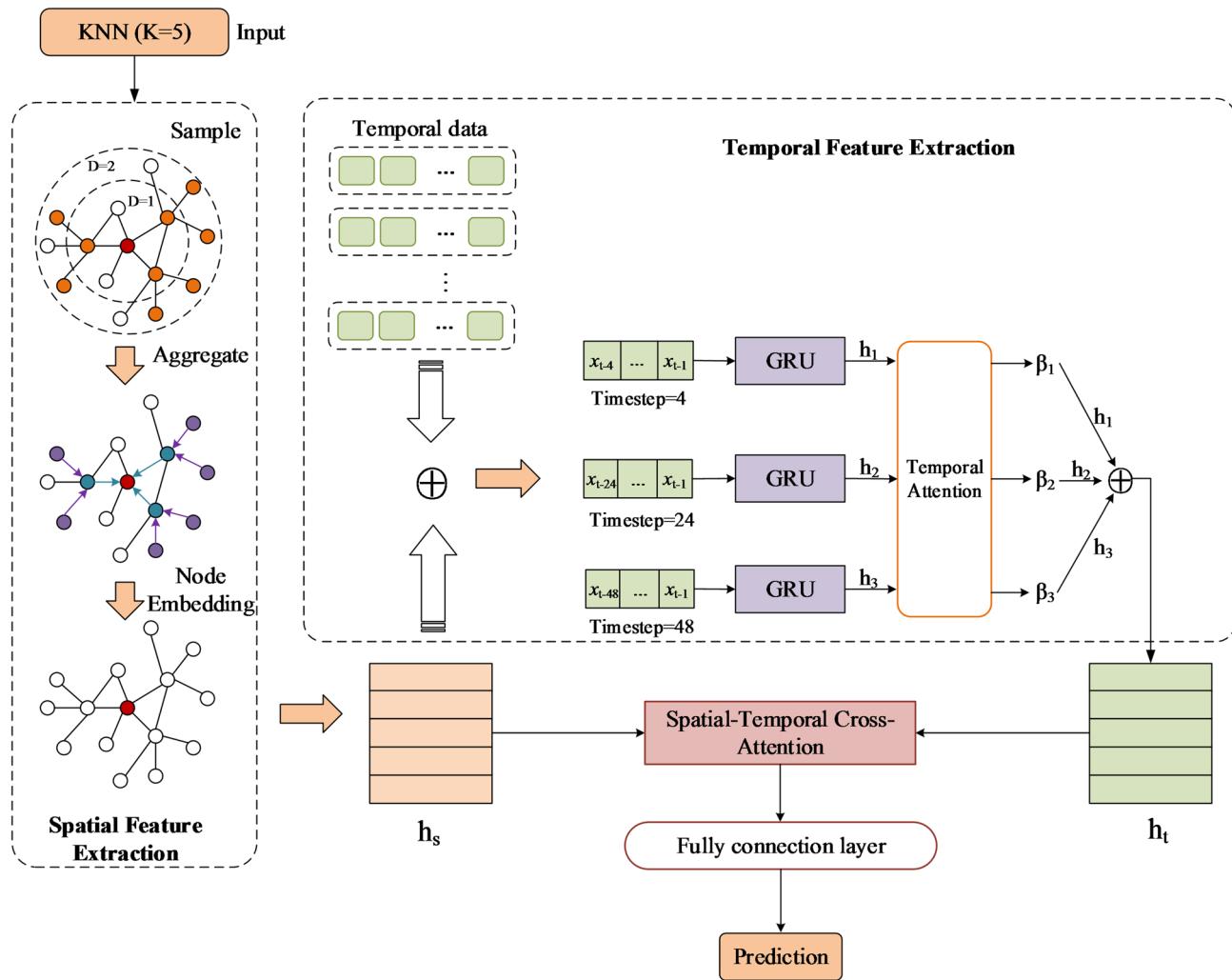


Fig. 4. Structure of the overall model.

Spatial feature extraction

This study utilized the GraphSAGE model to learn spatial feature representations of monitoring wells. The model effectively captured spatial dependencies between nodes by leveraging both the feature and structural information of nodes through neighbor sampling and feature aggregation mechanisms.

Node sampling: For each target node $\nu \in \mathcal{V}$, we employed a hierarchical sampling strategy to determine its multi-hop neighbor set $\mathcal{N}(\nu)$. The sampling process primarily focused on two parameters: the number of sampling layers D and the sampling size per layer. D represented the maximum hop count for neighbor aggregation. Experimental results demonstrated that the model achieved optimal performance when $D = 2$.

Node aggregation: The GraphSAGE model provided three aggregation functions: mean aggregation, LSTM aggregation, and pooling aggregation. Comparative experiments indicated that while both LSTM aggregation and pooling aggregation delivered good performance, the former exhibited significant computational inefficiency. Therefore, this study selected the pooling aggregation function, which operated by first applying a nonlinear transformation to the embedding of each neighbor node via a fully connected network, followed by the integration of neighborhood information to generate the target node embedding using max or mean pooling operations. The mathematical formulation was as follows:

$$\text{AGGREGATE}_d^{pool} = \max \left(\left\{ \sigma \left(\mathbf{W}_{pool} h_{u_i}^d + b \right), \forall u_i \in \mathcal{N}(\nu) \right\} \right). \quad (1)$$

Building upon these two processes, we first initialized the feature vector representation h_v for each node. For each node $\nu \in \mathcal{V}$, its neighbor nodes $\mathcal{N}(\nu)$ were obtained through node sampling. Subsequently, the aggregation function (Eq. 1) was employed to integrate feature information from neighboring nodes. Finally, the aggregated neighborhood features were combined with the node's own features through a nonlinear transformation to generate the updated node embedding representation, formulated as follows:

$$h_v^d = \sigma \left(\mathbf{W}^d \cdot \text{CONCAT} \left(h_v^{d-1}, h_{\mathcal{N}(\nu)}^d \right) \right). \quad (2)$$

Multi-scale temporal feature extraction

In the process of GWL prediction, the representation ability of temporal features is a critical factor influencing model accuracy. Inspired by Chen, et al.⁴³, this study employed a multi-branch GRU architecture that processed time-series data at different temporal scales in parallel, enabling joint modeling of both short-term fluctuations and long-term trends.

Considering the hydrological response characteristics of the karst aquifer system in Jinan, we defined three distinct sliding window lengths: short-term (one month), medium-term (6 months), and long-term (12 months) windows. The short-term window focused on recent GWL fluctuations. The medium-term window covered semi-annual hydrological cycles to model seasonal variation patterns, while the long-term window was dedicated to learning interannual trends. The original time series was partitioned into multiple subsequences according to these different window lengths. For example, considering a time series $T_w = \{x_1, x_2, \dots, x_n\}$ composed of the GWL observations from monitoring well w , to predict the GWL value at time t , if the sliding window was set to 3, the GWL values from the three preceding time steps were extracted, forming the input sequence $\{x_{t-3}, x_{t-2}, x_{t-1}\}$. These sub-sequences from the three distinct sliding windows were fed into three separate GRU branches, with each branch specifically processing the sub-sequence in a specific temporal scale. This parallel architecture enabled comprehensive modeling of both short-term perturbations (e.g., rainfall responses) and long-term evolutionary trends (e.g., seasonal cycles) in groundwater dynamics.

To further optimize feature fusion, this study introduced an attention mechanism to adaptively integrate multi-scale features. Let h_1 , h_2 , and h_3 denote the output feature vectors from the short-term, medium-term, and long-term GRU branches, respectively. The importance weights β_i of each branch were calculated through the attention mechanism. The feature fusion process based on attention weights can be expressed as:

$$h_t = \sum_{i=1}^3 \beta_i \cdot h_i \quad (3)$$

This mechanism dynamically adjusted the contribution weights of features across different temporal scales, generating more discriminative spatio-temporal feature representations. The design not only preserved scale-specific information but also enhanced predictive capability through synergistic feature interactions.

Spatio-temporal feature fusion

The core of spatiotemporal feature fusion lies in establishing coupled representations of spatial and temporal features. This study employed a cross-attention mechanism⁴⁶ to integrate temporal and spatial features, enabling more comprehensive feature representation. Through the aforementioned spatial and temporal feature extraction processes, supposed we obtain two feature sequences h_s and h_t , where h_s was the spatial feature sequence and h_t was the temporal feature sequence. The spatio-temporal cross-attention mechanism allowed one sequence (spatial features) to serve as Query, while the other sequence (temporal features) acted as both Key and Value. The Query, Key, and Value can be expressed as:

$$Q = h_s W_q, \quad K = h_t W_k, \quad V = h_t W_v, \quad (4)$$

where W_q , W_k , and W_v represented the projection matrices for Query, Key, and Value, respectively.

The cross-attention scores between spatial nodes and temporal steps were obtained by computing the similarity between Query and Key:

$$A = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \quad (5)$$

where d_k was the dimension of the Key, serving as a scaling factor for the dot product to prevent gradient vanishing. Each element $A(i, j)$ in the attention matrix quantified the dependency strength between the i -th monitoring well and the j -th timestep.

Finally, temporal features were aggregated to spatial nodes through a weighted sum:

$$Z = A \cdot V \quad (6)$$

Model optimization and evaluation

Experimental setup

The hardware and software environment configurations employed for model optimization and evaluation were detailed in Table 3.

To ensure the reproducibility of our proposed STGPM model, this subsection provided a comprehensive description of the specific architectural configurations used for each component. The final architecture was summarized in Table 4.

Hyperparameter optimization

To identify the optimal hyperparameter configuration for STGPM, a systematic grid search strategy was employed. This exhaustive method was selected due to the discrete and limited nature of the hyperparameter space, ensuring a comprehensive evaluation of all possible combinations to achieve globally optimal performance within the defined search domain, rather than settling for a computationally efficient but potentially local optimum.

CPU	Intel Core 11th (8 core), i9-11900, 2.50 GHz
GPU	NVIDIA GeForce GTX1660 12G
Hard disk	SSD 1 TB M.2 SATA
RAM	16G 2400 MHz DDR4 (2×8G)
Operating system	Microsoft Windows 10 Pro,64 bits
CUDA version	12.6
Programming language	Python 3.13.0
Deep learning framework	Pytorch 2.7.0

Table 3. Experiment environment.

Component	Parameter	Value
GraphSAGE	Number of Layers	2
	Embedding Dimension	64
	Aggregation Function	Max-pooling
Multi-branch GRU	Number of GRU Layers	1
	Hidden State Dimension (Short-term)	32
	Hidden State Dimension (Medium-term)	64
	Hidden State Dimension (Long-term)	64
	Dropout Rate	0.1
Attention Mechanism	Attention Dimension	128
	Attention Heads	4
Fully-Connected	Number of Layers	2
	Hidden Layer Dimension	64
	Output Dimension	1
	Activation Function (Hidden)	ReLU
	Activation Function (Output)	Linear

Table 4. Detailed architecture of the STGPM.

Hyperparameter	LSTM	GRU	STGCN	STGPM
Number of layers	2	2	2 (GCN)/1 (Temporal)	2 (GraphSAGE)/1 (GRU per branch)
Hidden dimension	128	128	64 (GCN)/128 (Temporal)	64 (Spatial)/32,64,64 (Temporal)
Temporal Window	12 (steps)	12 (steps)	12 (steps)	4, 24, 48 (steps)
Learning rate	0.001	0.001	0.001	0.001
Batch size	64	64	64	64
Optimizer	AdamW	AdamW	AdamW	AdamW
Dropout rate	0.2	0.2	0.1	0.1
Early stopping patience	15	15	15	15

Table 5. Final optimized hyperparameter configurations for all compared models.

Specifically, the grid search examined three critical parameters: learning rate, batch size, and the number of sampled neighbor nodes. The learning rate varied within the range of 10^{-4} to 10^{-2} , the batch size was tested at values of [16, 32, 64, 128] to balance computational efficiency and training stability, and generalization performance. The number of samples per hop was set to 1, 3, and 5 to determine the optimal amount of neighborhood information to aggregate for spatial feature extraction. The optimization objective was trained to minimize the Mean Squared Error (MSE) between its predictions and the truth groundwater level values. Each parameter combination underwent 100 training evaluations with early stopping patience to prevent overfitting. The training and validation loss curves were meticulously monitored to ensure convergence and assess generalization performance. This optimal configuration was subsequently used to train the final model on the combined training and validation sets for all subsequent performance evaluations reported in this study.

Similarly, we adopted a rigorous approach where each model underwent an independent hyperparameter optimization process using the same grid search strategy for each baseline model. The search space for each model included key architectural parameters: the number of layers [1, 2] and the number of hidden units [32, 64, 128]. Final optimized hyperparameter configurations for all compared models were shown in Table 5.

Evaluation metrics

To comprehensively evaluate model performance, this study employed a multi-dimensional metric system for quantitative analysis. The evaluation framework included the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). Each metric provided distinct insights: MAE measured the absolute deviation between predicted and observed values, RMSE quantified the dispersion degree of prediction errors, and R^2 assessed goodness-of-fit. The combination of these three indicators can objectively assess the prediction accuracy and model stability from different perspectives, providing a reliable quantitative basis for model comparison.

The MAE is the average absolute difference between predicted and observed values, quantifying the absolute magnitude of prediction errors. As the most intuitive metric, MAE is less sensitive to outliers due to the use of absolute values. Its formulation is given by:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (7)$$

where y_i and \hat{y}_i denote observed and predicted values, respectively, and n is the sample size.

The RMSE, calculated as the square root of the mean squared errors, provides greater sensitivity to prediction variability and extreme errors. The calculation method for RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (8)$$

R^2 quantifies the proportion of variability in the target variable explained by the model from a statistical perspective, serving as an important indicator of goodness-of-fit. An R^2 value closer to 1 indicates a better fit, while an R^2 close to 0 or negative suggests a poor model fit. The formula for calculating R^2 is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (9)$$

where \bar{y} is the mean of the observed values.

It is important to note that for model evaluation, the predictions were inverse-transformed back to the original scale (meters) before calculating the MAE, RMSE, and R^2 metrics to ensure their physical interpretability.

Model interpretability

Despite the superior predictive performance of machine learning and deep learning models in groundwater level prediction, their inherent “black box” nature limits the interpretability of the model decision-making process. Model interpretability aims to uncover the underlying mechanisms between input features (such as rainfall, evaporation, and groundwater extraction) and prediction outcomes, providing a scientific basis for water resource management decisions.

This study employed the SHapley Additive exPlanations (SHAP) framework, rooted in cooperative game theory, to quantify feature contributions to the model’s prediction results by calculating the SHAP values of each feature. The advantage of this approach was that SHAP values can simultaneously reveal both the polarity (positive/negative influence) and relative importance of each feature’s impact on predictions. This method transformed opaque model behavior into interpretable, physically consistent logic, thereby comprehensively assessing model behavior.

Results and discussion

Analysis of precipitation, temperature, evapotranspiration, and GWL distribution

The multi-source dataset in this study included four key hydrological variables: precipitation, temperature, evapotranspiration, and GWL. As shown in Fig. 5, linear trend (LT) analysis was applied to decompose and visualize the temporal trends of these variables from 2018 to 2023, revealing the dynamic characteristics of each variable. The results indicate that both precipitation and potential evapotranspiration exhibit significant seasonal cyclical variations and are synchronous, with higher values in summer months (June-August) and lower values in winter months (December-February). This pattern aligns closely with the study area’s typical monsoon climate. Similarly, temperature data displays marked annual cyclical fluctuations, peaking in the summer and reaching their lowest points in winter. GWL shows an overall upward trend during the observation period, superimposed with periodic fluctuations that are coupled with the seasonal pattern of precipitation: water levels rise during the wet season (summer) and decline during the dry season (winter). These observations confirm that seasonal variations of meteorological factors are key drivers of GWL fluctuations, providing crucial insights into the dynamic response mechanisms of the groundwater system in the study area.

Analysis of the model performance

Model training and error variation

This study employed systematic grid search⁴ to optimize key hyperparameters of the STGPM model (as detailed in section “Hyperparameter optimization”), which indicated that the model achieves optimal performance with a learning rate of 0.001, batch size of 64, and three samples per hop. As shown in Fig. 6, the model exhibits a rapid error reduction during the initial phases of training, followed by convergence to a stable low-loss region.

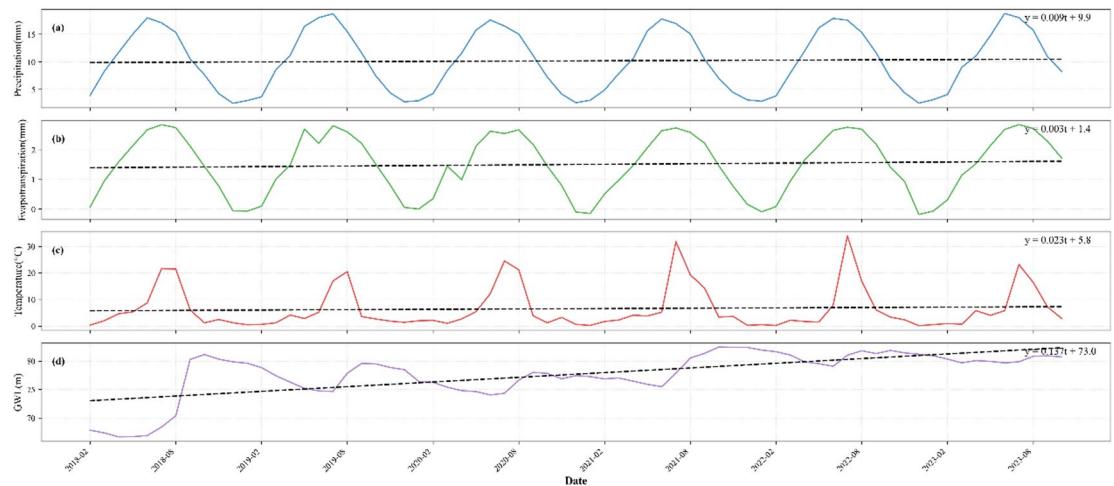


Fig. 5. Time series of monthly average precipitation, temperature, evapotranspiration, and GWL from 2018 to 2023.

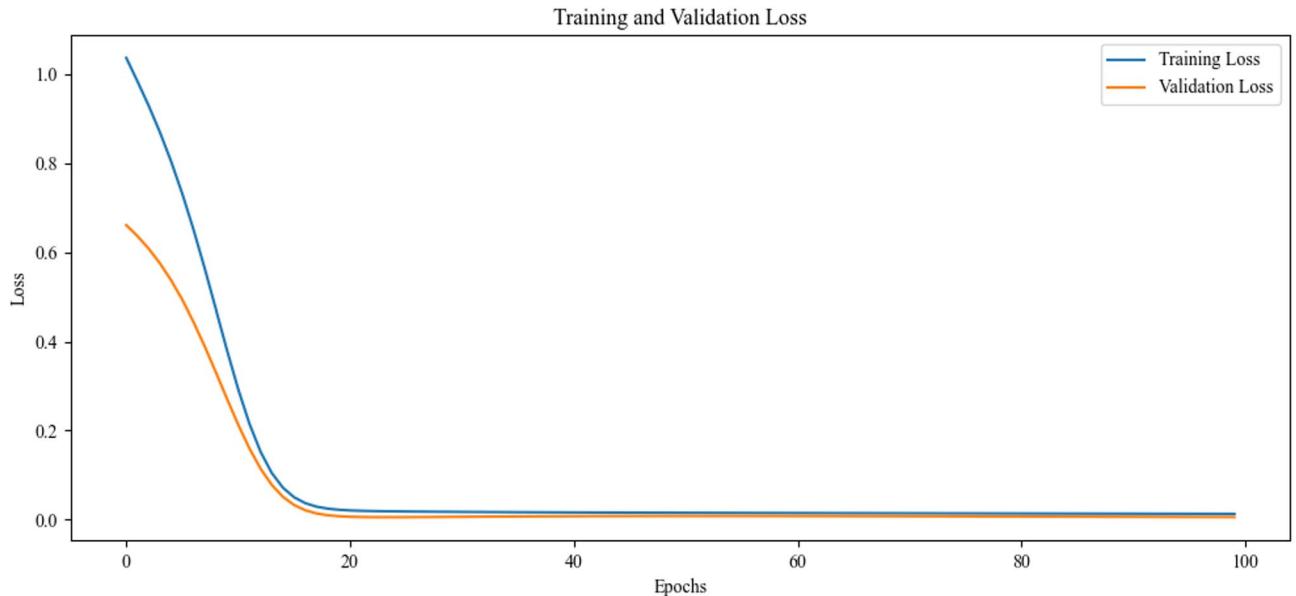


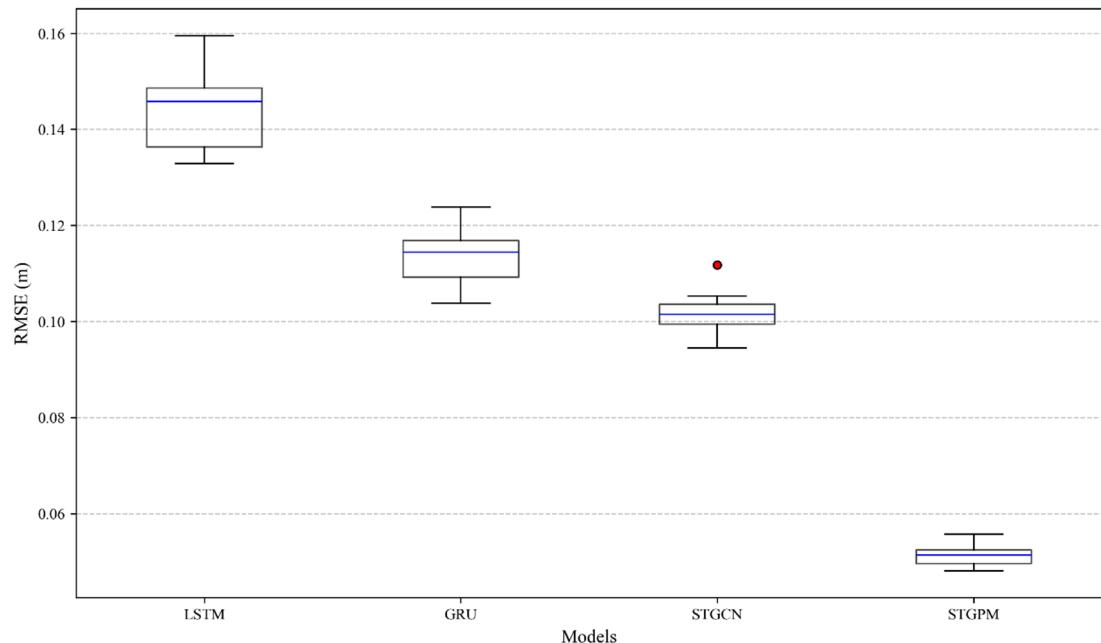
Fig. 6. Training and validation error curves.

This training dynamic not only confirms the rationality of the parameter configuration but also highlights the model's excellent generalization capability.

Performance comparison between STGPM and benchmarks

We conducted a systematic performance evaluation of STGPM by comparing it with three representative baseline models: the classical LSTM, GRU, and spatio-temporal graph convolutional network (STGCN). This comparative experiment aimed to validate the advantages of the STGPM model over existing mainstream methods in the task of GWL prediction. To ensure the fairness and comparability, all models adopted unified data partitioning strategies following the method in section “[Dataset partitioning](#)”, the data from 2023 were used as the test set, and the remaining data were divided into training and validation sets in an 8:2 ratio. Quantitative comparison was conducted using the evaluation metrics (MAE, RMSE, and R^2) defined in section “[Evaluation metrics](#)”. In terms of model training, a completely consistent hyperparameter setting was adopted: a learning rate of 0.001, 100 training epochs, a batch size of 64, with AdamW optimizer for parameter optimization and mean squared error (MSE) as the loss function. We did not employ identical network structures across models, as their fundamental operating principles differ (e.g., sequential processing vs. graph convolution). Instead, we adopted a rigorous approach where each model underwent an independent hyperparameter optimization process using the same grid search strategy. The final reported performance for each baseline model corresponds to its individually optimal configuration (Table 5) identified through this process. This strategy ensured that we

Model	Stage	MAE	RMSE	R^2
LSTM	Training	0.188 ± 0.012	0.164 ± 0.015	0.889 ± 0.011
	Testing	0.154 ± 0.011	0.139 ± 0.013	0.891 ± 0.009
GRU	Training	0.161 ± 0.009	0.151 ± 0.010	0.912 ± 0.012
	Testing	0.139 ± 0.008	0.121 ± 0.009	0.928 ± 0.009
STGCN	Training	0.095 ± 0.006	0.125 ± 0.008	0.956 ± 0.008
	Testing	0.095 ± 0.007	0.103 ± 0.006	0.965 ± 0.008
STGPM	Training	0.062 ± 0.004	0.064 ± 0.003	0.980 ± 0.001
	Testing	0.039 ± 0.002	0.052 ± 0.002	0.988 ± 0.001

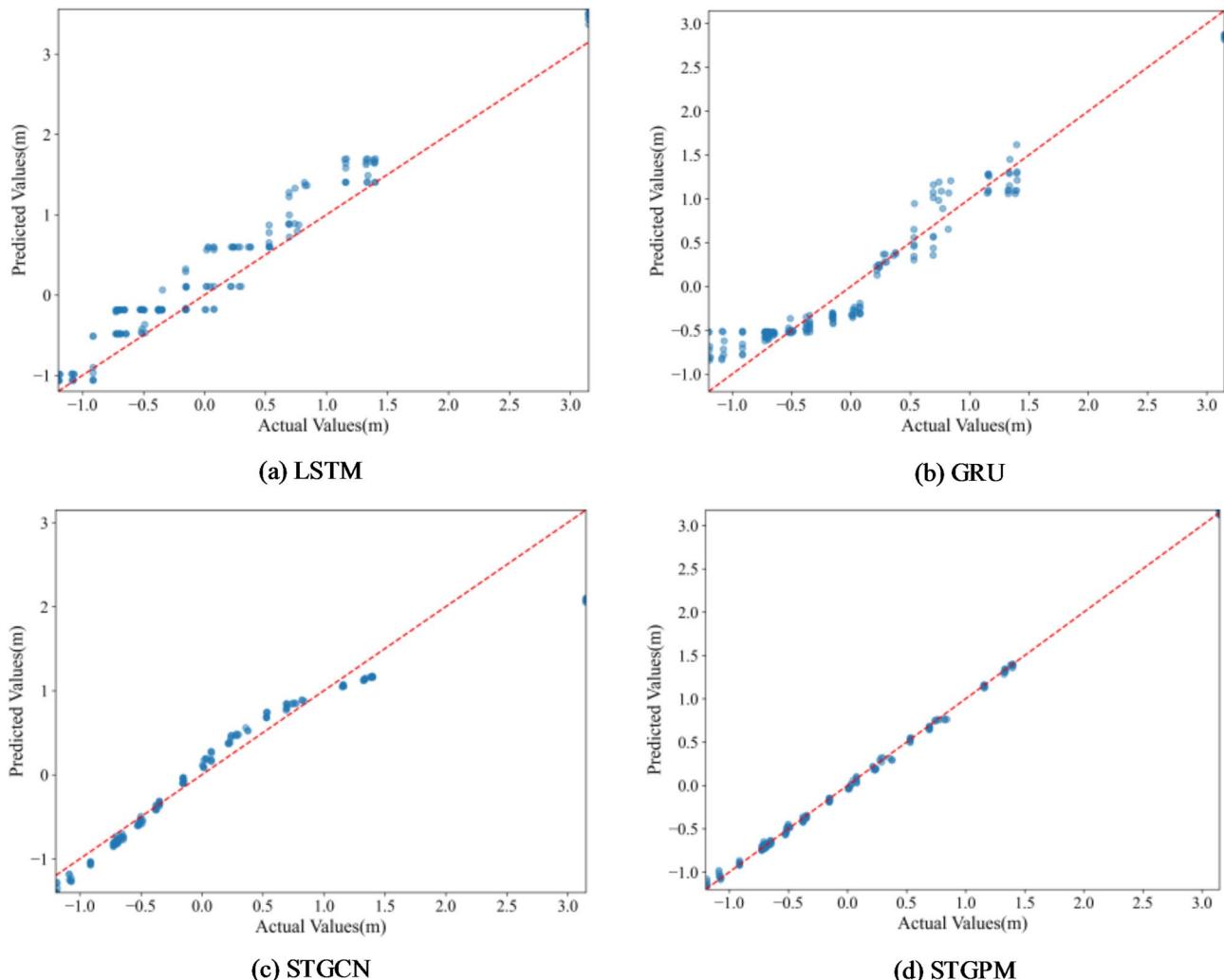
Table 6. Performance evaluation of different models.**Fig. 7.** The boxplot figure of the distribution of RMSE values on the conventional test set.

were comparing the best possible performance of each model architecture on our dataset, thereby attributing performance differences to the inherent efficacy of the models' inductive biases for spatio-temporal groundwater level prediction, rather than to arbitrary or suboptimal structural choices. To statistically validate the performance stability and robustness of the compared models, we conducted 10 independent training runs for each model with different random seeds.

Table 6 presents the evaluation results of four prediction models on the groundwater dataset. The experimental results demonstrate that STGPM achieves superior predictive performance, with the lowest errors on the test set ($MAE = 0.039$, $RMSE = 0.052$) and the highest R^2 (0.988), significantly outperforming the other benchmark models. Although STGCN shows slightly lower accuracy than STGPM, it still markedly exceeds traditional LSTM and GRU models. These findings confirm the critical role of spatial feature modeling in GWL prediction. Both STGPM and STGCN effectively capture spatial interactions between monitoring wells through graph neural networks, whereas traditional LSTM/GRU models, which rely solely on time series modeling, fail to represent this spatial dependencies, thereby resulting in limitations in predictive performance.

Figure 7 presents the distribution of the RMSE on the conventional test set across 10 independent runs using box plots. The results clearly indicate that the proposed STGPM model not only achieved the lowest median RMSE but also exhibited the most stable performance, as evidenced by its compact box and short whiskers. This signifies that the STGPM's superior performance is highly consistent and less sensitive to random initialization. In contrast, while the STGCN also shows relatively stable performance, its error distribution is significantly higher than that of the STGPM. The traditional LSTM and GRU models display both higher median errors and considerably larger variances. This statistical evidence reinforce the conclusion that the STGPM provides a more accurate and reliable solution for groundwater level prediction.

Model	MAE	RMSE	R^2
LSTM	0.158 ± 0.012	0.124 ± 0.015	0.898 ± 0.008
GRU	0.141 ± 0.005	0.125 ± 0.004	0.913 ± 0.006
STGCN	0.089 ± 0.003	0.112 ± 0.004	0.961 ± 0.002
STGPM	0.062 ± 0.002	0.087 ± 0.003	0.980 ± 0.001

Table 7. Performance evaluation results of predicted new monitoring wells.**Fig. 8.** Scatter plot comparison of predicted and actual values for new monitoring wells using different models: (a) LSTM; (b) GRU; (c) STGCN; (d) STGPM.

Predicted performance of new monitoring wells

The aforementioned experimental results demonstrate that existing models performed well in temporal prediction for trained monitoring wells. However, their capabilities for spatial extrapolation still need to be verified. To systematically evaluate the spatial generalization ability of the STGPM model, this study designed a dedicated prediction experiment using data from untrained monitoring wells.

Following the data partitioning scheme in Sect. 2.2.4, we randomly selected two monitoring wells to construct an unseen-well test set, with their data entirely excluded from model training. After the model was trained, the time-series data of these unseen monitoring wells were fed into the trained model for prediction. This experimental design enabled the assessment of the model's predictive ability for GWL at entirely new spatial locations, as the model must rely on the universal patterns it has learned rather than the memory of specific monitoring wells to make inferences.

Table 7 presents the average results over 10 independent runs. The scatter plot derived from the optimal result is shown in Fig. 8. From the results, STGPM maintains excellent prediction accuracy on untrained

wells, significantly outperforming the other models. This demonstrates that STGPM possesses strong spatial generalization capabilities and can effectively adapt to monitoring well data not involved in the training process. It is worth noting that the STGCN shows limited performance improvement, likely due to constraints in its ability to characterize features at unseen nodes.

Analysis of ablation experiments

To validate the effectiveness of the proposed methods, a series of ablation experiments were designed. Specifically, the impact of the following three modifications on model performance was evaluated: (1) Removing GraphSAGE (STGPM^{-G}): Retaining only temporal feature learning from the dataset; (2) Removing the multi-branch GRU (STGPM^{-M}): Using a single-branch GRU with a fixed time window; (3) Removing the spatio-temporal attention (STGPM^{-A}): Employing simple feature concatenation instead. The prediction results of these three experimental configurations on the test set are presented in Table 8.

The experimental results demonstrate that the complete STGPM model achieves the best on the test set, significantly outperforming the variant models in the ablation studies. This indicates that STGPM can make more accurate predictions when incorporating spatial feature correlations, multi-branch GRU, and spatio-temporal attention modules. Specifically, removing GraphSAGE (STGPM^{-G}) causes the most significant performance drop (MAE: 0.1382, RMSE: 0.3718, R²: 0.8658) compared to STGPM. This highlights the critical importance of spatial feature modeling, as GraphSAGE effectively captures spatial dependencies in the data. Without it, the model relies solely on temporal features and cannot fully utilize spatial information, leading to a substantial decrease in prediction accuracy. When the multi-branch GRU is removed (STGPM^{-M}), the performance decline is relatively smaller, but still inferior to the complete model. This suggests that the multi-branch GRU enhance temporal modeling capability by extracting features at different time steps, while the single-branch GRU fails to adequately capture multi-scale temporal patterns. Removing the spatio-temporal attention (STGPM^{-A}) also reduces performance, particularly in terms of RMSE and R². This highlights the attention mechanism plays a crucial role in feature fusion, as simple concatenation cannot adequately capture the interactions between spatial and temporal features, whereas attention dynamically weighted feature contributions to improve performance. These ablation experiments show that GraphSAGE, multi-branch GRU, and the spatio-temporal attention mechanism all contribute significantly to STGPM's performance. Their combined effect enable STGPM to achieve high accuracy and stability in prediction tasks.

Analysis of feature importance and correlation

Following the construction and training of STGPM, we employed SHAP analysis to quantitatively evaluate the predictive contributions of input features, as shown in Fig. 9. The results indicate that among the five input features (precipitation, temperature, evapotranspiration, GWL_{lag1}, and GWL_{lag2}), the SHAP values of the previous water level (GWL_{lag1}) and precipitation are relatively high, suggesting that they are the primary driving factors affecting water level variations. Notably, GWL_{lag1} has the highest SHAP value, reflecting the strong autocorrelation characteristics of GWLs. While GWL_{lag2}, temperature, and evapotranspiration all contribute to GWL prediction, their impacts are comparatively smaller. It is worth mentioning that while evapotranspiration generally tends to lower the water level, it may exhibit local positive correlation during the irrigation season due to artificial recharge.

The presence of correlated features among input variables can compromise model stability and increase sensitivity to uncertainties. To evaluate input stability, this study further quantified linear dependencies among features using the Pearson correlation coefficient. Figure 10 displays the correlation matrix among input features, where the size of each pixel reflects similarity between features. The higher the Pearson index, the stronger their correlation. Results demonstrate the correlation coefficients for all feature pairs are below 0.1, confirming no significant correlations among the input features. The input dataset thus meet the basic requirement of feature independence for machine learning models, which effectively avoids the risk of overfitting due to feature redundancy and ensures the reliability of the prediction results.

Conclusion

This study proposed a novel deep learning approach that significantly advanced groundwater level forecasting capabilities. By integrating GraphSAGE's spatial representation power with a multi-branch GRU architecture featuring attention mechanisms, our framework successfully captured both the hydrological connectivity between monitoring wells and multi-scale temporal dynamics of groundwater systems. The model's exceptional performance demonstrated the critical importance of explicitly incorporating spatial dependencies for accurate and generalizable predictions.

Through comprehensive ablation studies and benchmarking against state-of-the-art models, we established three key contributions to the field: First, the STGPM provided a novel architectural paradigm for spatio-

Model	MAE	RMSE	R ²
STGPM	0.0407	0.0535	0.9869
STGPM ^{-G}	0.1382	0.3718	0.8658
STGPM ^{-M}	0.0428	0.2069	0.9538
STGPM ^{-A}	0.0412	0.2021	0.9555

Table 8. Results of the ablation experiments.

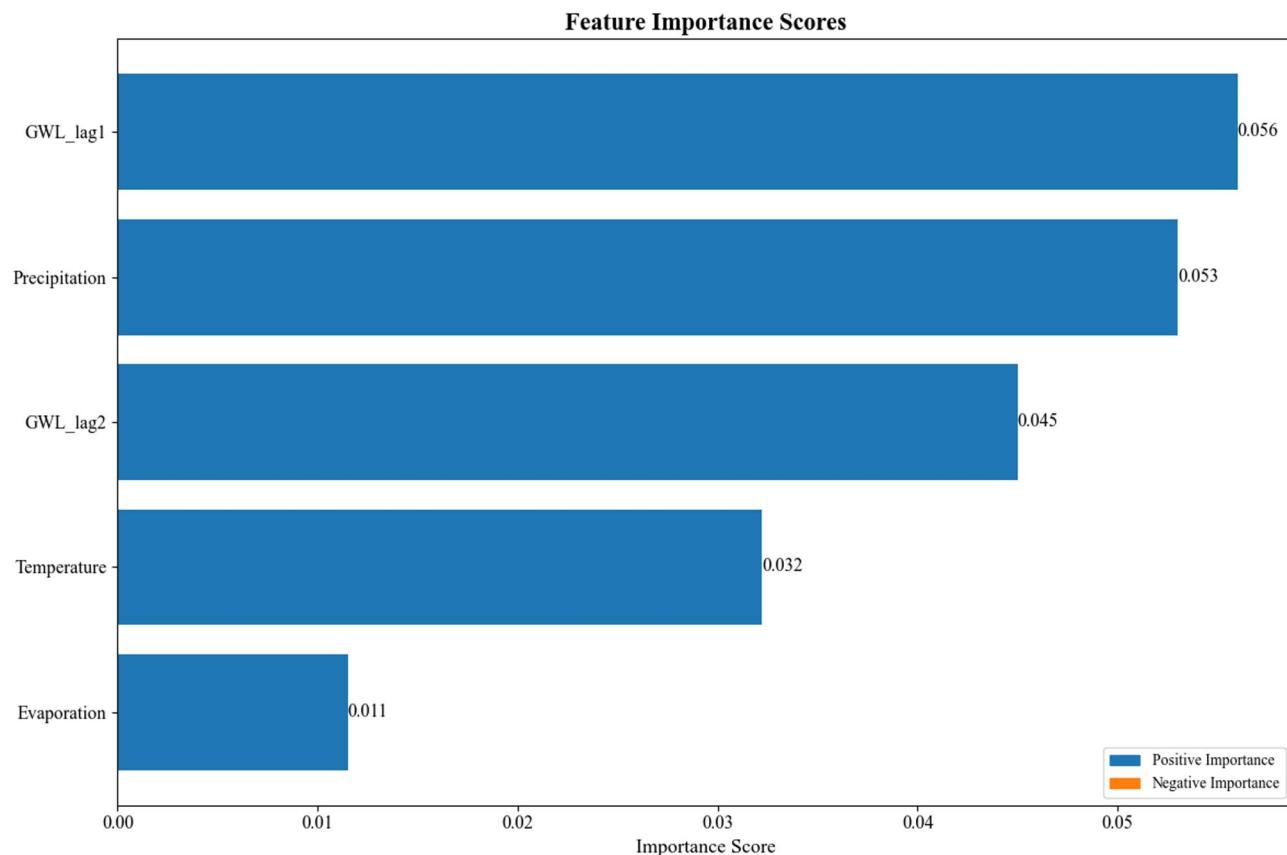


Fig. 9. The feature importance scores of input parameters in the STGPM.

temporal modeling in hydrogeology that effectively addressed the limitations of conventional time-series approaches. Second, through rigorous experimental validation, we quantitatively demonstrated the critical contribution of spatial features—specifically hydraulic connectivity between adjacent monitoring wells—to both prediction accuracy and model generalizability. Third, beyond providing a high-performance forecasting tool for groundwater dynamics, our methodology offered a valuable reference framework for addressing prediction challenges in other environmentally complex systems characterized by strong spatial heterogeneity, such as water quality forecasting and soil moisture prediction.

In future work, we will collect more data for model training to enhance its predictive capability and accuracy. By leveraging these improvements, researchers and decision-makers can advance their understanding and management of groundwater resources, ultimately contributing to the implementation of sustainable water management practices.

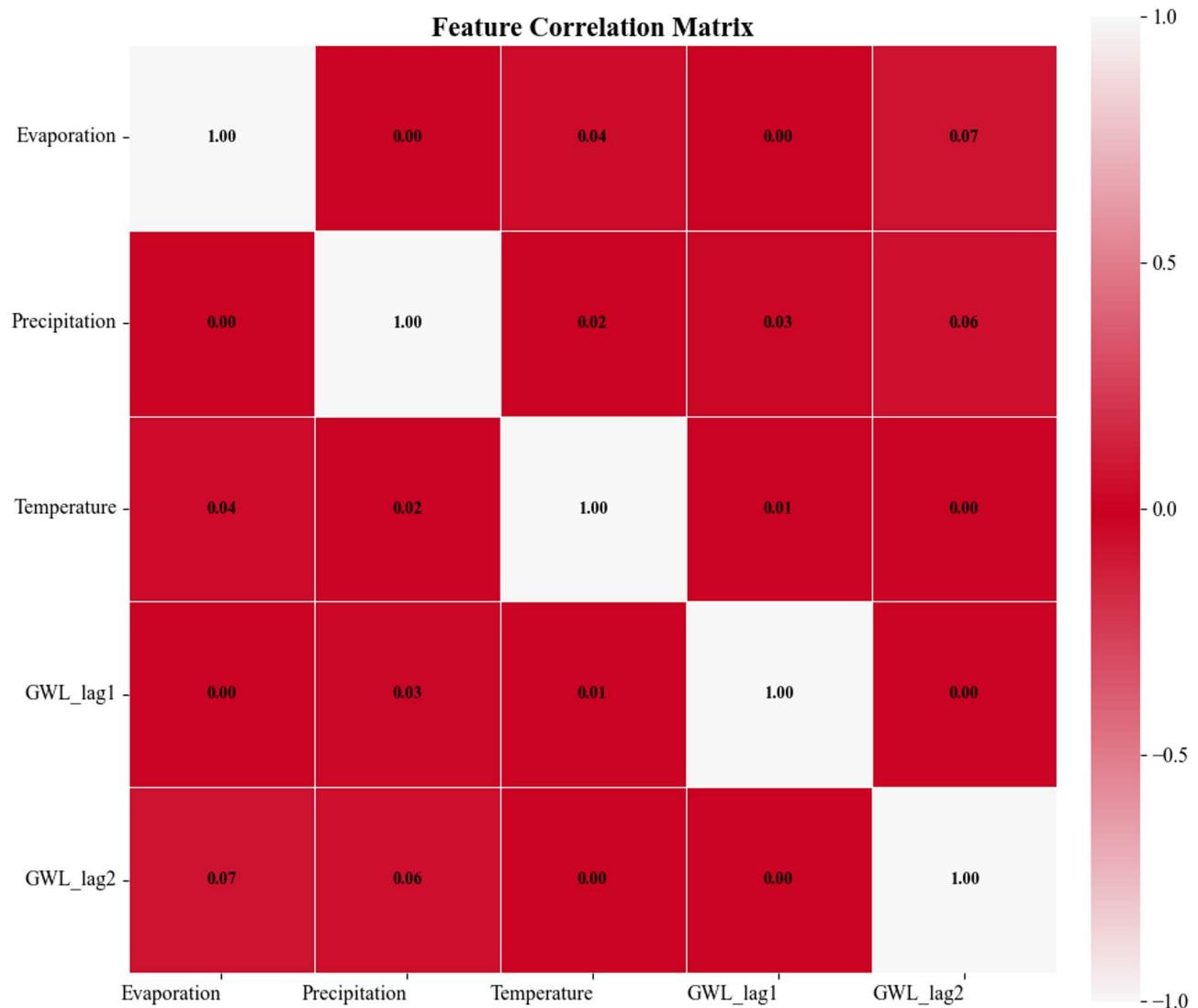


Fig. 10. The feature correlation estimates of input parameters in the STGPM.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 1 September 2025; Accepted: 8 November 2025

Published online: 24 December 2025

References

1. Abdi, E., Ali, M., Santos, C. A. G., Olusola, A. & Ghorbani, M. A. Enhancing groundwater level prediction accuracy using interpolation techniques in deep learning models. *Groundw. Sustain. Dev.* **26**, 101213. <https://doi.org/10.1016/j.gsd.2024.101213> (2024).
2. Ashofteh, P. S., Jalili, S. & Loáiciga, H. A. Assessment of climate change uncertainty effects on groundwater level prediction using bayesian analysis. *Theoret. Appl. Climatol.* **156**, 53. <https://doi.org/10.1007/s00704-024-05308-8> (2024).
3. Wu, H. et al. Assessing groundwater level variability in response to climate change: A case study of large plain areas. *J. Hydrol. Reg. Stud.* **57**, 102180. <https://doi.org/10.1016/j.ejrh.2025.102180> (2025).
4. Bo, Y. et al. Application of HP-LSTM models for groundwater level prediction in karst regions: A case study in Qingzhen City. *Water* **17**, 362. <https://doi.org/10.3390/w17030362> (2025).
5. Zhang, Z., Liu, Y. & Zhang, F. Prediction of groundwater table based on time series models in Baotu spring of Jinan. *J. China Inst. Water Resour. Hydropower Res.* **17**, 9. <https://doi.org/10.13244/j.cnki.jiwhr.2019.01.008> (2019).
6. Mahammad, S., Islam, A., Shit, P. K., Islam, A. R. M. T. & Alam, E. Groundwater level dynamics in a subtropical fan delta region and its future prediction using machine learning tools: sustainable groundwater restoration. *J. Hydrol. Reg. Stud.* **47** <https://doi.org/10.1016/j.ejrh.2023.101385> (2023).
7. Sugiyama, A. et al. Groundwater flow system and microbial dynamics of groundwater in a headwater catchment. *J. Hydrol.* **624**, 129881. <https://doi.org/10.1016/j.jhydrol.2023.129881> (2023).

8. Wang, F., Han, L., Liu, L., Wei, Y. & Guo, X. Prediction of groundwater level based on the integration of electromagnetic Induction, satellite data, and artificial intelligent. *Remote Sens.* **17**, 210. <https://doi.org/10.3390/rs17020210> (2025).
9. Banadkooki, F. B. & Haghighi, A. T. Groundwater level modeling using multiobjective optimization with hybrid artificial intelligence methods. *Environ. Model. Assess.* **29**, 45–65. <https://doi.org/10.1007/s10666-023-09938-6> (2024).
10. Tao, H. et al. Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing* **489**, 271–308. <https://doi.org/10.1016/j.neucom.2022.03.014> (2022).
11. Uc-Castillo, J. L., Marín-Celestino, A. E., Martínez-Cruz, D. A., Tuxpan-Vargas, J. & Ramos-Leal, J. A. A systematic review and meta-analysis of groundwater level forecasting with machine learning techniques: Current status and future directions. *Environ. Model. Softw.* **168**, 105788. <https://doi.org/10.1016/j.envsoft.2023.105788> (2023).
12. Chenjia, Z., Xu, T., Zhang, Y. & Ma, D. Deep learning models for groundwater level prediction based on delay penalty. *Water Supply*, **24**, 555–567. <https://doi.org/10.2166/ws.2024.009> (2024).
13. Ostad-Ali-Askari, K. & Shayannejad, M. Quantity and quality modelling of groundwater to manage water resources in Isfahan-Borkhar aquifer. *Environ. Dev. Sustain.* **23**, 15943–15959. <https://doi.org/10.1007/s10668-021-01323-1> (2021).
14. Li, J., Mao, X. & Li, M. Modeling hydrological processes in Oasis of Heihe river basin by landscape unit-based conceptual models integrated with FEFLOW and GIS. *Agric. Water Manag.* **179**, 338–351. <https://doi.org/10.1016/j.agwat.2016.09.007> (2017).
15. Omar, P. J., Gaur, S., Dwivedi, S. & Dikshit, P. Groundwater modelling using an analytic element method and finite difference method: An insight into lower Ganga river basin. *J. Earth Syst. Sci.* **128**, 195. <https://doi.org/10.1007/s12040-019-1225-3> (2019).
16. Jamin, P. et al. Direct measurement of groundwater flux in aquifers within the discontinuous permafrost zone: An application of the finite volume point Dilution method near Umiujaq (Nunavik, Canada). *Hydrogeol. J.* <https://doi.org/10.1007/s10040-020-02108-y> (2020).
17. Ukpaka, C., Adaobi, S. N. A. & Ukpaka, C. Development and evaluation of trans-amadi groundwater parameters: The integration of finite element techniques. *Chem. Int.* **3**, 406–413 (2018).
18. Boo, K. B. W. et al. Groundwater level forecasting with machine learning models: A review. *Water Res.* **252**, 121249. <https://doi.org/10.1016/j.watres.2024.121249> (2024).
19. Khan, J., Lee, E., Balobaid, A. S. & Kim, K. A. Comprehensive review of conventional, machine learning, and deep learning models for groundwater level (GWL) forecasting. *Appl. Sci.* **13**, 2743. <https://doi.org/10.3390/app13042743> (2023).
20. Chang, Y. W. et al. Advanced groundwater level forecasting with hybrid deep learning model: Tackling water challenges in taiwan's largest alluvial fan. *J. Hydrol.* **655**, 132887. <https://doi.org/10.1016/j.jhydrol.2025.132887> (2025).
21. Mohammed, K. S., Shabanlou, S., Rajabi, A., Yosefvand, F. & Izadbakhsh, M. A. Prediction of groundwater level fluctuations using artificial intelligence-based models and GMS. *Appl. Water Sci.* **13**, 54. <https://doi.org/10.1007/s13201-022-01861-7> (2022).
22. Pourmorad, S., Kabolizade, M. & Dimuccio, L. A. Artificial intelligence advancements for accurate groundwater level modelling: An updated synthesis and review. *Appl. Sci.* **14**, 7358. <https://doi.org/10.3390/app14167358> (2024).
23. Zhu, F. et al. A robust bayesian multi-machine learning ensemble framework for probabilistic groundwater level forecasting. *J. Hydrol.* **650**, 132567. <https://doi.org/10.1016/j.jhydrol.2024.132567> (2025).
24. Moghaddam, H. K., Milan, S. G., Kayhomayoon, Z. & Azar, N. A. The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. *Environ. Monit. Assess.* **193**, 1–20. <https://doi.org/10.1007/s10661-021-08961-y> (2021).
25. Yi, S., Kondolf, G. M., Solis, S. & Dale, L. S. Groundwater Level forecasting using machine learning: A case study of the Baekje Weir in Four Major Rivers Project, South Korea. *Water Resources Research* **60**, e2022WR032779 (2024). <https://doi.org/10.1029/2022WR032779>
26. Aderemi, B. A., Olwal, T. O., Ndambuki, J. M. & Rwanga, S. S. Groundwater levels forecasting using machine learning models: A case study of the groundwater region 10 at karst Belt, South Africa. *Syst. Soft Comput.* **5**, 200049. <https://doi.org/10.1016/j.sasc.2023.200049> (2023).
27. Sahoo, S. K. & Satapathy, D. P. An improved support vector machine model for groundwater level prediction: A case study. *Earth Sci. Inf.* **18**, 164. <https://doi.org/10.1007/s12145-024-01647-2> (2025).
28. Igwebuike, N., Ajayi, M., Okolie, C., Kanyere, T. & Halihan, T. Application of machine learning and deep learning for predicting groundwater levels in the West Coast aquifer system, South Africa. *Earth Sci. Inf.* **18**, 1–18. <https://doi.org/10.1007/s12145-024-01623-w> (2025).
29. Faruki Fahim, A. K., Kamal, A. S. M. M. & Shahid, S. Modeling Spatial groundwater level patterns of Bangladesh using physio-climatic variables and machine learning algorithms. *Groundw. Sustainable Dev.* **25**, 101142. <https://doi.org/10.1016/j.gsd.2024.101142> (2024).
30. Pham, Q. B. et al. Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Comput. Appl.* **34**, 10751–10773. <https://doi.org/10.1007/s00521-022-07009-7> (2022).
31. Roy, D. K. et al. Multiscale groundwater level forecasts with multi-model ensemble approaches: Combining machine learning models using decision theories and bayesian model averaging. *Groundw. Sustainable Dev.* **27**, 101347. <https://doi.org/10.1016/j.gsd.2024.101347> (2024).
32. Samantaray, S. & Sahoo, A. Groundwater level prediction using an improved ELM model integrated with hybrid particle swarm optimisation and grey Wolf optimisation. *Groundw. Sustainable Dev.* **26**, 101178. <https://doi.org/10.1016/j.gsd.2024.101178> (2024).
33. Singh, A., Patel, S., Bhadani, V., Kumar, V. & Gaurav, K. AutoML-GWL: Automated machine learning model for the prediction of groundwater level. *Eng. Appl. Artif. Intell.* **127**, 107405. <https://doi.org/10.1016/j.engappai.2023.107405> (2024).
34. Lee, E. H. Groundwater level prediction using modified recurrent neural network combined with meta-heuristic optimization algorithm. *Groundw. Sustainable Dev.* **28**, 101398. <https://doi.org/10.1016/j.gsd.2024.101398> (2025).
35. Saroughi, M. et al. A novel hybrid algorithms for groundwater level prediction. *Iran. J. Sci. Technol. Trans. Civil Eng.* **47**, 3147–3164. <https://doi.org/10.1007/s40996-023-01068-z> (2023).
36. Saroughi, M., Mirzania, E., Achite, M., Katipoglu, O. M. & Ehteram, M. Shannon entropy of performance metrics to choose the best novel hybrid algorithm to predict groundwater level (case study: Tabriz plain, Iran). *Environ. Monit. Assess.* **196**, 227. <https://doi.org/10.1007/s10661-024-12357-z> (2024).
37. Thakur, S., Karmakar, S. A. & Comparative Analysis Of ANN, LSTM and hybrid PSO-LSTM algorithms for groundwater level prediction. *Trans. Indian Natl. Acad. Eng.* **10**, 101–108. <https://doi.org/10.1007/s41403-024-00505-3> (2025).
38. Feng, F., Ghorbani, H. & Radwan, A. E. Predicting groundwater level using traditional and deep machine learning algorithms. *Front. Environ. Sci.* **12**, 1291327. <https://doi.org/10.3389/fenvs.2024.1291327> (2024).
39. Li, L., Sali, A., Liew, J. T., Saleh, N. L. & Ali, A. M. Machine learning for peatland ground water level (GWL) prediction via IoT system. *IEEE Access*, **12**, 89585–89598. <https://doi.org/10.1109/ACCESS.2024.3419237> (2024).
40. Thakur, A., Chandel, A. & Shankar, V. Prediction of groundwater levels using a long short-term memory (LSTM) technique. *J. Hydroinform.* **27**, 51–68. <https://doi.org/10.2166/hydro.2024.239> (2024).
41. Sun, W., Chang, L. C. & Chang, F. J. Deep dive into predictive excellence: Transformer's impact on groundwater level prediction. *J. Hydrol.* **636**, 131250. <https://doi.org/10.1016/j.jhydrol.2024.131250> (2024).
42. Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Adv. Neural. Inf. Process. Syst.* **30** <https://doi.org/10.48550/arXiv.1706.02216> (2017).
43. Chen, L. et al. Enhancing the accuracy of groundwater level prediction at different scales using spatio-temporal graph convolutional model. *Earth Sci. Inf.* **18**, 250. <https://doi.org/10.1007/s12145-025-01741-z> (2025).

44. Yu, M. et al. Study of large karst springs using the time series fractal method in Jinan. *Acta Geol. Sinica*. **94**, 2509–2519. <https://doi.org/10.19762/j.cnki.dizhixuebao.2020019> (2020).
45. Fan, X., Min, T. & Dai, X. The Spatio-Temporal dynamic patterns of shallow groundwater level and salinity: The yellow river Delta, China. *Water* **15**, 1426 (2023).
46. Vaswani, A. et al. Long Beach, CA, USA., in *31st Conference on Neural Information Processing Systems (NIPS 2017)* (2017).

Acknowledgements

The authors sincerely thank the editors and the anonymous reviewers for carefully reading this paper and their constructive comments.

Author contributions

All authors contributed to the study conception and design; analysis and interpretation of the data: Can Zhuang and Liangliang Cui; the drafting of the paper: Can Zhuang and Yi Cui; revising it critically for intellectual content: Can Zhuang and Liangliang Cui. All authors reviewed the final manuscript and agree to be accountable for all aspects of the work.

Funding

This work was jointly supported by the Natural Science Foundation of Shandong Province (ZR2025QC413) and the Shandong Province Science and Technology Small and Medium-sized Enterprises Innovation Ability Enhancement Project (2023TSGC0094).

Declarations

Ethical approval and consent to participate

Not applicable.

Consent for publication

Written informed consent for publication was obtained from all participants.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025