Environmental Science and Ecotechnology 28 (2025) 100632

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

journal homepage: www.journals.elsevier.com/environmental-science-and-ecotechnology/

Original Research

A hierarchical transformer and graph neural network model for high-accuracy watershed nitrate prediction

Jun Sun [a,b,c], Xuesong Gao [a,b,*], Zhiyong Deng [a,b], Yudong Zhao [a,b], Qi Wang [a], Xiyi Zhao [d], Xu Liu [e]

[a] College of Resources, Sichuan Agricultural University, Chengdu, 611130, China
[b] Key Laboratory of Investigation, Monitoring, Protection and Utilization for Cultivated Land Resources, Ministry of Natural Resources, Chengdu, 610045, China
[c] Key Laboratory of Mountain Surface Processes and Ecological Regulation, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, 610299, China
[d] Sichuan Institute of Land Science and Technology (Sichuan Center of Satellite Application Technology), Chengdu, 610041, China
[e] Suzhou Suji Fudi Environmental Technology Co., Ltd., Suzhou, 215000, China

ARTICLE INFO

Article history:
Received 8 April 2025
Received in revised form
24 October 2025
Accepted 24 October 2025

Keywords:
Non-point source pollution
Deep learning
Graph neural network
Hierarchical transformer
Watersheds

ABSTRACT

Non-point source pollution from agricultural activities poses a significant threat to water quality by introducing excess nutrients like nitrogen into aquatic ecosystems, leading to issues such as eutrophication and groundwater contamination. In agricultural watersheds, nitrate transport involves intricate physical, chemical, and biological processes influenced by meteorological conditions, hydrological features, and spatial topologies, making accurate short-term predictions challenging. Traditional data-driven deep learning models often fail to incorporate physical constraints and complex spatiotemporal dynamics, limiting their interpretability and predictive accuracy. Here we show a hierarchical transformer and graph neural network model that accurately predicts watershed nitrate concentrations by integrating multi-source data and simulating pollutant migration. The model captures nonlinear multivariate temporal patterns through hierarchical transformers, fuses global meteorological and local hydrological features via neural networks, and models runoff topologies with physically constrained graph neural networks. For predicting the concentration changes of pollutants discharged from watersheds, it outperforms baselines like multi-layer perceptrons, recurrent neural networks, and long short-term memory networks, with state-of-the-art performance in root mean square error, mean absolute error, and R^2. Ablation studies confirm the essential roles of multi-source data integration and watershed topological modeling in enhancing performance. This method of directly modeling physical processes by leveraging the characteristics of different neural network architectures opens up a new path for addressing the interpretability problem in neural earth system modeling, apart from the process-guided deep learning and differentiable modelling methods.

1. Introduction

The rapid and intensive progression of agricultural activities has led to an influx of nutrients into natural water bodies, significantly threatening water quality. This threat commonly materializes as a non-point source (NPS) of pollution [1–3], and it is one of the foremost environmental concerns driving the global decline in water quality. NPS pollution occurs when a diverse array of pollutants that originate from agricultural undertakings, with nitrogen and phosphorus being particularly prominent, gradually shift from land-based environments to water-based environments [4,5]. Nitrogen, an integral part of NPS pollution, is currently one of the most prevalent agricultural pollutants. NPS pollution originates from a broad spectrum of sources but is closely tied to human agricultural practices and is subject to the influence of numerous natural elements, leading to high intricacy, unpredictability, and variability [6–9]. The NPS pollution emanating from

* Corresponding author. College of Resources, Sichuan Agricultural University, Chengdu, 611130, China.
E-mail address: xuesonggao@sicau.edu.cn (X. Gao).

https://doi.org/10.1016/j.ese.2025.100632

agricultural fields, especially in nitrogenous form, gives rise to nitrogen loss within watersheds. This leads to soil acidification, the nitrate contamination of groundwater, atmospheric acid deposition, and eutrophication [10,11]. Consequently, precisely modeling the spread of NPS pollution and predicting its concentration levels are crucial for effective prevention and control.

In light of the challenges related to NPS pollution, efficacious strategies for its simulation and prediction merit attention. Therefore, significant strides have been made over the past few decades in the development of complex system modeling methodologies aimed at scrutinizing hydrological material cycles on a watershed scale, namely Earth system models (ESMs) [12]. These methods utilize mathematical equations to delineate the material migration and transformation processes within a watershed, with a particular focus on the intricacies of NPS pollution. By constructing process-based models [13–15] and amalgamating diverse subsystems into cohesive numerical frameworks, ESMs enable the simulation of a wide array of physical, chemical, and biological processes and their mutual interactions within the Earth system. This comprehensive approach has not only proven indispensable in the realm of Earth system science but has also substantially augmented our comprehension and prediction of how NPS pollution influences water quality dynamics. The ability of process-based models to encapsulate various influencing processes, such as water flow and biogeochemical cycles, accentuates their value in surmounting such environmental challenges. Nevertheless, the Earth system is extraordinarily complex and harbors a plethora of unknown processes. Researchers lack a thorough understanding of the underlying mechanisms of these processes, which is why the models developed fall short of accurately and comprehensively representing them. Further, procuring detailed data on the properties of diverse processes, including above- and below-ground dynamics, water flow, and biogeochemical processes, is no easy feat, and this curtails modeling capabilities over extended time periods. The limitations of process-based modeling become glaringly evident with nonlinear processes, especially in the context of high-dimensional interactions among multiple systems [16], which impede the predictive accuracy for abrupt changes.

In recent years, novel artificial intelligence (AI) tools have been extensively employed in Earth system modeling, enhancing traditional methodologies and leading to remarkable achievements [17,18]. AI techniques, usually also called machine learning or deep learning methodologies, built upon neural networks, have spurred the evolution of the neural Earth system model (NESM) [16], which couple AI with Earth system models. However, the limited interpretability and credibility of AI algorithms pose significant challenges to Earth system models, primarily stemming from the so-called “black box” quandary—the unknowability of the internal computation processes of AI models. Interpretable artificial intelligence (IAI) offers a promising pathway to mitigate these limitations, yet realizing its full potential will require overcoming substantial challenges through sustained research efforts.

Understanding the complex interactions between the various environmental factors in a watershed is essential for the effective prediction and management of water quality. Traditional models often struggle to capture the complex, nonlinear dependencies and spatial interactions inherent to these systems. Thus, there is a need for advanced methodologies that can integrate both global and local watershed features, providing a more comprehensive understanding of water quality dynamics. In this study, we developed a novel method called HTGNN-WNP—the hierarchical transformer and graph neural network model for watershed nitrate prediction—that addresses these modeling requirements by leveraging advanced machine learning techniques.

Our method adhered to the physical rules governing watershed runoff in the design phase of the NESM. We collected multi-year observational time series data on meteorology and pollution in the study area from official databases. AI algorithms were employed to mine hidden patterns of interactions among various variables from the data, which were then used to predict nitrate concentration levels in watershed runoff. This task required us to overcome three significant challenges: (i) extracting effective features and patterns from nonlinear, multivariate time series data, (ii) assimilating macro meteorological observation data with local runoff physical parameters within the model framework, and (iii) accurately modeling the complex topological relationships associated with runoff in watersheds.

We used the hierarchical transformer method to capture the interactive coupling relationships between the time series of watershed meteorological features and feature dimensions, which enabled us to extract complex patterns. We encoded and integrated global (meteorological features covering an entire small watershed) and local (hydrological features of a single sub-watershed) features through neural networks to obtain a comprehensive modeling perspective. Notably, we employed an innovative approach that involved the use of graph neural networks (GNNs) to model the topological relationships of runoff. This particular approach allowed us to mimic the physical translocation of pollutants within watersheds and capture spatial interactions beyond the limitations of traditional time series methods. We applied this proposed method to predict nitrate concentrations at the outlet of an agricultural watershed in northern Texas, the United States, thus obtaining distinctive perspectives and valuable insights into water quality prediction.

The key contributions of our study are as follows.

- The hierarchical transformer method was used to model the interactive coupling relationships between the time series and feature dimensions of watershed meteorological features, significantly enhancing the extraction of complex patterns from multivariate data.
- Our innovative watershed hydrological modeling method involved integrating macro meteorological data with specific local hydrological features of sub-watersheds. This facilitated a robust and coherent fusion of diverse data sources for improved watershed system modeling.
- To our knowledge, this was the first study to involve the integration of GNNs with physically driven message passing rules to model runoff topological relationships, enabling artificial neural networks to simulate the pollutant migration processes in watersheds.
- Our deep learning framework, HTGNN-WNP, demonstrated exceptional performance in watershed water quality prediction, as evidenced by experiments and tests conducted in the study area.

2. Materials and methods

2.1. Data collection and processing techniques

2.1.1. Nitrate monitoring data acquisition and processing

This research leveraged published monitoring data on nitrate concentrations in surface water from the national ecological observatory networks [19]. We selected observational data from a region in northern Texas, USA (33°20′–33°24′ N and 97°47′–97°53′ W), characterized by its subtropical humid climate with mean annual precipitation of approximately 890 mm (Fig. 1). The watershed covers about 40 km^2 and supports agricultural

activities, with the land primarily divided between ecological reserves and farmland. Nitrate (NO_3^--N) concentrations were measured downstream of the watershed using a SeaBird SUNA underwater ultraviolet nitrate analyzer at 15-min intervals, with results reported in μmol L^{-1}. We then calculated the daily mean concentrations of nitrate at the watershed outlet, covering the period from October 2017 to April 2024.

2.1.2. Hydrological data and spatial analysis

Hydrological data were obtained from the HydroSHEDS dataset [20], a high-resolution database provided by the United States Geological Survey and the World Wildlife Fund. We used the QGIS software to process high-resolution imagery sourced from Google Maps. This facilitated the detailed extraction of land cover types and spatial patterns and enabled us to accurately identify the physical characteristics of the runoff, such as the length, width, and head drop (Fig. 2a).

2.1.3. Runoff sinuosity calculation

Acknowledging sinuosity as a key hydrological indicator due to its capacity to moderate water flow and facilitate nitrogen transformation [21], we prioritized its accurate quantification. We used a vectorization process for each runoff path; curvatures were calculated at select points and subsequently averaged, which resulted in a mean curvature that succinctly represented the runoff's sinuosity.

2.1.4. Topological relationship mapping

To clarify the topological relationships within small watershed runoffs, binary images of the runoff patterns were prepared using QGIS. The RivGraph algorithm [22] was applied thereafter to draft comprehensive topological maps, which enabled the clear encoding of nodes and their connections (Fig. 2b). This technical approach ensured precise runoff connectivity mapping across the watershed.

2.1.5. Meteorological data processing

We collected meteorological data from the National Weather Service's (NWS) records for three adjacent airport weather stations. By calculating the arithmetic mean of observational data from three meteorological stations, we approximated the site-specific meteorological conditions accurately. For data processing, wavelet transforms were used to decompose the time series into components that reflected low-frequency long-term trends and abrupt high-frequency changes to enhance the model's ability to capture subtle and rapid weather alterations.

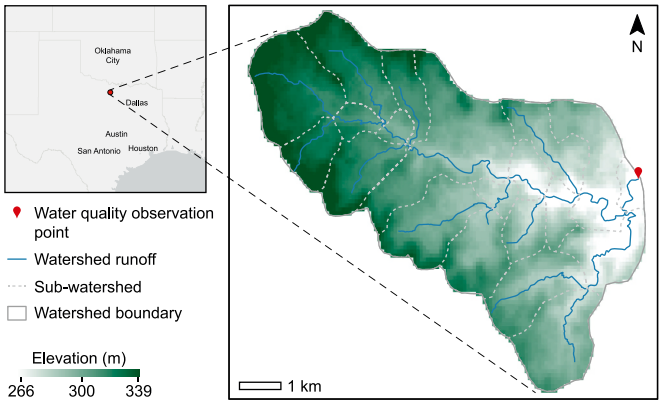

Fig. 1. Overview of the study area.

2.2. Overall Framework

We developed a comprehensive framework using transformer architectures and GNNs to model complex interactions across temporal and spatial dimensions (Fig. 3).

First, the collected watershed data was preprocessed and divided into a training set and a test set (Fig. 3a). We used a hierarchical transformer with dual transformer [23] modules to capture complex time series and intervariable patterns (Fig. 3b). We adopted feature compression methods to distill these features into global feature vectors that could reflect the impacts of meteorological conditions and historical pollution concentrations on the water quality of the watershed.

Neural networks were then employed to encode each sub-watershed and capture its distinct local features (Fig. 3c). The variables used for feature encoding included hydrological and land use metrics, spatial patterns, and other physical indicators. Following this, we broadcasted the watershed's global features to each sub-watershed's features to facilitate modeling from a global perspective (i.e., factors influencing the entire watershed) down to a local perspective (i.e., factors specific to individual sub-watersheds).

We utilized GNNs [24,25] for the spatial modeling of each sub-watershed (Fig. 3d). The aggregation and transmission rules of the graph adhered to the topological relationships of the runoff within the watershed and thus simulated the migration process of pollutants through the runoff network. This approach firmly anchored the model's feature transmission in the physical laws governing real-world environments, instead of relying purely on data-driven methods. Unlike conventional time series modeling [26–28], this graph-based approach could capture the spatial interactions of time series within the inherently complex, non-Euclidean watershed structures informed by runoff topologies. In the final phase of model construction, graph pooling was used to aggregate the features of each sub-watershed in the graph, and this information was integrated via a fully connected layer to predict nitrate concentration in the water at the watershed's outlet (Fig. 3e).

2.3. Watershed system modeling

2.3.1. Definition and notation

In this section, we introduce the key definitions and notations used throughout the paper. We first define the global time series within the watershed, including both univariate and multivariate time series and their respective notations. These provide the basis for understanding the overall characteristics of the time series data. We then specify the hydrological variables of the runoff within the watershed and their associated notations, which capture region-specific hydrological characteristics. Finally, we describe the watershed attribute graph and its notational definitions, representing the watershed's topological and structural properties.

Definition 1. Global variable time series. In watershed-scale neural geoscience modeling, global variable time series data cover observations across an entire watershed, such as daily temperature, humidity, and precipitation. A univariate time series is a sequence of scalar observations over time, denoted as $\boldsymbol{x} = \{x_1, x_2, \ldots, x_t\} \in \mathbb{R}^t$, and t represents the length of the time series. For our task, the model's input global time series data involved multiple variables over time, leading to the concept of a multivariate time series. This set of n-dimensional vector observations over time can be represented as $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^{n \times t}$, where n represents the number of variables.

Definition 2. Parameterization of hydrology. The hydrological

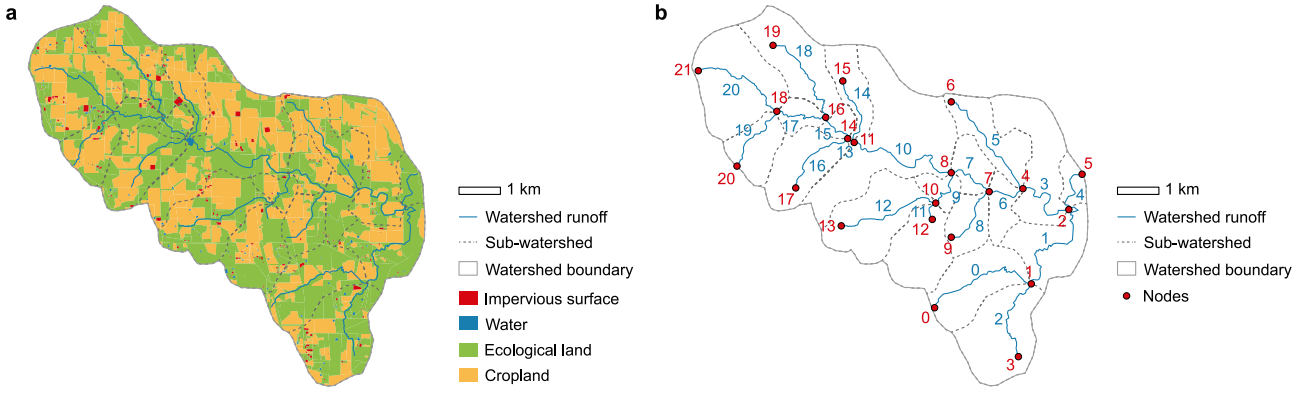

Fig. 2. **a**, Land use in the study area. **b**, Illustration of the topological relationships of runoff within a watershed. Red numbers represent the number of runoff nodes, and blue numbers represent the number of watershed runoff.

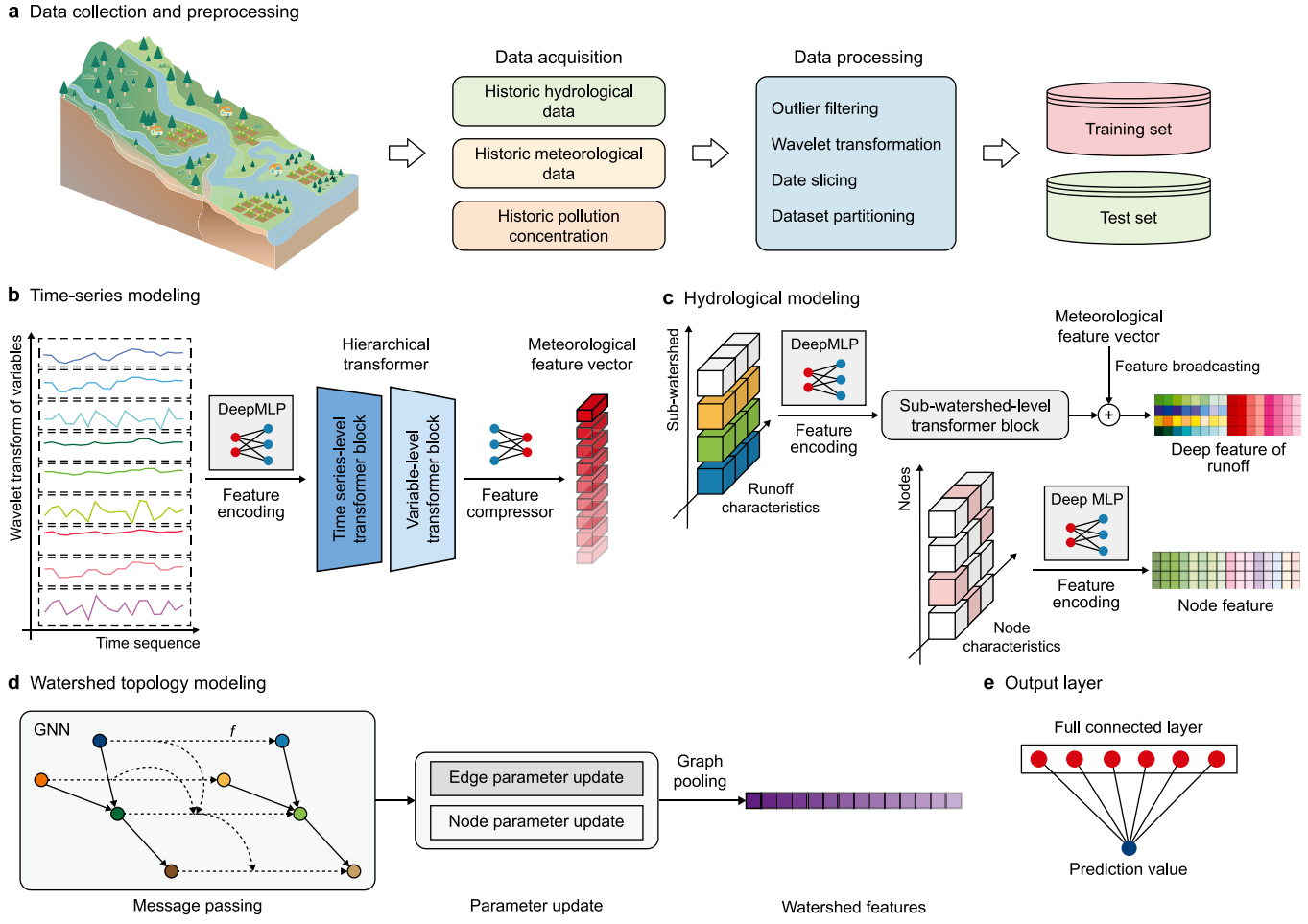

Fig. 3. Framework of the proposed method. **a**, Data collection and preprocessing. **b**, Hierarchical transformer for time-series modeling. **c**, Watershed hydrological modeling. **d**, Watershed topology modeling. **e**, Output layer. DeepMLP: deep multi-layer perceptron. GNN: graph neural networks.

parameters for each sub-watershed in the watershed were defined as $\boldsymbol{h} = \{h_1, h_2, \ldots, h_{12}\} \in \mathbb{R}^{12}$, where the dimension 12 represents the number of parameters. In our study, these 12 hydrological parameters specifically encompassed 12 indicators: length, width, sinuosity, runoff drop, and the area and proportion of the impervious surfaces, water bodies, ecological land, and arable land within each sub-watershed. The hydrological parameterization of the whole watershed was defined as $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_M\} \in \mathbb{R}^{M\times 12}$, where M represents the number of sub-watersheds.

Definition 3. ***Topological representation of the watershed.*** The topological relationships of surface runoff within a watershed are represented by flow paths and their connections. These are modeled using a directed graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V} = \{v_1, v_2, v_i, \ldots, v_j, v_N\} \in \mathbb{R}^{N\times 3}$ denotes the set of nodes (entrances, exits, and confluences of runoff) and $\boldsymbol{E} = \{e_1, e_2, \ldots, e_M\}$ denotes the set of

edges (runoff pathways). Here, N represents the number of nodes. M represents the number of edges, which is consistent with the number of sub-watersheds. The adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is defined as follows: for nodes $v_i, v_j \in \boldsymbol{V}$,

$$\boldsymbol{A}_{ij} = \begin{cases} 1, & \text{if there exists an edge } \left(v_i, v_j\right) \in \boldsymbol{E}, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Here, $\boldsymbol{A}_{ij}$ denotes the (i,j) th entry of $\boldsymbol{A}$, indicating whether runoff flows from node v_i to node v_j.

2.3.2. Hierarchical transformer for time series modeling

Meteorological factors, such as temperature, precipitation, wind speed, and humidity, have both direct and indirect effects on the nitrogen cycle process that tend to accumulate over time and manifest accordingly. These variables do not operate in isolation. Instead, they interact with each other and collectively influence the changes in water quality within a watershed.

To capture these complex patterns, we employed a hierarchical transformer that consisted of two transformer blocks: one for managing dependencies in the temporal domain and the other for addressing dependencies in the variable dimension (Fig. 4).

The raw data were enhanced using a wavelet transformation, which resulted in the original input data $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^{n\times t}$. This was then feature-encoded using a deep multi-layer perceptron (DeepMLP), which extended the features from low dimensions t to high dimensions $d/2$, resulting in $f_X \in \mathbb{R}^{n\times d/2}$ (equation (2)).

$$f_X^{B,n,d/2} = \text{Dropout}\left(\text{GELU}\left(\text{BN}\left(\text{Linear}_i\left(\boldsymbol{X}^{B,n,t}\right)\right)\right)\right) \quad (2)$$

Here, B denotes the batch size of the data processed each time, i.e., the number of samples processed simultaneously during each training step. The batch size balances computational efficiency with gradient update stability, thereby influencing both training speed and model generalization. $\text{Linear}_i(\cdot)$ indicates a linear transformation applied at layer i; $\text{BN}(\cdot)$ refers to the batch normalization applied after each linear transformation; and $\text{GELU}(\cdot)$ is the Gaussian error linear unit activation function applied after batch normalization. $\text{Dropout}(\cdot)$ denotes dropout with probability 0.5 applied after the GELU activation in all layers, which was used to mask some neurons in the neural network to prevent overfitting. $\text{Linear}_i(\cdot)$ was essentially a linear neural network layer that was combined with $\text{BN}(\cdot)$, $\text{GELU}(\cdot)$, and $\text{Dropout}(\cdot)$, to form a DeepMLP set. $\text{Linear}_i(\cdot)$ was defined by a weight matrix $W \in \mathbb{R}^{t\times(d/2)}$ that mapped the input $\boldsymbol{X}^{B,n,t}$ to $\mathbb{R}^{B,n,d/2}$. This transformation projects the original input sequence from the time dimension t to the high-dimensional feature space $d/2$, enabling the transformer model to capture the feature dependencies of water quality in the time dimension. In general, when the input feature dimension is low, the attention algorithm of the transformer may yield nearly uniform attention weights across time steps (i.e., minimal differences in weights), due to the limited numerical dimension. After high-dimensional expansion, the attention calculation results are more dispersed, allowing the model to assign weights that more accurately reflect the intensity of the temporal dependencies. For example, with an input time series length $t = 10$ (e.g., 10 days of meteorological data) and high-dimensional feature dimension $d = 256$, the DeepMLP would map the input $\boldsymbol{X}^{B,n,10}$ to $f_X^{B,n,128}$. For the purpose of our study, the dimension was set at $d/2$, as the remaining half was to be used for concatenating the features from the sub-watersheds in the subsequent processing stage.

We introduced a hierarchical transformer model to effectively capture patterns in both the temporal domain and variable dimensions within $f_X \in \mathbb{R}^{B,n,d/2}$. The transformer [23] included a self-attention mechanism that enabled each element in the input sequence to dynamically attend to and integrate information from all other elements, thereby capturing long-range dependencies and contextual relationships. To compensate for the inherent lack of sequential order information in self-attention, explicit positional encodings were injected into the input embeddings, which allowed the model to effectively leverage the sequence structure. The input variable $f_X \in \mathbb{R}^{B,n,d/2}$ integrated n meteorological variables (variable dimension) and $d/2$-dimensional temporal features. To capture positional dependencies in both dimensions, we independently apply positional encoding to the meteorological variable and temporal feature dimensions (Fig. 4). Subsequently, we used transformer-based encoders for hierarchical feature extraction, allowing the model to learn the complex relationships embedded in the meteorological variables and high-dimensional temporal dynamics.

In the first level of the hierarchical transformer model, positional encoding was applied to the temporal dimension $d/2$ as follows: $f_X^{B,d/2,n} = \text{PE}(f_X^{B,n,d/2}, \text{permute} = (0, 2, 1); \text{dim} = d/2)$. Specifically, as a fundamental component of the transformer architecture, the $\text{PE}(\cdot)$ function first transposed the original input tensor into the shape of $(B, d/2, n)$, and then injected positional encoding into the dimension $d/2$ to construct sequential information. This served the transformer block in capturing the sequential dependencies in the $d/2$ dimension. The encoded tensor was then inputted into the first transformer encoder block T^t. In the second level, $\text{PE}(\cdot)$ was applied to the variable dimension n after transposing the output tensor from the first level to $(B, d/2, n)$: $f_X^{B,n,d/2} = \text{PE}(f_X^{B,d/2,n}, \text{permute} = (0, 2, 1); \text{dim} = n)$. The encoded tensor was then input into the second transformer encoder block T^n. The hierarchical transformer model was implemented using the two-level operations given above, in line with the following equations:

$$f_X^{B,d/2,n} = T^t\left(f_X^{B,d/2,n}\right) \quad (3)$$

$$f_X^{'B,n,d/2} = T^n\left(f_X^{B,n,d/2}\right) \quad (4)$$

After the hierarchical transformer, we employed a feature compression method to compress a tensor of shape $(B, n, d/2)$ into a feature vector f_g of shape $(B, 1, d/2)$ and thereby obtain a meteorological feature vector that covered the entire watershed. Specifically, the feature compression was implemented via a fully connected (FC) layer-based approach. A linear transformation with a weight matrix $W \in \mathbb{R}^{n\times 1}$ was applied to the variable dimension n, which projected the n-dimensional meteorological variables onto a single comprehensive feature. This process was mathematically formulated as follows: $f_g^{B,1,d/2} = \text{FC}(f_X^{'B,n,d/2}) = f_X^{'B,n,d/2} \times W^{n\times 1}$; the resulting $f_g^{B,1,d/2}$ integrated the spatial dependencies across all meteorological variables (e.g., precipitation, temperature, humidity) into a unified representation. The fully connected compression enabled the model to learn the adaptive weights for different variables and emphasize those with stronger contributions to watershed nitrate concentration dynamics. This mechanism reduced the feature dimension while preserving the critical temporal patterns in the $d/2$ dimension and also generated a watershed scale meteorological feature vector.

2.3.3. Watershed hydrological modeling

In this section, we detail our hydrological modeling approach

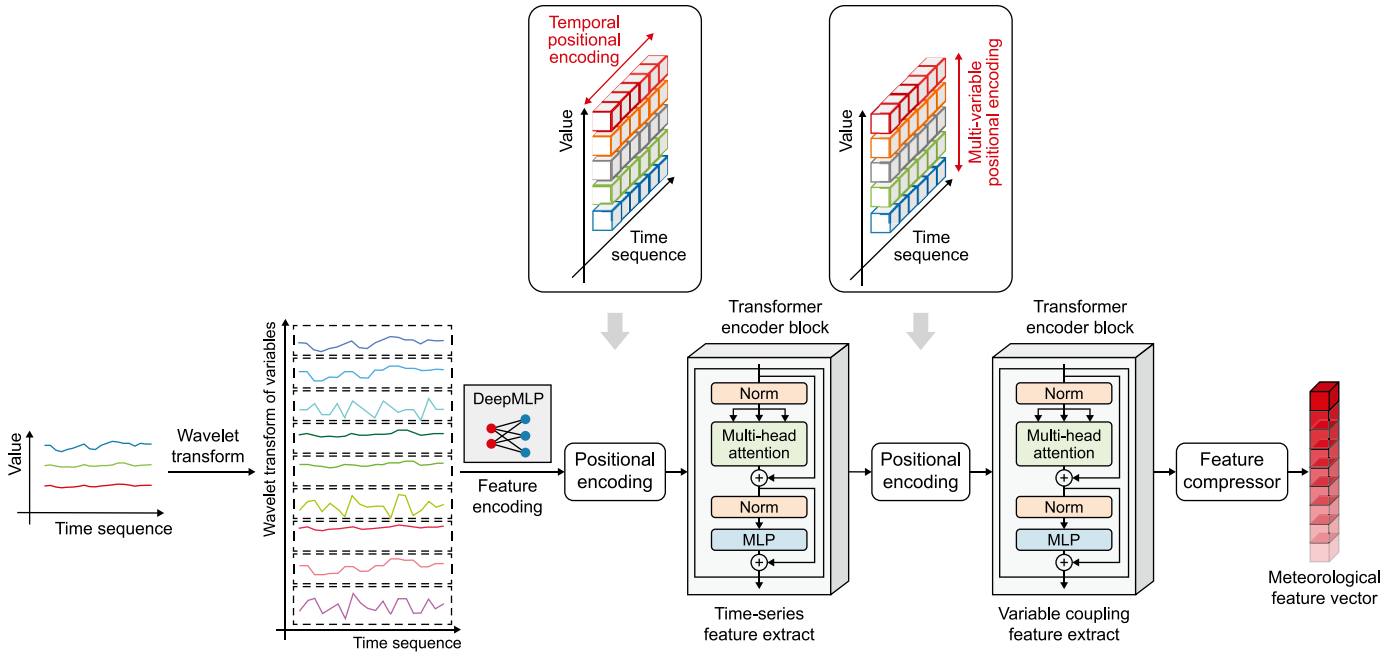

Fig. 4. Hierarchical transformer module. This module adopts a two-layer transformer structure, where one layer is used to model dependencies in the temporal domain and the other to address dependencies in the variable dimension. DeepMLP: deep multi-layer perceptron; Norm: layer normalization.

based on deep neural networks in three parts: sub-watershed hydrological feature extraction, global feature broadcasting, and node feature extraction. Through these steps, we achieved the fine modeling of hydrological data to capture the watershed's hydrological characteristics. The process unfolded as follows (Fig. 5):

Sub-watershed hydrological modeling and feature extraction. We extracted the sub-watershed features from the hydrological observational data $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_M\} \in \mathbb{R}^{M\times 12}$. Sub-watershed features were initially extracted from the data using the DeepMLP (equation (5)).

$$f_H^{B,M,d/2} = \mathrm{Dropout}\left(\mathrm{GELU}\left(\mathrm{BN}\left(\mathrm{Linear}\left(\boldsymbol{H}^{B,M,12}\right)\right)\right)\right) \tag{5}$$

The watershed observational data $\boldsymbol{H}^{B,M,12}$ with batch size B served as the input for Linear$(\cdot)$—a linear transformation defined by the weight matrix $W \in \mathbb{R}^{m\times(d/2)}$, which mapped the input $\boldsymbol{H}^{B,M,12}$ to $\mathbb{R}^{B,M,d/2}$. The parameter B was consistent with that in equation (2). Based on Linear$(\cdot)$, BN$(\cdot)$, GELU$(\cdot)$, and Dropout$(\cdot)$ were combined to form a DeepMLP, and a feature matrix of the sub-watershed hydrology was extracted. Specifically, high-dimensional hydrological features $f_H \in \mathbb{R}^{B,M,d/2}$ of each sub-watershed were extracted from the input hydrological data. This process was used to model the interrelationships among various hydrological characteristics, such as the length and width of the runoff or the cultivated land area and its proportion in a sub-watershed.

Subsequently, positional encoding was applied to dimension M of the sub-watersheds and expressed as follows: $f_H^{B,M,d/2} = \mathrm{PE}(f_H^{B,M,d/2}, \mathrm{dim} = M)$. A transformer encoder block T^M was then utilized to perform sequential modeling of the high-dimensional hydrological features, with the aim of capturing the potentially complex interactions between runoffs in the various sub-watersheds and to further refine the features (equation (6); Fig. 5a).

$$f_H^{'B,M,d/2} = T^M\left(f_H^{B,M,d/2}\right) \tag{6}$$

Global feature broadcasting. As detailed in Section 2.3.2, the global meteorological features f_g exerted their influence on the entire watershed. To conduct a thorough analysis, these features were integrated with the hydrological characteristics specific to the sub-watersheds. We utilized a feature broadcasting method, to effectively merge the meteorological features $f_g \in \mathbb{R}^{B,1,d/2}$ and the sub-watershed hydrological features $f'_H \in \mathbb{R}^{B,M,d/2}$. This involved an initial expansion phase, during which f_g underwent expansion to align its dimensions with those of the sub-watershed features, resulting in $f'_g \in \mathbb{R}^{B,M,d/2}$ through M repetitions (equation (7)). This ensured dimensional consistency between the global meteorological features and the sub-watershed hydrological features. Then, in the concatenation phase, the expanded meteorological features were combined with the hydrological features to form $\boldsymbol{\varepsilon} \in \mathbb{R}^{B,M,d}$ (equation (8)). Finally, the global feature broadcasting of f_g to the features of each sub-watershed was completed.

$$f_g^{'B,M,d/2} = f_g \oplus f_g \oplus \cdots \oplus f_g \ (M \text{ times}) \tag{7}$$

$$\boldsymbol{\varepsilon} = \mathrm{Concat}\left(f'_H, f'_g\right) \in \mathbb{R}^{B,M,d} \tag{8}$$

Node feature extraction. For nodes describing inputs, outputs, and confluences within the watershed, $\boldsymbol{V} = \{v_1, v_2, v_i, \ldots, v_j, v_N\} \in \mathbb{R}^{N\times 3}$, we initialized feature extraction using a DeepMLP to extract high-dimensional features of the nodes (equation (9); Fig. 5b).

$$\boldsymbol{v} = \mathrm{Dropout}\left(\mathrm{GELU}\left(\mathrm{BN}\left(\mathrm{Linear}\left(\boldsymbol{V}^{B,N,3}\right)\right)\right)\right) \in \mathbb{R}^{B,N,d} \tag{9}$$

Through the above operations, we obtained the node feature matrix $\boldsymbol{v} \in \mathbb{R}^{B,N,d}$ and the edge feature matrix $\boldsymbol{\varepsilon} \in \mathbb{R}^{B,M,d}$, which

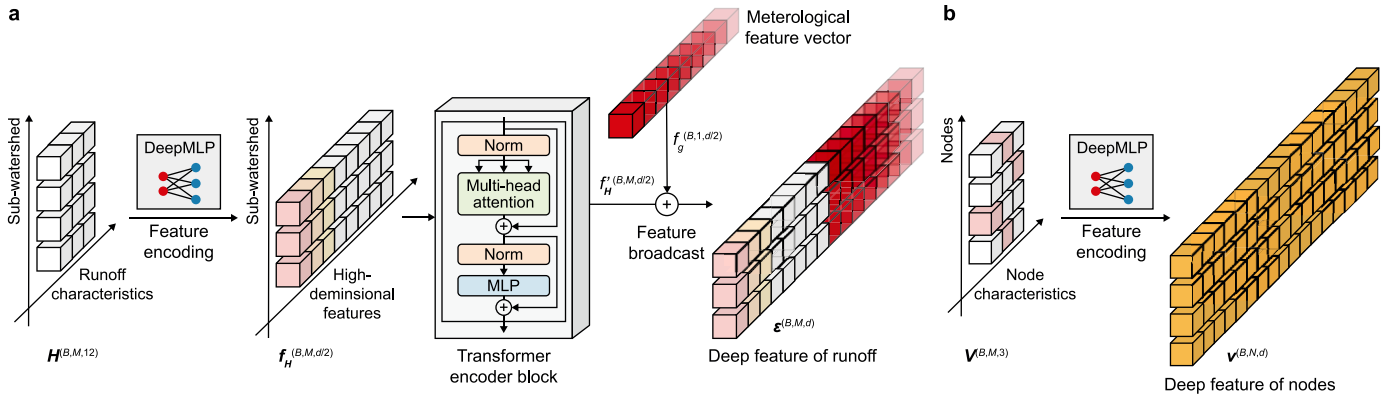

Fig. 5. Watershed hydrological modeling module. **a**, Sub-watershed feature extraction block. This block extracts runoff features through a set of feature encodings and transformers, and integrates meteorological features. **b**, Node feature extraction block. The deep features of the nodes are extracted by a feature encoder. DeepMLP: deep multi-layer perceptron; Norm: layer normalization.

contained the meteorological information of the watershed. Here, d represents the feature dimension. The node feature dimension was kept consistent with that of the edge feature matrix for the next step of processing. This dimension was determined based on the hidden layer size of the DeepMLP used in feature extraction. The dimension d served as a hyperparameter, and its design principle was to balance out the model's performance and computational efficiency.

2.3.4. Watershed topology modeling

By maintaining physical consistency between the parameter propagation process of the GNNs and nutrient migration with the water flow within the watershed, the nutrient migration and transformation processes were simulated in the topological model of the watershed. As given in Definition 3, the topological relationship of watershed runoff was defined as a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$. The adjacency matrix $\boldsymbol{A}$ described the topological relationships. As described in Sections 2.3.2 and 2.3.3, we extracted each sub-watershed runoff feature $\boldsymbol{\varepsilon} \in \mathbb{R}^{B,M,d}$ and the node features $\boldsymbol{v} \in \mathbb{R}^{B,N,d}$.

Inspired by Sanchez-Lengeling et al. [29], we constructed a GNNs block that achieved the topological modeling of the watershed through two key stages: message passing and information updating. The transition from layer L to layer $L+1$ simulated the process of material migration within the watershed (Fig. 6a). During this message passing stage, information was exchanged between the nodes and edges. The adjacency matrix $\boldsymbol{A}$ played a key role in constraining the transmission rules for the watershed features; for instance, if the element in the ith row and jth column was 1, it indicated that the feature of node v_i was first transmitted to edge e_{ij} ($v_i \rightarrow e_{ij}$), and then information was conveyed from edge e_{ij} to node v_j ($e_{ij} \rightarrow v_j$). The feature of edge e_{ij} was updated by aggregating the features from node v_i and edge e_{ij}, and node v_j's feature v'_j was subsequently updated using the newly updated edge feature e'_{ij}, along with v_j's original feature (equation (10); Fig. 6b).

$$
\begin{aligned}
&\text{Given} \\
&\quad v_i, v_j \subseteq \boldsymbol{v}[B,:,d] \\
&\quad e_{ij} \subseteq \boldsymbol{\varepsilon}[B,:,d] \\
&\text{if} \\
&\quad A_{ij} = 1 \\
&\text{then} \\
&\quad f_{\boldsymbol{\varepsilon}} : e'_{ij} = \sigma\left(W_e e_{ij} + W_v v_i + b_e\right);\ \text{Node}\rightarrow\text{Edge} \\
&\quad f_{\boldsymbol{v}} : v'_j = \sigma\left(W_v v_j + W_e e'_{ij} + b_v\right);\ \text{Edge}\rightarrow\text{Node}
\end{aligned}
\tag{10}
$$

Following the message passing stage, the updating stage took place, during which the node and edge features were refined through specific update functions, f_v and f_ε, which effectively mapped the features from layer L to layer $L+1$ (Fig. 6c). In this GNNs framework, the trainable parameters were explicitly defined within the update functions f_v and f_ε and specifically encompassed the weight matrices W_e and W_v and the bias terms b_e and b_v (equation (10)). These parameters were optimized during the training process to enable the model to learn the dynamic material migration patterns within the watershed. Notably, the adjacency matrix $\boldsymbol{A}$, which quantified the topological relationships of the runoff, was fixed a priori, serving as a physical constraint that dictated the order of parameter transmission between the nodes and edges. This fixed topology ensured that the message passing process adhered to the inherent hydrological connectivity of the watershed and that the trainable parameters within the update functions adapted to model the nuanced material transformation processes driven by the water flow. The update functions were implemented using a set of DeepMLPs. This processing enabled the

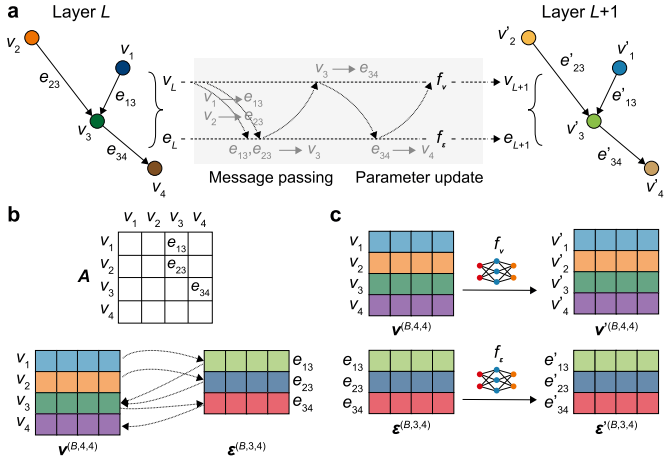

Fig. 6. Message passing and parameter updating mechanism of the graph neural network (GNN) layer (four-node and three-edge simplified version). **a**, GNN layer. **b**, The message passing process based on the adjacency matrix. **c,** Parameter update. The features of nodes $\mathcal{V}$ and edges $\mathcal{E}$ implement parameter passing and parameter updating in the GNNs according to the rules of the adjacency matrix $\boldsymbol{A}$.

GNNs to capture the topological relationships and the dynamic changes of material migration within the watershed, thereby mimicking the water flow-driven processes to provide an accurate representation of nutrient migration and transformation. The mathematical representation of message passing and feature updating is shown in equation (10).

To ultimately predict the nitrate discharge concentration in the watershed at the final stage of the entire model, a set of graph pooling functions was used to aggregate the edge features $\boldsymbol{\varepsilon}' \in \mathbb{R}^{B,M,d}$ updated by the function f_ε into a feature vector $\boldsymbol{f} \in \mathbb{R}^{B,1,d}$ to describe the nitrate discharge concentration in the watershed (Fig. 3d). This vector was then fed into a fully connected network and transformed via linear mapping and activation functions into a scalar prediction $y' \in \mathbb{R}^{B,1}$ that represented the estimated nitrate concentration at the watershed outlet (Fig. 3e). The loss (mean squared error) between the predicted value y' and the observed nitrate concentration y was calculated, and the model parameters were iteratively optimized to minimize this loss and thereby make the predicted values approach the actual observed values.

The novelty of our use of GNNs to model the migration of pollutants lies in their task-specific design. We model the watershed topology with the material transport paths of runoff within the watershed as physical constraints, integrating the hydrological parameters (such as sinuosity and land use conditions) and meteorological features into the node and edge characteristics. We then use message passing and parameter update to simulate the dynamics of pollutant transport.

3. Experiments and results

3.1. Dataset

We collected meteorological data, daily nitrate concentration data from the watershed outlet of the study area, hydrological data from each sub-watershed, and the topological relationships of the watershed (Section 2.1). Because nitrate migration and transformation exhibit time lag effects, it was necessary to balance the use of sufficient historical information with the risk of redundancy [30]. Consequently, the time series length of past data for nitrate concentration prediction was set to 10 days, i.e., the global variable time series length $t = 10$. The watershed comprised 21 sub-watersheds ($M = 21$) and 22 nodes ($N = 22$). We used nitrate emission monitoring data from the study area between October 2017 and April 2024 to train and evaluate our model. The training set comprised 75% of the overall dataset, and the test set comprised 25%. The global variable data structure of a single batch input is shown in Table 1, and Local runoff hydrological data is shown in Table 2. Predictions were generated sequentially along the timeline via a sliding window approach until the dataset was fully traversed.

3.2. Implementation details

Stage One. To predict the nitrate emission concentration on day i, we used watershed-scale time series meteorological data together with historical nitrate concentrations from day $i-10$ to day $i-1$. These inputs were represented as the global variables $\boldsymbol{X}^{20,18,10}$ and processed by the DeepMLP model designed based on equation (2) to extract high-dimensional features: $f_X^{20,18,128} \leftarrow \boldsymbol{X}^{20,18,10}$, where the batch size was $B = 20$ (i.e., 20 batches of data processed at a time); $n = 18$ (16 meteorological indicators, one historical nitrate concentration indicator, and one day of year [DOY]); the time series length was $t = 10$; and the high-dimensional feature was $d/2 = 128$. The hierarchical transformer model (equations (3) and (4)) was then applied for feature encoding, followed by feature compression, yielding a global feature vector $f_g^{20,1,128}$ that covered the entire watershed: $f_g^{20,1,128} \leftarrow f_X^{'20,18,128} \leftarrow f_X^{20,18,128}$.

Stage Two. To model the watershed hydrological information, equation (5) was first used to extract high-dimensional features $f_H^{20,21,128}$ from the hydrological parameters $\boldsymbol{H}^{20,21,12}$ of the 21 sub-watersheds in the study area. Feature encoding was then applied using equation (6), yielding: $f_H^{'20,21,128} \leftarrow f_H^{20,21,128} \leftarrow \boldsymbol{H}^{20,21,12}$. Subsequently, based on equations (7) and (8), the global features were broadcasted to each sub-watershed, producing: $\boldsymbol{\varepsilon}^{20,21,256} \leftarrow (f_H^{'20,21,128}, f_g^{20,1,128})$. For node information $\boldsymbol{v}^{20,22,3}$, high-dimensional features were extracted using equation (9): $\boldsymbol{v}^{20,22,256} \leftarrow \boldsymbol{v}^{20,22,3}$.

Stage Three. Under the constraint of the topological relationship (adjacency matrix $\boldsymbol{A}$) and in line with equation (10), a GNNs was used to simulate the migration process of nutrients with the runoff in the watershed via message passing: $\boldsymbol{\varepsilon}^{20,21,256}, \boldsymbol{v}^{20,22,256} \leftarrow \mathrm{GNN}(\boldsymbol{\varepsilon}^{20,21,256}, \boldsymbol{v}^{20,22,256})$. Subsequently, graph pooling and a FC layer were applied to map the sub-watershed features $\boldsymbol{\varepsilon}^{20,21,256}$ to a scalar value y': $y' \in \mathbb{R}^{20} \leftarrow \mathrm{FC}(\boldsymbol{\varepsilon}^{20,1,256}) \leftarrow \mathrm{pooling}(\boldsymbol{\varepsilon}^{20,21,256})$. The loss between the predicted and observed values was iteratively minimized to achieve accurate prediction, expressed as: $\min \mathrm{Loss}(y', y)$.

We constructed the HTGNN-WNP model using PyTorch. Model training employed the Adadelta [31] as the optimizer, with $\rho = 0.9$ and an initial learning rate of 0.1, which was decayed by 1% every 20 epochs. We employed L2 regularization in the optimizer, with a coefficient of 1×10^{-4}, to mitigate overfitting and enhance generalization. A large batch size ($B > 20$) was used to approximate the dataset's overall gradient more accurately, enabling the model to capture global feature distributions while reducing sensitivity to noise in small batches. All experiments, including model training and testing, were conducted on a single NVIDIA GeForce RTX 3070 GPU. The model contained 0.59 million parameters and 2.35 billion FLOPs. The training process required approximately 1.3 h for the entire cycle, and the inference time was 0.0057 s per sample on the same GPU hardware.

3.3. Evaluation metrics

To evaluate our model's performance in predicting nitrate emission concentrations, we used the mean absolute error (*MAE*), root mean square error (*RMSE*), and the coefficient of determination (R^2) as the assessment metrics:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i' - y_i\right)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i' - y_i| \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{13}$$

Here, n represents the total number of samples, y_i denotes the ith observed value, y_i denotes the ith corresponding predicted value, and $\bar{y}$ is the average of all true values. *RMSE* reflected the precision of the model and the scale of the errors, while *MAE*

Table 1
Model input parameters of global variable data.

Variable name	Variable description
T_Max	Daily maximum temperature (°F)
T_Max_D	Low-frequency signal of wavelet transform of daily maximum temperature (°F)
T_Max_A	High-frequency signal of wavelet transform of daily maximum temperature (°F)
T_Min	Daily minimum temperature (°F)
T_Min_D	Low-frequency signal of wavelet transform of daily minimum temperature (°F)
T_Min_A	High-frequency signal of wavelet transform of daily minimum temperature (°F)
T_Avg	Daily average temperature (°F)
T_Avg_D	Low-frequency signal of wavelet transform of daily average temperature (°F)
T_Avg_A	High-frequency signal of wavelet transform of daily average temperature (°F)
AWND	Daily average wind speed (mph)
AWND_D	Low-frequency signal of wavelet transform of daily average wind speed (mph)
AWND_A	High-frequency signal of wavelet transform of daily average wind speed (mph)
RH	Daily relative humidity (%)
RH_D	Low-frequency signal of wavelet transform of daily relative humidity (%)
RH_A	High-frequency signal of wavelet transform of daily relative humidity (%)
PRCP	Daily precipitation (inches)
Nitrate	Nitrate concentration (micromolar, μM)
DOY	One day of the year (dimensionless)

Note: Global variable data includes meteorological data, pollutant concentration at the watershed outlet, and one day of the year. The time span of the data input into the model at one time is $t = 10$, indicating the sliding window length.

indicated the magnitude of the average error, with smaller values indicating better performance. R^2 was used to measure the proportion of variation in the dependent variable that was explained by the independent variable, reflecting the strength of the linear correlation between predicted and observed variables. Its range is [0, 1], with values closer to 1 indicating stronger predictive performance.

3.4. Model's performance for nitrate concentration

This study focused on predicting nitrate concentrations in surface water using the HTGNN-WNP. Unlike conventional approaches that directly simulate runoff dynamics, HTGNN-WNP incorporates runoff topological relationships as structural inputs to the GNNs. The model was evaluated using data from April 2023 to April 2024, and its predictions were compared with observed nitrate concentrations (Fig. 7). A quantitative performance assessment is summarized in Table 3.

The HTGNN-WNP demonstrated exceptional predictive performance over the entire prediction period. The model yielded an *RMSE* of 2.29 and an *MAE* of 1.67, indicating that prediction errors were small and that the predicted values closely tracked the observed trends. Notably, the coefficient of determination ($R^2 = 0.92$) indicated that 92% of the variance in observed nitrate concentrations was explained by the model, reflecting a strong goodness-of-fit and robust predictive accuracy.

The model effectively captured sudden changes in nitrate levels throughout the prediction period (Fig. 7), demonstrating robustness to sudden environmental fluctuations. Such adaptability is critical for forecasting water quality in dynamic systems, where variations are often frequent and unpredictable. Moreover, the HTGNN-WNP model exhibited excellent generalization and representation capabilities, effectively integrating diverse data inputs and modeling complex watershed relationships. These strengths underscore its potential for broader application in environmental monitoring and management, where it may support water quality assessment and policymaking across similar watersheds. Overall, the results confirm the model's reliability for nitrate concentration prediction.

3.5. Comparison of the HTGNN-WNP with other methods

We compared the performance of the HTGNN-WNP against several advanced methods, such as multi-layer perceptron (MLP) [32], recurrent neural network (RNN) [33], long short-term memory (LSTM) [34], support vector regression (SVR) [35], gated recurrent unit (GRU) [36], and the transformer method [23] (Table 3). We reproduced these classic models and strictly followed their original architectural designs (Supplementary Texts S1–S6). To ensure fairness, all models were trained on the same preprocessed input data (Tables 1 and 2) and evaluated using an identical dataset split (75% training, 25% testing). Model performance was assessed through unified evaluation metrics (*RMSE*, *MAE*, and R^2) to ensure comparability and reproducibility. Across all key metrics, HTGNN-WNP consistently achieved the best performance.

Predicted versus observed nitrate concentrations aligned closely for HTGNN-WNP, with points clustering around the 1:1 line, in contrast to the more dispersed patterns of other methods (Fig. 8). This agreement is consistent with the model's high R^2 value and underscores its accuracy in capturing underlying data trends. The outstanding performance of the HTGNN-WNP model

Table 2
Model input parameters of local runoff hydrological data.

Variable name	Variable description	Variable category
Len	The length of the sub-watershed runoff (m)	Hydrological Data ($M = 21$, indicating the number of sub-watersheds.)
Wid	The width of the sub-watershed runoff (m)	
Sin	Sinuosity of the sub-watershed runoff (dimensionless)	
Elev_Diff	Elevation difference within the sub-watershed (m)	
Imp_Area	Impervious surface area within the sub-watershed (m^2)	
Imp_Ratio	Percentage of impervious surface area relative to the total area of the sub-watershed (%)	
Water_Area	Area of water bodies within the sub-watershed (m^2)	
Water_Ratio	Percentage of water body area relative to the total area of the sub-watershed (%)	
Eco_Area	Ecological land area within the sub-watershed (m^2)	
Eco_Ratio	Percentage of ecological land area relative to the total area of the sub-watershed (%)	
Agri_Area	Agricultural land area within the sub-watershed (m^2)	
Agri_Ratio	Percentage of agricultural land area relative to the total area of the sub-watershed (%)	
RN_Vec	Runoff node vector. A vector [inlet, outlet, confluence] represents the status of a runoff node, with 0 for 'No' and 1 for 'Yes' with each element.	Runoff node data ($N = 22$, indicating the number of nodes)

stems from its innovative design, comprising both hierarchical transformers and GNNs. This approach provided a significant advantage in modeling the topological relationships within the watershed's physical runoff layer. By jointly capturing complex spatial and temporal dependencies, the model achieved substantial gains in predictive accuracy.

4. Discussion

4.1. Ablation experiment

We conducted two sets of ablation experiments: one to evaluate the contributions of different model components to water quality prediction, and the other to assess the impact of alternative data preprocessing methods on model performance.

4.1.1. Ablation experiment one: impact of various model components

To systematically evaluate the impact of each module, we developed three model variants constructed from the distinct components outlined in Section 2.2. Specifically, HT-WNP consisted of the hierarchical transformer module (Fig. 4); HTH-WNP extended HT-WNP by incorporating hydrological modeling (Figs. 4 and 5a); and HTGNN-WNP represented the full model architecture (Fig. 3). These ablation models enabled a structured analysis of the individual components' contributions within the proposed framework, with the quantitative comparison presented in Supplementary Table S1.

The HT-WNP variant achieved an *RMSE* of 2.68, an *MAE* of 1.95, and an R^2 value of 0.89, indicating that the basic transformer architecture captured substantial patterns in the nitrate concentration dynamics (Table 4). Incorporating hydrological features further improved performance: HTH-WNP reduced *RMSE* to 2.56, lowered *MAE* to 1.87, and increased R^2 to 0.90. These results indicate that the integration of hydrological features enhanced the model's ability to replicate the observed concentration trends and its consistency between predictions and measured values.

The complete model, HTGNN-WNP, demonstrated the best performance, with an *RMSE* of 2.29, *MAE* of 1.67, and R^2 of 0.92. Incorporating complex hydrological features and runoff topology substantially reduced the discrepancy between predictions and observations. These results emphasize the importance of considering hydrological, topological, temporal, and feature aspects, which together enhanced predictive accuracy. Such a comprehensive approach demonstrates the value of integrating multiple perspectives when addressing water quality prediction in complex environments.

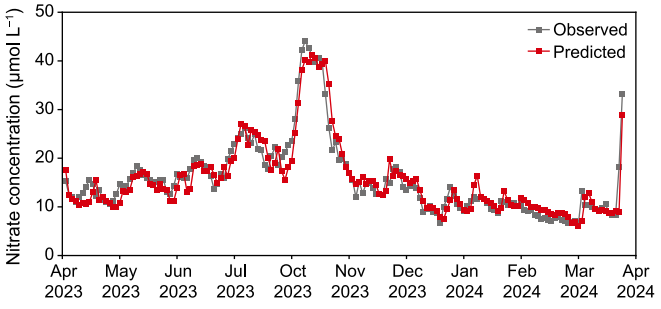

Fig. 7. The prediction results of the hierarchical transformer and graph neural network model for the watershed nitrate prediction model. Note that some months have missing observations.

Table 3
Results of different methods for nitrate concentration prediction.

Method	*RMSE*	*MAE*	R^2
MLP	4.31	3.59	0.72
RNN	4.27	3.36	0.72
LSTM	3.74	2.59	0.79
GRU	3.62	2.63	0.80
SVR	3.54	2.52	0.81
Transformer	3.34	2.74	0.83
HTGNN-WNP	2.29	1.67	0.92

Note: MLP, multi-layer perceptron; RNN, recurrent neural network; LSTM, long short-term memory; GRU, gated recurrent unit; SVR, support vector regression; Transformer, the transformer method; HTGNN-WNP, hierarchical transformer and graph neural network model for watershed nitrate prediction. Smaller root mean square error (*RMSE*) and mean absolute error (*MAE*), and an R^2 closer to 1 denote better model performance.

4.1.2. Ablation experiment two: impact of data preprocessing methods

To investigate the impact of data preprocessing strategies on model performance, we compared the use of raw input data with data preprocessed by wavelet transform (WT). WT decomposes time series into low-frequency trends and high-frequency abrupt components, which enhances the extraction of nonlinear temporal features. The original data and WT-preprocessed data were used as inputs and compared, allowing a controlled comparison of preprocessing effects.

The quantitative results show that WT preprocessing reduced the *RMSE* from 2.66 to 2.29 and the *MAE* from 2.02 to 1.67, and R^2 improved from 0.89 to 0.92 (Table 5). These findings indicate that WT effectively mitigated noise and enhanced the model's capability to capture dynamic pollution patterns. By separating the frequency components of the time series, WT enabled the model to better distinguish changing trends (e.g., fertilization cycles) from short-term abrupt changes (e.g., rainfall-induced runoff), making it a valuable preprocessing complex hydrological and meteorological data in watershed studies.

4.2. Advantages of HTGNN-WNP in watershed water quality modeling

Human activities and climate change-induced NPS pollution are the main reasons for water quality changes in watersheds [37]. Accurate predictions of watershed water quality is therefore critical for managing agricultural water, soil resources, and the environment. In this context, the HTGNN-WNP has several advantages.

4.2.1. Superior feature extraction and pattern recognition

Extracting effective features and patterns from nonlinear, multivariate time series data is the key to accurate watershed water quality prediction. Interactions and couplings among different variables add significant complexity to prediction models, reflecting the underlying environmental processes and feedback mechanisms that can profoundly influence changes in water quality [38,39].

The HTGNN-WNP addressed this challenge using a hierarchical transformer constructed with dual transformer modules. Latent patterns were independently captured from both temporal sequences and multivariate dimensions, and these features were then integrated to enhance predictive performance. This approach fully leveraged the strengths of each dimension while enabling the model to comprehensively uncover the complex underlying dynamics of water quality variations.

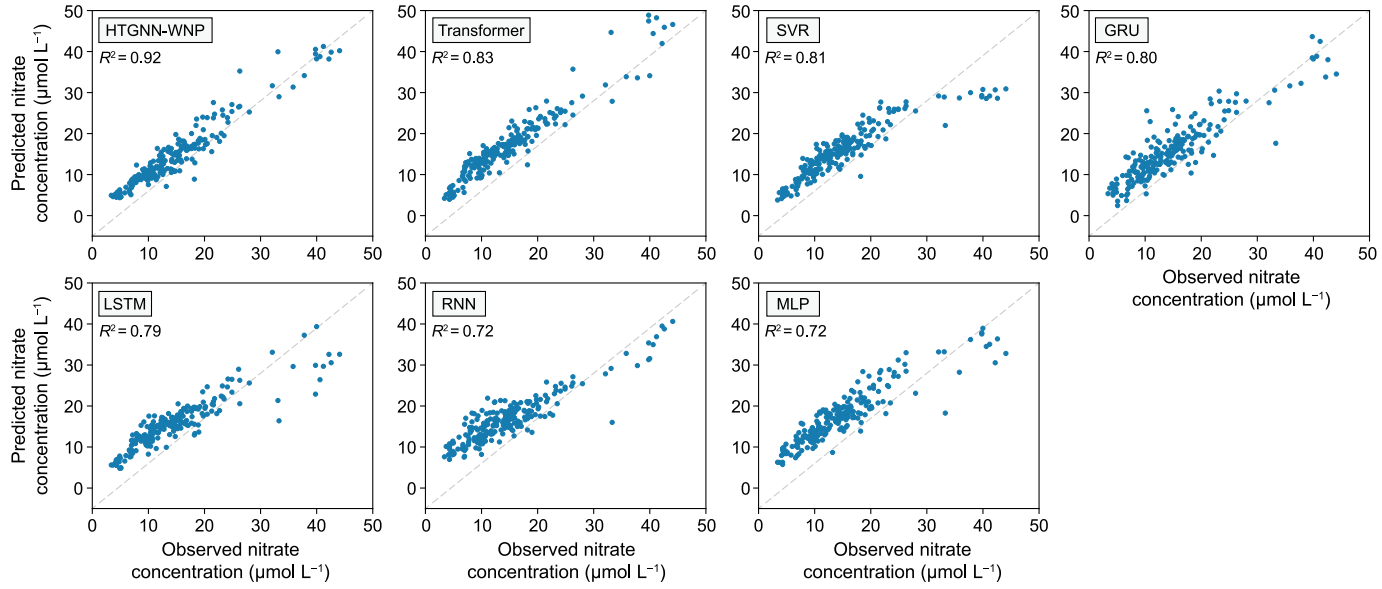

Fig. 8. Scatter comparison plots of nitrate concentration prediction results of hierarchical transformer and graph neural network model for watershed nitrate prediction (HTGNN-WNP) and multiple methods. Transformer: the transformer method; SVR: support vector regression; GRU: gated recurrent unit; LSTM: long short-term memory; RNN: recurrent neural network; MLP: multi-layer perceptron.

Table 4
Prediction performance of ablation models for nitrate concentration.

Method	*RMSE*	*MAE*	R^2
HT-WNP	2.68	1.95	0.89
HTH-WNP	2.56	1.87	0.90
HTGNN-WNP	2.29	1.67	0.92

Note: HT-WNP, hierarchical transformer for watershed nitrate prediction; HTH-WNP, hierarchical transformer incorporating hydrological modeling for watershed nitrate prediction; HTGNN-WNP, hierarchical transformer and graph neural network model for watershed nitrate prediction. Smaller root mean square error (*RMSE*) and mean absolute error (*MAE*), and an R^2 closer to 1 denote better model performance.

Table 5
Quantitative performance comparison of the HTGNN-WNP model under two data preprocessing schemes for nitrate concentration prediction.

Model input	*RMSE*	*MAE*	R^2
Original data	2.66	2.02	0.89
Wavelet transform-preprocessed data	2.29	1.67	0.92

Note: Smaller root mean square error (*RMSE*) and mean absolute error (*MAE*), and an R^2 closer to 1 denote better model performance.

4.2.2. Effective assimilation of multisource data

Runoff characteristics and local landscapes are crucial factors determining NPS pollution formation and spread [42–44]. As a link between terrestrial and aquatic environments, runoff directly affects the migration rate and diffusion range of pollutants through characteristics such as flow velocity, discharge, and path direction [40,41].

Data integration is a significant aspect that allows for understanding and addressing NPS pollution. The HTGNN-WNP demonstrated outstanding capabilities in this regard, particularly in integrating macro meteorological observation data with local runoff physical parameters. Watershed water quality is influenced by a combination of global and local factors. Our methodology involved encoding and combining global (meteorological features spanning an entire small watershed) and local (hydrological features of a single sub-watershed) watershed characteristics via neural networks. This approach offered a comprehensive perspective of the watershed system, considering both the large-scale climatic impacts and the specific traits of each sub-watershed. The use of multiscale data resulted in more accurate predictions than models relying on a single data type, as it captured the intricate interplay among the diverse factors affecting water quality.

4.2.3. Accurate modeling of topological relationships

The incorporation of GNNs within the HTGNN-WNP framework was a pivotal innovation. During model design, watershed topological relationships were incorporated as physical constraints in simulating material migration, enabling the model to characterize runoff connectivity across sub-watersheds with high fidelity. Previous studies have shown that the connectivity delineated by the topological architecture of a watershed plays an essential role in determining the migration and transformation of pollutants, such as nitrates, among diverse subregions [45,46].

Previous studies on water quality analysis that involved GNNs and topological relationships [24,46] mainly focused on topological feature extraction and did not integrate the connection attributes of the topology as physical constraints into the GNNs model. For example, Xia et al. [46] used topological relationships to analyze nutrient retention in small water bodies but did not couple the topological structure with time series pollution dynamics. By jointly integrating hydrological parameters, meteorological time series, and watershed topology, HTGNN-WNP provided end-to-end predictions and demonstrated strong performance in capturing abrupt nitrate fluctuations (Table 3). Using GNNs to embed topological relationships into the model enabled a more accurate representation of runoff connectivity and improved quantification of nutrient retention and transfer among interconnected water bodies. Moreover, the integration of GNNs enabled the model to inherently assimilate non-Euclidean spatial traits and interactions, often overlooked in conventional time series models [17,18,47]. Consequently, this strategy facilitated the capture of elaborate spatio-temporal dynamics in a manner consistent with the physical processes of runoff and pollutant transport. Grounding feature transmission in physical principles not only enhanced predictive accuracy but also addressed limitations inherent in purely data-driven methods.

4.2.4. Feature importance analysis

To address the interpretability of the HTGNN-WNP model and explore the mechanistic relationships between the inputs and the nitrate concentration predictions, we employed permutation feature importance (PFI) [48], a model-agnostic method that quantifies the impact of each feature on model performance. PFI operates by randomly permuting the values of a feature and measuring the decrease in prediction accuracy (e.g., reduction in R^2), with larger drops indicating higher feature importance. Each feature was permuted 50 times to ensure statistical reliability, and the results were visualized to highlight the key drivers of nitrate dynamics.

Meteorological and historical concentration feature analysis. Historical nitrate concentration (Nitrate) emerged as the most critical factor (Fig. 9), underscoring the strong temporal autocorrelation of nitrate levels in watershed runoff, where past concentrations directly influence current predictions. This reflects the persistence and cumulative effects characteristic of nutrient transport processes in watershed systems. Temperature-related features, such as daily maximum temperature and its low-frequency component, were also found to be significantly impactful, likely because temperature regulates microbial activity and nutrient transformation rates in soil and water. Among the meteorological features, the wavelet-transformed components (e.g., T_Max_D, T_Min_A, RH_A) consistently ranked among the most important, indicating that temperature changes affect the migration of nitrate through the hydrological cycle. Similarly, RH_A (high-frequency component of relative humidity) captures abrupt changes in atmospheric moisture, which affect soil water retention and subsequent nitrate leaching. The prominence of these wavelet-transformed features validates our preprocessing strategy, which involved decomposing time series into trend and fluctuation components and subsequently enabled the model to capture both slow- and fast-varying environmental impacts on nitrate dynamics.

Hydrological and land use features analysis. Land use metrics, particularly Agricultural_Ratio (%) and Impervious_Ratio (%), showed the largest declines in R^2 (Fig. 10), highlighting the critical role of spatial land use patterns in nitrate dynamics. Agricultural areas directly contribute nitrate via fertilizer runoff, while impervious surfaces enhance surface runoff velocity, accelerating nitrate transport to water bodies [50]. Notably, impervious surfaces primarily accelerate nitrate transport at the sub-catchment scale, whereas agricultural dominance drives nitrate accumulation at the watershed scale. Hydrological features such as sinuosity and Elevation_Diff (m) also exhibited statistically significant contributions (Fig. 10). Runoff sinuosity moderates water flow and nutrient retention, while elevation differences influence gravitational-driven transport. Although the absolute importance values for hydrological features appeared modest, their collective impact was substantial in modeling the topological transport of nitrates across sub-watersheds. For example, Water_Area (m^2) demonstrated a consistent negative R^2 drop, indicating that water body distribution affects nitrate dilution and retention.

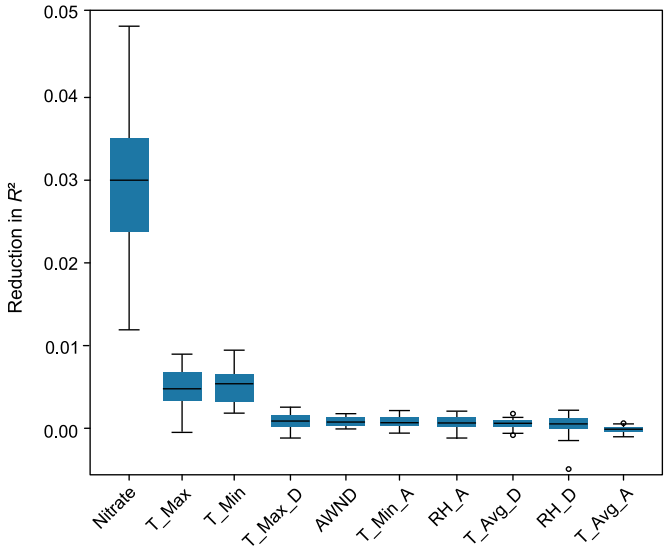

Fig. 9. Impact of meteorological and historical concentration features on model performance. The circles represent test outliers, and the full names of the abbreviations for the *x*-axis variables are provided in Table 1.

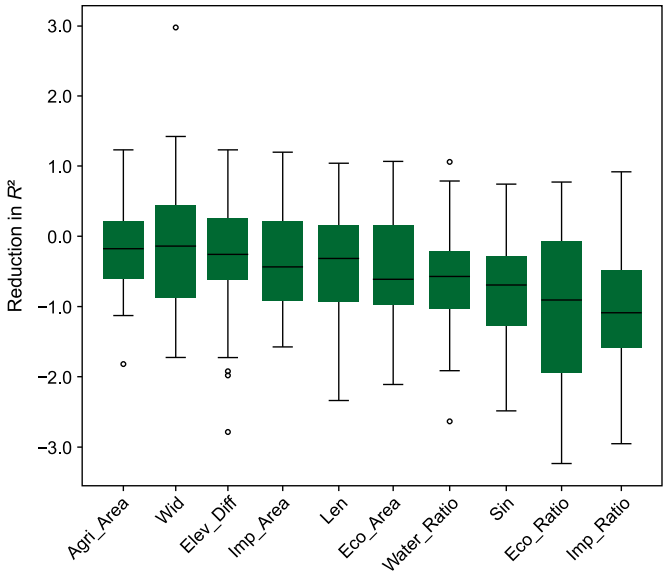

Fig. 10. Influence of hydrological and land use characteristics on model performance. The circles represent test outliers, and the full names of the abbreviations for the *x*-axis variables are provided in Table 2.

The feature–importance analysis confirmed that HTGNN-WNP not only predicts nitrate concentrations with high accuracy but also captures mechanistic relationships. The analysis indicates that historical pollution loads and land use patterns are the primary factors influencing the concentration of non-point source pollutants in the study area. Meanwhile, the wavelet-transformed meteorological features lead to the model's sensitivity to both long-term climatic trends and short-term fluctuations. These findings align with prior studies [18,49], which show that nitrate dynamics are governed by a combination of legacy pollution, land management practices, and hydrometeorological conditions. By linking data-driven predictions with process-based interpretation, the PFI results positioned HTGNN-WNP as a systematic tool for identifying the key environmental factors that influence watershed water quality.

4.3. Applicability and prospects

In the context of water quality research, the HTGNN-WNP exhibits remarkable applicability. First, our model offers significant advantages due to the universality of its data sources: it relies on readily accessible datasets, thereby reducing the need for complex data acquisition procedures and enabling implementation across diverse regions. Second, the model demonstrated exceptional proficiency in short-term pollutant concentration prediction,

particularly in evaluating the impact of extreme climatic phenomena on pollution. Its ability to accurately capture abrupt nitrate concentration fluctuations induced by extreme rainfall or drought events can provide critical support for emergency response and risk assessment in water quality management. The HTGNN-WNP model can also interpolate missing data when estimating pollutant fluxes, effectively filling observational gaps. Third, our method is flexible in its input data requirements. Anthropogenic factors within a watershed—such as the extent, scale, and spatial configuration of cultivated land—can be incorporated into the model's input features to predict the concentration of pollutant emissions. Through feature influence analysis, HTGNN-WNP can assist in the formulation and implementation of agricultural management measures and policies.

Process-based models depict the migration and transformation of substances within a watershed through mathematical equations and rely on detailed data of hydrological and biogeochemical processes [12,13]. However, a large number of unknown processes in the Earth system, such as the nonlinear interactions between soil, water, and vegetation, are difficult to accurately describe with equations, which limits the predictive ability for sudden pollution events [16]. Process-based models rely on high-quality observational data and require complex processing through geographic information systems. The HTGNN-WNP is a data-driven deep learning model that learns implicit patterns without preset physical equations. Moreover, it uses graph neural networks to simulate the physical migration path of pollutants along with runoff. The two types of models are complementary in their application fields: process-based models focus on the analysis of mechanisms and processes, while HTGNN-WNP offers rapid, accurate predictions that are more suitable for efficient management and decision-making.

Our method exhibited scalability and upgradability. Since NPS pollution arises from complex physical, chemical, and biological processes, future work should focus on systematically embedding these mechanisms into the GNNs' message-passing operations to develop a more interpretable framework for pollution assessment and prediction. Further, meteorological or hydrological model predictions can extend the model's prediction horizon to weekly or monthly timescales. Finally, the model's application scenarios can be expanded through extensive applied research on urban, forested, and cold-region watersheds.

5. Conclusion

In this study, we developed the HTGNN-WNP to predict nitrate concentrations in watersheds, with the aim of addressing the challenges posed by NPS pollution. By combining hierarchical transformers with graph neural networks, the model effectively captures complex patterns and topological relationships in a watershed system. Extensive experiments and comparisons demonstrated its superior performance over several state-of-the-art methods. Ablation experiments revealed the significance of each component in the model and highlighted the importance of considering hydrological and topological interactions along with temporal and feature dependencies. The HTGNN-WNP exhibited a distinct advantage in predicting water quality through its ability to extract effective features from nonlinear time series, assimilate multisource data, and accurately model topological. The model's applicability was evidenced by its data accessibility, short-term prediction accuracy, and handling of missing data. Although validated in a single watershed, broader testing across diverse climates and larger systems is needed to establish generalizability. Future research should explore integration with process-based models by incorporating physical, chemical, and biological processes into the framework, thereby enhancing interpretability, adaptability, and applicability to other pollutants and watersheds. Overall, the HTGNN-WNP model is an innovative solution for advancing watershed nitrate prediction and supporting water quality management.

CRediT authorship contribution statement

Jun Sun: Writing - Original Draft, Methodology, Investigation, Conceptualization. **Xuesong Gao:** Validation, Supervision, Resources, Project Administration, Funding Acquisition. **Zhiyong Deng:** Visualization, Data Curation. **Yudong Zhao:** Visualization, Data Curation. **Qi Wang:** Investigation, Formal Analysis. **Xiyi Zhao:** Software, Data Curation. **Xu Liu:** Visualization, Software.

Data availability

The data and the related code for HTGNN-WNP are publicly available. All resources will be released at https://github.com/sj6440/HTGNN-WNP.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Ministry of Science and Technology of China (Grant No. 2022YFD1901400). This work is supported by the Open Fund of Key Laboratory of Investigation, Monitoring, Protection and Utilization for Cultivated Land Resources, Ministry of Natural Resources (Grant No. CLRKL2024GP10). The authors would like to express their sincere gratitude to Dr. Minghua Zhou. His valuable insights inspired the research and its innovative content. Finally, we would like to express our gratitude to the anonymous reviewers and editors for their hard work, which has enabled the article to be presented in its optimal form.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2025.100632.

References

[1] N.K. Shrestha, R.P. Rudra, P. Daggupati, P.K. Goel, R. Shukla, A comparative evaluation of the continuous and event-based modelling approaches for identifying critical source areas for sediment and phosphorus losses, J. Environ. Manag. 277 (2021) 111427, https://doi.org/10.1016/j.jenvman.2020.111427.

[2] V. Varekar, V. Yadav, S. Karmakar, Rationalization of water quality monitoring locations under spatiotemporal heterogeneity of diffuse pollution using seasonal export coefficient, J. Environ. Manag. 277 (2021) 111342, https://doi.org/10.1016/j.jenvman.2020.111342.

[3] P. Rao, S. Wang, A. Wang, D. Yang, L. Tang, Spatiotemporal characteristics of nonpoint source nutrient loads and their impact on river water quality in Yancheng city, China, simulated by an improved export coefficient model coupled with grid-based runoff calculations, Ecol. Indic. 142 (2022) 109188, https://doi.org/10.1016/j.ecolind.2022.109188.

[4] T. Chen, J. Lu, T. Lu, X. Yang, Z. Zhong, H. Feng, M. Wang, J. Yin, Agricultural non-point source pollution and rural transformation in a plain river network: insights from Jiaxing city, China, Environ. Pollut. 333 (2023) 121953, https://doi.org/10.1016/j.envpol.2023.121953.

[5] M. Luo, X. Liu, N. Legesse, Y. Liu, S. Wu, F.X. Han, Y. Ma, Evaluation of agricultural non-point source pollution: a review, water, air, Soil Pollut 234 (2023) 657, https://doi.org/10.1007/s11270-023-06686-x.

[6] K. Peng, J.K. Li, G.R. Hao, Y.W. Liu, X. Zhou, W.F. Xie, Characteristics of non-

point source pollution based on monitoring experiment in the Yingwugou small watershed, China, Ecohydrol. & Hydrobiol 23 (2023) 1–14, https://doi.org/10.1016/j.ecohyd.2022.09.001.
[7] X. Tong, Y. Zhou, J. Liu, P. Qiu, Y. Shao, Non-point source pollution loads estimation in Three Gorges Reservoir Area based on improved observation experiment and export coefficient model, Water Sci. Technol. 85 (2022) 27–38, https://doi.org/10.2166/wst.2021.508.
[8] S. Wang, Y. Wang, Y. Wang, Z. Wang, Assessment of influencing factors on non-point source pollution critical source areas in an agricultural watershed, Ecol. Indic. 141 (2022) 109084, https://doi.org/10.1016/j.ecolind.2022.109084.
[9] L. Zou, Y. Liu, Y. Wang, X. Hu, Assessment and analysis of agricultural non-point source pollution loads in China: 1978–2017, J. Environ. Manag. 263 (2020) 110400, https://doi.org/10.1016/j.jenvman.2020.110400.
[10] D. Tian, S. Niu, A global analysis of soil acidification caused by nitrogen addition, Environ. Res. Lett. 10 (2015) 024019, https://doi.org/10.1088/1748-9326/10/2/024019.
[11] R. Dupas, M. Delmas, J.-M. Dorioz, J. Garnier, F. Moatar, C. Gascuel-Odoux, Assessing the impact of agricultural pressures on N and P loads and eutrophication risk, Ecol. Indic. 48 (2015) 396–407, https://doi.org/10.1016/j.ecolind.2014.08.007.
[12] R.G. Prinn, Development and application of earth system models, Proc. Natl. Acad. Sci. USA 110 (Suppl. 1) (2013) 3673–3680, https://doi.org/10.1073/pnas.1107470109.
[13] K.R. Douglas-Mankin, R. Srinivasan, J.G. Arnold, Soil and Water Assessment Tool (SWAT) model: current developments and applications, Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 53 (2010) 1423–1431, https://doi.org/10.13031/2013.34915.
[14] R.A. Young, C.A. Onstad, D.D. Bosch, W.P. Anderson, AGNPS: a nonpoint-source pollution model for evaluating agricultural watersheds, J. Soil Water Conserv. 44 (2) (1989) 168–173.
[15] R.A. Leonard, W.G. Knisel, Selection and application of models for nonpoint source pollution and resource conservation, Dev. Environ. Model. 10 (1986) 213–229, https://doi.org/10.1016/B978-0-444-99505-6.50015-0.
[16] C. Irrgang, N. Boers, M. Sonnewald, et al., Towards neural Earth system modelling by integrating artificial intelligence in Earth system science, Nat. Mach. Intell. 3 (2021) 667–674, https://doi.org/10.1038/s42256-021-00374-3.
[17] H. Wan, R. Xu, M. Zhang, Y. Cai, J. Li, X. Shen, A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods, J. Hydrol. 612 (2022) 128081, https://doi.org/10.1016/j.jhydrol.2022.128081.
[18] R. Xu, S. Hu, H. Wan, Y. Xie, Y. Cai, J. Wen, A unified deep learning framework for water quality prediction based on time-frequency feature extraction and data feature enhancement, J. Environ. Manag. 351 (2024) 119894, https://doi.org/10.1016/j.jenvman.2023.119894.
[19] NEON (National Ecological Observatory Network). Nitrate in surface water (DP1.20033.001), RELEASE-2024. https://doi.org/10.48443/mvdb-k902. Dataset accessed from https://data.neonscience.org/data-products/DP1.20033.001/RELEASE-2024 on May 1, 2024.
[20] B. Lehner, K. Verdin, A. Jarvis, New global hydrography derived from space-borne elevation data, Eos Trans. Am. Geophys. Union 89 (10) (2008) 93–94, https://doi.org/10.1029/2008EO100001.
[21] X. Li, X. Yan, H. Han, G. Luo, X. Yan, Y. Xia, The trade-off effects of water flow velocity on denitrification rates in open channel waterways, J. Hydrol. 637 (2024) 131374, https://doi.org/10.1016/j.jhydrol.2024.131374.
[22] J. Schwenk, J. Hariharan, RivGraph: automatic extraction and analysis of river and delta channel network topology, J. Open Source Softw. 6 (2021), https://doi.org/10.21105/joss.02952. LA-UR-21–20218.
[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017), https://doi.org/10.48550/arXiv.1706.03762.
[24] Y. Liu, M. Jin, S. Pan, C. Zhou, Y. Zheng, F. Xia, P.S. Yu, Graph Self-Supervised Learning: a Survey, IEEE Trans. Knowl. Data Eng. 35 (6) (2023) 5879–5900, https://doi.org/10.1109/TKDE.2022.3172903.
[25] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, IEEE Transact. Neural Networks Learn. Syst. 32 (1) (2021) 4–24, https://doi.org/10.1109/TNNLS.2020.2978386.
[26] B. Zhao, H. Lu, S. Chen, J. Liu, D. Wu, Convolutional neural networks for time series classification, J. Syst. Eng. Electron. 28 (1) (2017) 162–169, https://doi.org/10.21629/JSEE.2017.01.18.
[27] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, G. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction, Proc. Twenty-Sixth Int. Joint Conf. Artif. Intell. (IJCAI-17) (2017), https://doi.org/10.48550/arXiv.1704.02971.
[28] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, L. Sun, Transformers in time series: a survey, Proc. Thirty-Second Int. Joint Conf. Artif. Intell. (2023), https://doi.org/10.24963/ijcai.2023/759. Macao, P.R.China.
[29] B. Sanchez-Lengeling, E. Reif, A. Pearce, A.B. Wiltschko, A gentle introduction to graph neural networks [WWW Document] (2021). URL, https://distill.pub/2021/gnn-intro/, 5.1.2024.
[30] S. Shin, Y. Her, R. Muñoz - Carpena, X. Yu, C. Martinez, A. Singh, Climate change impacts on water quantity and quality of a watershed - lake system using a spatially integrated modeling framework in the Kissimmee River-Lake Okeechobee system, J. Hydrol. Reg. Stud. 47 (2023) 101408, https://doi.org/10.1016/j.ejrh.2023.101408.
[31] M.D. Zeiler, Adadelta: an adaptive learning rate method, arXiv 1212.5701 (2012), https://doi.org/10.48550/arXiv.1212.5701.
[32] A. Lapedes, R. Farber, Nonlinear signal processing using neural networks: prediction and system modelling, Los Alamos Natl, Lab. Rep. LA-UR-87-2662; CONF-8706130-4 (1987).
[33] R.J. Williams, D. Zipser, A learning algorithm for continually running fully recurrent neural networks, Neural Comput. 1 (2) (1989) 270–280, https://doi.org/10.1162/neco.1989.1.2.270.
[34] F.A. Gers, N.N. Schraudolph, J. Schmidhuber, Learning precise timing with LSTM recurrent networks, J. Mach. Learn. Res. 3 (1) (2003) 115–143, https://doi.org/10.1162/15324430376896613.
[35] K.-j. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1–2) (2003) 307–319, https://doi.org/10.1016/S0925-2312(03)00372-2.
[36] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv 1406.1078v3 (2014), https://doi.org/10.48550/arXiv.1406.1078.
[37] EPA, Basic information about nonpoint source (NPS) pollution [WWW Document]. URL, https://www.epa.gov/nps/basic-information-about-nonpoint-source-nps-pollution, 2022. (Accessed 22 November 2024).
[38] J. Monteith, M. Unsworth, Principles of Environmental Physics: Plants, Animals, and the Atmosphere, Academic Press, Cambridge, MA, 2013.
[39] J.A. Navarro Alberto, Multivariate Statistical Methods: a Primer, fourth ed., Chapman and Hall/CRC, New York, 2016 https://doi.org/10.1201/9781315382135.
[40] M. Noguchi, T. Hiwatashi, Y. Mizuno, M. Minematsu, Pollutant runoff from non-point sources and its estimation by runoff models, Water Sci. Technol. 46 (11–12) (2002) 407–412, https://doi.org/10.2166/wst.2002.0770.
[41] Q.D. Zhu, J.H. Sun, G.F. Hua, Runoff characteristics and non-point source pollution analysis in the Taihu Lake Basin: a case study of the town of Xueyan, China, Environ. Sci. Pollut. Res. 22 (2015) 15029–15036, https://doi.org/10.1007/s11356-015-4709-y.
[42] Q. Yao, L. Sun, D. Chen, J. Liao, L. Tang, Q. Sun, The response of the migration of non-point source pollution to land use change in a typical small watershed in a semi-urbanized area, Sci. Total Environ. 785 (2021) 147387, https://doi.org/10.1016/j.scitotenv.2021.147387.
[43] J. Yang, J. Liang, G. Yang, Y. Feng, G. Ren, C. Ren, X. Han, X. Wang, Characteristics of non-point source pollution under different land use types, Sustainability 12 (5) (2020) 2012, https://doi.org/10.3390/su12052012.
[44] Z.-y. Shen, Q. Hong, H. Yu, J.-f. Niu, Parameter uncertainty analysis of non-point source pollution from different land use types, Sci. Total Environ. 408 (8) (2010) 1971–1978, https://doi.org/10.1016/j.scitotenv.2009.12.007.
[45] P. Kumar, A. Anders, E. Bauer, N.E. Blair, M. Cain, A. Dere, J. Druhan, T. Filley, C. Giannopoulos, A.E. Goodwell, D. Grimley, D. Karwan, L.L. Keefer, J. Kim, L. Marini, M. Muste, A.N.T. Papanicolaou, B.L. Rhoads, L.C. Hernandez Rodriguez, S. Roque-Malo, S. Schaeffer, A. Stumpf, A. Ward, L. Welp, C.G. Wilson, Q. Yan, S. Zhou, Emergent role of critical interfaces in the dynamics of intensively managed landscapes, Earth Sci. Rev. 244 (2023) 104543, https://doi.org/10.1016/j.earscirev.2023.104543.
[46] Y. Xia, D. Zhao, X. Yan, W. Hu, J. Qiu, X. Yan, A new framework to model the distributed transfer and retention of nutrients by incorporating topology structure of small water bodies, Water Res. 238 (2023) 119991, https://doi.org/10.1016/j.watres.2023.119991.
[47] W. Liu, T. Liu, Z. Liu, H. Luo, H. Pei, A novel deep learning ensemble model based on two-stage feature selection and intelligent optimization for water quality prediction, Environ. Res. 224 (2023) 115560, https://doi.org/10.1016/j.envres.2023.115560.
[48] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[49] X. Bai, W. Shen, P. Wang, X. Chen, Y. He, Response of non-point source pollution loads to land use change under different precipitation scenarios from a future perspective, Water Resour. Manag. 34 (2020) 3987–4002, https://doi.org/10.1007/s11269-020-02626-0.
[50] X. Tong, Q. Liang, G. Sander, G. Wang, X. Lai, A physically based model for non-point source pollutant wash-off process over impervious surfaces, Water Resour. Res. 61 (6) (2025) e2024WR038791, https://doi.org/10.1029/2024WR038791.