https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

A Hierarchical Hydrological Knowledge-guided Attention Network for Groundwater Depth Prediction: Insights from Multi-regional Model Interpretation

Jing Xu[1*], Yuming Mo[2], Senlin Zhu[1], Chengji Shen[3], Xinli Zhu[4], Chenming Zhang[5], Qihao Jiang[6], Ling Li[7]

[1]College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou, China
[2]School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang, China
[3]Ministry of Education Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes, Hohai University, Nanjing, China
[4]School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou, China
[5]School of Civil Engineering, The University of Queensland, Brisbane, Australia
[6]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing, China
[7]Key Laboratory of Coastal Environment and Resources of Zhejiang Province (KLaCER), School of Engineering,Westlake University, Hangzhou, China

Correspondence to: Jing Xu (xujing7503@yzu.edu.cn)

Abstract. Given the intensive influence of climate change and anthropogenic activities, accurate groundwater depth (GWD) prediction is essential for sustainable groundwater management. However, existing models struggle to capture spatiotemporal dependencies from complex factors. This study develops a novel Hierarchical Hydrological knowledge-guided Attention Network (HHA-Net) that processes multi-source heterogeneous data through physics-guided encoders, employs adaptive weight allocation and spatiotemporal attention to achieve fourteen-step GWD prediction, and provides insights into groundwater dynamics. Three distinct hydroclimatic and geographical regions in China (128 sites with 233,728 observations) serve as case studies, including the Yanshan-Taihang Mountain Region (YTMR), North China Plain (NCP), and North Jiangsu Plain (NJP). Results show that HHA-Net outperforms baseline models across different sites (natural, agricultural, and urban), with MAPE ranging from 1.02% to 5.95% and R^2 ranging from 0.71 to 0.98. The model demonstrates improved performance under droughts but slightly weaker predictive capability during rainfall events, particularly at natural sites in the YTMR. The geographical encoder dominates GWD in the mountainous YTMR (35.6%), while the human activity encoder and historical encoder control it in the NCP (32.5%) and the NJP (36.7%), respectively. The GWD exhibits prolonged memory effects (25 days) and delayed responses to rainfall (7.5 days) in the YTMR, whereas the over-exploited NCP shows rapid decay (3 days) with negative rainfall thresholds (-0.16) and anthropogenic-dominated patterns. The humid NJP demonstrates low-positive thresholds (0.07) and balanced natural-anthropogenic effects. These findings demonstrate the broad applicability of HHA-Net for GWD prediction and response pattern interpretation across diverse regions, providing scientific support for groundwater management.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

1 Introduction

As an essential subsurface component of hydrological processes, groundwater accounts for about 30% of global freshwater reserves, providing drinking water for over 2 billion people and supporting 40% of global agricultural irrigation demands (Wada et al., 2012; Loaiciga and Doh, 2024). Accelerating climate change and intensive anthropogenic activities, including frequent extreme events and aquifer overexploitation, have exacerbated the complexity and instability of groundwater systems, posing persistent challenges to water supply security and aquifer protection (Kuang et al., 2024; Fan et al., 2023). Groundwater depth (GWD), as a fundamental indicator reflecting groundwater system status, directly influences groundwater accessibility, ecosystem stability, and the risk of geological hazards such as land subsidence (Bai et al., 2024; Herrera-García et al., 2021). Given these increasing environmental and anthropogenic pressures, accurate prediction of spatiotemporal variations in GWD has become a significant challenge due to its nonlinear responses and complex interactions among multiple factors, yet it remains essential for achieving sustainable groundwater utilization and protection (Jasechko and Perrone, 2021; Richey et al., 2015).

Physics-based numerical models can effectively simulate groundwater transport processes and provide detailed mechanistic understanding of aquifer dynamics; however, they are constrained by difficult parameter acquisition, intensive computational requirements, and complex calibration procedures that often require extensive field data and experimental measurements (Tripathy and Mishra, 2024; Yang and Mccoy, 2023). Statistical methods such as time series analysis and empirical regression approaches, while easy to implement due to their reliance on linear assumptions or simplified non-linear relationships, are inadequate for capturing the complex dynamic behaviors of groundwater systems driven by multiple factors (Kratzert et al., 2018).

The rapid development of deep learning algorithms has provided new technical pathways for GWD prediction (Yi et al., 2024). Specifically, temporal sequence models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), along with spatial feature extraction methods including Convolutional Neural Networks (CNN) and Graph Neural Networks, have demonstrated considerable prediction capabilities across different hydroclimatic environments, as validated in regions such as Laizhou Bay in China, Erzurum Province in Turkey and Guadiana basin in Spain (Zhang et al., 2023a; Gezici et al., 2024; Collados-Lara et al., 2023). However, these data-driven approaches lack physical constraints and tend to achieve high accuracy by exploiting statistical patterns while prioritizing prediction accuracy over physical consistency, thus potentially yielding predictions that violate fundamental hydrological laws (Heudorfer et al., 2024). Moreover, spatial heterogeneity and differences in response mechanisms among various aquifers increase the difficulty of accurate GWD prediction (Rafik et al., 2023; Arshad et al., 2024), while the “black-box” nature of deep learning models poses challenges for interpreting physical mechanisms and hydrological processes in groundwater systems (Jung et al., 2024).

Previous studies have demonstrated the effectiveness of multi-head attention in addressing the spatial heterogeneity and non-stationarity of groundwater systems (Hou et al., 2024; Zhou, 2025). The multiple attention heads can capture different types of hydrogeological relationships by automatically identifying critical features, thereby enhancing understanding of complex

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

dependencies in aquifer systems and improving model interpretability (Serrano and Smith, 2019; Pang et al., 2024). More specifically, spatiotemporal attention mechanisms can adaptively identify time-varying correlations and spatial interactions, providing powerful tools for modeling long-term dependencies, short-term variations, and spatial heterogeneity characteristics in groundwater systems (Bai and Tahmasebi, 2023; Zhang et al., 2023a). With these attention mechanisms, hydrological-knowledge guided deep learning approaches can further bridge the gap between data-driven models and physical mechanisms by incorporating physical constraints including physics-informed loss functions, constraint-based regularization, and penalty terms that enforce physical consistency, which help to improve model performance and robustness and enhance model interpretability (Karniadakis et al., 2021; Read et al., 2019). However, effectively integrating physical constraints with attention-based temporal modeling in groundwater systems remains challenging, particularly for capturing both long-term trends and short-term fluctuations while maintaining physical consistency (Shen et al., 2018; Wang et al., 2024a).

Groundwater systems are subject to complex influences from multi-source heterogeneous factors including climate change, human activities, topographic variations, and historical memory effects (Reichstein et al., 2019). Traditional feature combination methods struggle to adequately capture the complex relationships and interactions among these diverse data sources (Bengio et al., 2013; Yosinski et al., 2014). To address this limitation, encoder networks can effectively learn unified representations from heterogeneous data by capturing complex non-linear relationships and transforming diverse inputs into standardized feature representations (Li et al., 2020). Furthermore, as a connected hydrological system, groundwater systems exhibit inherent similarities due to shared geological conditions, hydrometeorological influences, and environmental factors (Haaf and Barthel, 2018). Previous studies have typically quantified spatial similarity based on geographic proximity, using distance-based metrics such as Euclidean distance or inverse distance weighting to measure the spatial relationships between monitoring sites (Rinderer et al., 2017; Ohlert et al., 2023). However, diverse factors including aquifer characteristics, human activities, topographic variations, and boundary conditions create complex spatiotemporal heterogeneity in groundwater dynamic patterns (Varouchakis et al., 2019; Fan et al., 2013). Therefore, relying solely on spatial distance-based similarity measures is inadequate for accurately capturing the complex interdependencies within groundwater systems, as they neglect the heterogeneous factors that control groundwater dynamics (Sun et al., 2021). Accordingly, we introduce a novel four-dimensional spatial similarity matrix considering geographic distance, aquifer type, site types, and watershed affiliation in this study. Subsequently, we develop a spatial attention network that generates adaptive weights for aggregating historical GWD sequences. By incorporating information from highly similar stations rather than relying solely on single-site data, this approach substantially enhances the feature representation of target sites, effectively expanding the feature space and capturing broader hydrogeological patterns, thereby improving prediction accuracy and generalization performance of deep learning models (Ali et al., 2024a).

This study develops a Hierarchical Hydrological knowledge-guided Attention Network (HHA-Net) framework that processes historical GWD, meteorological, geographical, and socio-economic data through hierarchical encoder architectures integrated with physical constraints, achieving effective representation and fusion of multi-source heterogeneous features,

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

and employs adaptive spatiotemporal multi-head attention mechanisms to capture complex dependencies in groundwater systems, thereby obtaining multi-step high-precision prediction of GWD and providing interpretable physical insights. This study selects three distinct hydroclimatic and geographic regions as case studies in China, including the Yanshan-Taihang Mountain Region (YTMR), the North China Plain (NCP), and the North Jiangsu Plain (NJP), to assess model generalization capability across different hydrogeological conditions and analyze regional variations in groundwater system responses to various influencing factors. This research provides a comprehensive modeling framework for GWD prediction and response pattern identification, offering scientific insights into groundwater system dynamics and making important contributions to sustainable groundwater management and global water security.

2 Methodology

2.1 Framework of the HHA-Net model

This study develops an HHA-Net model following an encoding-prediction-interpretation architecture (Figure 1). Considering that medium-range (7-14 day) forecasts are a critical timescale for water resources management and early warning systems (Orozco López et al., 2024; Olsson et al., 2024), this study aims to generate fourteen-day multi-step GWD predictions. The model provides interpretable physical insights by extracting multi-source data features through hierarchical specialized encoders, employing adaptive weight allocation and multi-head attention mechanisms for spatiotemporal dependency modeling, and integrating hydrological constraints within the prediction framework.

2.1.1 Enhanced historical features

The historical encoder processes 30-day historical GWD observations from different site types to extract long-term trends and short-term fluctuations with site-specific adaptations, which reduces data requirements and enhances applicability for implementation in data-limited areas. The physical characteristics of groundwater variations, including delayed responses, and nonlinear-threshold responses, were captured through a physics-constrained LSTM network. LSTM effectively addresses gradient vanishing problems in traditional RNNs through three gating mechanisms including input, forget, and output gates, enabling effective modeling of long-term temporal dependencies (Vu et al., 2021; Zou et al., 2023). The model employs an LSTM-Convolution architecture to simultaneously capture long-term temporal dependencies and short-term local variation patterns in GWD, generating comprehensive temporal representations through feature fusion layers (Shelar et al., 2023). The detailed design of the physics-constrained LSTM and LSTM-convolution structure is presented in Appendix A1.

The hydraulic connectivity of groundwater systems results in significant spatial autocorrelation in GWD, which are controlled not only by geographical distance but also by aquifer types, site types, and watershed boundaries (Rinderer et al., 2019). These factors were incorporated into a four-dimensional spatial similarity matrix, with detailed design presented in Appendix A2. Spatial attention weights are obtained through SoftMax normalization of the similarity matrix, enabling

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

weighted aggregation of historical features from all sites to prioritize information from those with high similarity (Ali et al., 2024b).

Multi-head attention mechanisms model interactions between original historical features and spatially aggregated features (Li et al., 2024). Each attention head focuses on different correlation patterns, generating spatially enhanced historical features through concatenation and linear transformation of multi-head outputs (Wang et al., 2024b). Residual connections preserve temporal information integrity by adding the aggregated features to multi-head attention outputs, followed by further enhancement through fully connected layers to ensure effective spatial information integration while preventing temporal information loss.

2.1.2 Meteorological features

The meteorological encoder processes meteorological variables and extreme events. Meteorological variables include maximum/average/minimum temperature and daily rainfall. Extreme event indicators use binary variables to identify extreme rainfall and drought events (Frame et al., 2020).

Temporal correlations and periodicity in meteorological variables are captured by LSTM. Considering that extreme events primarily influence groundwater systems through cumulative effects across prediction windows rather than gradual daily changes (Le Brocque et al., 2018), discrete event indicators are processed using temporal averaging to convert 30-day binary sequences into single intensity values, subsequently encoded through dual-layer fully connected networks. Considering the complex interactions between meteorological variables and extreme events that significantly affect groundwater systems, multi-head attention mechanisms learn these conditional interactions and then dynamically modify meteorological feature representations through attention weights to achieve adaptive feature encoding.

2.1.3 Geographical features

The geographical encoder incorporates topographic features (elevation, slope, and distance to rivers) and multi-scale land use patterns. These topographic variables serve as controlling factors for groundwater distribution and flow paths (Gleeson and Manning, 2008) , while land use encompasses area proportions of five land use types (cropland, forest, grassland, waters, and artificial surfaces) within four spatial scales (500m, 1000m, 2000m, and 3000m radii) around monitoring sites.

Topographic features undergo nonlinear encoding through dual-layer fully connected networks, with the first layer mapping inputs to hidden dimensions and the second layer generating encoded feature representations. Land use features undergo scale-specific encoding through independent dual-layer fully connected networks for each land use type, with multi-scale features subsequently integrated to maintain spatial scale independence. Considering interactions between topography and land use (e.g., flat terrain tends to be cropland with strong anthropogenic disturbance) (Wang et al., 2025; Mao et al., 2024), the model captures feature interdependence by concatenating these features for fusion encoding through fully connected layers.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

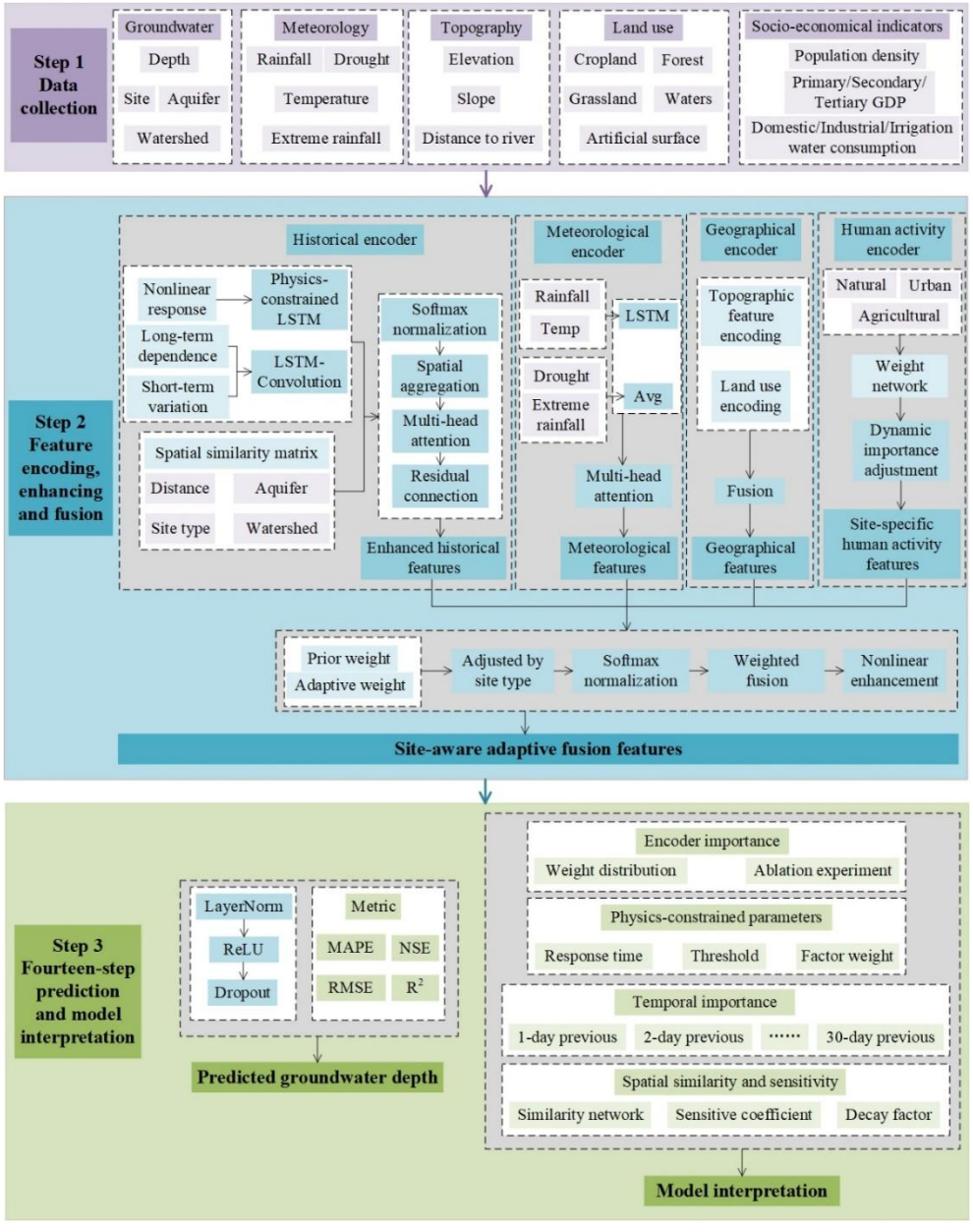


Figure 1: Framework of the Hierarchical Hydrological knowledge-guided Attention Network (HHA-Net) for GWD prediction and model interpretation.

2.1.4 Site-specific human activity features

The human activity encoder processes socioeconomic indicators including agricultural, secondary, and tertiary industry Gross Domestic Product (GDP), population density, and irrigation/domestic/industrial water consumption. Site types are

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

defined based on land use proportions within 500m buffers, specifically agricultural sites (>60% cropland), urban sites (>60% artificial surfaces), and natural sites (neither agricultural nor urban sites) (Table S1, S2).

Different site types exhibit distinct patterns of anthropogenic activities, requiring site-specific weighting strategies for socioeconomic indicators (Scanlon et al., 2017). Agricultural sites prioritize agricultural GDP and irrigation water consumption; urban sites focus on domestic and industrial water consumption, secondary and tertiary industry GDP, and population density; natural sites emphasize population density and domestic water consumption (Cetin, 2019; Flörke et al., 2013). To adaptively adjust factor importance based on specific anthropogenic activity intensities, each site type employs independent dynamic weight calculation networks that use corresponding socioeconomic indicators as inputs, calculating 0-1 weight coefficients through fully connected layers and sigmoid activation functions, to produce weighted features (Reichstein et al., 2019). After obtaining weighted features, corresponding encoding networks are selected based on site types for feature extraction.

2.1.5 Site-aware adaptive fusion features

Due to heterogeneity in key influencing factors for GWD variations across different site types, encoder feature weights require adaptive configuration. This study proposes a combined weighting approach that combines domain knowledge-based prior weights and data-driven adaptive weights. Prior weights are developed based on hydrological theory and empirical knowledge, with preset bias weights for four encoders across different station types shown in Table S3 (Zhang et al., 2023b; Liu et al., 2023; Zhang et al., 2021). Adaptive weights are learned from data through neural networks, dynamically adjusting encoder feature importance based on specific input features. Prior and adaptive weights are linearly combined with coefficients of 0.4 and 0.6, respectively.

The adaptive weight network processes features from four encoders through fully connected layers and SoftMax normalization to automatically determine relative feature importance based on input conditions. The learned fusion weights are then applied to combine the four feature types through linear weighting. Finally, the fused features undergo nonlinear transformation via fully connected layers and ReLU activation to capture complex feature interactions (Gomez et al., 2024).

2.1.6 Fourteen-step prediction of GWD

The prediction network employs a four-layer fully connected architecture with reducing hidden dimensions, incorporating LayerNorm to accelerate convergence and improve generalization. ReLU activation functions introduce nonlinearity for learning complex feature transformations, while dropout with a rate of 0.3 prevents overfitting. The final layer directly outputs 14-dimensional results corresponding to fourteen-step GWD predictions without activation functions to allow unrestricted numerical outputs. Detailed training configurations and regularization parameters are provided in Appendix A3.

To enhance prediction accuracy and physical plausibility, this study develops a composite loss function integrating multiple constraint mechanisms including temporal continuity and physical consistency in groundwater systems (Zhong et al., 2023; Narasimhan, 2018). The total loss function employs weighted linear combinations, combining mean squared error loss, mean

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

absolute error loss, Huber loss, and consistency constraint loss according to different weight coefficients to form multi-level adaptive loss optimization objectives (detailed design in Appendix A4), thereby avoiding local optimization problems caused by single loss functions.

Model performance evaluation employs four widely recognized metrics in hydrological modeling assessment, including mean absolute percentage error (MAPE), root mean square error (RMSE), Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970) and coefficient of determination (R^2). The formulas are presented in Eq. 1-4:

$$MAPE = \sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right| \frac{100\%}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(P_i - A_i\right)^2}{n}} \quad (2)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(A_i - P_i\right)^2}{\sum_{i=1}^{n} \left(A_i - \overline{A_i}\right)^2} \quad (3)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n} \left(A_i - \overline{A_i}\right)\left(P_i - \overline{P_i}\right)\right]^2}{\sum_{i=1}^{n} \left(A_i - \overline{A_i}\right)^2 \sum_{i=1}^{n} \left(P_i - \overline{P_i}\right)^2} \quad (4)$$

where A_i and P_i denote actual and predicted GWD at the ith observation; $\overline{A_i}$ and $\overline{P_i}$ represent the mean values of actual and predicted GWD respectively; n denotes the total number of GWD observations.

To account for differences in groundwater monitoring station types and prediction steps, prediction uncertainty is quantified using stratified Bootstrap, a resampling-based statistical inference technique that estimates distribution characteristics and uncertainty through repeated sampling with replacement (Lan et al., 2023; Singh and Sharma, 2022). Specifically, 1000 Bootstrap iterations are performed for each combination of site type (agricultural, natural, and urban) and prediction step (T+1 to T+14). Bootstrap standard errors (BSEs) serve as uncertainty measures by evaluating result variability across repeated predictions.

2.1.7 Ablation experiments

To systematically validate the necessity of each component in HHA-Net and quantify their individual contributions to model performance, we designed hierarchical ablation experiments. At the encoder level, we sequentially removed the four independent encoders (historical, meteorological, geographical, and human activity) to evaluate the contributions of hierarchical processing of multi-source data across various hydrogeological regions.

At the mechanism level, we further validated the internal encoder designs through three sets of experiments. First, the spatial attention module was disabled to verify the necessity of four-dimensional similarity aggregation and adaptive spatial weight allocation. Second, the site-type-aware adaptive fusion layer was replaced with simple average fusion to assess the impact of site-type differentiation on model generalization capability. Third, we evaluated the LSTM-convolution dual architecture by separately removing the CNN branch and the LSTM branch to verify the design rationale through performance degradation

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

analysis. The convolutional branch captures short-term local variation patterns, while the LSTM branch models long-term temporal dependencies. All ablation experiments were conducted using identical training datasets, hyperparameter configurations, and test datasets. The contribution of each component was quantified by calculating the percentage increase in RMSE relative to the full model for each ablation experiment.

2.1.8 Model interpretation

This study evaluates encoder importance through ablation experiments and fusion weight analysis. Ablation experiments systematically remove specific encoders (setting target encoder output features to zero) and compare prediction errors before and after ablation to quantify component importance (Li and Janson, 2024). Fusion weight analysis utilizes site-aware fusion mechanisms within the model framework to directly extract encoder weight distributions, identifying dominant factors for GWD variations across different site types.

Temporal dimension importance analysis identifies critical time windows for prediction by comparing model performance after masking different time steps. Spatial similarity weight networks analyze comprehensive similarity between groundwater locations and spatial attention sensitivity analysis adjusts distance decay factors to identify optimal spatial weight configurations. Physical insights into groundwater dynamics and responses can be derived from analyzing time-delay gating, threshold activation, and factor weight parameters, which reveal how the physics-constrained LSTM captures underlying physical processes.

2.2 Benchmark Models

This study selects five representative baseline models for comparison with the novel HHA-Net model, including Multi-Layer Perceptron (MLP), Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), Transformer, Graph Neural Networks (GNN) and Spatio-Temporal Graph Neural Networks (ST-GNN). The MLP represents a feedforward neural network architecture that learns complex relationships between inputs and outputs through nonlinear transformations across multiple fully connected layers (Riedmiller and Lernen, 2014). The MLP model flattens time series data and employs three hidden layers for GWD prediction, with each layer followed by ReLU activation for nonlinear transformation and dropout (with a rate of 0.3) for overfitting prevention. The output layer contains 14 neurons to generate fourteen-step predictions.

The CNN-LSTM represents a hybrid neural network architecture that combines the local feature extraction capabilities of CNNs with the sequential modeling advantages of LSTMs for enhanced time series forecasting (Zhao et al., 2025). The CNN-LSTM model employs a hierarchical processing strategy with two one-dimensional convolutional layers for spatiotemporal feature extraction, each followed by ReLU activation and dropout (with a rate of 0.3). The convolutional outputs undergo dimensional transformation before being fed into the LSTM layer, which models long-term temporal dependencies in the sequence. The output layer produces 14-dimensional predictions for fourteen-step GWD prediction.

The Transformer represents a neural network architecture that relies entirely on attention mechanisms to model sequence dependencies, enabling simultaneous focus on relationships between any two positions without distance limitations (Han et

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

al., 2021). The Transformer model employs an encoder architecture with 4 layers and 8 attention heads per layer for time series prediction, utilizing sine and cosine positional encoding to capture temporal information. Each encoder layer incorporates multi-head self-attention mechanisms and feedforward networks with residual connections, layer normalization, and dropout (with a rate of 0.3) for training stability, with the final output layer generating 14-dimensional predictions for fourteen-step GWD prediction. All baseline models adopt identical training strategies and hyperparameter settings as HHA-Net to ensure fair comparative experiments.

GNNs are designed to overcome the limitations of traditional neural networks that only handle Euclidean data, and their core mechanism involves each node updating its representation by aggregating feature information from neighboring nodes (Bai and Tahmasebi, 2023; Asif et al., 2021). This process operates layer by layer, enabling the model to capture both local structural characteristics and broader global information, thereby effectively extracting feature representations at the nodes, edges, and entire graph level. The GNN model constructed in this study employs a hidden layer dimension of 64. Following graph convolution computation, node features are compressed into a 64-dimensional vector through global average pooling, then sequentially passed through a linear layer, ReLU activation function, and Dropout layer (dropout rate of 0.3), before finally being mapped to 14-dimensional predictions by the output layer.

Building upon GNNs, ST-GNNs can simultaneously model spatial dependencies and temporal dynamics (Taccari et al., 2024; Martinez-Ruiz et al., 2025). ST-GNNs typically employ a hybrid architecture where GNN layers capture spatial topological structures while LSTMs or Temporal Convolutional Networks (TCNs) characterize temporal evolution patterns. This dual modeling capability enables the model to understand both spatial interactions between nodes and how these interactions evolve over time, demonstrating superior performance in complex dynamic system prediction. The key distinction of the ST-GNN model lies in the feature fusion stage, where extracted temporal features are concatenated with spatial features to form a 128-dimensional hybrid vector. This vector is then processed through a fully connected layer and ReLU activation function (dropout rate of 0.3) before the output layer generates 14-dimensional predictions.

3 Study area and data source

This study focuses on three regions with distinct meteorological and hydrogeological characteristics in China, including the semi-arid YTMR, the semi-humid NCP, and the humid NJP (Figure 2). YTMR is located in the transition zone between the Inner Mongolia Plateau and northern mountainous areas, with Zhangjiakou and Chengde cities forming the core area of this region. The region has complex terrain that slopes from northwest to southeast, characterized primarily by mountains and hills, with elevations reaching 1,400 to 1,600 meters (Sun et al., 2024). It experiences a temperate continental monsoon climate with annual average temperatures of approximately 2-10°C and annual rainfall of around 400 mm mainly during June to September (Guo et al., 2019). Rivers in the region originate primarily from highland sources and exhibit pronounced seasonal variations in flow. Due to relatively low precipitation and high evaporation, the region has limited water resources with highly uneven spatial and temporal distribution (Zhang et al., 2024).

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

The NCP is the second-largest plain in China, and Baoding city is in its northern section. The region has flat topography with elevations between 20 and 100 m. Baoding experiences a warm temperate semi-humid monsoon climate, with annual average temperatures between 11-13°C and annual precipitation of approximately 500-600 mm which over 60% is concentrated mainly in July and August (Liu et al., 2022a). Baoding is a water-scarce region where groundwater serves as an important water source; however, prolonged over-extraction has led to continuously declining groundwater levels, creating extensive groundwater depression zones and intensifying water supply-demand conflicts(Yang et al., 2024).

The NJP is located in northern Jiangsu Province and represents one of the important alluvial plains in eastern China. The terrain is characterized by low, flat topography with elevations predominantly below 50 m. The NJP has a subtropical monsoon climate, with annual average temperatures between 13-15°C and annual precipitation of approximately 800-1,000 mm, concentrated mainly in summer, with dense river networks and well-developed water systems, including the Huaihe River and Hongze Lake (Xu et al., 2025). Although the NJP has abundant groundwater resources, intensive groundwater extraction for agricultural irrigation and industrial operations, along with contamination from industrial discharge and agricultural runoff, has put considerable stress on regional groundwater resources (Mo et al., 2025).

The study collected daily GWD data between January 1, 2018, and December 31, 2022, from 128 monitoring stations (total 233,728 observations, Table 1) via the National Groundwater Monitoring Network in China (https://geocloud.cgs.gov.cn), operated by the Institute of Geological Environment Monitoring, China Geological Survey (Yao et al., 2025). This nationally standardized monitoring system provides uniform daily continuous observations of high quality, with fewer than 1% missing records primarily due to routine equipment maintenance. Therefore, simple linear interpolation was employed to fill these data gaps, a method commonly applied in processing hydrological time series data (Lehr and Lischeid, 2020; Bikše et al., 2023). Daily rainfall records were acquired from the National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn) with 0.1° spatial resolution (Han Jingya et al., 2023). The digital elevation model (DEM) at 90 m resolution and daily temperature data were obtained from the Resource and Environmental Science Data Platform (RESDC, http://www.resdc.cn/). Land use data for 2020 at 30 m resolution were obtained from the Globeland30 dataset provided by the National Geomatics Center of China (https://www.ngcc.cn). Five land use types were considered in this study, including cropland, forest, grassland, waters, and artificial surfaces. Area percentages of land use types were calculated across multi-scale buffers (500m, 1000m, 2000m, and 3000m radii) around each monitoring site. All groundwater monitoring site coordinates and related spatial data were processed and analyzed using the Gauss-Kruger projection system within ArcGIS 10.5 software (ESRI Company, Redlands, California, USA). Socioeconomic indicators in the study regions were gathered from statistical yearbooks of Hebei and Jiangsu provinces (http://tjj.hebei.gov.cn/; https://www.jiangsu.gov.cn/) as well as local statistical bureaus (e.g. https://baoding.gov.cn/; https://tjj.yancheng.gov.cn/), including GDP (primary, secondary, and tertiary), population density, and irrigation/domestic/industrial water consumption. These datasets were cross-verified with additional materials from different departments such as local water resources bulletins and municipal development reports (e.g. https://www.chengde.gov.cn/; http://slj.huaian.gov.cn/) to minimize potential data bias.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Table 1. Summary of datasets across the three study regions.

Attributes	NJP	NCP	YTMR
Record number	116864	58432	58432
Site numbers	64	32	32
Agricultural sites	23	19	14
Natural sites	17	5	11
Urban sites	24	8	7
Groundwater depth (m)	15.4±10.4	22.7±8.1	21.9±22.9
Rainfall (mm/day)	2.7±9.0	1.5±5.8	1.2±4.4
Average temperature (℃)	14.7±9.8	12.9±11.4	7.2±13.0
Maximum temperature (℃)	19.5±9.8	18.5±11.5	14.0±13.3
Minimum temperature (℃)	10.8±10.0	8.0±11.4	1.6±12.8
Elevation (m)	13.2±11.9	32.0±14.6	776.8±358.7
Slope (°)	0.6±0.3	6.3±4.2	17.4±21.0
Watershed numbers	3	4	5
Aquifer types	2	2	2
Distance to the nearest river (m)	571.9±442.7	2162.1±2320.5	4406.3±8612.2
Population density (capita/km^2)	503.2±164.6	1098.3±708.3	196.4±185.5
GDP (10^8 yuan)	2031.4±2252.2	915.3±1367.5	147.1±28.1
Primary industry GDP (10^8 yuan)	206.1±200.4	104.3±154.5	32.9±23.3
Secondary industry GDP (10^8 yuan)	856.0±958.6	309.9±483.0	42.9±20.7
Tertiary industry GDP (10^8 yuan)	969.8±1102.2	492.1±735.8	71.3±28.3
Domestic water (L/(capita·d))	131.7±19.8	58.0±19.4	139.7±133.8
Industrial water (m^3/(10^4 yuan))	15.9±8.2	20.9±14.6	13.1±10.6
Irrigation water (m^3/mu)	405.1±69.6	120.8±56.9	21.7±10.7
Cropland in 500m buffer (%)	45.7±31.5	66.0±30.2	49.3±30.1
Forest in 500m buffer (%)	0	0.1±0.2	0.8±1.8
Grassland in 500m buffer (%)	0.1±0.2	0.1±0.3	9.0±14.8
Waters in 500m buffer (%)	3.2±8.99	0.1±0.2	1.3±4.7
Artificial surface in 500m buffer (%)	50.9±31.8	33.7±30.3	39.5±24.8

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

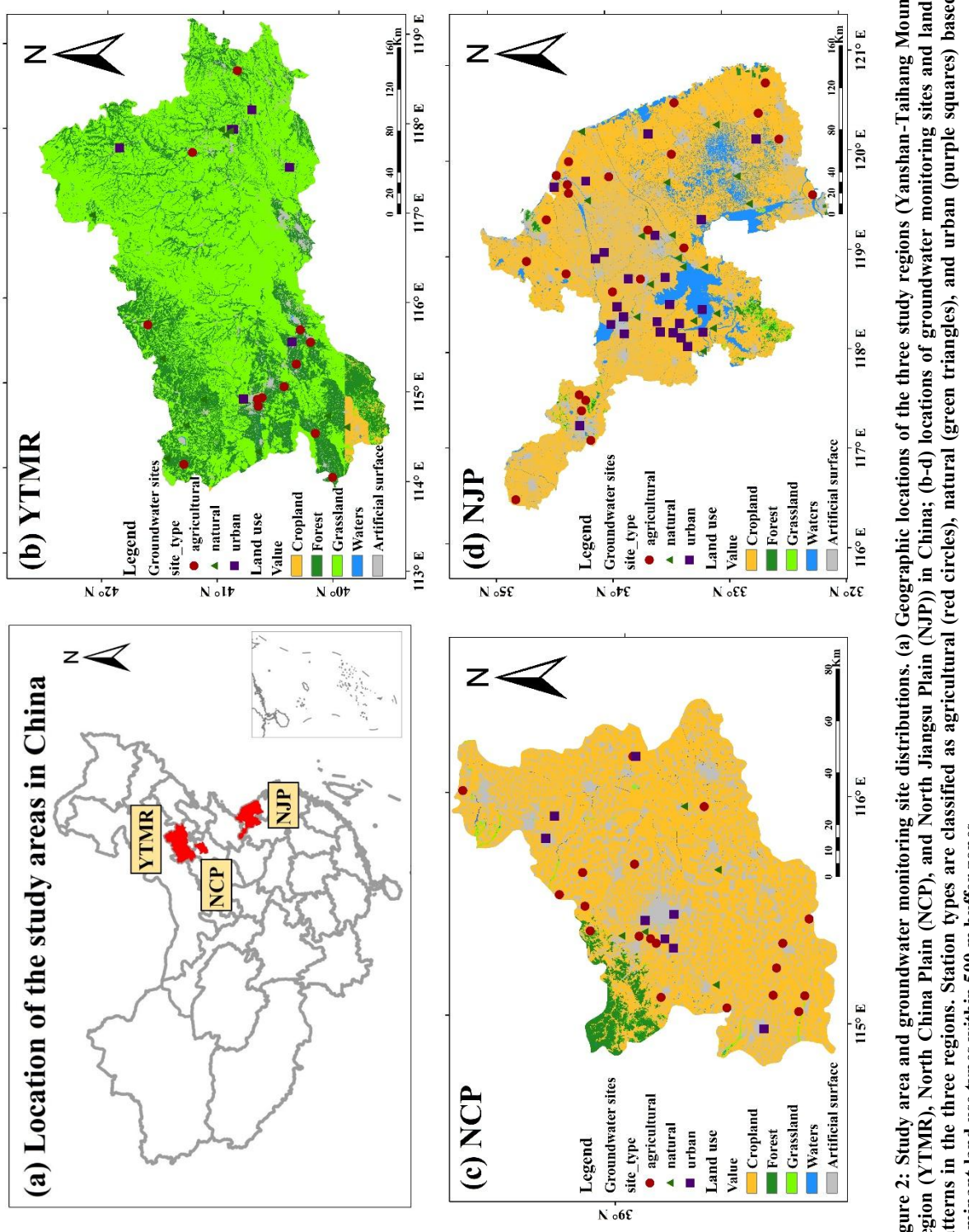


Figure 2: Study area and groundwater monitoring site distributions. (a) Geographic locations of the three study regions (Yanshan-Taihang Mountain Region (YTMR), North China Plain (NCP), and North Jiangsu Plain (NJP)) in China; (b-d) locations of groundwater monitoring sites and land use patterns in the three regions. Station types are classified as agricultural (red circles), natural (green triangles), and urban (purple squares) based on dominant land use types within 500 m buffer zones.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

4 Results

4.1 Model performance and comparisons

4.1.1 HHA-Net performance evaluation and benchmark comparison

HHA-Net exhibits excellent predictive capability across various hydrological conditions (Figure 3 and Figure S1-S3), effectively capturing complex seasonal oscillations, gradual/abrupt change trends, and extreme values. Scatter plots demonstrate that HHA-Net achieves better alignment along the 1:1 diagonal line with tighter clustering, while baseline models exhibit greater scatter and larger prediction errors across different GWD ranges. HHA-Net exhibits different performance across site types (Table 2). At urban sites, HHA-Net achieves impressive accuracy with MAPE of 1.03%–3.34% and RMSE of 0.30–0.64, reaching its optimal performance in the NCP (R^2 = 0.984, NSE = 0.975). For natural sites, the model maintains MAPE within 2.50%–5.95%, RMSE between 0.31–0.46, and R^2 ranging from 0.676 to 0.943. The generally stable performance at these sites stems from minimal anthropogenic interference and relatively natural groundwater variation patterns, though the slightly higher MAPE (5.95%) in the NCP suggests regional hydrogeological complexity can still pose challenges. Agricultural sites maintain competitive prediction accuracy, with MAPE ranging from 1.10% to 4.72%. HHA-Net reaches its highest agricultural MAPE (4.72%) in the NJP region. This specific challenge may be attributed to multiple factors including rainfall infiltration, irrigation withdrawal, and soil permeability changes induced by agricultural activities, resulting in high spatiotemporal heterogeneity in groundwater systems at agricultural sites (Zhang et al., 2019a).

HHA-Net outperforms baseline models across most site types and regions, particularly in the YTMR and NCP. At agricultural sites, HHA-Net significantly reduces prediction errors compared to the best-performing baseline models (e.g., Transformer with 1.62% MAPE in the NCP, and MLP with 10.185% MAPE in the YTMR). Notably, in the YTMR agricultural sites, the MAPE values of Transformer (28.74%) and ST-GNN (31.89%) are approximately 7.5 and 8.2 times that of HHA-Net (3.86%). This indicates that despite their capacity to handle spatiotemporal dependencies, deep learning models struggle to effectively capture groundwater dynamics in regions with complex spatial relationships without hydrological knowledge guidance. At natural sites, HHA-Net also exhibits significant advantages in high-accuracy prediction; for instance, in the YTMR, baseline models like GNN and Transformer show MAPE values roughly 6 to 8 times higher than HHA-Net (2.50%). Similarly, HHA-Net maintains the highest fitting accuracy (R^2 = 0.943) and lowest RMSE (0.459) at natural sites in the NCP. At urban sites in the YTMR, HHA-Net substantially outperforms all baseline models, with the MAPE of ST-GNN (25.99%) reaching 7.8 times that of HHA-Net (3.34%). However, in the NJP region, ST-GNN demonstrates exceptional performance across all site types, achieving a MAPE of 1.97% at urban sites compared to HHA-Net's 2.53%. This likely stems from densely distributed wells and regular groundwater extraction patterns in the NJP, which create well-defined spatial association patterns that ST-GNN can effectively capture through its graph structure, thereby

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

leveraging its strengths in handling spatially correlated data. Nevertheless, the R^2 of HHA-Net in the NJP remains highly comparable to that of ST-GNN (e.g., 0.892 vs. 0.908 at urban sites), indicating that its overall prediction capability maintains a robust and high standard.

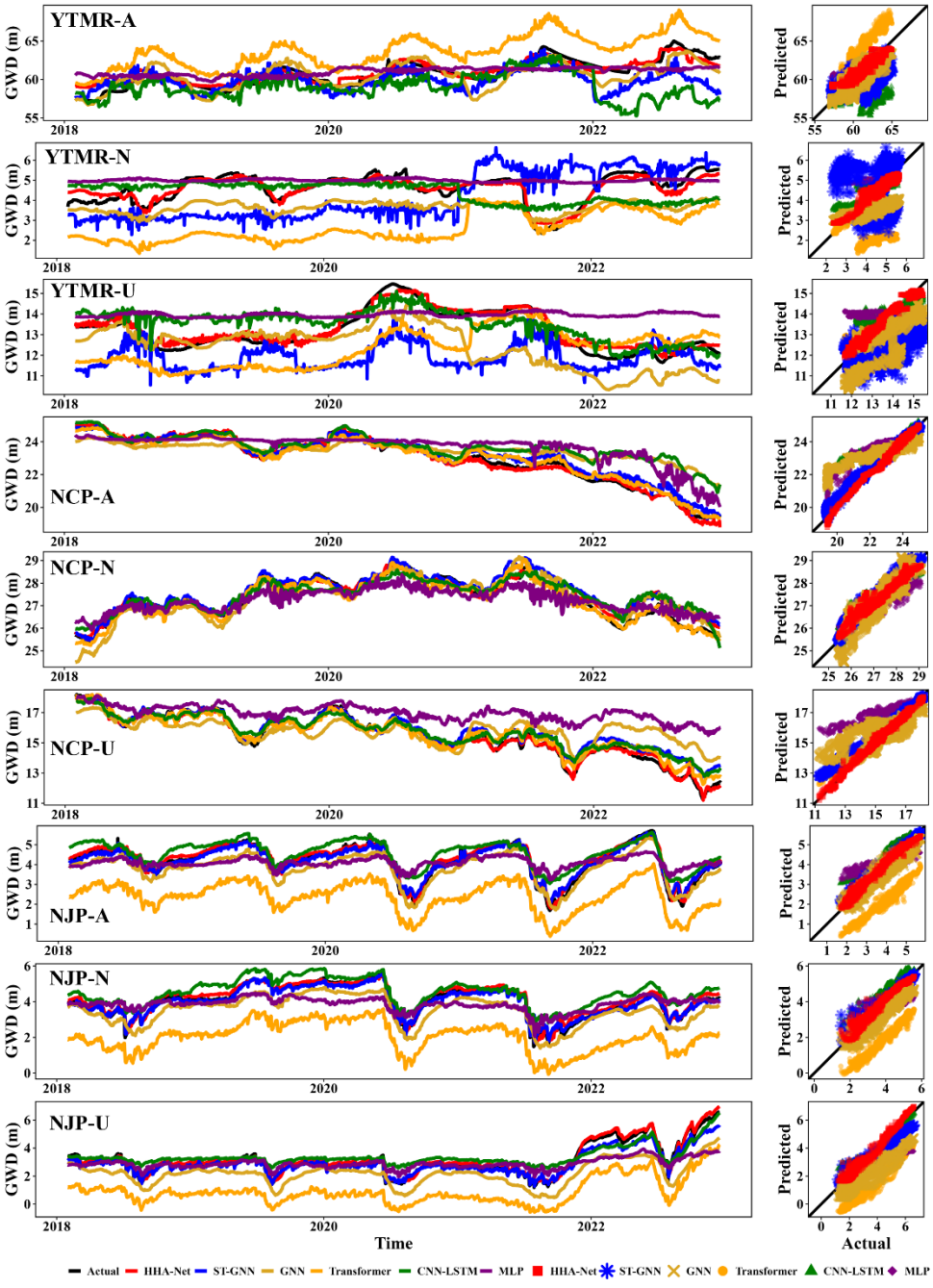
Given the inherent uncertainty in GWD arising from aquifer heterogeneity and spatial variability, we selected an additional 35 sites in the Lower Yangtze River Plain (LYRP), which is a region characterized by intensive human activities, high urbanization levels, and significant industrialization, to validate the model's generalizability and performance (Song et al., 2016). Site information and data are summarized in Table S4, with the corresponding model results presented in Table S5. The results demonstrate that HHA-Net consistently provides high-precision predictions across three site types, achieving MAPE values ranging from 2.25% to 4.92% and R^2 values ranging from 0.704 to 0.961 (Figure S4). Overall, HHA-Net achieves consistently excellent performance across diverse hydrogeological scenarios by effectively integrating multi-source heterogeneous data through physics-guided specialized encoders and capturing complex spatiotemporal dependencies via attention mechanisms based on four-dimensional spatial similarity matrices. Moreover, we have further validated our model on the benchmark dataset GEMS-GER, which includes GWD from 3,207 groundwater sites in Germany (Ohmer et al., 2026). Using 52 weeks of historical GWD to predict 1 week ahead (with input-output configurations identical to those in the original GEMS-GER study), our model achieved a predictive performance (mean RMSE = 0.18, mean NSE = 0.65, mean R^2 = 0.91) that significantly surpasses the performance of CNN and LSTM models reported in the original study (mean RMSE = 0.40, mean NSE = 0.41, mean R^2 = 0.49). More importantly, the HHA-Net demonstrates remarkable data efficiency by providing comparably high-precision predictions using only 7 weeks of historical data to predict 1 week ahead (mean RMSE = 0.19, mean NSE = 0.59, mean R^2 = 0.91). The distributions of model metrics and examples of scatter plots comparing predicted and observed GWD are shown in the Figure S5 and S6.

Table 2. Performance comparison between HHA-Net and baseline models across agricultural, natural, and urban sites in three study regions.

		Agricultural sites				Natural sites				Urban sites			
		MAPE(%)	RMSE	NSE	R^2	MAPE(%)	RMSE	NSE	R^2	MAPE(%)	RMSE	NSE	R^2
YTMR	GNN	22.038	1.920	-12.646	0.279	16.466	1.784	-41.960	0.259	20.572	2.915	-38.774	0.303
	CNN_LSTM	12.131	3.075	-53.012	0.273	14.432	2.930	-129.523	0.166	10.361	2.068	-170.05	0.121
	MLP	10.185	1.521	-12.961	0.234	7.189	1.010	-26.948	0.214	9.838	1.185	-1.662	0.228
	ST_GNN	31.888	2.808	-30.210	0.246	19.897	2.205	-37.312	0.209	25.998	3.743	-93.805	0.203
	Transformer	28.744	2.887	-32.335	0.391	22.149	1.953	-31.032	0.228	23.280	3.053	-30.302	0.384
	HHA_Net	**3.855**	**0.611**	**0.437**	**0.714**	**2.503**	**0.306**	**0.336**	**0.676**	**3.336**	**0.639**	**0.038**	**0.725**
NCP	GNN	4.716	1.017	0.593	0.852	15.006	1.2417	-1.345	0.648	5.088	1.215	0.578	0.825
	CNN_LSTM	4.314	0.966	0.604	0.882	8.254	0.839	0.457	0.790	4.248	0.991	0.700	0.941
	MLP	6.693	1.437	0.594	0.815	11.848	1.321	0.112	0.760	5.710	1.250	0.401	0.839
	ST_GNN	2.068	0.459	0.927	0.962	3.8153	0.596	0.788	0.826	1.743	0.452	0.938	0.970
	Transformer	1.623	0.404	0.913	0.964	21.008	0.976	-3.247	0.782	1.499	0.398	0.950	0.976
	HHA_Net	**1.097**	**0.296**	**0.966**	**0.978**	**5.947**	**0.459**	**0.691**	**0.943**	**1.027**	**0.302**	**0.975**	**0.984**
NJP	GNN	11.659	0.580	-0.382	0.799	5.277	0.512	0.543	0.901	5.568	0.590	0.130	0.785
	CNN_LSTM	9.384	0.498	0.185	0.737	5.400	0.630	0.546	0.824	5.829	0.630	-0.588	0.783
	MLP	10.346	0.592	0.089	0.535	5.773	0.760	0.265	0.721	5.961	0.746	-0.691	0.614
	ST_GNN	**3.429**	**0.166**	**0.851**	**0.940**	**1.310**	**0.149**	**0.953**	**0.975**	**1.972**	**0.219**	**0.743**	**0.908**
	Transformer	42.122	1.374	-14.114	0.862	16.975	0.977	-3.664	0.933	14.569	1.042	-4.395	0.873
	HHA_Net	***4.722***	***0.332***	***0.733***	***0.883***	***2.854***	***0.432***	***0.799***	***0.923***	***2.525***	***0.394***	***0.525***	***0.892***

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026



390 **Figure 3: Comparison of GWD prediction performance across the three site types (agricultural (A), natural (N), and urban sites (U)) in the three study regions. Left panels show time series predictions for typical sites in YTMR, NCP, and NJP (prediction performance for all sites is shown in Figure S1, S2 and S3). Right panels display scatter plots comparing predicted versus actual GWD.**

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

4.1.2 Temporal stability and spatial distribution of model performance

To evaluate the spatiotemporal stability of HHA-Net performance, multi-step prediction results (Figure 4) and spatial distribution patterns of HHA-Net (Figure 5) are analyzed. The YTMR region demonstrates varying temporal stability across different site types. Natural sites exhibit the lowest prediction errors and highest stability, with MAPE ranging from approximately 2.0% at t+1 to 2.8% at t+14. Agricultural and urban sites show relatively higher prediction errors, with the MAPE of agricultural sites increasing from ~3.8% to ~4.6%. In the NCP, agricultural and urban sites show highly comparable and excellent prediction performance, maintaining low MAPE (increasing gently from ~1.0% to ~1.8%) and high NSE and R^2 values (mostly above 0.9). In contrast, natural sites display significantly higher prediction errors and temporal variability, reflecting the challenges posed by complex, unmanaged groundwater systems in natural environments. In the NJP, model performance degradation is most pronounced at agricultural sites with high MAPE increasing from ~4.7% to ~6.4%. This temporal performance degradation at urban and agricultural sites may be attributed to the cumulative effects of complex factors such as intense human activities, seawater intrusion, and tidal variations in coastal regions. In contrast, natural sites in the NJP maintain the highest NSE and R^2, demonstrating better relative stability in capturing groundwater trends (Mo et al., 2025).

In terms of spatial distribution, the YTMR region exhibits relatively uniform model performance, except for two urban sites located in foothill regions and transition zones, which exhibit slight degradation (RMSE up to 2.09 m). In the NCP, the spatial distribution of performance is influenced by the intensity of human activities, with agricultural sites demonstrating the highest GWD prediction accuracy. However, one natural site located at the boundary between agricultural and urban areas shows abnormal performance with MAPE reaching 24.98%. This may be attributed to the spatial heterogeneity in human activities, where agricultural drainage systems and urban development pressures create complex influences that deviate from natural hydrological circulation patterns (Li et al., 2023b). Model performance in the NJP region is consistently excellent, except for one agricultural site in the western inland area that shows relatively higher errors (MAPE of 23.06%). Overall, HHA-Net demonstrates excellent spatial robustness across all three regions, indicating strong capabilities and reliable performance for practical groundwater management.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

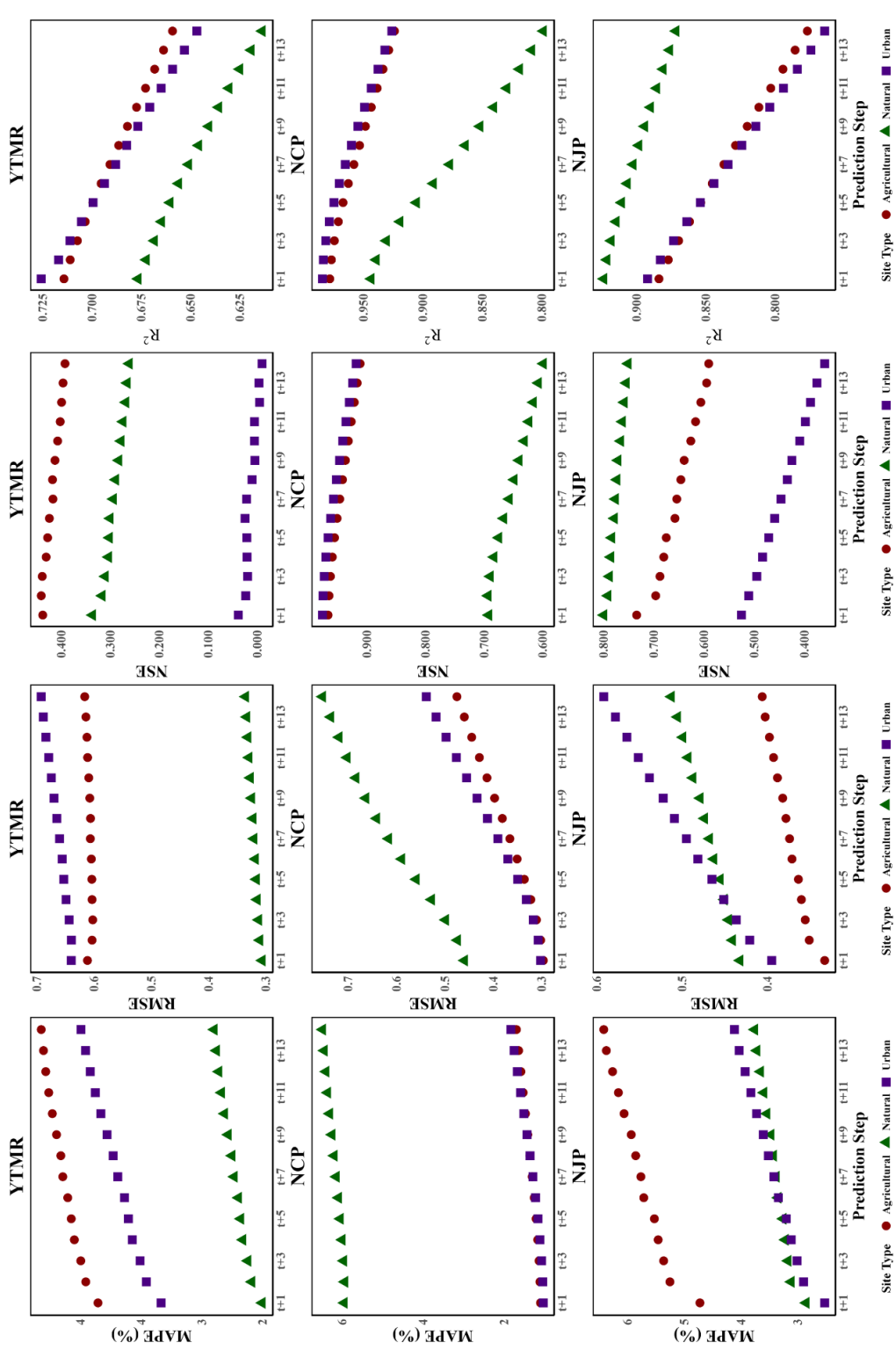


Figure 4: Fourteen-step prediction performance evaluation across different site types and study regions. Performance metrics include mean absolute percentage error (MAPE), root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), and coefficient of determination (R^2) for prediction steps from t+1 to t+14.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

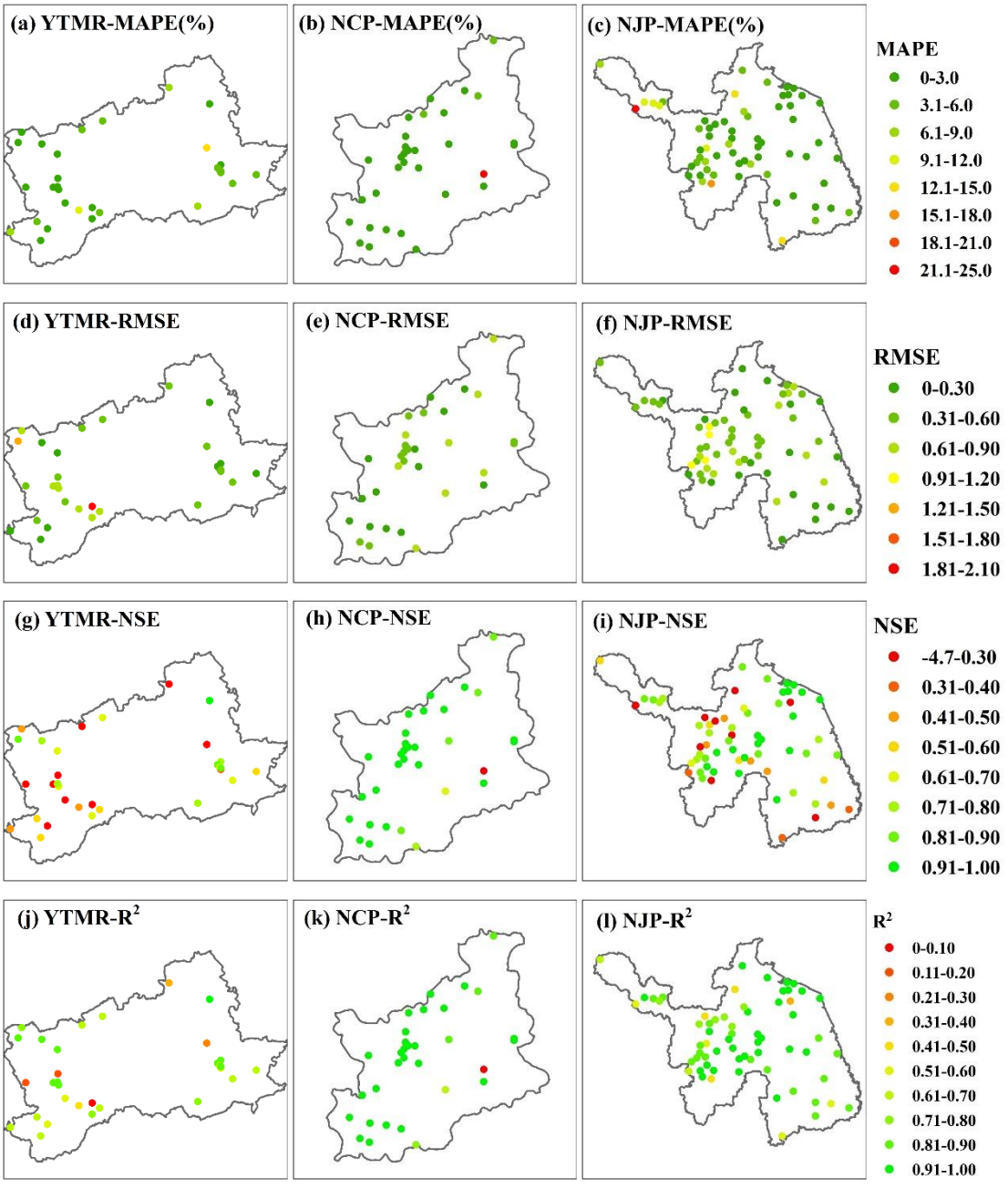


Figure 5: Spatial distribution of model performance metrics (MAPE (%), RMSE (m), NSE, and R^2) across groundwater monitoring sites in the three study regions.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

4.1.3 Ablation experiment analysis

Encoder-level ablation reveals substantial performance degradation when removing individual encoders, with the historical encoder demonstrating the most critical role across all regions (Table S6). Removing the historical encoder results in RMSE increases of 2213.2%, 1490.2%, and 1398.5% in the YTMR, NCP, and NJP, respectively, indicating that historical GWD states serve as the foundation for accurate predictions. Removal of the meteorological encoder causes greater performance degradation in the NCP (117.1%) and NJP (157.3%) plain regions compared to the YTMR mountainous area (53.0%). In contrast, removal of the geographical encoder causes the most severe degradation in the YTMR (581.2%), while demonstrating only an 11.8% increase in RMSE in the NJP.

Mechanism-level ablation experiments reveal that removing the site-type-aware fusion layer causes the most severe performance degradation across all regions, with RMSE increasing by 5103.7%, 2436.1%, and 1390.0% in the YTMR, NCP, and NJP, respectively. This demonstrates that the site-type-specific processing for natural, agricultural, and urban sites is crucial for model generalization across heterogeneous hydrogeological regions. Similarly, disabling the spatial attention module leads to substantial degradation, with RMSE increases exceeding 1000% in all three study regions (reaching 2591.3% in the YTMR, 1095.8% in the NCP, and 1420.9% in the NJP), which validates the essential role of adaptive spatial weighting based on four-dimensional similarity in capturing complex spatial dependencies in groundwater systems. Ablation of the dual-branch architecture within the historical encoder reveals distinct regional patterns; specifically, removing the CNN branch causes significantly greater performance degradation in the YTMR (1639.4%) and NJP (97.7%) than removing the LSTM branch (48.9% and 540.3%, respectively), while the opposite trend emerges in the NCP (removing the LSTM branch causes 618.4% increase versus 97.6% for the CNN branch).

4.2 Model performance under drought and extreme rainfall conditions

Figure 6 presents model performance during extreme events including droughts and extreme rainfalls. Compared to normal conditions, model prediction accuracy was substantially improved during drought events, with MAPE decreasing by 67.6% and 50.5% for agricultural and natural sites in the NCP, and by 51.3% for agricultural sites in the YTMR. In contrast, model prediction accuracy decreased during extreme rainfall events, with MAPE for natural sites in the YTMR increasing by 41.9%, and for agricultural sites in the NJP by 38.6% compared to normal conditions. Notably, the model maintains stable performance in the NCP under extreme rainfall, with MAPE increases of only 7.7% and 10.6% for agricultural and urban sites, while natural sites even show slight improvement. Analysis of coefficient of variation versus mean MAPE reveals that model performance consistently improved during drought events, with most scenarios achieving low error and low variability. Conversely, model performance consistently degraded during extreme rainfall events, with most site types in the YTMR and NJP exhibiting high error and high variability.

Analysis of meteorological variable importance provides potential explanations for the deteriorated model performance under extreme rainfall events. Rainfall-related variables exhibit the highest importance across the three regions, especially in

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

the NCP, acting as the primary driving factor for groundwater systems. Temperature variables show relatively low and stable importance, while drought events generally have minimal influence, though importance increases to 0.467 in the NJP similar to the importance level of the rainfall indicator. During drought periods, groundwater dynamics are characterized by gradual, monotonic recession processes governed primarily by evapotranspiration and baseflow (Cadol et al., 2012; Hameed et al., 2023). These processes exhibit strong temporal autocorrelation and relatively simple linear relationships with meteorological forcing, thereby enabling the model to learn long-term recession patterns and achieve accurate predictions (Soleimani Motlagh et al., 2017; Zhao et al., 2022). In contrast, heavy rainfall, as the dominant driver of groundwater fluctuations, triggers complex hydrological processes during extreme events, including sudden infiltration surges, altered flow pathways, and disrupted hydraulic gradients, which generate rapid nonlinear dynamics that substantially increase prediction difficulty (Wittenberg et al., 2019; Shamsudduha and Taylor, 2020).

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

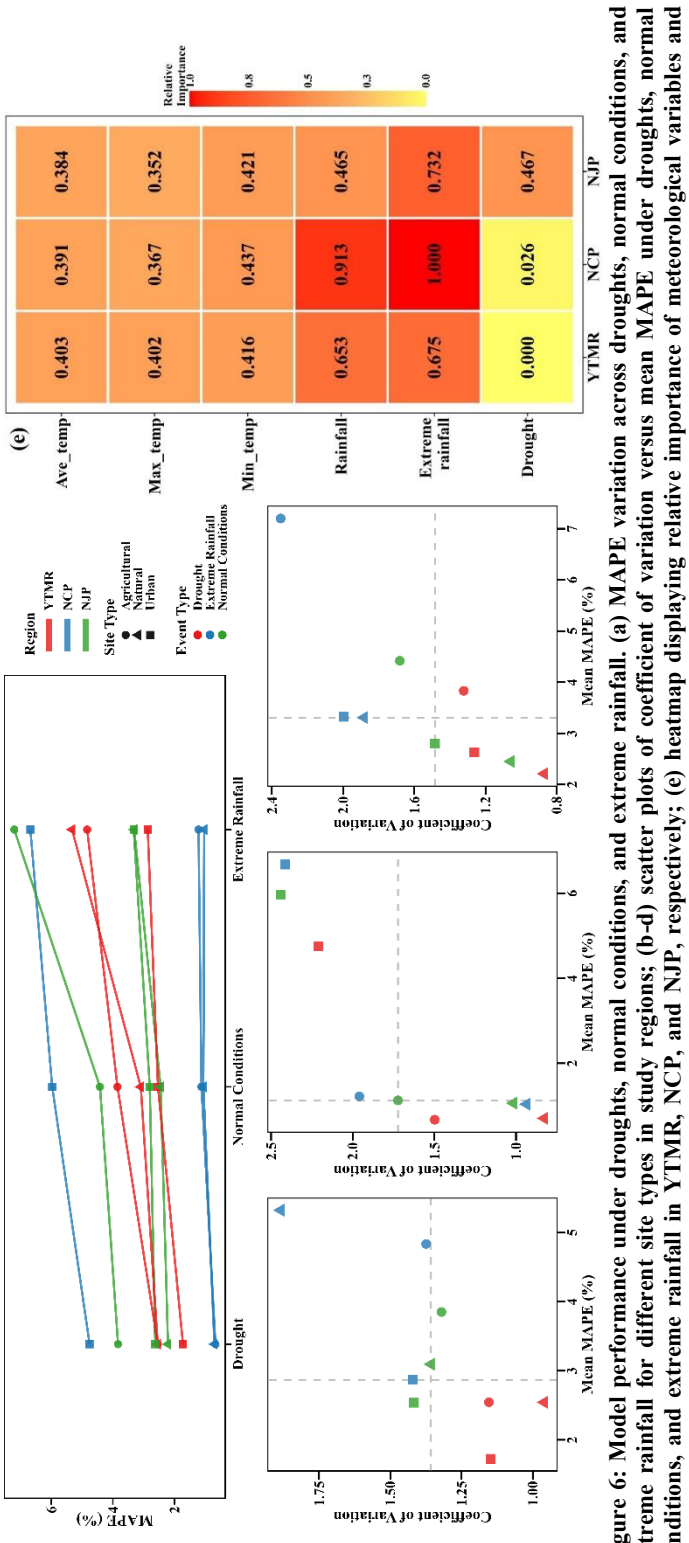


Figure 6: Model performance under droughts, normal conditions, and extreme rainfall. (a) MAPE variation across droughts, normal conditions, and extreme rainfall for different site types in study regions; (b-d) scatter plots of coefficient of variation versus mean MAPE under droughts, normal conditions, and extreme rainfall in YTMR, NCP, and NJP, respectively; (e) heatmap displaying relative importance of meteorological variables and extreme events across the three study regions.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

4.3 Uncertainty analysis and prediction stability assessment

This study employs stratified bootstrap methods to analyze model uncertainty (Figure 7). The NCP shows the highest prediction uncertainty, particularly at urban sites (BSE 0.013), with uncertainty metrics increasing rapidly from T+1 to T+6 before remaining stable. The YTMR region, as a relatively natural mountainous area, demonstrates moderate uncertainty levels (BSE 0.003-0.011), with performance across the three site types exhibiting a relatively balanced triangular distribution. The NJP region shows the lowest prediction uncertainty (0.004) and maintains consistently low and stable uncertainty throughout the multi-step predictions.

Although the NJP region demonstrates the highest stability in uncertainty assessment, prediction accuracy at natural sites deteriorates with increasing time steps. This primarily stems from long-term cumulative effects of complex marine hydrological processes such as seawater intrusion and tidal influence in coastal environments (Michael et al., 2017; Heiss et al., 2022). These systematic errors gradually amplify with prediction steps, but random variations remain relatively small, resulting in characteristics of low uncertainty but declining accuracy. In contrast, both prediction accuracy and stability at NCP urban sites decrease significantly with time steps, primarily related to intensive human interventions including large-scale groundwater extraction and external water input from water diversion projects (Zhu et al., 2020; Yang et al., 2020). The combined effects of these multiple factors increase the complexity and prediction difficulty of groundwater dynamics, particularly as urban environmental factors such as impervious surface expansion and underground pipeline leakage further exacerbate groundwater prediction uncertainty (Zeydalinejad et al., 2024).

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

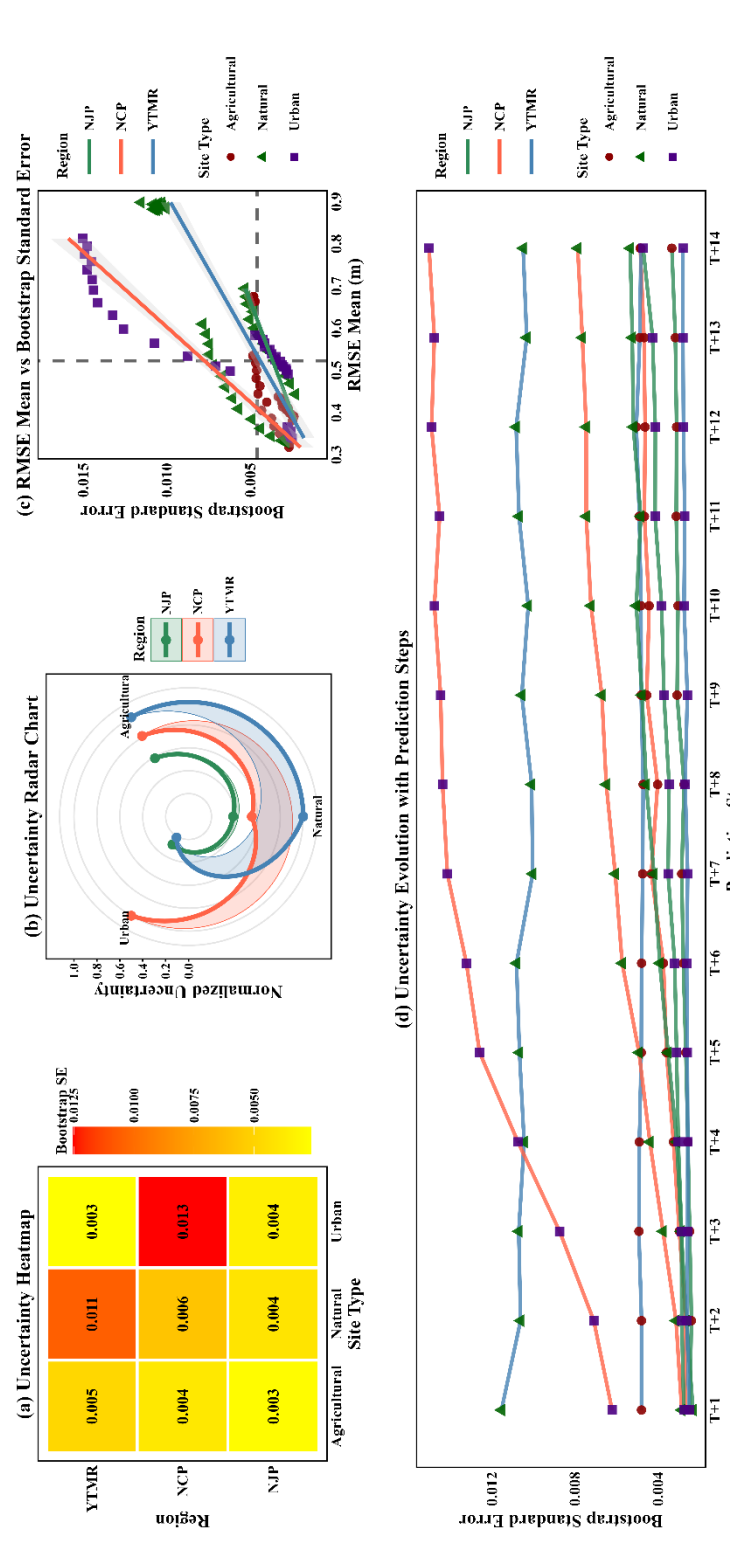


Figure 7: Uncertainty analysis across the three study regions and site types. (a) Uncertainty heatmap showing average bootstrap standard errors for different combinations of regions and site types. (b) Uncertainty radar chart displaying normalized uncertainty patterns across site types for each region, with data normalized to [0,1] scale for comparative analysis. (c) Scatter plot of RMSE mean estimates versus bootstrap standard errors, with region-specific regression lines and 95% confidence intervals (shaded areas). Dashed lines indicate median values for both axes, creating four quadrants for uncertainty categorization. (d) Temporal patterns of bootstrap standard errors across prediction steps from T+1 to T+14.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

5 Discussions

5.1 Encoder importance analysis and regional variations

The importance of the four encoders is analyzed through fusion weight evaluation (Figure 8). In the YTMR, the geographical encoder exhibits the highest weight (35.6%), followed by the meteorological and human activity encoders, which is consistent with ablation results showing that removal of the geographical encoder caused the highest performance degradation in the YTMR predictions. The dominance of the geographical encoder indicates that complex topographical variations in this mountainous terrain fundamentally control groundwater recharge mechanisms through elevation-dependent precipitation gradients, slope-regulated infiltration rates, and valley-confined aquifer geometries that create distinct hydrogeological compartments (Fan et al., 2019; Welch and Allen, 2012). In these semi-arid mountainous areas where natural recharge is limited and spatially heterogeneous, concentrated agricultural irrigation in valley bottoms (with irrigation water and domestic water consumption showing dominant weights of 0.655 at agricultural sites and 0.746 at natural sites, respectively) and domestic extraction in populated zones create strong localized drawdowns that alter natural flow patterns, making anthropogenic perturbations a critical driver of short-term GWD fluctuations (Dalin et al., 2017; Wang et al., 2021a). Although the historical encoder has the lowest weight and did not serve as a primary driving factor, historical state information provides essential baseline calibration that significantly enhances model prediction accuracy (Li et al., 2023a).

In the NCP, the human activity encoder achieves the highest weight (32.5%), followed by the historical encoder (25.6%), while the meteorological encoder shows relatively weaker influence (18.9%). This reveals the profound impact of intensive anthropogenic activities on groundwater dynamics in this heavily exploited agricultural region. Agricultural sites exemplify this pattern, with primary industry GDP exhibiting the highest weight (0.750) followed by irrigation water consumption (0.522); even natural sites show irrigation water as the dominant factor (0.742), nearly equivalent to domestic water consumption (0.738), underscoring the region-wide intensity of agricultural extraction. The unsustainable extraction for irrigation has fundamentally altered natural flow regimes and exacerbated large-scale depression cones across the NCP (Li et al., 2014; Yuan et al., 2013). The substantial weight of the historical encoder indicates that the system retains strong temporal memory effects, attributable to the thick unconsolidated sediment sequences in the plain and confined aquifer conditions, even under intensive exploitation (Nadeau et al., 2018). The relatively lower meteorological encoder weight reflects the limited natural recharge capacity in this semi-arid region, where annual precipitation provides insufficient recharge to significantly modulate groundwater dynamics that are already dominated by extraction-induced drawdown (Cheng et al., 2017; Yaraghi et al., 2019).

The NJP region presents distinct characteristics with the historical encoder playing the dominant role (36.7%), followed by comparable contributions from the meteorological (22.8%) and human activity (22.4%) encoders, while the geographical encoder shows the lowest weight (18.2%). The dominance of the historical encoder reflects the inherent memory effects in plain groundwater systems, where slow flow velocities, high storage capacity, and strong hydraulic connectivity make

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

antecedent states the most reliable predictors of future GWD dynamics (Kratzert et al., 2019). The substantial meteorological encoder weight underscores the critical role of rainfall as the primary recharge source in this humid region, particularly during the wet season when concentrated rainfall provides rapid and direct recharge to shallow aquifers through efficient vadose zone infiltration (Wang et al., 2009). Cuthbert et al. (2019a) demonstrated through global analysis that humid regions exhibit higher groundwater sensitivity to climate variability compared to arid regions, consistent with the strong meteorological dependence observed in the NJP. The nearly equivalent importance of the human activity encoder reveals significant anthropogenic influence on groundwater dynamics in this densely populated and economically developed coastal plain (Li et al., 2025b), as evidenced by urban sites where secondary industry GDP and population density dominate with weights of 0.772 and 0.631, respectively. The relatively low geographical encoder weight indicates that, unlike in the topographically heterogeneous YTMR, the gentle terrain and horizontally continuous aquifer systems in the coastal plain reduce spatial heterogeneity.

The comparative encoder importance analysis across the three regions reveals scale-dependent groundwater control mechanisms governed by the interplay between hydrogeological characteristics and human pressures. The transition from topography-dominated controls in mountainous terrain (YTMR) to temporal memory-anthropogenic coupling in plains (NCP and NJP) reflects fundamental differences in aquifer response timescales and storage capacities. Notably, the divergence between the semi-arid inland NCP and humid coastal NJP, despite both being plain systems, demonstrates that climatic regime and extraction intensity jointly determine whether groundwater dynamics are overwhelmed by anthropogenic perturbations or retain sensitivity to natural recharge processes. The finding challenges simplistic climate-based or geology-based classifications and highlights that effective groundwater management requires understanding the coupled effects of natural hydrogeological conditions and regional human activity patterns.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

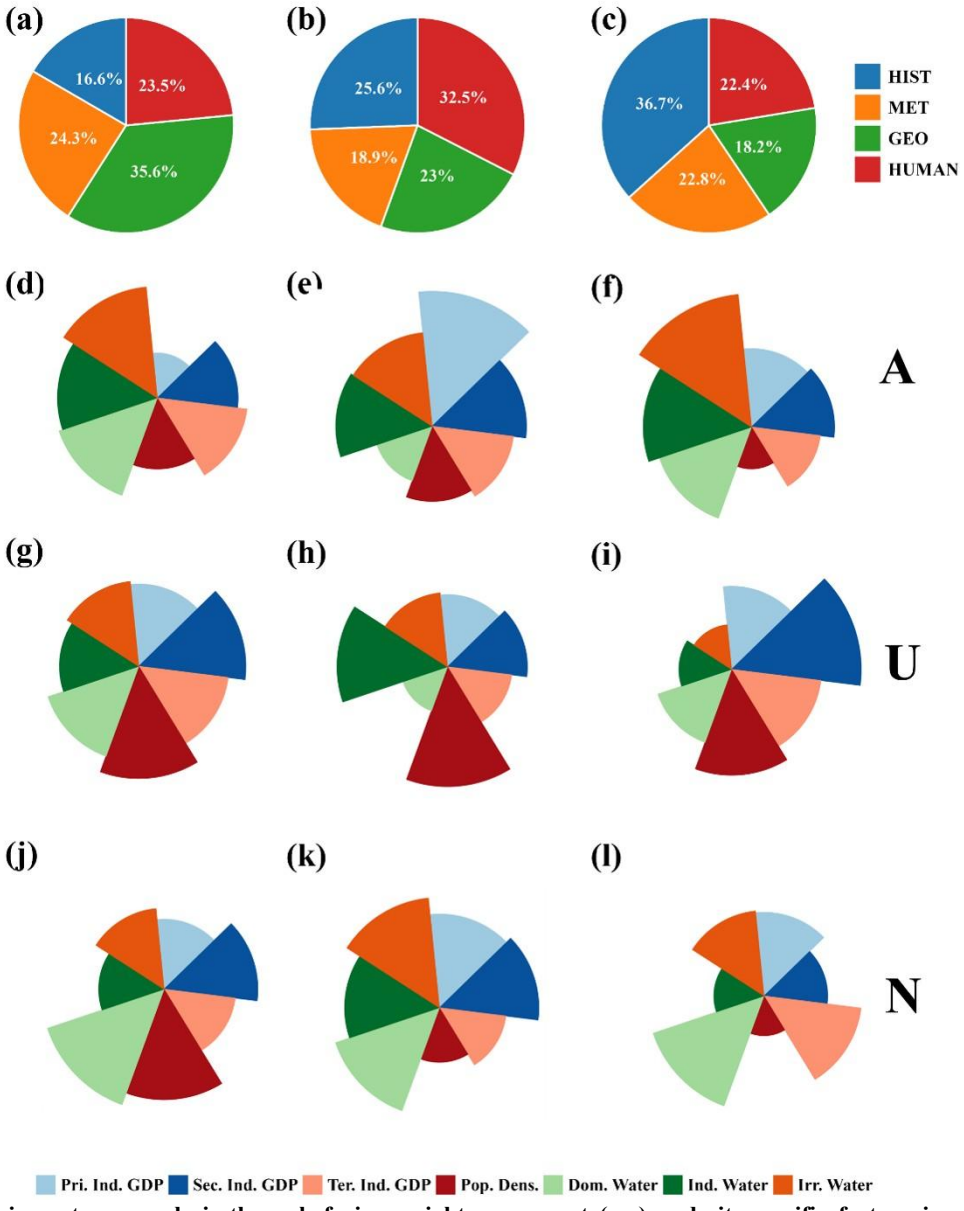


Figure 8: Encoder importance analysis through fusion weight assessment (a-c) and site-specific feature importance of seven socioeconomic indicators across Agricultural, Urban, and Natural site types (d-l) for YTMR, NCP and NJP. In panels (a-c), the legends denote the Historical (HIST), Meteorological (MET), Geographical (GEO), and Human Activity (HUMAN) encoders. Panels (d-l) illustrate the specific contributions of the seven socioeconomic indicators including Population Density (Pop. Dens.), GDP across Primary (Pri. Ind. GDP), Secondary (Sec. Ind. GDP), and Tertiary (Ter. Ind. GDP) industries, and Water Consumption for Domestic (Dom. Water), Industrial (Ind. Water), and Irrigation (Irr. Water) purposes.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

5.2 Temporal memory analysis and response patterns of GWD

The analysis of temporal step importance and response patterns of GWD in the HHA-Net model is presented in Figure 9. In the YTMR, temporal step importance exhibits sustained high-memory characteristics, with agricultural sites maintaining high importance (>0.75) for 25 days before gradually declining. This pattern may be attributed to agricultural water management practices in mountainous areas, particularly periodic irrigation and water retention in terrace systems (Zhang et al., 2019b; Wei et al., 2019). In contrast, urban sites show intermediate and relatively stable importance (0.3~0.5) across the historical 30 days, while natural sites display the lowest importance values (<0.2). The YTMR exhibits the longest response times of GWD to temperature (9.15 days) and rainfall (7.47 days). This is because rainwater must percolate through substantial vadose zones to recharge the water table in the semi-arid region, requiring longer vertical transmission times compared to humid regions (Sun et al., 2025; Manna et al., 2019). Moreover, being characterized by low precipitation and high evaporation, the region generates effective infiltration only after satisfying soil moisture deficits and evaporative demands, thereby explaining why the YTMR exhibits the highest rainfall response threshold (0.70). Cuthbert et al. (2019b) reported that rainfall thresholds tend to rise with increasing aridity, where groundwater recharge becomes more episodic and increasingly dominated by focused recharge through losses from ephemeral overland flows. Furthermore, bedrock fractures tend to partially close or become clogged with fine particles during droughts, and these fractures require time to undergo wetting-swelling-reopening after rainfall events, further delaying the groundwater response (Hahm et al., 2022; Rathay et al., 2018).

Temporal step importance of groundwater in the NCP presents rapid decay patterns, with importance for all site types peaking at the previous 3 days, then declining sharply, and approaching zero by day 10. This is attributable to the high porosity and excellent permeability of thick unconsolidated aquifers in the NCP, resulting in rapid and sensitive response characteristics of GWD to external disturbances, as further confirmed by the shorter response time in this region (approximately 5.76 days) (Cao et al., 2016).

The NCP exhibits a pronounced negative threshold response to rainfall (-0.16), which means that significant GWD fluctuations can be induced regardless of rainfall magnitude (with an activation level of 0.83 even when rainfall equals zero, calculated by Equations A.3 and A.4). This anomalous sensitivity reflects fundamental transformations of the aquifer system. Decades of over-exploitation have created extensive “groundwater depression cones” in the area with dramatically steeper hydraulic gradients compared to regions with adequate groundwater resources (Wang et al., 2021b; Yuan et al., 2013). These structural changes together with intensive pumping have fundamentally altered aquifer hydraulic properties. Li et al. (2025a) pointed out that excessive groundwater extraction for agricultural irrigation in the NCP resulted in a maximum subsidence rate of up to 120 mm/yr from 2015 to 2019 based on Interferometric Synthetic Aperture Radar (InSAR) and Global Navigation Satellite System (GNSS) data, which triggered aquifer system compaction, causing long-term inelastic deformation and the irreversible loss of groundwater storage capacity. Han and Cao (2018) and Su et al. (2021) also reported that severe land subsidence in the NCP due to groundwater over-exploitation had led to the compaction of aquifers and the

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

compression of clay layers. Liu et al. (2022b) reported that most monitoring sites in the NCP showed continuous subsidence from 2003 to 2020, based on GRACE gravity satellite and GNSS geodetic measurements, indicating ongoing groundwater depletion, which was especially evident in spring and summer. As a result, the aquifer has departed from its natural equilibrium state due to severe compromise of the natural attenuation mechanisms typically provided by unsaturated zones and aquifer storage, leading to essential loss of the groundwater system's natural buffering capacity. In this stressed state, the same magnitude of recharge or extraction induces larger GWD fluctuations than in equilibrium systems, with this sensitivity being especially pronounced at the bottom of depression cones where marginal water flux changes can trigger disproportionately large hydraulic head variations.

More critically, even substantial recharge and supply-demand balancing may be insufficient to restore system stability, posing serious threats to ecological security and management strategies (Du et al., 2024). Similar over-exploited aquifer systems exist globally, including the Central Valley of California and the Toluca Valley Basin in Mexico, where comparable threshold behaviors may emerge as extraction intensifies under growing water demands and climate change (Lach et al., 2006; Calderhead et al., 2012). Understanding these negative threshold dynamics is crucial for developing early warning systems that can identify approaching tipping points before irreversible damage occurs, emphasizing the urgent need for proactive rather than reactive groundwater management frameworks.

The NJP shows similar temporal importance decay patterns to the NCP, with GWD at natural sites being more strongly controlled by historical states than at urban and agricultural sites. GWD in the NJP exhibits the shortest response time to rainfall (3.6 days), which is primarily attributed to the dense river network system that forms tight hydraulic connections with shallow groundwater, enabling rainfall to quickly infiltrate into the aquifer system through enhanced surface water-groundwater exchange (Gao et al., 2019). This rapid response is further facilitated by the humid coastal plain environment where soils typically maintain high moisture content, which reduces vadose zone storage requirements and allows rapid transmission of infiltrated water to the water table. Unlike arid regions that require high rainfall intensity to overcome soil interception and evaporative losses, the moisture-rich soils in the NJP enable groundwater recharge from relatively small rainfall events, as evidenced by the low rainfall threshold (0.07). Besides, rainfall consistently exerts strong effects on GWD across all three regions (0.90~0.95), reflecting its dominant role in groundwater dynamics regardless of hydrogeological conditions.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

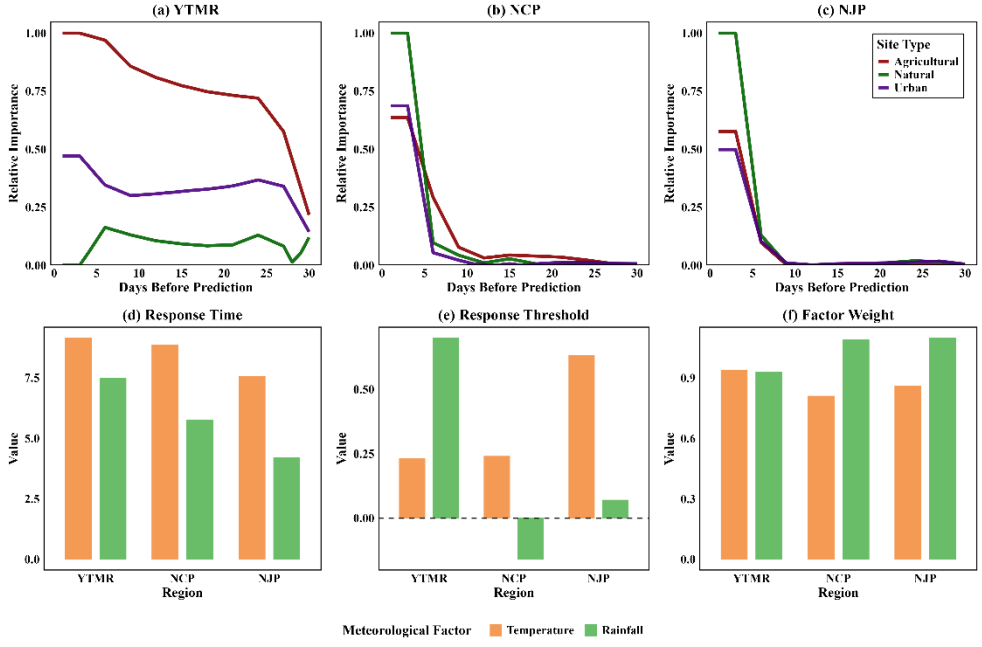


Figure 9: Temporal importance analysis and physics-constrained parameter interpretation.

5.3 Sensitivity coefficients of spatial attention components and spatial similarity networks

Based on the optimal spatial attention coefficients learned by the model, including distance decay factors and similarity coefficients for aquifer type, site type, and watershed, this study constructed spatial similarity networks (Figure S7) and conducted spatial attention sensitivity analysis (Table S7) to understand the spatial connectivity of groundwater systems across different regions.

The YTMR exhibits a distance decay factor of 0.311, indicating that spatial similarity among groundwater sites decays gradually with distance, thus enabling both proximate and relatively distant sites to maintain meaningful similarity weights. Meanwhile, the site-type similarity coefficient (0.471) demonstrates the highest sensitivity among all spatial attention components, suggesting that groundwater dynamics are primarily governed by local environmental conditions associated with the surrounding land use (agricultural, natural, and urban) rather than geographical proximity. This is further evidenced by the moderate aquifer similarity coefficient (0.295) and low watershed similarity coefficient (0.179), such that the spatial similarity network in this region exhibits site-type-based clustering with relatively weak spatial correlations.

The NCP displays an optimal distance decay factor of 0.077, the lowest among the three regions, along with the highest spatial attention sensitivity, indicating extremely weak distance decay patterns and forming a highly dense and complex similarity network. Notably, the site-type similarity coefficient (0.743) demonstrates the highest value, similar to the YTMR. However, the watershed similarity coefficient (0.038) is the lowest among all regions, revealing that natural groundwater recharge-discharge processes and cross-watershed hydraulic connections have been severely disrupted, with groundwater

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

dynamics in different watersheds becoming hydrologically decoupled. This low watershed connectivity combined with high site-type similarity suggests that groundwater dynamics are primarily driven by anthropogenic activities rather than natural hydrological processes in the NCP. As one of the most critically over-exploited aquifer systems, intensive agricultural pumping (>70% of total groundwater use), synchronized irrigation in winter wheat-summer maize rotation cycles, large-scale water diversion projects, and coordinated water resource management have fundamentally altered natural recharge-discharge patterns, creating anthropogenically governed groundwater fluctuations across the region (Chen et al., 2019; Wu et al., 2019). Liu et al. (2021) indicates that soil water storage showed a slight increase after 2015 in the Hebei Province, attributing this to supplementary irrigation from water diversions through the South-to-North Water Diversion Middle Route Project. Long et al. (2025) extends this analysis to groundwater systems, demonstrating that the water diversion project has driven large-scale groundwater recovery in the NCP after 2020, with the water table in shallow aquifers rising at a rate of 0.7 m/year; they also note that, coupled with a 30% decrease in irrigation intensity (reduction of double-cropping agricultural areas), groundwater in some rural areas had begun recovering as early as 2013.

The NJP presents an optimal distance decay factor of 1.107, the highest among the three regions, indicating rapid spatial decay and relatively localized spatial dependencies. The site-type similarity coefficient (0.402) remains the dominant factor, though notably lower than in the NCP. Together with a comparable watershed coefficient (0.246) and aquifer similarity coefficient (0.221), this suggests a more balanced influence between anthropogenic activities and natural hydrological processes in this region. Abundant rainfall recharge, well-developed river networks and canal systems, and favorable groundwater storage conditions homogenize aquifer responses and create a resilient groundwater system in this humid coastal plain, where local hydrological processes dominate over distant spatial dependencies. The highest watershed similarity coefficient in the NJP further confirms that natural surface water-groundwater interactions, including river-aquifer exchanges and tidal influences in coastal areas, maintain strong control over groundwater dynamics (Xiong et al., 2022). This reflects the inherent stability and strong self-regulating capacity of groundwater systems in the region, standing in marked contrast to the NCP, where natural watershed controls have been overridden by anthropogenic extraction (Zheng et al., 2021).

5.4 Limitations and future work

Although HHA-Net demonstrates excellent performance in multi-step GWD prediction across three distinct hydroclimatic regions and successfully reveals physical insights into groundwater response patterns, this study still presents several limitations. The selection of a 30-day input window represents a trade-off between model generalizability and the description of long-term hydrogeological processes (e.g., groundwater recharge cycles and aquifer compaction), with priority given to model applicability in globally data-scarce regions. We will incorporate long-term aquifer storage changes (e.g., GRACE) as features to capture the influences of these slow hydrological processes on GWD dynamics at yearly scales. In terms of physical constraint mechanisms, although the current HHA-Net model incorporates basic physical principles, it lacks explicit representation of fundamental hydrogeological processes, such as groundwater flow equations or aquifer hydraulic

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

parameters. For example, the maximum lag time for groundwater response (5 days) is based on empirical synthesis from regional studies and may not be universally applicable across different hydrogeological environments or extreme conditions. Furthermore, when the prediction accuracy decreases when sites are located in transition zones between different land use types or affected by multiple anthropogenic activities, their hydrological processes often exceed typical patterns that single classification systems can describe. For instance, a natural-type site located in the agricultural-urban transition zone of the NCP region showed a MAPE of 24.98%, significantly exceeding the prediction accuracy of other sites. Additionally, the current dual-weight fusion strategy employs fixed linear weighting with prior and adaptive weights, which may limit model performance in scenarios requiring different optimal weight balances.

Future research directions should focus on extending the temporal memory window based on long-term hydrogeological records, with seasonal attention introduced to dynamically adjust weight allocation for information at different time scales (Darlow et al., 2023). Furthermore, fundamental physical constraints including groundwater flow equations (the Richards equation, Darcy's law) and aquifer hydraulic parameters should be directly embedded into neural network architecture as hard constraints rather than soft penalty terms in loss functions. For example, designing physics-informed neural network (PINN) hybrid architectures that combine data-driven HHA-Net with physics equation solvers can satisfy data fitting while strictly adhering to physical laws (Tartakovsky et al., 2020; Haruzi and Moreno, 2023). For complex environment modeling, dynamic site type identification systems and mixed environment adaptive modeling methods should be developed, utilizing spatiotemporal changes in land use and human activity intensity gradients to identify transition regions, developing soft classification mechanisms to assign multiple type probabilities, and establishing spatial heterogeneity-aware attention mechanisms to adaptively handle mixed influence effects in complex environments. Regarding the weight fusion mechanism, future improvements should focus on developing adaptive weight allocation strategies that dynamically learn optimal balances between prior and adaptive weights based on uncertainty measures and hierarchical feature dimensions. Furthermore, integrating IPCC climate scenarios (SSPs and RCPs) with CMIP6 and regional climate models (RegCM4, WRF) would enable the development of climate-aware neural architectures that adapt to evolving precipitation, temperature, and extreme event patterns, enhancing long-term model applicability for groundwater sustainability assessment (Kayhomayoon et al., 2023; Furusho-Percot et al., 2022).

6 Conclusions

This study presents a novel HHA-Net that processes multi-source heterogeneous data through hydrologically-guided hierarchical encoders, employs adaptive weight allocation and multi-head spatiotemporal attention mechanisms, and integrates hydrological knowledge constraints to achieve high-precision multi-step GWD prediction while providing interpretable physical insights into groundwater dynamics. Three distinct hydroclimatic and geographic regions in China were selected as case studies, and results demonstrate that HHA-Net consistently outperforms baseline models, with fourteen-step prediction MAPE ranging from 1.02% to 5.95% and R^2 ranging from 0.71 to 0.98. The model maintains

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

excellent performance across different site types and spatiotemporal prediction scales, particularly showing superior predictive ability in the YTMR, while the NCP and NJP regions exhibit slightly decreased precision at specific sites due to anthropogenic disturbances and coastal hydrological processes. The model demonstrates significantly improved accuracy under drought events, while performance decreases during extreme rainfall events, particularly at natural sites in the YTMR. Model interpretation reveals distinct groundwater system response patterns across the three regions, with the geographical encoder demonstrating the highest contribution to GWD prediction in the mountainous YTMR (35.6%), the human activity encoder dominating in the inland NCP (32.5%), and the historical encoder governing the coastal NJP (36.7%). Temporally, the groundwater system in the YTMR demonstrates prolonged memory with temporal attention weights remaining >0.75 for the past 25 days at agricultural sites, the highest rainfall threshold (0.70), and the longest response times to temperature (9.15 days) and rainfall (7.47 days) due to delayed infiltration in semi-arid conditions. In contrast, both the NCP and NJP show rapid decay with importance peaking at 3 days, though they exhibit contrasting rainfall responses, with the NCP exhibiting a negative threshold (-0.16) and moderate response time (5.76 days) reflecting unstable dynamics from decades of over-exploitation, while the NJP shows the shortest response time (3.6 days) and a lower positive threshold (0.07) due to dense river networks and high soil moisture in the humid coastal environment. Spatially, the NCP exhibits the most interconnected network (decay factor 0.077) with the highest site-type similarity (0.743) but lowest watershed similarity (0.038), indicating that anthropogenic activities have fundamentally overridden natural hydrological processes. The YTMR shows moderate connectivity (decay factor 0.311) dominated by site-type similarity (0.471), while the NJP demonstrates the most localized dependencies (decay factor 1.107) with comparable influence among spatial similarity components, maintaining integrated control by both anthropogenic and natural processes. These findings highlight the potential of HHA-Net to transform groundwater science by bridging the gap between advanced machine learning and hydrological understanding, offering a pathway toward more informed decision-making for sustainable water resource management under growing environmental uncertainty and increasing water stress.

Appendix A: Details of the HHA-Net model

A1. Design of Physics-constrained LSTM and LSTM-convolution structure

This study designs a physics-constrained LSTM to simulate groundwater response characteristics to influencing factors. Three physical mechanisms are integrated, including a time-delay gating mechanism that simulates delayed responses of groundwater to rainfall infiltration through learnable parameters; a threshold activation mechanism that captures nonlinear system responses through learnable threshold parameters; and a dynamic factor weighting mechanism that adaptively adjusts contributions of different influencing factors under different hydrological conditions.

These delay parameters are constrained to the 0-1 interval through sigmoid functions (Equation A.1) and multiplied by maximum delay days to obtain GWD lagged response times (Equation A.2). Mo et al. (2025) indicated that shallow groundwater in coastal Jiangsu exhibits 1-4 day lag responses to rainfall; Luo et al. (2024) studying the Yellow River Delta

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

aquifer responses to three consecutive days of extreme rainfall, and found that groundwater levels began rising 12 hours after rainfall initiation, reaching peaks within 1-2 days after rainfall cessation; Lee et al. (2006) reported that GWD in shallow areas where the unsaturated zone is less than 18 m thick responded extremely rapidly (within 0–2 days) to rainfall, while it showed significantly delayed responses (12 to 18 days) in thicker unsaturated zones (<50 meters) in the Chalk aquifer in southern England; Yu et al. (2017) demonstrated by numerical simulations that groundwater systems do not respond instantaneously to rainfall events, exhibiting a direct response lag of 4 to 7 days, while the peak response time extended to 7~13 days with increasing vadose zone thickness. Based on these studies, we set the maximum rainfall lag time at 15 days to balance capturing the rainfall-groundwater response process and maintaining sufficient historical sequence length for model training, given that we use 30-day historical data to predict GWD over the subsequent 14 days.

$$\mathrm{delay\_weight} = \sigma(\tau) \tag{A.1}$$

$$\mathrm{response\_time} = \mathrm{delay\_weight} \times 15 \tag{A.2}$$

where τ is the learnable delay parameter and σ is the Sigmoid activation function that constrains τ to the (0,1) interval to obtain normalized delay weights.

To simulate nonlinear threshold response characteristics, we introduce learnable threshold parameters, whereby system responses are significantly amplified when the absolute values of input signals exceed the thresholds. The modulation factor M is defined as:

$$M = \mathrm{Sigmoid}[10 \times (x - \tau)] \tag{A.3}$$

$$\mathrm{Sigmoid}(z) = \frac{1}{1 + e^{-z}} \tag{A.4}$$

where x represents the meteorological input variables (temperature and rainfall in the study), τ is the response threshold parameter learned through model optimization, and the coefficient 10 serves as a scaling factor that steepens the sigmoid transition, enabling the model to better capture threshold-driven nonlinear responses. This formulation ensures that meteorological forcing has significant effects on GWD only when input magnitudes surpass the response threshold.

Weight vectors for current time steps are calculated based on previous time step hidden states through fully connected layers and SoftMax functions, dynamically adjusting contribution weights of each influence factor under different hydrological conditions. To simultaneously capture long-term temporal dependencies and short-term local variation patterns, an LSTM-convolution dual-branch structure is employed. The LSTM branch specializes in extracting long-term temporal dependencies, leveraging its advantages in memorizing and transmitting hydrological information across extended time spans. The one-dimensional convolution branch focuses on capturing short-term local temporal patterns, using convolution kernels of 3, 5, and 7-day time scales to identify GWD fluctuation patterns. Through feature fusion layers, LSTM global features are linearly combined with convolutional local features to obtain comprehensive features incorporating both long-term memory and local variations.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

A2. Design of four-dimensional spatial similarity matrix

This study proposes a four-dimensional spatial similarity measurement method that comprehensively considers geographical distance, aquifer type, site type, and watershed affiliation similarities. Based on Tobler's First Law of Geography, exponential decay functions measure geographical distance similarity (Goodchild, 2009), reflecting rapid decay characteristics of spatial autocorrelation with increasing distance (Equation A.5). Aquifer, site type, and watershed similarity coefficients are predefined to serve as initial weights in the attention mechanism. For each similarity type, identical categories receive weights of 1.0, while different categories receive weights of 0.3, 0.5, and 0.2 for aquifer, site type, and watershed similarity, respectively, based on regional characteristics and empirical knowledge. The model learns optimal region-specific values for these coefficients during training, allowing the similarity structure to adapt to actual hydrogeological conditions rather than relying solely on categorical assignments. Spatial similarities are evaluated through a four-dimensional similarity coefficient matrix (Equation A.6).

$$\mathrm{distance\_similarity} = \exp(-\mathrm{dist}_{[j,k]} \times \lambda) \tag{A.5}$$

$$\mathrm{spatial\_sim}_{[j,k]} = \mathrm{diatance\_sim}_{[j,k]} \times \mathrm{aquifer\_sim}_{[j,k]} \times \mathrm{site\_sim}_{[j,k]} \times \mathrm{watershed\_sim}_{[j,k]} \tag{A.6}$$

where *j,k* represents two monitoring sites for spatial similarity calculation, *dist* represents the Euclidean distance between stations, and λ represents the distance decay factor (learnable parameter) that adaptively adjusts distance decay rates according to groundwater system characteristics in different regions.

A3. Training configuration and regularization

The dataset is divided into training, validation, and test sets with an 8:1:1 ratio following the temporal sequence of observations (Jahangir and Quilty, 2025). The training process employs a warmup-decay learning rate scheduling strategy. To avoid gradient oscillations and convergence instability during early training phases caused by random initialization, the warmup phase begins at 10% of the target learning rate, linearly increasing to the target learning rate (0.003) over the first 10 epochs. After warmup completion, a *ReduceLROnPlateau* adaptive decay strategy is adopted, multiplying the learning rate by a decay factor of 0.5 when validation loss shows no improvement for 5 consecutive epochs. To prevent training stagnation due to excessive learning rate decay, the minimum learning rate is set to 1×10^{-6}.

Gradient clipping techniques prevent gradient explosion problems during temporal data processing, with the maximum gradient norm set to 1. Weight decay (L2 regularization) prevents overfitting with a regularization coefficient of 0.01. L2 regularization constrains model complexity by adding parameter sum-of-squares terms to the loss function, promoting simple and generalizable parameter configurations. All fully connected and attention layers employ Dropout regularization with a deactivation probability set to 0.3. Dropout increases training randomness by randomly deactivating portions of neurons, preventing models from over-relying on specific neuron combinations and enhancing generalization capability and robustness. The AdamW optimizer provides superior regularization effects. To ensure timely training termination at optimal

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

generalization performance and prevent further optimization on training sets from harming generalization capability, early stopping mechanisms are adopted with patience set to 15 epochs. Training automatically stops when validation loss fails to decrease for 15 consecutive epochs.

A4. Detailed design of multi-objective loss function

Mean squared error loss is more sensitive to large errors, mean absolute error loss is more robust to outliers, and Huber loss combines the advantages of both, exhibiting quadratic function characteristics in small error intervals and converting to linear functions in large error intervals, providing more balanced optimization guidance for models. Temporal consistency loss includes first-order and second-order difference losses, while the former constrains prediction sequence continuity by calculating differences between adjacent time step predictions, while the latter constrains smoothness by calculating curvature changes in prediction sequences, ensuring prediction results conform to physical process continuity requirements and avoiding physically unreasonable predictions from purely data-driven methods.

Physical consistency constraint loss primarily includes variation magnitude constraints and continuity constraints. Variation magnitude constraints determine maximum reasonable daily variation magnitudes based on historical data (set to 0.5 m in this model), with changes exceeding this threshold considered unreasonable and subject to additional penalties. Continuity constraints ensure reasonable continuity between initial prediction sequence values and final historical observation values, preventing unrealistic abrupt changes. In the total loss function, the primary loss MSE weight is set to 1.0 as the main optimization objective, the MAE loss weight is set to 0.5 for balancing and stabilizing effects, the temporal consistency constraint weight is set to 0.3 to ensure prediction temporal reasonableness, and the physical constraint weight is set to 0.2 to ensure physical reasonableness without excessively constraining model flexibility.

Code and data availability

The daily GWD data are available at https://geocloud.cgs.gov.cn. The code used in this study can be found on Zenodo (https://doi.org/10.5281/zenodo.18130111). The model evaluation and detailed figures and metrics for the GEMS-GER dataset can also be found on Zenodo (https://doi.org/10.5281/zenodo.19230562).

Author contributions

Jing Xu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Software, Validation, Visualization, Writing - original draft. Yuming Mo: Methodology, Funding acquisition, Investigation, Writing - review &editing. Senlin Zhu: Methodology, Software, Writing - review &editing. Chengji Shen: Validation, Writing - review &editing. Xinli Zhu: Methodology, Software. Chenming Zhang: Methodology, Software. Qihao Jiang: Validation, Writing - review &editing. Ling Li: Supervision, Writing - review & editing.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Financial support

This research was supported by the Natural Science Foundation of Jiangsu Province, China (BK20240937), the Basic Research Program of Jiangsu (BK20251017), and the Belt and Road Special Foundation of the National Key Laboratory of Water Disaster Prevention (2022491411, 2021491811).

References

Ali, A. S. A., Jazaei, F., Clement, T. P., and Waldron, B.: Physics-informed neural networks in groundwater flow modeling: Advantages and future directions, Groundwater for Sustainable Development, 25, 101172, https://doi.org/10.1016/j.gsd.2024.101172, 2024a.

Ali, A. S. A., Jazaei, F., Babakhani, P., Ashiq, M. M., Bakhshaee, A., and Waldron, B.: An overview of deep learning applications in groundwater level modeling: Bridging the gap between academic research and industry applications, Applied Computational Intelligence and Soft Computing, 2024, 9480522, 2024b.

Arshad, A., Mirchi, A., Vilcaez, J., Akbar, M. U., and Madani, K.: Reconstructing high-resolution groundwater level data using a hybrid random forest model to quantify distributed groundwater changes in the Indus Basin, Journal of Hydrology, 628, 10.1016/j.jhydrol.2023.130535, 2024.

Asif, N. A., Sarker, Y., Chakrabortty, R. K., Ryan, M. J., Ahamed, M. H., Saha, D. K., Badal, F. R., Das, S. K., Ali, M. F., and Moyeen, S. I.: Graph neural network: A comprehensive review on non-euclidean space, Ieee Access, 9, 60588-60606, 2021.

Bai, T. and Tahmasebi, P.: Graph neural network for groundwater level forecasting, J. Hydrol., 616, 128792, 2023.

Bai, T., Wang, X.-S., and Han, P.-F.: Controls of groundwater-dependent vegetation coverage in the yellow river basin, china: Insights from interpretable machine learning, J. Hydrol., 631, 130747, https://doi.org/10.1016/j.jhydrol.2024.130747, 2024.

Bengio, Y., Courville, A., and Vincent, P.: Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence, 35, 1798-1828, 2013.

Bikše, J., Retike, I., Haaf, E., and Kalvāns, A.: Assessing automated gap imputation of regional scale groundwater level data sets with typical gap patterns, J. Hydrol., 620, 129424, https://doi.org/10.1016/j.jhydrol.2023.129424, 2023.

Cadol, D., Kampf, S., and Wohl, E.: Effects of evapotranspiration on baseflow in a tropical headwater catchment, J. Hydrol., 462, 4-14, 2012.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Calderhead, A. I., Martel, R., Garfias, J., Rivera, A., and Therrien, R.: Pumping dry: an increasing groundwater budget deficit induced by urbanization, industrialization, and climate change in an over-exploited volcanic aquifer, Environmental Earth Sciences, 66, 1753-1767, 2012.

Cao, G., Scanlon, B. R., Han, D., and Zheng, C.: Impacts of thickening unsaturated zone on groundwater recharge in the North China Plain, J. Hydrol., 537, 260-270, 2016.

Cetin, M.: Agricultural water use, in: Water resources of Turkey, Springer, 257-302, 2019.

Chen, X., Thorp, K. R., Ouyang, Z., Hou, Y., Zhou, B., and Li, Y.: Energy consumption due to groundwater pumping for irrigation in the North China Plain, Sci. Total Environ., 669, 1033-1042, 2019.

Cheng, Y., Zhan, H., Yang, W., Dang, H., and Li, W.: Is annual recharge coefficient a valid concept in arid and semi-arid regions?, Hydrology and Earth System Sciences, 21, 5031-5042, 2017.

Collados-Lara, A. J., Pulido-Velazquez, D., Ruiz, L. G. B., Pegalajar, M. C., Pardo-Igúzquiza, E., and Baena-Ruiz, L.: A parsimonious methodological framework for short-term forecasting of groundwater levels, Sci. Total Environ., 881, 10.1016/j.scitotenv.2023.163328, 2023.

Cuthbert, M., Gleeson, T., Moosdorf, N., Befus, K. M., Schneider, A., Hartmann, J., and Lehner, B.: Global patterns and dynamics of climate–groundwater interactions, Nature Climate Change, 9, 137-141, 2019a.

Cuthbert, M. O., Taylor, R. G., Favreau, G., Todd, M. C., Shamsudduha, M., Villholth, K. G., MacDonald, A. M., Scanlon, B. R., Kotchoni, D. O. V., Vouillamoz, J.-M., Lawson, F. M. A., Adjomayi, P. A., Kashaigili, J., Seddon, D., Sorensen, J. P. R., Ebrahim, G. Y., Owor, M., Nyenje, P. M., Nazoumou, Y., Goni, I., Ousmane, B. I., Sibanda, T., Ascott, M. J., Macdonald, D. M. J., Agyekum, W., Koussoubé, Y., Wanke, H., Kim, H., Wada, Y., Lo, M.-H., Oki, T., and Kukuric, N.: Observed controls on resilience of groundwater to climate variability in sub-Saharan Africa, Nature, 572, 230-234, 10.1038/s41586-019-1441-7, 2019b.

Dalin, C., Wada, Y., Kastner, T., and Puma, M. J.: Groundwater depletion embedded in international food trade, Nature, 543, 700-704, 2017.

Darlow, L. N., Joosen, A., Asenov, M., Deng, Q., Wang, J., and Barker, A.: FoldFormer: sequence folding and seasonal attention for fine-grained long-term FaaS forecasting, Proceedings of the 3rd Workshop on Machine Learning and Systems, 71-77,

Du, J., Laghari, Y., Wei, Y.-C., Wu, L., He, A.-L., Liu, G.-Y., Yang, H.-H., Guo, Z.-Y., and Leghari, S. J.: Groundwater depletion and degradation in the North China Plain: Challenges and mitigation options, Water, 16, 354, 2024.

Fan, X. Y., Peterson, T. J., Henley, B. J., and Arora, M.: Groundwater Sensitivity to Climate Variations Across Australia, Water Resour. Res., 59, 10.1029/2023wr035036, 2023.

Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339, 940-943, 10.1126/science.1229881, 2013.

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., and Grant, G.: Hillslope hydrology in global change research and earth system modeling, Water Resour. Res., 55, 1737-1772, 2019.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environmental Change, 23, 144-156, 2013.

Frame, D. J., Rosier, S. M., Noy, I., Harrington, L. J., Carey-Smith, T., Sparrow, S. N., Stone, D. A., and Dean, S. M.: Climate change attribution and the economic costs of extreme weather events: a study on damages from extreme rainfall and drought, Climatic Change, 162, 781-797, 2020.

Furusho-Percot, C., Goergen, K., Hartick, C., Poshyvailo-Strube, L., and Kollet, S.: Groundwater model impacts multiannual simulations of heat waves, Geophysical research letters, 49, e2021GL096781, 2022.

Gao, C., Zhang, Y., Zhou, Y., Gu, C., Qing, D., and Wang, Z.: Evaluation of river network planning layout in plain city consideration for combining water quality and flood control, Desalination and Water Treatment, 168, 224-234, 2019.

Gezici, K., Katipoglu, O. M., and Sengül, S.: Hybrid machine learning models for groundwater level prediction in a snow-dominated region: An evaluation of EEMD, VMD and EWT decomposition techniques, Hydrological Processes, 38, 10.1002/hyp.15169, 2024.

Gleeson, T. and Manning, A. H.: Regional groundwater flow in mountainous terrain: Three-dimensional simulations of topographic and hydrogeologic controls, Water Resour. Res., 44, 2008.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Gomez, M., Nölscher, M., Hartmann, A., and Broda, S.: Assessing groundwater level modelling using a 1-D convolutional neural network (CNN): linking model performances to geospatial and time series features, Hydrology and Earth System Sciences, 28, 4407-4425, 2024.

Goodchild, M.: First law of geography, 2009.

Guo, J., Ren, G., Xiong, M., and Huang, H.: The spatiotemporal pattern of rainy-season precipitation in the Haihe River Basin, North China, Hydrology, 6, 73, 2019.

Haaf, E. and Barthel, R.: An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs, J. Hydrol., 559, 222-237, 2018.

Hahm, W. J., Dralle, D. N., Sanders, M., Bryk, A. B., Fauria, K. E., Huang, M.-H., Hudson-Rasmussen, B., Nelson, M. D., Pedrazas, M. A., and Schmidt, L.: Bedrock vadose zone storage dynamics under extreme drought: Consequences for plant water availability, recharge, and runoff, Water Resour. Res., 58, e2021WR031781, 2022.

Hameed, M., Nayak, M. A., and Ahanger, M. A.: Event-Based Recession Analysis for Estimation of Basin-Wide Characteristic Drainage Timescale and Groundwater Storage Trends, Water Resour. Res., 59, e2023WR035829, 2023.

Han, D. and Cao, G.: Phase difference between groundwater storage changes and groundwater level fluctuations due to compaction of an aquifer-aquitard system, J. Hydrol., 566, 89-98, 2018.

Han Jingya, Gou Jiaojiao, and Chiyuan, M.: A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations, National Tibetan Plateau Data Center [dataset], 10.11888/Atmos.tpdc.300523, 2023.

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., and Wang, Y.: Transformer in transformer, Advances in neural information processing systems, 34, 15908-15919, 2021.

Haruzi, P. and Moreno, Z.: Modeling water flow and solute transport in unsaturated soils using physics-informed neural networks trained with geoelectrical data, Water Resour. Res., 59, e2023WR034538, 2023.

Heiss, J. W., Mase, B., and Shen, C.: Effects of Future Increases in Tidal Flooding on Salinity and Groundwater Dynamics in Coastal Aquifers, Water Resour. Res., 58, e2022WR033195, https://doi.org/10.1029/2022WR033195, 2022.

Herrera-García, G., Ezquerro, P., Tomás, R., Béjar-Pizarro, M., López-Vinielles, J., Rossi, M., Mateos, R. M., Carreón-Freyre, D., Lambert, J., Teatini, P., Cabral-Cano, E., Erkens, G., Galloway, D., Hung, W.-C., Kakar, N., Sneed, M., Tosi, L., Wang, H., and Ye, S.: Mapping the global threat of land subsidence, Science, 371, 34-36, 10.1126/science.abb8549, 2021.

Heudorfer, B., Liesch, T., and Broda, S.: On the challenges of global entity-aware deep learning models for groundwater level prediction, Hydrol. Earth Syst. Sci., 28, 525-543, 10.5194/hess-28-525-2024, 2024.

Hou, M., Wei, J., Shi, Y., Hou, S., Zhang, W., Xu, J., Wu, Y., and Wang, H.: Hydroformer: Frequency domain enhanced multi-attention transformer for monthly lake level reconstruction with low data input requirements, Water Resour. Res., 60, e2024WR037166, 2024.

Jahangir, M. S. and Quilty, J.: Hierarchical Deep Learning for Consistent Multi-Timescale Hydrological Forecasting, Water Resour. Res., 61, e2024WR038105, https://doi.org/10.1029/2024WR038105, 2025.

Jasechko, S. and Perrone, D.: Global groundwater wells at risk of running dry, Science, 372, 418-421, 10.1126/science.abc2755, 2021.

Jung, H., Saynisch-Wagner, J., and Schulz, S.: Can eXplainable AI Offer a New Perspective for Groundwater Recharge Estimation?-Global-Scale Modeling Using Neural Network, Water Resour. Res., 60, 10.1029/2023wr036360, 2024.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, Nature Reviews Physics, 3, 422-440, 10.1038/s42254-021-00314-5, 2021.

Kayhomayoon, Z., Jamnani, M. R., Rashidi, S., Milan, S. G., Azar, N. A., and Berndtsson, R.: Soft computing assessment of current and future groundwater resources under CMIP6 scenarios in northwestern Iran, Agricultural Water Management, 285, 108369, 2023.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005-6022, 10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resour. Res., 55, 11344-11354, 2019.

Kuang, X., Liu, J., Scanlon, B. R., Jiao, J. J., Jasechko, S., Lancia, M., Biskaborn, B. K., Wada, Y., Li, H., and Zeng, Z.: The changing nature of groundwater in the global water cycle, Science, 383, eadf0630, 2024.

Lach, D., Ingram, H., and Rayner, S.: You never miss the water till the well runs dry: crisis and creativity in California, in: Clumsy solutions for a complex world: Governance, politics and plural perceptions, Springer, 226-240, 2006.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Lan, P., Guo, L., Sun, H., Zhang, Y., and Jiang, Y.: Modeling stream baseflow nitrate concentration in an agricultural watershed using neural network and bootstrap method, Ecol. Indic., 156, 111097, 2023.
Le Brocque, A. F., Kath, J., and Reardon-Smith, K.: Chronic groundwater decline: A multi-decadal analysis of groundwater trends under extreme climate cycles, J. Hydrol., 561, 976-986, https://doi.org/10.1016/j.jhydrol.2018.04.059, 2018.
Lee, L., Lawrence, D., and Price, M.: Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England, J. Hydrol., 330, 604-620, 2006.
Lehr, C. and Lischeid, G.: Efficient screening of groundwater head monitoring data for anthropogenic effects and measurement errors, Hydrol. Earth Syst. Sci., 24, 501-513, 10.5194/hess-24-501-2020, 2020.
Li, F., Ma, G., Ju, C., Chen, S., and Huang, W.: Data-driven forecasting framework for daily reservoir inflow time series considering the flood peaks based on multi-head attention mechanism, J. Hydrol., 645, 132197, 2024.
Li, J., Yuan, X., and Ji, P.: Long-lead daily streamflow forecasting using Long Short-Term Memory model with different predictors, Journal of Hydrology: Regional Studies, 48, 101471, 2023a.
Li, L., Fang, Y., Wu, J., Wang, J., and Ge, Y.: Encoder–decoder full residual deep networks for robust regression and spatiotemporal estimation, IEEE transactions on neural networks and learning systems, 32, 4217-4230, 2020.
Li, M. and Janson, L.: Optimal ablation for interpretability, Advances in Neural Information Processing Systems, 37, 109233-109282, 2024.
Li, M., Sun, J., Xue, L., Shen, Z.-K., Li, Y., Zhao, B., and Hu, L.: Characterizing Aquifer Properties and Groundwater Storage at North China Plain Using Geodetic and Hydrological Measurements, Water Resour. Res., 61, e2024WR037425, https://doi.org/10.1029/2024WR037425, 2025a.
Li, Y., Zhang, F., Han, Z., Wang, P., Chen, H., and Zhang, Z.: Evolution characteristics and influence factors of deep groundwater depression cone in North China Plain, China—a case study in Cangzhou region, Journal of Earth Science, 25, 1051-1058, 2014.
Li, Y., Mi, W., Ji, L., He, Q., Yang, P., Xie, S., and Bi, Y.: Urbanization and agriculture intensification jointly enlarge the spatial inequality of river water quality, Sci. Total Environ., 878, 162559, 2023b.
Li, Z., Lu, C., Zhang, Y., Yu, X., Gao, Q., Yuan, Z., Wu, C., Liu, B., and Shu, L.: Seasonal groundwater variations in the coastal plain under intense human activities: Hydrochemical evolution, quality degradation, and health risk assessment, J. Environ. Manage., 388, 126062, 2025b.
Liu, B.-Y., Liu, W.-S., Lin, B.-J., Liu, W.-X., Han, S.-W., Zhao, X., and Zhang, H.-L.: Sustainable management practices to improve the water use efficiency of winter wheat in the North China Plain: a meta-analysis, Agronomy for Sustainable Development, 42, 33, 2022a.
Liu, J., Jiang, L., Zhang, X., Druce, D., Kittel, C. M. M., Tøttrup, C., and Bauer-Gottwein, P.: Impacts of water resources management on land water storage in the North China Plain: Insights from multi-mission earth observations, Journal of Hydrology, 603, 126933, https://doi.org/10.1016/j.jhydrol.2021.126933, 2021.
Liu, R., Zhong, B., Li, X., Zheng, K., Liang, H., Cao, J., Yan, X., and Lyu, H.: Analysis of groundwater changes (2003–2020) in the North China Plain using geodetic measurements, Journal of Hydrology: Regional Studies, 41, 101085, https://doi.org/10.1016/j.ejrh.2022.101085, 2022b.
Liu, Y., Shan, F., Yue, H., Wang, X., and Fan, Y.: Global analysis of the correlation and propagation among meteorological, agricultural, surface water, and groundwater droughts, J. Environ. Manage., 333, 117460, 2023.
Loaiciga, H. A. and Doh, R.: Groundwater for people and the environment: a globally threatened resource, Groundwater, 62, 332-340, 2024.
Long, D., Xu, Y., Cui, Y., Cui, Y., Butler, J. J., Dong, L., Wang, L., Liu, D., Wada, Y., Hu, L., Bai, G., Li, B., Wang, S., Nong, X., Cai, Y., Cheng, C., Mu, Y., Qiao, Y., Wang, J., Wang, H., and Scanlon, B. R.: Unprecedented large-scale aquifer recovery through human intervention, Nature Communications, 16, 7296, 10.1038/s41467-025-62719-5, 2025.
Luo, F., Wang, D., Tian, X., Bi, X., Zheng, Q., Zhou, Z., and Tang, Z.: Estuarine groundwater level response to and recovery from extreme precipitation events: Typhoon Lekima in the Yellow River Delta, J. Hydrol., 632, 130918, https://doi.org/10.1016/j.jhydrol.2024.130918, 2024.
Manna, F., Murray, S., Abbey, D., Martin, P., Cherry, J., and Parker, B.: Spatial and temporal variability of groundwater recharge in a sandstone aquifer in a semiarid region, Hydrol. Earth Syst. Sci., 23, 2187-2205, 10.5194/hess-23-2187-2019, 2019.

Mao, M., Wei, L., Gong, W., Wu, G., and Liu, T.: Impacts of Spatio-Temporal Changes in Anthropogenic Disturbances on Landscape Patterns in the Nandu River Basin, China, Sustainability, 16, 2724, 2024.

Martinez-Ruiz, A., Gomez-Gil, P., and Fonseca-Delgado, R.: An Analysis of Spatio-Temporal Graph Neural Networks Based on Synthetic Time Series with Known Structural Dependencies, Mexican International Conference on Artificial Intelligence, 313-324,

Michael, H. A., Post, V. E. A., Wilson, A. M., and Werner, A. D.: Science, society, and the coastal groundwater squeeze, Water Resour. Res., 53, 2610-2617, https://doi.org/10.1002/2017WR020851, 2017.

Mo, Y., Xu, J., Zhu, S., Xu, B., Wu, J., Jin, G., Wang, Y.-G., and Li, L.: Spatial heterogeneity of groundwater depths in coastal cities and their responses to multiple factors interactions by interpretable machine learning models, Geoscience Frontiers, 16, 102033, https://doi.org/10.1016/j.gsf.2025.102033, 2025.

Nadeau, S., Rosa, E., and Cloutier, V.: Stratigraphic sequence map for groundwater assessment and protection of unconsolidated aquifers: A case example in the Abitibi-Témiscamingue region, Québec, Canada, Canadian water resources journal/Revue canadienne des ressources hydriques, 43, 113-135, 2018.

Narasimhan, H.: Learning with complex loss functions and constraints, International Conference on Artificial Intelligence and Statistics2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10, 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Ohlert, P. L., Bach, M., and Breuer, L.: Accuracy assessment of inverse distance weighting interpolation of groundwater nitrate concentrations in Bavaria (Germany), Environmental Science and Pollution Research, 30, 9445-9455, 2023.

Ohmer, M., Liesch, T., Habbel, B., Heudorfer, B., Gomez, M., Clos, P., Nölscher, M., and Broda, S.: GEMS-GER: a machine learning benchmark dataset of long-term groundwater levels in Germany with meteorological forcings and site-specific environmental features, Earth Syst. Sci. Data, 18, 77-95, 10.5194/essd-18-77-2026, 2026.

Olsson, F., Moore, T. N., Carey, C. C., Breef-Pilz, A., and Thomas, R. Q.: A Multi-Model Ensemble of Baseline and Process-Based Models Improves the Predictive Skill of Near-Term Lake Forecasts, Water Resour. Res., 60, e2023WR035901, https://doi.org/10.1029/2023WR035901, 2024.

Orozco López, E., Kaplan, D., and Linhoss, A.: Interpretable Transformer Neural Network Prediction of Diverse Environmental Time Series Using Weather Forecasts, Water Resour. Res., 60, e2023WR036337, https://doi.org/10.1029/2023WR036337, 2024.

Pang, M., Du, E., and Zheng, C.: Contaminant transport modeling and source attribution with attention-based graph neural network, Water Resour. Res., 60, e2023WR035278, 2024.

Rafik, A., Ait Brahim, Y., Amazirh, A., Ouarani, M., Bargam, B., Ouatiki, H., Bouslihim, Y., Bouchaou, L., and Chehbouni, A.: Groundwater level forecasting in a data-scarce region through remote sensing data downscaling, hydrological modeling, and machine learning: A case study from Morocco, Journal of Hydrology: Regional Studies, 50, 101569, https://doi.org/10.1016/j.ejrh.2023.101569, 2023.

Rathay, S. Y., Allen, D., and Kirste, D.: Response of a fractured bedrock aquifer to recharge from heavy rainfall events, J. Hydrol., 561, 1048-1062, 2018.

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., Karpatne, A., Hansen, G. J. A., Hanson, P. C., Watkins, W., Steinbach, M., and Kumar, V.: Process-Guided Deep Learning Predictions of Lake Water Temperature, Water Resour. Res., 55, 9173-9190, https://doi.org/10.1029/2019WR024922, 2019.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 10.1038/s41586-019-0912-1, 2019.

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., and Rodell, M.: Quantifying renewable groundwater stress with GRACE, Water Resour. Res., 51, 5217-5238, https://doi.org/10.1002/2015WR017349, 2015.

Riedmiller, M. and Lernen, A.: Multi layer perceptron, Machine learning lab special lecture, University of Freiburg, 24, 11-60, 2014.

Rinderer, M., McGlynn, B., and Van Meerveld, H.: Groundwater similarity across a watershed derived from time-warped and flow-corrected time series, Water Resour. Res., 53, 3921-3940, 2017.

Rinderer, M., Van Meerveld, H., and McGlynn, B.: From points to patterns: using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics, Water Resour. Res., 55, 5784-5806, 2019.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Scanlon, B. R., Ruddell, B. L., Reed, P. M., Hook, R. I., Zheng, C., Tidwell, V. C., and Siebert, S.: The food-energy-water nexus: Transforming science for society, Water Resour. Res., 53, 3550-3556, 2017.
Serrano, S. and Smith, N. A.: Is attention interpretable?, arXiv preprint arXiv:1906.03731, 2019.
Shamsudduha, M. and Taylor, R. G.: Groundwater storage dynamics in the world's large aquifer systems from GRACE: uncertainty and role of extreme precipitation, Earth System Dynamics, 11, 755-774, 2020.
Shelar, Y., Sharma, P., and Rawat, C. S. D.: Image forgery detection using integrated convolution-LSTM (2D) and convolution (2D), International Journal of Electrical and Electronics Research, 11, 631-638, 2023.
Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., and Fang, Z.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrology and Earth System Sciences, 22, 5639-5656, 2018.
Singh, U. and Sharma, P. K.: Seasonal uncertainty estimation of surface nuclear magnetic resonance water content using bootstrap statistics, Water Resources Management, 36, 2493-2508, 2022.
Soleimani Motlagh, M., Ghasemieh, H., Talebi, A., and Abdollahi, K.: Identification and analysis of drought propagation of groundwater during past and future periods, Water resources management, 31, 109-125, 2017.
Song, S., Xu, Y., Zhang, J., Li, G., and Wang, Y.: The long-term water level dynamics during urbanization in plain catchment in Yangtze River Delta, Agricultural Water Management, 174, 93-102, 2016.
Su, G., Wu, Y., Zhan, W., Zheng, Z., Chang, L., and Wang, J.: Spatiotemporal evolution characteristics of land subsidence caused by groundwater depletion in the North China plain during the past six decades, J. Hydrol., 600, 126678, 2021.
Sun, A. Y., Jiang, P., Mudunuru, M. K., and Chen, X.: Explore Spatio-Temporal Learning of Large Sample Hydrology Using Graph Neural Networks, Water Resour. Res., 57, e2021WR030394, https://doi.org/10.1029/2021WR030394, 2021.
Sun, J., Hu, L., Zhang, J., and Yin, W.: Providing enhanced insights into groundwater exchange patterns through downscaled grace data, Remote Sensing, 16, 812, 2024.
Sun, J., Bao, Y., Liu, T., Duan, L., Wang, Z., Guo, X., and Singh, V. P.: Study on the response mechanism of groundwater to rainfall in different geomorphic units of semi-arid regions, Journal of Hydrology: Regional Studies, 62, 102971, https://doi.org/10.1016/j.ejrh.2025.102971, 2025.
Taccari, M. L., Wang, H., Nuttall, J., Chen, X., and Jimack, P. K.: Spatial-temporal graph neural networks for groundwater data, Scientific Reports, 14, 24564, 2024.
Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., and Barajas-Solano, D.: Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems, Water Resour. Res., 56, e2019WR026731, 2020.
Tripathy, K. P. and Mishra, A. K.: Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions, J. Hydrol., 628, 130458, https://doi.org/10.1016/j.jhydrol.2023.130458, 2024.
Varouchakis, E. A., Theodoridou, P. G., and Karatzas, G. P.: Spatiotemporal geostatistical modeling of groundwater levels under a Bayesian framework using means of physical background, J. Hydrol., 575, 487-498, https://doi.org/10.1016/j.jhydrol.2019.05.055, 2019.
Vu, M., Jardani, A., Massei, N., and Fournier, M.: Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network, J. Hydrol., 597, 125776, 2021.
Wada, Y., van Beek, L. P., and Bierkens, M. F.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48, 2012.
Wang, C., Jiang, S., Zheng, Y., Han, F., Kumar, R., Rakovec, O., and Li, S.: Distributed hydrological modeling with physics-encoded deep learning: A general framework and its application in the Amazon, Water Resour. Res., 60, e2023WR036170, 2024a.
Wang, H., Wang, K., Zhou, Y., and Bai, X.: Effects of human disturbance on the landscape pattern in the mountainous area of western Henan, China, J. Environ. Manage., 380, 124959, 2025.
Wang, S., Song, X., Wang, Q., Xiao, G., Liu, C., and Liu, J.: Shallow groundwater dynamics in North China plain, Journal of Geographical Sciences, 19, 175-188, 2009.
Wang, W., Zhang, Z., Yin, L., Duan, L., and Huang, J.: Topical Collection: Groundwater recharge and discharge in arid and semi-arid areas of China, Hydrogeology Journal, 29, 521-524, 10.1007/s10040-021-02308-0, 2021a.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Wang, Y.-C., Chiang, C.-H., Su, C.-J., Wang, K.-C., Tseng, W.-L., Chen, C.-T., and Liang, H.-C.: Using Multi-Head Attention Deep Neural Network for Bias Correction and Downscaling for Daily Rainfall Pattern of a Subtropical Island, EGUsphere, 2024, 1-35, 2024b.
Wang, Y., Chen, Z., Chen, J., and Wei, W.: Effects of ultrafiltration on salt migration and isotopic composition of pore water in groundwater depression area in Hengshui, NCP, Hydrological Processes, 35, e14201, 2021b.
Wei, W., Feng, X., Yang, L., Chen, L., Feng, T., and Chen, D.: The effects of terracing and vegetation on soil moisture retention in a dry hilly catchment in China, Sci. Total Environ., 647, 1323-1332, https://doi.org/10.1016/j.scitotenv.2018.08.037, 2019.
Welch, L. and Allen, D.: Consistency of groundwater flow patterns in mountainous topography: Implications for valley bottom water replenishment and for defining groundwater flow boundaries, Water Resour. Res., 48, 2012.
Wittenberg, H., Aksoy, H., and Miegel, K.: Fast response of groundwater to heavy rainfall, J. Hydrol., 571, 837-842, 2019.
Wu, D., Fang, S., Li, X., He, D., Zhu, Y., Yang, Z., Xu, J., and Wu, Y.: Spatial-temporal variation in irrigation water requirement for the winter wheat-summer maize rotation system since the 1980s on the North China Plain, Agricultural water management, 214, 78-86, 2019.
Xiong, G., Chen, G., Wu, J., Fu, T., Yang, Y., Xu, X., Zhu, X., Yu, H., Liu, S., and Gao, M.: Seawater intrusion-retreat processes and groundwater evolution in intruded coastal aquifers with land reclamation: A case study of Eastern Jiangsu, China, Lithosphere, 2021, 1308487, 2022.
Xu, J., Mo, Y., Jiang, Q., Kong, L., Wu, J., Ding, Z., Jin, G., and Li, L.: Rainfall-dependent influence of water parameter interactions and land use on lake water quality: A hybrid ensemble approach and management implications, J. Hydrol., 662, 134019, https://doi.org/10.1016/j.jhydrol.2025.134019, 2025.
Yang, C., Li, H.-Y., Fang, Y., Cui, C., Wang, T., Zheng, C., Leung, L. R., Maxwell, R. M., Zhang, Y.-K., and Yang, X.: Effects of Groundwater Pumping on Ground Surface Temperature: A Regional Modeling Study in the North China Plain, Journal of Geophysical Research: Atmospheres, 125, e2019JD031764, https://doi.org/10.1029/2019JD031764, 2020.
Yang, G. and McCoy, K.: Modeling groundwater-level responses to multiple stresses using transfer-function models and wavelet analysis in a coastal aquifer system, J. Hydrol., 627, 130426, https://doi.org/10.1016/j.jhydrol.2023.130426, 2023.
Yang, Q., Zhang, X., Hu, J., Gui, R., and Yang, L.: Estimation of Land Deformation and Groundwater Storage Dynamics in Shijiazhuang–Baoding–Cangzhou–Hengshui Using Multi-Temporal Interferometric Synthetic Aperture Radar, Remote Sensing, 16, 1724, 2024.
Yao, Y., Guo, S., Andrews, C. B., Zhang, F., Lancia, M., Kuang, X., and Zheng, C.: Seeing China's Invisible Groundwater: Advances and Challenges, Water Resour. Res., 61, e2024WR038980, https://doi.org/10.1029/2024WR038980, 2025.
Yaraghi, N., Ronkanen, A.-K., Darabi, H., Kløve, B., and Haghighi, A. T.: Impact of managed aquifer recharge structure on river flow regimes in arid and semi-arid climates, Sci. Total Environ., 675, 429-438, 2019.
Yi, S., Kondolf, G. M., Solis, S. S., and Dale, L.: Groundwater Level Forecasting Using Machine Learning: A Case Study of the Baekje Weir in Four Major Rivers Project, South Korea, Water Resources Research, 60, 10.1029/2022wr032779, 2024.
Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, Advances in neural information processing systems, 27, 2014.
Yu, X., Xin, P., Lu, C., Robinson, C., Li, L., and Barry, D.: Effects of episodic rainfall on a subterranean estuary, Water Resour. Res., 53, 5774-5787, 2017.
Yuan, R., Song, X., Han, D., Zhang, L., and Wang, S.: Upward recharge through groundwater depression cone in piedmont plain of North China Plain, J. Hydrol., 500, 1-11, 2013.
Zeydalinejad, N., Javadi, A. A., and Webber, J. L.: Global perspectives on groundwater infiltration to sewer networks: A threat to urban sustainability, Water Res., 262, 122098, https://doi.org/10.1016/j.watres.2024.122098, 2024.
Zhang, H., Ding, J., Wang, Y., Zhou, D., and Zhu, Q.: Investigation about the correlation and propagation among meteorological, agricultural and groundwater droughts over humid and arid/semi-arid basins in China, J. Hydrol., 603, 127007, 2021.
Zhang, K., Xie, X., Zhu, B., Meng, S., and Yao, Y.: Unexpected groundwater recovery with decreasing agricultural irrigation in the Yellow River Basin, Agricultural Water Management, 213, 858-867, 2019a.
Zhang, M., Luo, G., Cao, X., Hamdi, R., Li, T., Cai, P., Ye, H., and He, H.: Numerical Simulation of the Irrigation Effects on Surface Fluxes and Local Climate in Typical Mountain-Oasis-Desert Systems in the Central Asia Arid Area, Journal of Geophysical Research: Atmospheres, 124, 12485-12506, https://doi.org/10.1029/2019JD030507, 2019b.

https://doi.org/10.5194/egusphere-2026-2379
Preprint. Discussion started: 2 June 2026

Zhang, P., Duan, Q., Dong, J., Piao, L., and Cui, Z.: Ecological Importance Evaluation and Ecological Function Zoning of Yanshan-Taihang Mountain Area of Hebei Province, Sustainability, 16, 10233, 2024.
Zhang, X., Dong, F., Chen, G., and Dai, Z.: Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks, Hydrology and Earth System Sciences, 27, 83-96, 10.5194/hess-27-83-2023, 2023a.
Zhang, X., Gao, S., Wu, Q., Li, F., Wu, P., Wang, Z., Wu, J., and Zeng, J.: Buffer zone-based trace elements indicating the impact of human activities on karst urban groundwater, Environ. Res., 220, 115235, 2023b.
Zhao, A., Xiang, K., Zhang, A., and Zhang, X.: Spatial-temporal evolution of meteorological and groundwater droughts and their relationship in the North China Plain, J. Hydrol., 610, 127903, 2022.
Zhao, Y., Yang, L., Pan, H., Li, Y., Shao, Y., Li, J., and Xie, X.: Spatio-temporal prediction of groundwater vulnerability based on CNN-LSTM model with self-attention mechanism: A case study in Hetao Plain, northern China, J. Environ. Sci., 153, 128-142, 2025.
Zheng, Y., Yu, C., Zhou, H., and Xiao, J.: Spatial variations and influencing factors of river networks in river basins of China, International Journal of Environmental Research and Public Health, 18, 11910, 2021.
Zhong, L., Lei, H., and Gao, B.: Developing a physics-informed deep learning model to simulate runoff response to climate change in alpine catchments, Water Resour. Res., 59, e2022WR034118, 2023.
Zhou, R.: Multi-scale dynamic spatiotemporal graph attention network for forecasting karst spring discharge, J. Hydrol., 659, 133289, 2025.
Zhu, L., Gong, H., Chen, Y., Wang, S., Ke, Y., Guo, G., Li, X., Chen, B., Wang, H., and Teatini, P.: Effects of Water Diversion Project on groundwater system and land subsidence in Beijing, China, Engineering Geology, 276, 105763, https://doi.org/10.1016/j.enggeo.2020.105763, 2020.
Zou, Y., Wang, J., Lei, P., and Li, Y.: A novel multi-step ahead forecasting model for flood based on time residual LSTM, J. Hydrol., 620, 129521, 2023.