Predicting effects of non-point source Pollution emission control schemes Based on VMD-BiLSTM and MIKE21

Xianqi Zhang
Water Conservancy College, North china University of Water Resources and Electric Power
Yu Qi
978665082@qq.com
Water Conservancy College, North china University of Water Resources and Electric Power
Fang Liu
Water Conservancy College, North china University of Water Resources and Electric Power
Haiyang LI
Water Conservancy College, North china University of Water Resources and Electric Power
Shifeng Sun
Water Conservancy College, North china University of Water Resources and Electric Power

Research Article

Keywords: Non-point source pollution, Simulation prediction, VMD-BiLSTM, MIKE21, Weihe river

Posted Date: May 30th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2965021/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Environmental Modeling & Assessment on February 26th, 2024. See the published version at https://doi.org/10.1007/s10666-024-09959-9.

Predicting effects of non-point source Pollution emission control schemes Based on VMD-BiLSTM and MIKE21

Xianqi Zhang[a,b,c], Yu Qi[a,*], Fang Liu[a], Haiyang Li[a], Shifeng Sun[a]

[a]*Water Conservancy College, North China University of Water Resources and Electric Power, Zhengzhou 450046, China*

[b]*Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Zhengzhou 450046, China*

[c]*Technology Research Center of Water Conservancy and Marine Traffic Engineering, Henan Province, Zhengzhou 450046, China*

***Corresponding authors**: Yu Qi. Email: 978665082@qq.com; phone: +86-176-3429-5082

Authors:

Xianqi Zhang. E-mail: zxqi@163.com; phone: +86-158-3719-7937

Yu Qi. Email: 978665082@qq.com; phone: +86-176-3429-5082

Fang Liu. Email: 3543471502@qq.com; phone: +86-131-4760-0501

Haiyang Li. Email: 2415623753@qq.com; phone: +86-156-7058-9873

Shifeng Sun. Email: 1035136450@qq.com; phone: +86-157-3856-1668\

Abstract

Controlling non-point source (NPS) pollution is crucial for implementing water environment management, and simulating the water quality response to NPS pollution emission control schemes is of great importance. Variational mode decomposition (VMD) can overcome endpoint effects and modal aliasing issues, effectively separating intrinsic mode components. Bidirectional long short-term memory (BiLSTM) can fully mine the information contained in time series and has good predictive performance. MIKE21, when coupled with the Ecolab module, can well simulate the diffusion process of NPS pollution. The Weihe River water environment prediction model was constructed using VMD-BiLSTM and MIKE21, with ammonia nitrogen (NH_3-N), total phosphorus (TP), and chemical oxygen demand (COD) as pollution indicators, showing the water quality response of the Weihe River within a few years after the implementation of agricultural and urban NPS pollution emission control schemes. Among them, the COD concentration decreased by up to 71.3%, the NH3-N concentration decreased by up to 31.4%, and the TP concentration decreased by up to 43.1%. The results show that the water quality of the Weihe River can be significantly improved by controlling NPS pollution emission, and reducing agricultural NPS pollution emission is key to decreasing ammonia nitrogen and total phosphorus concentrations and improving water quality.

Key Words: **Non-point source pollution; Simulation prediction; VMD-BiLSTM; MIKE21; Weihe river**

1.Introduction

NPS pollution has become a major factor leading to various water environment issues, such as river pollution, deterioration of aquatic ecosystems, and severe eutrophication (Xiang et al., 2017; Giri et al., 2018; Fu et al., 2020). According to the "Second National Pollution Source Census Bulletin" released by the Chinese government in 2020, it is estimated that NPS contributes to 84% of total water pollution. The deterioration of the water environment has caused serious damage to aquatic environments and organisms and has limited the sustainable development of society and economy (Sun et al., 2018). However, existing policies mainly focus on controlling point source (PS) pollution, such as improving sanitation facilities and sewage treatment plants in urban and rural areas, while the dispersed NPS pollution has been somewhat overlooked (Yang et al., 2011; Tong et al., 2017). It has been proven that the implementation of emission control schemes, including low impact development (LID) and best management practices (BMPs), can effectively reduce NPS pollution from urban and agricultural land use (Taghizadeh et al., 2021). Before implementing NPS pollutant control schemes, simulating water quality response predictions under emission control plans is crucial for decision-making in water environment management. With the help of field monitoring and model predictions, the study of NPS pollution characteristics and aquatic environment responses has attracted widespread attention from scholars in related fields internationally, helping to draft pollution control plans (Chen et al., 2018). In previous research, researchers typically used historical hydrological data to simulate non-point source pollutants, for example, Ji et al. (2022) built a water environment model for the Beijing sub-center using MIKE11, selecting measured hydrological data input of historical significant rainfall events to highlight the impact of emission control schemes on water quality. Ouyang et al. (2008) used SWAT to construct a water environment model for Bazhong City, Sichuan Province, and used daily meteorological data from 1996 to 2005 to simulate basin climate conditions. Hou et al. (2021) simulated nitrogen transport through MIKE SHE and MIKE11 coupled with Ecolab. However, since runoff has a significant impact on the water environment carrying capacity, and the annual variability of runoff in some areas is considerable, using historical data cannot adequately demonstrate the changes in water quality over a period after the implementation of emission control schemes. As a result, the support provided for decision-making on water pollution control schemes is not sufficiently robust.Data-driven models aim to derive linear or nonlinear relationships between explanatory variables and target variables based on a large amount of input data, making them a type of black-box model. Deep learning is a kind of data-driven model that has been proven to be a valuable tool in various hydrological applications. Initially applied to speech-to-text conversion, machine translation, and sequence data processing, Long Short-Term Memory (LSTM) neural networks have been used for numerous hydrological applications, including soil moisture prediction, flood forecasting, and groundwater level prediction. BiLSTM, consisting of two oppositely directed LSTM, has recently been demonstrated to outperform conventional unidirectional LSTM. For example, it was found that, during calibration and validation periods, BiLSTM could predict groundwater level fluctuations better than unidirectional LSTM. This study aims to simulate the river water quality response over several years following the implementation of pollution emission control schemes through the coupling of VMD-BiLSTM and MIKE21. This approach allows for the prediction and evaluation of the effectiveness of nonpoint source pollution emission control schemes.

In order to achieve simulation predictions of water quality response within a certain period after the implementation of emission control schemes, we have applied deep learning techniques to the water environment model. Deep learning is a type of data-driven model that has been proven to be a valuable tool in various applications (Lee et al., 2020). For example, it was initially applied to speech-to-text conversion, machine translation, and sequence data processing. Among them, LSTM networks have been used in many hydrological applications, including soil moisture prediction (Fang et al., 2017), flood forecasting (Hu et al., 2018), and groundwater level prediction (Zhang et al., 2018). BiLSTM consists of two LSTMs with opposite directions and has recently been shown to outperform traditional unidirectional ones. For example, Ghasemlounia et al. (2021) found that, during calibration and validation periods, BiLSTM can predict groundwater level fluctuations better than unidirectional LSTM. Zhang et al. (2023) found that BiLSTM has higher accuracy in rainfall prediction compared to LSTM. In this study, we couple VMD-BiLSTM with MIKE21 to simulate the river water quality response within one year after the implementation of pollution emission control schemes, thus predicting and evaluating the effectiveness of non-point source pollution emission control schemes.

The Weihe River is one of the most important tributaries of the Yellow River, originating from the Niaoshu Mountain in Gansu Province, passing through the provinces of Shaanxi, Gansu, and Ningxia, and finally joining the Yellow River in Tongguan County, Shaanxi Province. The Weihe River Basin is bordered by the millennium imperial capital of Luoyang to the east, connected to Nanyang City by the Funiu Mountain to the south, and overlooks the ancient city of Xi'an to the west. It has been the center of various economic and cultural exchanges and integration since ancient times. The Weihe River Basin is located in the northern Loess Plateau region of China and is one of the most severely affected areas by soil and water loss in the Yellow River Basin. It is also one of China's important grain, cotton, and oil production areas and industrial production bases. The Weihe River serves as the ultimate receiving water body for pollutants in the basin (Yan et al., 2022), and its water quality directly affects the aquatic environment of the Yellow River Basin. In the past decade, Shaanxi Province has become the strictest province in the Yellow River Basin for water pollution control to address severe water pollution issues. In this study, we selected the key river sections in the Weihe River for future water pollution control as the research area. According to data from the China National Environmental Monitoring Center, the current river water quality in this area is all below Grade V, indicating that the water environment urgently needs improvement. We established a water environment model for the Weihe River based on the coupled VMD-BiLSTM and MIKE model, set different emission control measures for NPS pollution, and used COD, NH3-N, and TP as indicators to analyze the water quality of the Weihe River after the implementation of emission control schemes and evaluate the effectiveness of these plans. This research provides support for water environment management in the Weihe River and the optimization of non-point source pollution control schemes.

2. Study area

The Weihe River flows 502.4 km within Shaanxi Province, passing through Xianyang City, Xi'an City, and joining the Yellow River in Tongguan County, Weinan City. We selected a 136.6 km long section of the Weihe River, which lies within the area demarcated by 34° 08′ N, 34° 42′ N, 108° 39′ E, and 109° 55′ E for our study. The terrain in the study area is relatively flat, sloping from southwest to northeast, with elevations ranging from 393 to 332 meters and slope gradients between 0.2‰ and 0.6‰. The average annual rainfall is 740mm, with significant interannual variability, and more than 80% of the total annual precipitation occurring between July and September (Qiu et al., 2022). The Weihe River is the largest tributary of the Yellow River, with an average annual runoff of 75.7

billion cubic meters, accounting for approximately 13% of the Yellow River's average annual runoff.

As a center of cultural convergence and an essential grain, cotton, and oil-producing area and industrial production base, high-quality water environments are a strategic necessity for promoting the sustainable development of industry, agriculture, and tourism in the Weihe River Basin. However, due to upstream and internal water pollution, the water quality is currently unsatisfactory. Approximately 90% of the annual rainfall and 70% of the urban sewage in the upstream counties of Shaanxi Province are received by the Weihe River, Additionally, PS and NPS pollution within the study area exacerbate the pollution problem (Jiake et al., 2011). Both organic and inorganic pollutants have damaged the surface water environment in the study area. According to data from the China National Environmental Monitoring Center, more than 90% of the water quality in the Weihe River section within the study area is currently classified as Class V or below, with the main pollutants exceeding the standard limits being NH_3-N, TP, and COD. Therefore, there is an urgent need to improve the water environment in the Weihe River Basin.

3. Material and methods

3.1 Model description

3.1.1 VMD

Variational Mode Decomposition (VMD) is an adaptive, fully non-recursive mode variation and signal processing method that can determine the number of decomposition components based on actual conditions. While overcoming the endpoint effects and mode mixing issues present in traditional Empirical Mode Decomposition (Liu et al., 2018), it effectively separates the intrinsic mode functions and decomposes them into multiple frequency-scale and relatively stationary subsequences (Ma et al., 2020). The core of VMD is to construct and solve the variational problem. Assuming the original signal f is decomposed into K finite-width mode components with central frequencies, and the sum of the estimated bandwidths of each mode is minimized, the constraint condition is that the sum of all modes is equal to the original signal. The corresponding constrained variational expression is as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[(\delta(t) + j / \pi t)^* u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

$$s.t \sum_{k=1}^{k} u_k = f \quad (2)$$

Here, K represents the number of decomposed modes, $\{u_k\}$, $\{\omega_k\}$ correspond to the k-th mode component and the central frequency after decomposition, respectively. $\delta(t)$ is the Dirac delta function, * denotes the convolution operator, and f is the original time series.

3.1.2 BiLSTM

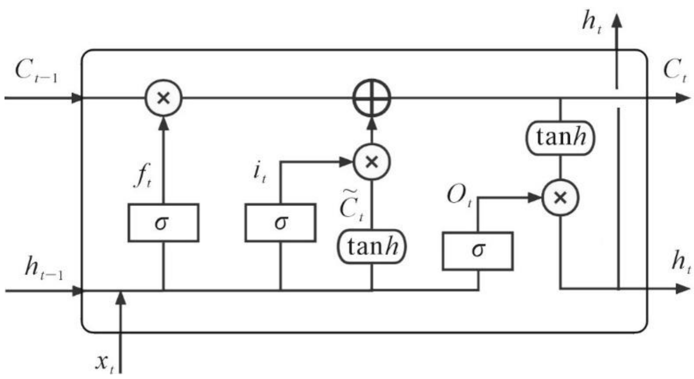


Fig.1 Structure of a single LSTM neuron

LSTM is a special type of Recurrent Neural Network (RNN) that not only inherits the advantages of RNN but also addresses the gradient vanishing and gradient exploding problems present in RNNs (Zhou et al., 2019). It is widely used in time series processing and prediction. Different from regular RNNs, the core concept of LSTM lies in the cell state and gate structure. The cell state ensures that information can be passed down the sequence, while the gate structure determines which information should be saved or forgotten during the training process. This unique network structure enables LSTM to capture long-term dependencies in time series while learning current information, making it more suitable for processing and predicting events with longer intervals and delays in time series. The structure of a single LSTM neuron is shown in **Fig.2**. Below are the formal representations of the three gates in LSTM:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

$$O_t = \sigma(W_O [h_{t-1}, x_t] + b_O) \tag{7}$$

$$\mathrm{h_t} = \mathrm{O_t} * \tanh(\mathrm{C_t}) \tag{8}$$

Here, W_f, W_i, W_C and W_O are the weight matrices; b_f, b_i, b_C and b_O are the bias vectors; h_{t-1} and h_t are the input at the previous time step and the output at the current time step, respectively; C_{t-1} and C_t are the cell states at the previous and current time steps, respectively; $\tilde{C}_t$ is the information state passing through the input gate; and σ represents the sigmoid function.

Traditional neural networks are always trained through forward propagation, which cannot fully exploit the inherent information in daily runoff time series and results in low data utilization. The structure of the BiLSTM network model is shown in **Fig.2**. Compared with the traditional unidirectional LSTM network, its prominent feature is the construction of a bidirectional recurrent neural network that includes both forward and backward propagation, overcoming the deficiency of insufficient data mining in the unidirectional LSTM neural network.

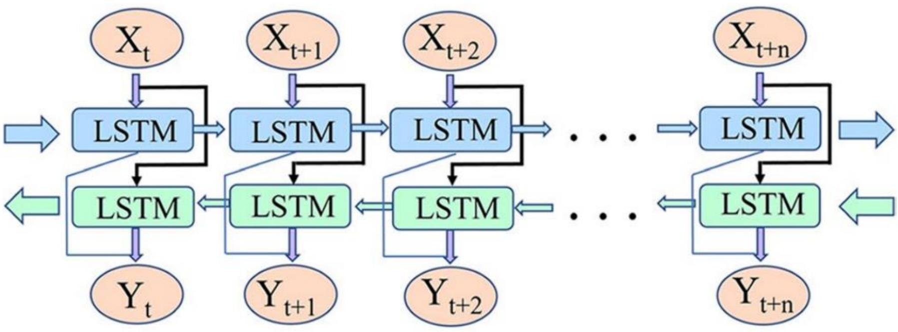


Fig.2 BiLSTM model structure

To use VMD-BiLSTM for daily runoff prediction, the runoff sequence needs to be decomposed by VMD first, thus obtaining multiple stationary intrinsic mode functions ($\mathrm{IMF}_{(i=1\sim n)}$) and a trend component (r). Then, the obtained components are normalized and the input and output of the LSTM model are finally determined. This model has good predictive performance and can be used for runoff prediction at different time scales.

3.1.1 MIKE21

MIKE21 is a numerical hydrodynamic model developed by the Danish DHI company, which can be used for simulating and analyzing processes such as water movement, water quality, and sediment transport (Ramteke et al., 2020). MIKE21 is one of the widely used hydrodynamic models, and it can be adjusted for different levels of model accuracy and complexity according to various application scenarios and requirements. The MIKE21 model is based on numerical methods and physical equations and can simulate and analyze the hydrodynamic and water quality characteristics of different water bodies such as oceans, rivers, lakes, and reservoirs. The core of the model is the description of physical processes such as convection, diffusion, turbulence, material transport, and sedimentation, which can accurately calculate the changes in water parameters such as flow velocity, water level, water temperature, salinity, dissolved oxygen, nutrients, and suspended matter. In terms of environmental protection, MIKE21 can be used for predicting and analyzing pollutant diffusion and transport, providing a basis for environmental monitoring and emergency response. When coupled with the Ecolab module, it can effectively simulate the diffusion process of non-point source pollution.

MIKE21 uses the Hydrodynamic (HD) module for water movement simulation. The HD module is the core and basic computational module of MIKE21. It is based on the control equation of three-dimensional flow, which can be integrated along the water depth, averaged along the water depth, and can obtain the two-dimensional shallow water flow averaged along the water depth. The equation is represented as:

$$\frac{\partial \zeta}{\partial t}+\frac{\partial p}{\partial x}+\frac{\partial q}{\partial y}=\frac{\partial d}{\partial t} \quad (9)$$

$$\frac{\partial p}{\partial t}+\frac{\partial}{\partial x}\left(\frac{p^2}{h}\right)+\frac{\partial}{\partial y}\left(\frac{pq}{h}\right)+gh\left(\frac{\partial \xi}{\partial x}\right)+\frac{gp\sqrt{p^2+q^2}}{C^2h^2} \quad (10)$$

$$-\frac{1}{\rho}\left[\frac{\partial}{\partial x}(h\tau_{xx})+\frac{\partial}{\partial y}(h\tau_{xx})\right]-\Omega q-fVV_x+\frac{h}{\rho}\frac{\partial}{\partial x}p_a=0 \quad (11)$$

$$\frac{\partial p}{\partial t}+\frac{\partial}{\partial y}\left(\frac{p^2}{h}\right)+\frac{\partial}{\partial x}\left(\frac{pq}{h}\right)+gh\left(\frac{\partial \xi}{\partial y}\right)+\frac{gp\sqrt{p^2+q^2}}{C^2h^2} \quad (12)$$

$$-\frac{1}{\rho}\left[\frac{\partial}{\partial y}(h\tau_{xx})+\frac{\partial}{\partial y}(h\tau_{xx})\right]-\Omega q-fVV_y+\frac{h}{\rho}\frac{\partial}{\partial y}p_a=0 \quad (13)$$

Here, $f=2\omega\sin\phi$, ξ is the free water level, m. t is time, d. x, y are spatial coordinates, m. p, q is the flow densities in the x and y direction, respectively, $m^2\,/\,s$. d is the water depth, m. h is the water depth, m. g is the acceleration of gravity, $m^2\,/\,s$, g=9.8$m^2\,/\,s$. C is the Chézy coefficient, $m^{1/2}\,/\,s$. ρ is the density of water, kg/m3. τ_{xx}, τ_{xy}, τ_{yy} are the horizontal shear stress in the x-direction, the vertical shear stress in the x-direction, and the vertical shear stress in the y-direction, P_a. Ωq is the Coriol coefficient. f is the wind resistance coefficient. V, Vx, Vy is the wind speed and the wind speed components in the x and y directions, m/s. P is atmospheric pressure, P_a.

The water quality simulation is carried out using the Ecolab module, which is a new water quality and aquatic ecology tool developed by the Danish Hydraulic Research Institute (DHI) based on the concept of traditional water quality modules. The transport equation in the water body is as follows:

$$\frac{\partial h\bar{c}}{\partial t}+\frac{\partial h\bar{u}\bar{c}}{\partial x}+\frac{\partial h\bar{u}\bar{c}}{\partial y}=h\left[\frac{\partial}{\partial \chi}\left(D_h\frac{\partial}{\partial x}\right)+\frac{\partial}{\partial y}\left(D_h\frac{\partial}{\partial y}\right)\right]\bar{C}-hk_P\bar{C}+hC_s \qquad (14)$$

Here, C_s is the concentration of the scalar variable. D_h is the horizontal diffusion coefficient. *t* is the time. k_P is the scalar solution coefficient.

The calculation formula for the water quality model is as follows:

$$\frac{\partial c}{\partial t}+u\frac{\partial c}{\partial x}+v\frac{\partial c}{\partial y}+w\frac{\partial c}{\partial z}=D_x\frac{\partial^2 c}{\partial z^2}+D_y\frac{\partial^2 c}{\partial z^2}+D_z\frac{\partial^2 c}{\partial z^2}+S_c+P_c \qquad (15)$$

Here, c—concentration of Ecolab state variable. u, v, w—The flow velocity component of the convection term. D_x, D_y, D_z—The dispersion coefficient of the diffusion term. S_c—Source and sink items. P_c—Biochemical reactions of Ecolab.

3.2 Data input and model setup

MIKE21 simulates the water environment through two modules: the hydrodynamic (HD) module and the water quality (Ecolab) module. In this study, we obtained 30m-resolution DEM data for the research area through the Geospatial Data Cloud (www.gscloud.cn) and combined it with remote sensing imagery for analysis and processing, resulting in topographic data for the Weihe River basin from Xianyang City to Weinan City in Shaanxi Province. To predict hydrological information for a period in the future, we obtained historical hydrological measurement data from the Xianyang and Hua County hydrological stations through the Shaanxi Provincial Water Resources Bureau, which was strictly processed and used to train the VMD-BiLSTM model. We used the predicted values from the model as boundary data input to the MIKE21 hydrodynamic (HD) module. We also obtained monitoring data at Xianyang Tieqiao, Xinfeng Town Bridge, and Shawangdu, three China's National Surface Water Environmental Quality Monitoring Points, from the China National Environmental Monitoring Center (http://www.cnemc.cn), to analyze the water quality situation in the Weihe River within the research area and to input it as upstream water quality boundary conditions for the MIKE21 water quality (Ecolab) module.

Based on the main non-point source pollutants and the water quality status of the Weihe River in the research area, this study selected BOD, NH_3-N, and TP as the main water quality indicators in the water quality (Ecolab) module. Using the data provided by the China National Environmental Monitoring Center, we set the initial water quality conditions for the Weihe River in the study area to the Class V water quality standard in the Surface Water Environmental Quality Standards of the People's Republic of China, with the pollutant concentrations for each standard shown in **Table.1**.

Table 1

Surface water environmental quality standard (GB3838-2002) unit: mg/L

Serial number	State variables	I	II	III	IV	V
1	DO⩾	7.5	6	5	3	2
2	NH_3-N⩽	0.15	0.5	1	1.5	2
3	TP⩽	0.02	0.1	0.2	0.3	0.4
4	COD	15	15	20	30	40

In the research area, where all wastewater sources are treated by wastewater treatment plants, this model focuses on NPS pollution and excludes the interference of incoming water pollution. The locations where non-point source pollution enters the river are determined by the intersection of the river channel and the sub-catchments along the riverbank, while the sub-catchments are delineated by the elevation and stormwater pipes within the research area. The model identifies 137 NPS pollution sources, which can be classified into urban pollution sources and rural pollution sources. We applied the EcoHydrological Assessment Tools (EcoHAT) model to generate NPS pollutant loads (Dong et al., 2014), including COD, NH_3-N, and TP, for each pollution source. The calculated pollution load results are used as input data for water quality modeling in MIKE21. Other water quality parameters were determined from previous relevant research and measured data for the Weihe River basin. Coupling catchment pollutant loads and water environment models ensures the accuracy of input data compared to empirical formulas and assumptions. A description of the two simulation schemes can be found in this study's flowchart **Fig.3**.

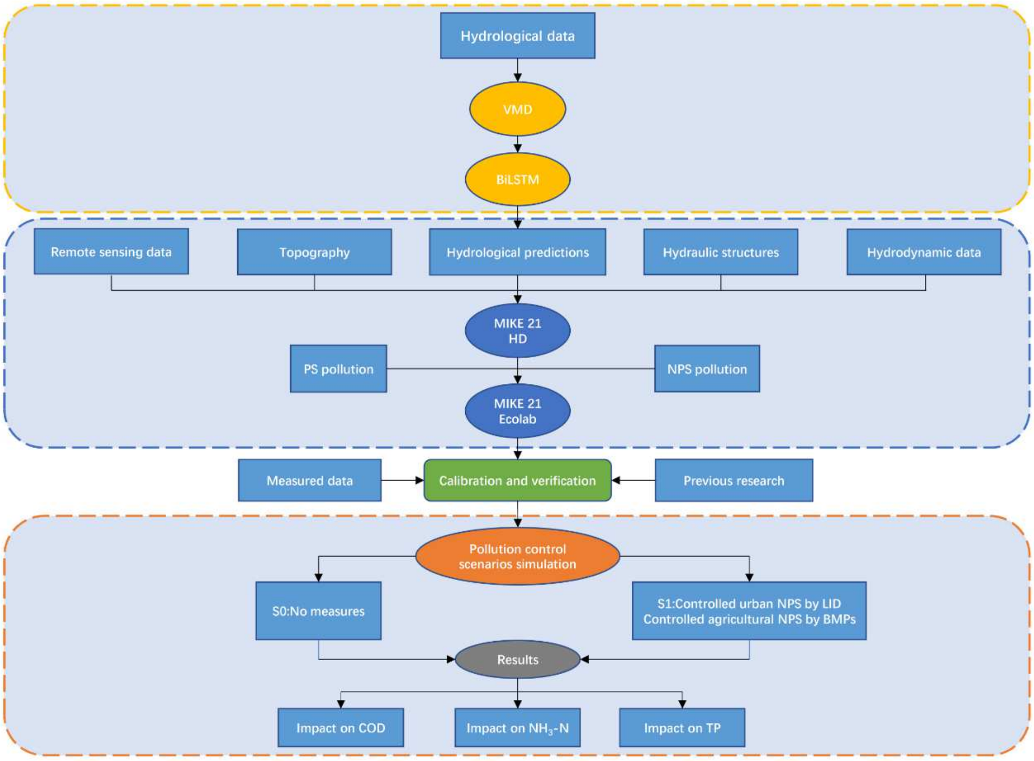


Fig.3 The flowchart of this study

3.3 Evaluation criteria of model performance

In this study, we assess the model's performance using three statistical indicators: R^2, MAPE, and NSE, to ensure the accuracy of the model output data. These indicators have been widely used in previous research and can effectively represent the relationship between the model output values and the actual values. The specific descriptions of the indicators are as follows:

3.3.1 Correlation index（R^2）

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}(M_i - S_i)^2}{\Sigma_{i=1}^{N}(M_i - \bar{M}_i)^2 \Sigma_{i=1}^{N}(S_i - \bar{S}_i)^2} \qquad (16)$$

Where M is the measured value, S is simulated value, N is the number of data sequence, I is the mean value of the measured value.$\bar{S}_i$ is mean value of simulated value. The value of R^2 ranges from 0 to 1 with model accuracy are considered to be excellent for R^2 values above 0.85, perfect for R^2 values between 0.65 an 0.85, good for values between 0.5 to 0.65, and poor for values below 0.5(Henriksen et al., 2003)

3.3.2 Nash-Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{N}(M_i - S_i)^2}{\Sigma_{i=1}^{N}(M_i - \bar{M}_i)^2} \qquad (18)$$

Where all variables are defined as in eq1. The value of NSE ranges from minus infinity to 1 with model accuracy increasing as the NSE value approaches 1. A negative NSE value indicates that the simulated value is far from the measured value and unsatisfactory simulation result(Yang et al., 2018).

3.3.3 Mean absolute percentage error (MAPE)

$$MAPE = \sum_{i=1}^{N} \frac{|M_i - S_i|}{M_i} \times \frac{100}{N} \qquad (17)$$

Where all variables are defined as in Eq1. Model accuracy increases as MAPE value approaches 0.

3.4 Data decomposition and noise reduction processing

We applied VMD to decompose a total of 4748 daily runoff time series data of the Weihe River in the study area from 2010 to 2022, and the decomposition results are shown in **Fig.4**.

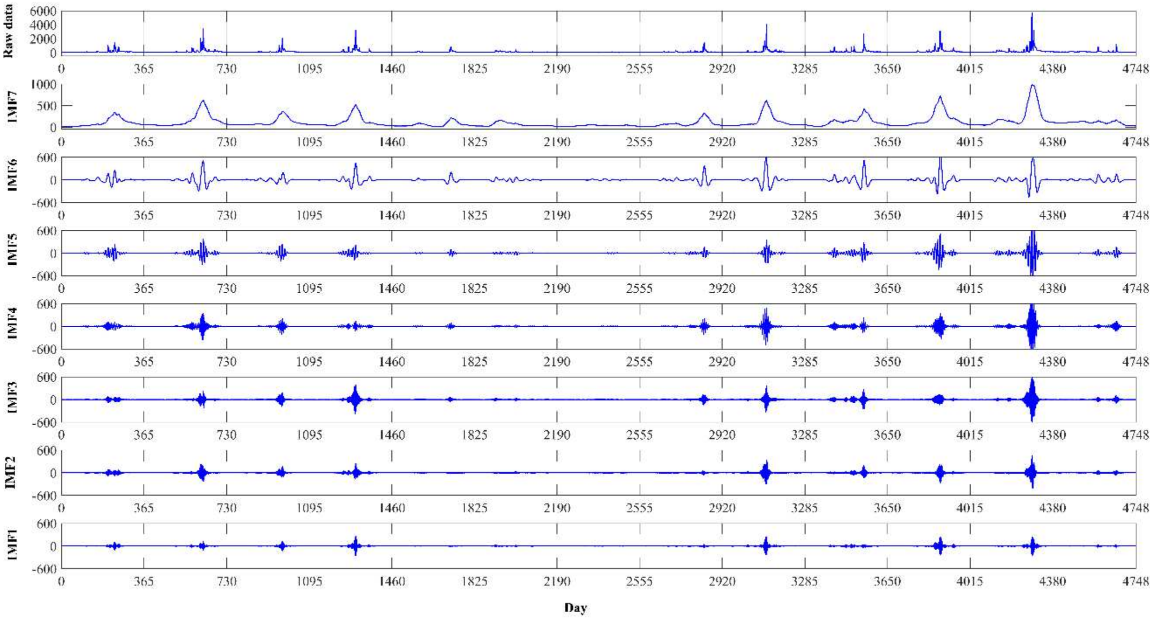


Fig.4 VMD of runoff sequence

From the **Fig.4**, it can be seen that the original daily runoff time series was decomposed into seven components ranging from high to low frequency, i.e., IMF1, IMF2, ..., IMFn. With the increase of data volume, it not only reduces noise but also identifies the hidden periodic oscillation changes in the runoff sequence, which helps the model to better analyze the intrinsic transformation rules of the time series and improve the prediction effect.

3.5 Runoff forecast

When using BiLSTM to predict runoff, we processed a total of 4,380 days of daily runoff data for the Weihe River from 2010 to 2021 as the training set, and 365 days of daily runoff data for 2022 as the test set. First, we apply the wavelet signal denoising(WSD) algorithm to denoise the high-frequency components obtained after decomposition. Then, the denoised results and low-frequency components are jointly input into the BiLSTM model to obtain the predicted values of each component. After extensive testing, we set the number of hidden units to 200, the gradient threshold to 1, and the initial learn rate to 0.005. After 125 rounds of training, we set the learn rate drop factor to 0.2 to reduce the learning rate. The prediction results of each component were then reconstructed to obtain the daily runoff prediction data for the Weihe River in 2022, and the comparison between the prediction data and the original data of the test set is shown in **Fig.5**. Evaluating the VMD-BiLSTM prediction results for the test set, we obtained an NSE of 0.89, R^2 of 0.92, and MAPE of 8.39%. As can be seen from **Fig.5**, VMD-BiLSTM cannot accurately predict flood peak flows during the flood season; nevertheless, the overall prediction effect for annual runoff is quite good. The R^2, MAPE, and NSE indicators demonstrate the accuracy of the prediction results, suggesting that using VMD-BiLSTM to predict the Weihe River runoff is feasible. Finally, we used a total of 4,748 days of daily runoff data from 2010 to 2022 as the training set, predicted 365 days of daily runoff data for the Weihe River in 2023, and used it as the upstream boundary data for the water environment model to simulate non-point source pollution for the next year.

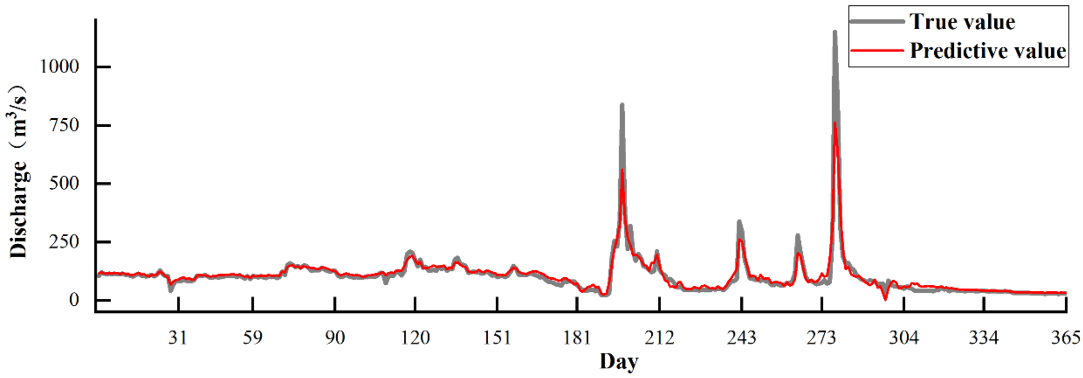
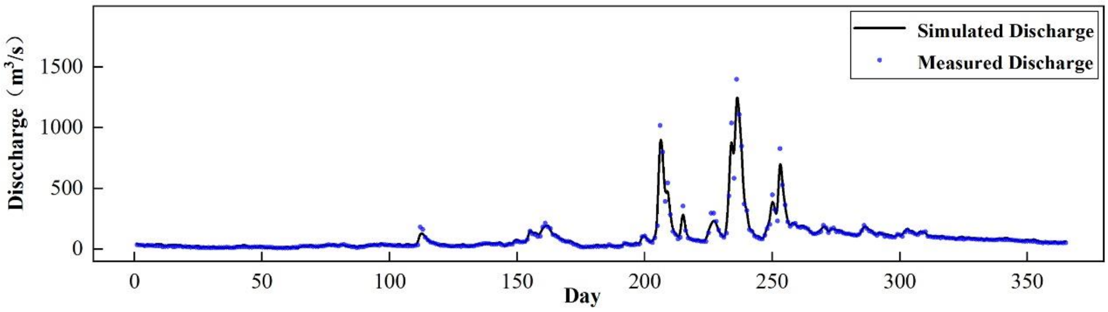


Fig.5 Model prediction curve

3.6 MIKE21 model calibration and verification

In this study, the model is calibrated by comparing its output data with the measured hydrological data from 2010. The roughness Manning coefficient is selected within the range of 0.022-0.03 sm1/3 using the trial-and-error method. The initial water depth and initial flow rate are determined through measured data. The simulation process starts from January 1, 2010, and ends on December 31, 2010, with a time step of 30 seconds. The accuracy of the model is evaluated based on R^2, MAPE, and NSE, as shown in Table 2. The comparative analysis of the model output data and the measured hydrological data is presented in **Fig.6**.

Xianyang hydrological station



Huaxian hydrological station

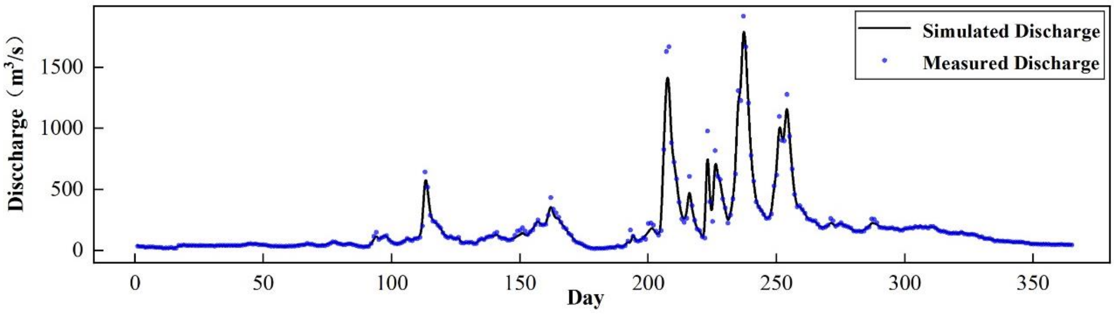
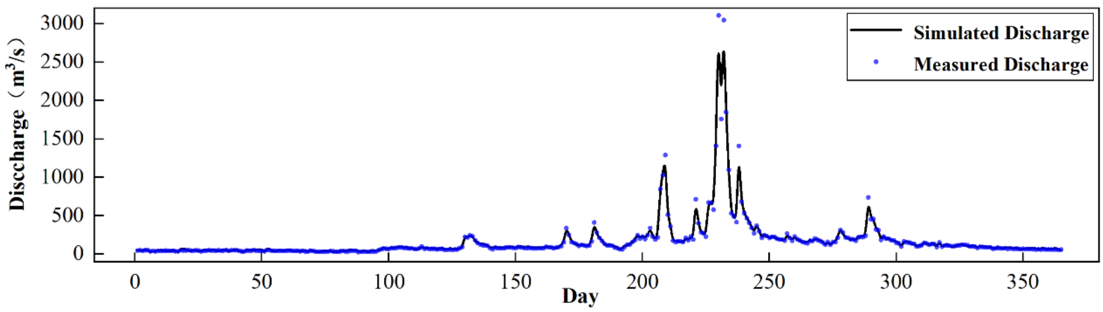


Fig.6 Measured and simulated values of hydrodynamic parameters for calification: (a) discharge at Xianyang station and (b) discharge at Huaxian station.

As shown in the **Fig.6**, the variables output by the model are similar to the measured data. From the comparison chart, it can be seen that the model cannot accurately predict flood peak flow during the flood season. Despite this, the R^2, MAPE, and NSE values of the model reach 0.91, 0.93, and 14.9%, respectively, indicating that the model has relatively high accuracy.

The hydrodynamic simulation performance of the model is verified by comparing the measured hydrological data from 2020 and the model output results. The simulation period is from January 1, 2018, to December 31, 2018, with a time step set to 20 seconds. The values of R^2, MAPE, and NSE are shown in **Table 2**. The comparative analysis of measured values and model output results can be seen in **Fig.7**.

Xianyang hydrological station



Huaxian hydrological station

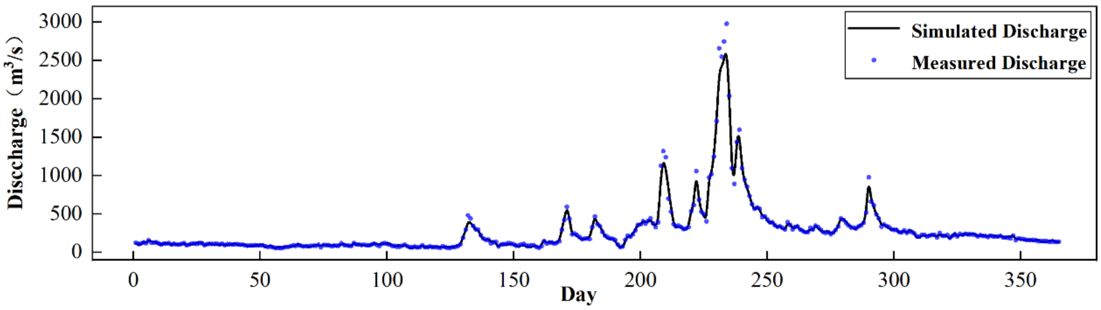


Fig.7 Measured and simulated values of hydrodynamic parameters for model verification: (a) discharge at Xianyang station and (b) discharge at Huaxian station.

As shown in the **Fig.7**, the model output results and the measured values show the same trend. Similar to the model calibration, the model cannot accurately predict peak flow rates. Although there is some mismatch between the model output results and the measured values, the R^2, MAPE, and NSE indicators demonstrate the reliability of the model.

Table 2

Statistics for model calibration and verification

Sampling station	Period	Hydrodynamic Parameters	*NSE*	R^2	*MAPE*
Xianyang	Calibration	Discharge	0.93	0.94	23.5%
Xianyang	Verification	Discharge	0.89	0.91	19.1%
Huaxian	Calibration	Discharge	0.87	0.89	14.9%
Huaxian	Verification	Discharge	0.91	0.93	11.2%

With the limitations of limited available data, the overall performance of the model in reproducing the water environment in the study area is acceptable. Due to the lack of long-term water quality monitoring stations and historical measured data in the study area, water quality parameters were not calibrated or validated. We referred to previous relevant studies on the Weihe River Basin and selected the relevant water quality parameters. The decay coefficients of COD, NH_3-N, and TP were set to 0.43, 0.3, and 0.037 per day, respectively, and the diffusion coefficient was set to 15.

3.7 NPS pollution control scenarios

The land use analysis of the study area is shown in **Fig.8**. The western part of the study area is highly urbanized, with densely populated residential areas, while the eastern part consists of large areas of farmland mixed with a small portion of urban residential areas. This study simulates and compares

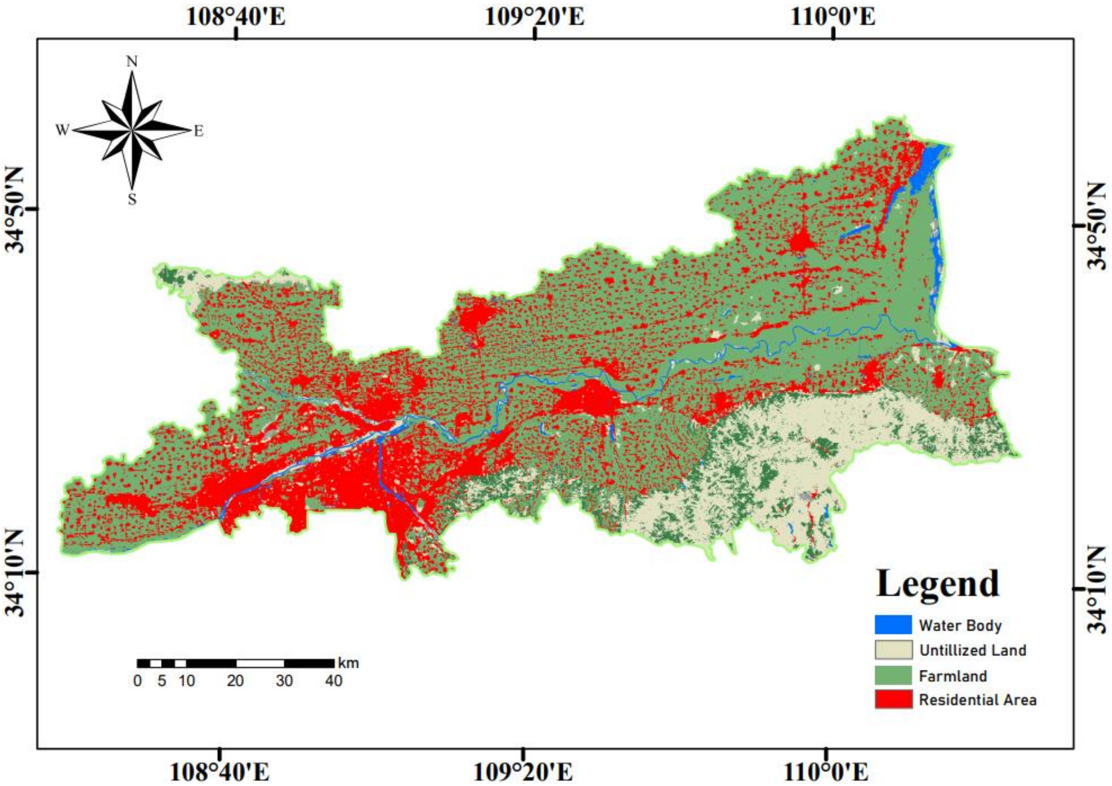


Fig.8 Land use in the study area

the results of two different scenarios to predict the water quality response after implementing the NPS pollution emission control measures, including agricultural NPS pollution and urban NPS pollution. Urban roads, residential areas, and unused land are the main causes of urban NPS pollution, and we assume that only NPS pollution generates COD. Previous research reports have shown that Low Impact Development (LID) facilities can reduce the concentrations of COD, NH3-N, and TP from urban pollution sources by 35%, 35%, and 40%, respectively. Therefore, in scenario S1, LID facilities such as bioretention ponds, permeable pavements, and grass swales are set up to regulate urban NPS pollution. On the other hand, farmland and grassland are the main sources of agricultural nonpoint source pollution; currently, Best Management Practices (BMPs) including soil and water conservation, reducing fertilizer use, and putting vegetative filter strips in places are widely used for agricultural NPS pollution control to reduce NH3-N and TP concentrations in pollution sources. Previous studies have shown that BMPs can reduce NH3-N and TP concentrations by 35% and 40%, respectively. In S1 (scenario 1), we consider the combined effects of LID facilities and BMPs. In S0(scenario 0), no measures are taken to control NPS pollution, and by comparing S0 and S1, we can more clearly see the water quality response after implementing NPS pollution emission control measures. In addition, all wastewater discharges in the study area are treated by wastewater treatment plants, which can eliminate the interference from other pollution sources.

4 Result and discussion

We used the water quality predictions for a flood season in 2023 generated by the VMD-BiLSTM model to highlight the impact of pollution emission control schemes on river pollutant concentrations under different scenarios. We selected two sections near the main NPS pollution sources to analyze river pollution under different land use types. Section A is located near the Xinfeng Town Bridge in Xi'an City, surrounded by highly urbanized land and densely populated residential areas. Section B is located near the Shawangdu in Weinan City, mainly surrounded by cultivated land. Both sections have measured water quality data from China's National Surface Water Environmental Quality Monitoring Points available for analysis.

4.1 Effect of different scenarios on COD

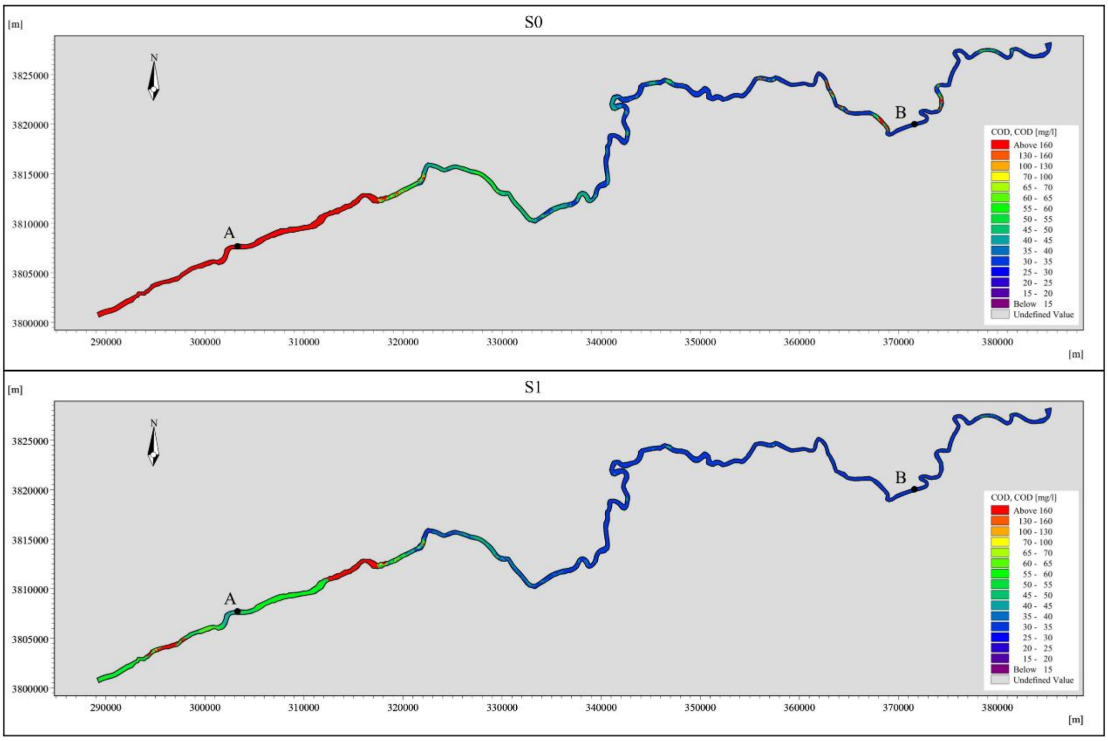


Fig.9 Effects of different scenarios on COD

The impact of different scenarios on COD is shown in **Fig.9**. Since we assume that only urban NPS pollution generates COD, the impact of different scenarios on the COD concentration near Section A is more significant. In S0, the part of the Weihe River passing through densely populated residential areas is severely polluted, with COD concentrations exceeding 160 mg/L. At the same time, the downstream part of the river is also polluted. In the river sections surrounded by cultivated land, the COD concentration is mainly below 40 mg/L. In S1, some parts of the Weihe River surrounded by urban areas still experience continuous severe pollution, but for the most part, especially near Section A, the COD concentration is significantly lower than in Scenario 0. Compared to Scenario 0, the water quality of the river sections surrounded by cultivated land is higher, with the majority of COD concentrations below 30 mg/L. The implementation of Scenario 1 results in a 71.3% reduction in COD concentration at Section A compared to Scenario 0, and a 22.1% reduction at Section B.

4.2 Effect of different scenarios on NH_3-N

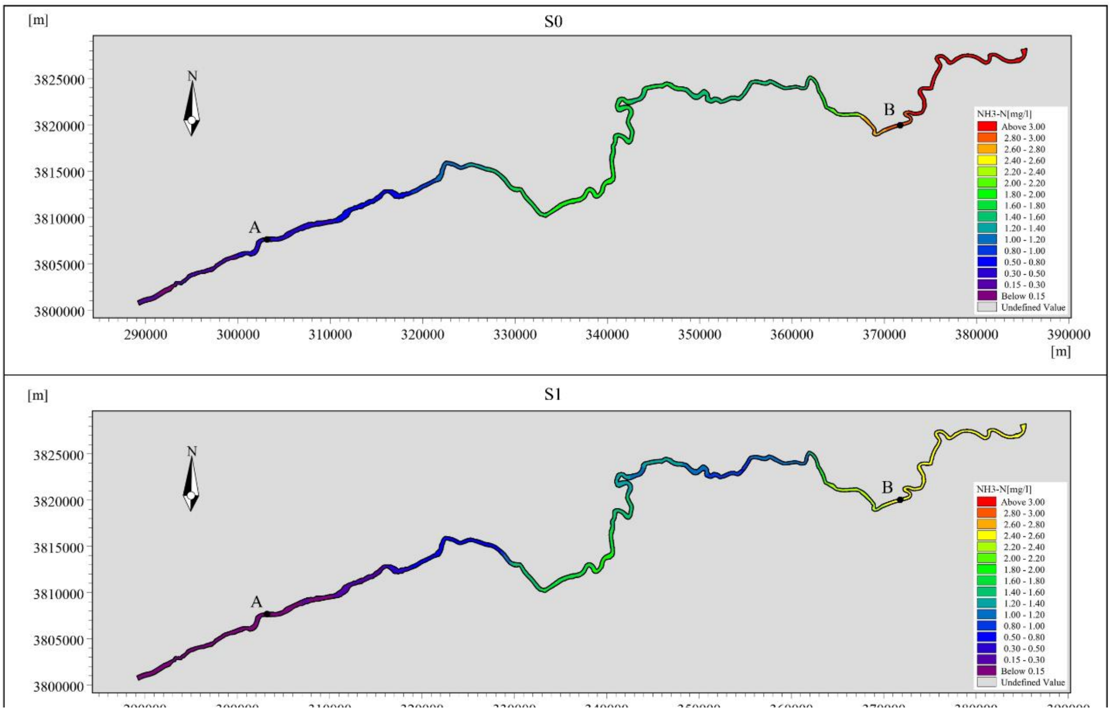


Fig.10 Effects of different scenarios on NH_3-N

The water quality response of NH_3-N under different scenarios is shown in **Fig.10**. In S0, the water quality pollution is most severe in the northeastern part of the study area, where the river is surrounded by a large amount of cultivated land, with NH_3-N concentrations exceeding 3 mg/L. In S1, due to the reduction of agricultural non-point source pollution, the water quality of the northeastern river has improved significantly compared to S0, with NH_3-N concentrations in most areas being lower than in S0. Compared to S0, the NH_3-N concentration at Section A decreased by 17.1%, and the concentration at Section B decreased by 31.4%. These results indicate that controlling non-point source pollution emissions can have a significant effect on improving river water quality, especially for agricultural non-point source pollution. Therefore, when considering economic value, controlling agricultural pollution has a greater impact on the treatment of NH_3-N in rivers.

4.3 Effect of different scenarios on TP

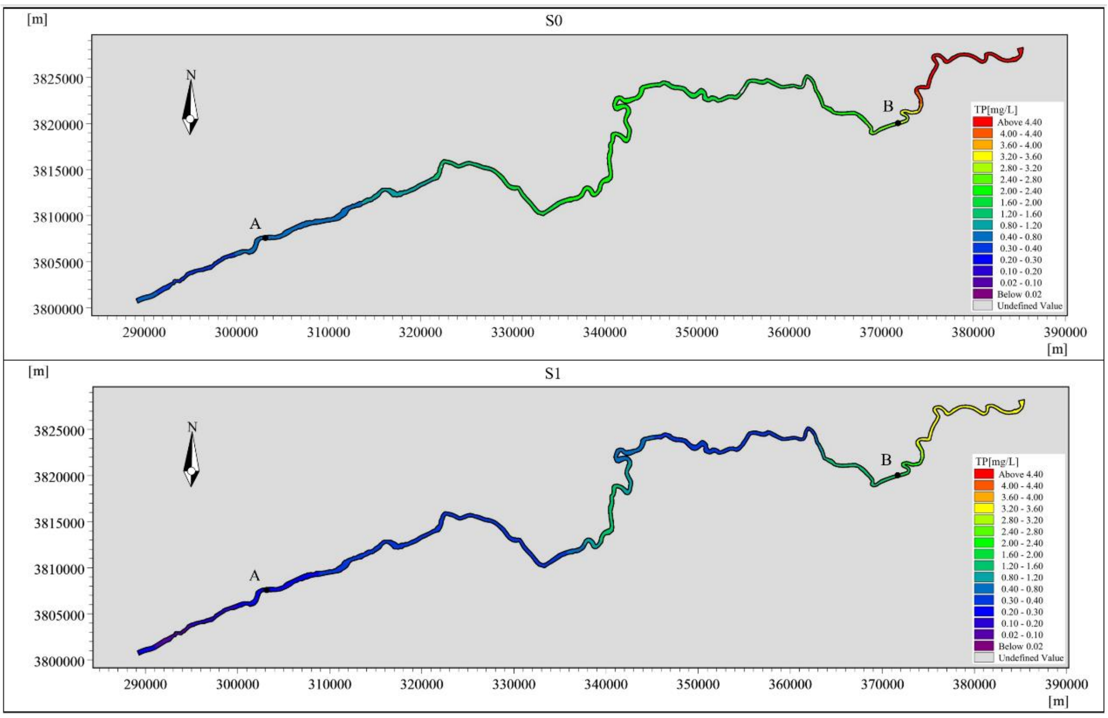


Fig.11 Effects of different scenarios on TP

The impact of different scenarios on TP in rivers is shown in **Fig.11**. In Scenario 0, most of the river areas are severely polluted. In Scenario 1, due to the reduction of agricultural NPS pollution, the TP concentration in the northeastern river surrounded by cultivated land in the study area is significantly lower than in Scenario 0. At the same time, the water quality around Section A also improves to some extent, with a decrease in TP concentration. However, in almost all areas of the two scenarios, TP concentrations exceed 0.3 mg/L, indicating severe river pollution. By comparing Scenario 0 with Scenario 1, we can conclude that controlling agricultural NPS pollution can generally improve water quality, as agricultural activities are the main source of TP. If measures to control both urban and agricultural non-point source pollution emissions are implemented simultaneously, the water quality in the study area will continue to improve. However, as the river water quality in most areas is below Class IV, more effective measures should be implemented to reduce TP.

Compared to Scenario 0, in Scenario 1, the TP concentration at Section A decreased by 23.7%, while the concentration at Section B decreased by 43.1%. These results indicate that TP concentrations in rivers vary with different land use types and the implementation of emission reduction measures. TP concentrations in river areas surrounded by cultivated land are significantly higher than those in areas surrounded by densely populated residential areas, and the effect of controlling agricultural NPS pollution on reducing TP concentrations in rivers is more significant than controlling urban NPS pollution.

4.4 Discussion

Previous studies have widely documented the ability of water environment-water dynamics coupling models to simulate and control NPS pollution and support local environmental policies. However, these studies lack attention to predicting hydrological conditions and analyzing water quality. This study fills this gap and designs a comprehensive framework for runoff prediction-water dynamics-pollution load-water environment simulation to evaluate the impact of NPS pollution control scenarios on the water quality of the Weihe River from Xianyang City to Weinan City. During the calibration and validation phase, the NSE values for simulating flow and water level were both higher than 0.87, and the R^2 values were higher than 0.89, indicating good performance in simulating hydrological and water dynamics processes. Due to limited available data, the overall performance of the model in the study area is acceptable.

Numerous studies have shown the effectiveness of LID and BMPs in influencing the transport of NPS pollutants and improving water quality. In this study, scenario analysis is used to explore the water quality response to implementing LID and BMPs measures to control urban and agricultural NPS pollution. In scenario S0 without pollution control measures, the concentration distribution patterns of NH_3-N and TP are relatively consistent, while that of COD is different, especially at section A. This may be due to the different characteristics of pollutant emissions in different land use areas. For section B surrounded by agriculture-dominated areas, farmland, livestock, poultry, fertilizer application, and household wastewater may be the main sources of NH_3-N and TP. The correlation between NH_3-N and COD cannot be determined.

The results of two design scenarios show that reducing agricultural NPS pollution through BMPs is key to reducing NH_3-N and TP concentrations. However, controlling COD pollution through LID to regulate urban NPS pollution sources is crucial, consistent with previous research conclusions. Residential and road construction sites are mainly located in the southwest of the study area, so it is recommended to strengthen the control of urban non-point source pollution from the source, migration, and terminal aspects. Proposed protective measures include controlling initial rainfall runoff by arranging distributed sponges, wetlands, and permeable roads, and setting up ecological embankments and buffer zones on both sides of the river. On the other hand, farmland and rural areas are mainly located in the northeast of the study area. To reduce agricultural non-point source pollution, it is recommended to improve the fertilizer application structure of farmland and regulate the operation of livestock and poultry farming. In addition, ecological embankments and buffer zones should also be set up on both sides of farmland ditches or ecological ditches should be constructed.

In the VMD-LSTM, MIKE21, and EcoHAT coupling system, we believe that only urban NPS pollution produces COD, while in reality, agricultural activities also produce COD. Because there are few water quality monitoring stations in the study area and lack long-term water quality monitoring data, the water quality parameters in the model have not been calibrated and rely on previous research results from the same or adjacent areas. These deficiencies need to be further improved.

5. Conclusions

In this study, we used VMD-LSTM coupled with MIKE21 to construct a water environment model for the Weihe River from Xianyang City to Weinan City. We predicted the water quality response within the next year after the implementation of urban and agricultural NPS pollution reduction measures and evaluated the effectiveness of these measures in improving water quality.

We predicted runoff data for a future period using VMD-LSTM and input it into the water environment model, enabling the model to better predict the water quality response within a future period after the implementation of emission reduction measures. In Scenario 0, we did not set any measures to control NPS emissions, while in Scenario 1, we implemented LID and BMPs to control

NPS emissions. During the flood season, the highest reduction in COD concentrations was 71.3%, NH3-N concentrations decreased by up to 31.4%, and TP concentrations decreased by up to 43.1%. This indicates that the water quality of the Weihe River from Xianyang City to Weinan City can be significantly improved by controlling NPS pollution emissions. However, in Scenario 1, the water quality in most regions of the river still falls below the Chinese Water Quality Standard Level IV, suggesting that more effective measures are needed to reduce TP.

The results also show that the water quality response varies with different land-use types and the implementation of reduction measures. Controlling agricultural NPS pollution emissions has a more significant effect on improving river water quality. Therefore, when considering economic value, priority can be given to controlling agricultural NPS pollution emissions through BMPs and other means.

This study provides a basis for decision-making in water environment management, especially in controlling NPS pollution emissions, in the Weihe River Basin of Shaanxi Province.

Declarations

Ethical Approval

Not applicable.

Competing interest

The authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Authors' contributions

Xianqi Zhang and Yu Qi wrote the main manuscript text. Fang Liu, Haiyang Li and Shifeng Sun prepared all figures. All authors reviewed the manuscript.

Funding

No funding received.

Availability of data and materials

All datasets used can be accessed.

Reference

[1] Xiang, C., Wang, Y., & Liu, H. (2017). A scientometrics review on nonpoint source pollution research. Ecological Engineering, 99, 400-408.

[2] Giri, S. , Qiu, Z. , & Zhang, Z. . (2018). Assessing the impacts of land use on downstream water quality using a hydrologically sensitive area concept. Journal of Environmental Management, 213(MAY1), 309.

[3] Fu, Q., Zhu, Y., & Huang, S. (2020). Regionalization of agricultural nonpoint source pollution over China with a combination of qualitative and quantitative method. *Sustainability*, *12*(1), 405.

[4] Sun, D., Xu, S., Jin, X., Feng, P., & Chang, C. (2018). Water flow regulation and scheme optimization in the Haihe River. *Environmental Engineering Science*, *35*(6), 627-644.

[5] Taghizadeh, S., Khani, S., & Rajaee, T. (2021). Hybrid SWMM and particle swarm optimization model for urban runoff water quality control by using green infrastructures (LID-BMPs). *Urban Forestry & Urban Greening*, *60*, 127032.

[6] Ji, H., Peng, D., Fan, C., Zhao, K., Gu, Y., & Liang, Y. (2022). Assessing effects of non-point source pollution emission control schemes on Beijing's sub-center with a water environment model. *Urban Climate*, *43*, 101148.

[7] Ouyang, W., Hao, F. H., Wang, X. L., & Cheng, H. G. (2008). Nonpoint source pollution responses simulation for conversion cropland to forest in mountains by SWAT in China. *Environmental management*, *41*, 79-89.

[8] Lee, J., Lee, J. E., & Kim, N. W. (2020). Estimation of hourly flood hydrograph from daily flows using artificial neural network and flow disaggregation technique. *Water*, *13*(1), 30.

[9] Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophysical Research Letters*, *44*(21), 11-030.

[10] Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*, *10*(11), 1543.

[11] Zhang, D., Lindholm, G., & Ratnaweera, H. (2018). Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *Journal of hydrology*, *556*, 409-418.

[12] Ghasemlounia, R., Gharehbaghi, A., Ahmadi, F., & Saadatnejadgharahassanlou, H. (2021). Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional Long Short-Term Memory (BiLSTM) deep neural network. *Computers and Electronics in Agriculture*, *191*, 106568.

[13] Zhang, X., Chen, H., Wen, Y., Shi, J., & Xiao, Y. (2023). A new rainfall prediction model based on ICEEMDAN-WSD-BiLSTM and ESN. *Environmental Science and Pollution Research*, *30*(18), 53381-53396.

[14] Yan, B., Wang, Y., Li, G., & Ding, X. (2022). Comprehensive evaluation of the water environment carrying capacity of a river basin: a case study of the Weihe River Basin in China. *Water Policy*, *24*(1), 31-48.

[15] Qiu, D., Wu, C., Mu, X., Zhao, G., & Gao, P. (2022). Changes in extreme precipitation in the Wei River Basin of China during 1957–2019 and potential driving factors. *Theoretical and Applied Climatology*, *149*(3-4), 915-929.

[16] Jiake, L. I., Huaien, L. I., Bing, S. H. E. N., & Yajiao, L. I. (2011). Effect of non-point source pollution on water quality of the Weihe River. *International Journal of Sediment Research*, *26*(1), 50-61.

[17] Tong, Y., Zhang, W., Wang, X., Couture, R. M., Larssen, T., Zhao, Y., ... & Lin, Y. (2017). Decline in Chinese lake phosphorus concentration accompanied by shift in sources since 2006. *Nature Geoscience*, *10*(7), 507-511.

[18] Yang, S., Dong, G., Zheng, D., Xiao, H., Gao, Y., & Lang, Y. (2011). Coupling Xinanjiang model and SWAT to simulate agricultural non-point source pollution in Songtao watershed of Hainan, China. *Ecological Modelling*, *222*(20-22), 3701-3717.

[19] Chen, L., Dai, Y., Zhi, X., Xie, H., & Shen, Z. (2018). Quantifying nonpoint source emissions and

their water quality responses in a complex catchment: a case study of a typical urban-rural mixed catchment. *Journal of Hydrology*, *559*, 110-121.

[20] Hou, C., Chu, M. L., Botero-Acosta, A., & Guzman, J. A. (2021). Modeling field scale nitrogen non-point source pollution (NPS) fate and transport: Influences from land management practices and climate. *Science of the Total Environment*, *759*, 143502.

[21] Liu, H., Duan, Z., Han, F. Z., & Li, Y. F. (2018). Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. *Energy Conversion and Management*, *156*, 525-541.

[22] Ma, Z., Chen, H., Wang, J., Yang, X., Yan, R., Jia, J., & Xu, W. (2020). Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. *Energy Conversion and Management*, *205*, 112345.

[23] Zhou, Q. C., Shen, H. H., Zhao, J., & Liu, X. C. (2019). Bearing fault diagnosis based on improved stacked recurrent neural network. *J. Tongji Univ*, *47*(10), 1500-1507.

[24] Ramteke, G., Singh, R., & Chatterjee, C. (2020). Assessing impacts of conservation measures on watershed hydrology using MIKE SHE model in the face of climate change. *Water Resources Management*, *34*, 4233-4252.

[25] Dong, G., Yang, S., Gao, Y., Bai, J., Wang, X., & Zheng, D. (2014). Spatial evaluation of phosphorus retention in riparian zones using remote sensing data. *Environmental earth sciences*, *72*, 1643-1657.

[26] Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., & Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. *Journal of Hydrology*, *280*(1-4), 52-71.

[27] Yang, X., Warren, R., He, Y., Ye, J., Li, Q., & Wang, G. (2018). Impacts of climate change on TN load and its control in a River Basin with complex pollution sources. *Science of the Total Environment*, *615*, 1155-1163.