Environmental Science and Ecotechnology 18 (2024) 100320

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

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

Original Research

Machine learning parallel system for integrated process-model calibration and accuracy enhancement in sewer-river system

Yundong Li [a, b], Lina Ma [a], Jingshui Huang [b], Markus Disse [b], Wei Zhan [a], Lipin Li [a], Tianqi Zhang [a], Huihang Sun [a], Yu Tian [a, *]

[a] State Key Laboratory of Urban Water Resource and Environment (SKLUWRE), School of Environment, Harbin Institute of Technology, Harbin, 150090, China
[b] Chair of Hydrology and River Basin Management, Technical University Munich, Arcisstrasse 21, 80333, Munich, Germany

ARTICLE INFO

Article history:
Received 4 December 2022
Received in revised form
13 September 2023
Accepted 14 September 2023

Keywords:
Integrated sewer–river model
LSTM
ACO
Sewer–WWTP–river system
Water pollution control strategy

ABSTRACT

The process-based water system models have been transitioning from single-functional to integrated multi-objective and multi-functional since the worldwide digital upgrade of urban water system management. The proliferation of model complexity results in more significant uncertainty and computational requirements. However, conventional model calibration methods are insufficient in dealing with extensive computational time and limited monitoring samples. Here we introduce a novel machine learning system designed to expedite parameter optimization with limited data and boost efficiency in parameter search. MLPS, termed the machine learning parallel system for fast parameter search of integrated process-based models, aims to enhance both the performance and efficiency of the integrated model by ensuring its comprehensiveness, accuracy, and stability. MLPS was constructed upon the concept of model surrogation + algorithm optimization using Ant Colony Optimization (ACO) coupled with Long Short-Term Memory (LSTM). The optimization results of the Integrated sewer network and urban river model demonstrate that the average relative percentage difference of the predicted river pollutant concentrations increases from 1.1 to 6.0, and the average absolute percent bias decreases from 124.3% to 8.8%. The model outputs closely align with the monitoring data, and parameter calibration time is reduced by 89.94%. MLPS enables the efficient optimization of integrated process-based models, facilitating the application of highly precise complex models in environmental management. The design of MLPS also presents valuable insights for optimizing complex models in other fields.

1. Introduction

Water environment process-based models, such as sewer networks and river models, have emerged as crucial tools for investigating urban water environment problems. These models enable the quantification of complex pollution processes, identification of the main sources, and migration pathways of pollutants and provide valuable insights for decision-making in urban watershed planning and management [1,2]. However, existing models are functionally independent and exhibit limited connectivity, thereby failing to reflect the response relationships among the various elements of the sewer–wastewater treatment plant (WWTP)–river system [3–5]. To address this limitation, the coupling of multiprocess-based models can facilitate comprehensive simulations of the entire water resource utilization process within the sewer-WWTP-river system. This integrated approach offers managers a more comprehensive and systematic understanding of watershed hydrology, hydrodynamics, and water quality [6,7].

Nevertheless, integrated process-based models encounter challenges related to data accessibility and high computational load associated with model parameter optimization. On the one hand, constructing integrated process-based models requires sufficient basic data for model building, calibration, and validation. The quality and quantity of data largely determine the quality of model applications. Unfortunately, practical limitations often lack accurate long-term monitoring data [8,9]. On the other hand, the model operation process requires numerous parameters to describe different pollution processes, including mainly rainfall–runoff, soil

* Corresponding author.
E-mail address: hit_tianyu@163.com (Y. Tian).

https://doi.org/10.1016/j.ese.2023.100320

erosion, sediment re-suspension, pollutant transport, and transformation processes. These processes exert individual and interactive effects on model response variables. and the value of each parameter affects the final model accuracy [10].

Traditional parameter optimization methods involve repeated fitting iterations of integrated process-based models, and different parameter combinations encompass various independent calculations and subsequent unified storage and analysis of all parameter combination scenarios. Consequently, this approach incurs a significant computational burden and notable time consumption [10–12]. Thus, developing an efficient global parameter optimization method that can yield accurate predictions based on limited data can enhance the accuracy of integrated sewer networks and urban river models.

The incorporation of machine learning algorithms, known as data-driven model simulation techniques, is a key strategy for constructing comprehensive datasets and accelerating the optimization of process-based model parameters [13,14]. Machine learning simulation, a concept pioneered by Blanning and Kleijnen [15], represents a significant advancement in studying complex mathematical models over the past decade. It involves establishing response surfaces and ignoring the internal process dynamics of process-based models to achieve efficient and fast computational processes [16–20]. Extensive studies have demonstrated the application of machine learning in simulation. For instance, Moreno-Rodenas et al. [21] used a polynomial expansion simulator to simulate dissolved oxygen in the Dommel River in the southern Netherlands instead of an urban water quality model. Yin et al. [22] employed surface regression models, artificial neural networks, and support vector machines in saltwater intrusion prediction and saltwater purification design to replace high-fidelity solute transport models. Machine learning also addresses the challenge of missing data, enhancing the fitting accuracy of process-based models when data is insufficient or hard to obtain. Manley et al. [23] demonstrated the ability of machine learning to bridge data gaps in sparse data environments by collecting multiple machine learning sample datasets. Arriagada et al. [24] showcased the effectiveness of the machine learning algorithm MissForest in performing accurate and reliable daily flow time series simulations in regions with sparse data and high climate variability. Despite the good performance of model speed and dataset compatibility, the "data-based" nature of machine learning models blurs the simulation of the pollutant transport and transformation process, making them hardly replace the process-based model in this research. However, model surrogation can still be employed to reduce the time consumption of process-based model optimization. In regard to parameter optimization, Cho et al. [25] applied a fast optimization process to the parameters of the QUAL2K river water quality model based on the influence coefficient algorithm and genetic algorithm (GA). Liu et al. [26] investigated a new support vector machine agent modeling method. They combined it with the generalized likelihood uncertainty estimation (GLUE) method for optimization to obtain optimal model parameter intervals. Based on the above research, we infer that integrating the surrogate model approach with machine learning optimization may enable the development of an efficient global parameter optimization system for the process-based model, thereby advancing the application of integrated process-based models in the environmental field.

As a time-series neural network, the long short-term memory (LSTM) technique is beneficial for solving data series prediction problems due to its memory properties [27]. Recent studies have demonstrated the effectiveness of LSTM in practical applications within the water environment. For example, Vu et al. [28] applied LSTM modeling to bridge groundwater-level data gaps over many years, extending the existing time series. Chen et al. [29] combined the advantages of LSTM and migration learning techniques to solve the problem of missing large-scale continuous dissolved oxygen data at water quality monitoring stations based on limited data. These studies have confirmed the advantages and reliability of the LSTM technique in practical applications in the water environment [30,31], surpassing other machine learning methods commonly used in water quality prediction. In water quality prediction, the back propagation (BP) and radial basis function (RBF) neural networks are often constrained by the problem of insufficient training [32]. The hybrid mechanism and artificial neural network (ANN) model cannot perform extreme value prediction due to the difficulty in learning state characteristics between time series data [33]. Other studies have shown that LSTM is more suitable for time series data prediction than machine learning methods [34]. Nevertheless, the LSTM technique cannot optimize parameters, limiting its applicability to integrated process-based models of sewage networks and urban rivers. Ant colony optimization (ACO) is considered one of the most widely applicable and efficient optimization algorithms due to its excellent multiconstraint adaptability, efficient multiobjective optimization capability, and robustness in solving highly nonlinear problems [11,35,36]. Afshar [37] applied the ACO algorithm to optimize the key decision variable of the sewer network node elevation. Zhang et al. [38] developed a new ant colony optimization and support vector machine model (ACO-SVM) using the ACO algorithm. The results indicated that the algorithm could optimize the model parameters and provide favorable application prospects. However, due to the complexity of the integrated process-based model, applying ACO for global parameter optimization poses significant challenges. Therefore, using the LSTM technique for model simulator construction and employing the ACO algorithm for global parameter optimization could effectively improve the parameter calibration efficiency for integrated process-based models.

In this study, we proposed a parallel LSTM-ACO parameter search system for an integrated process-based model. The study encompasses the following key steps: (1) constructing an integrated process-based model of the sewer network and urban rivers based on the stormwater management model (SWMM) and water quality analysis simulation program (WASP); (2) leveraging the ACO algorithm in conjunction with an LSTM mode to obtain the optimal parameter set, which is subsequently utilized to enhance the integrated process-based model; (3) evaluating the water quality prediction results of the optimized, integrated process-based model during dry and rainy seasons and throughout the year to evaluate the model accuracy; and (4) investigating the reasons for model accuracy improvement by comparing the parameters before and after optimization. This study could promote the integration of machine learning and process-based models, offering a general and efficient parameter optimization solution for complex process-based models in the urban water pollution control and earth sciences fields.

2. Materials and methods

2.1. Study area and data acquisition

The study area encompassed an urban river in southern China, the Jiuqu River, which is located in the central Sichuan Basin. The geographical coordinates of the river range from 104°31′–104°38′ E and 30°06′–30°12′ N. The mainstream section of the river spans a length of 38.3 km, encompassing a watershed area of 216.78 km^2. The topography of the region exhibits higher elevations in the northwest, gradually decreasing towards the southeast, with elevations ranging from 346 to 482 m. The Jiuqu River watershed is

predominantly characterized by hilly terrain, accounting for approximately 94% of the total area. The geological structure of the watershed is stable, and the soil types mainly include purple sandstone, weathered shale, sand, and clay. Land usage within the area predominantly comprises arable land, accounting for 56.72% of the total area (40.99% dry land and 15.73% paddy fields), along with woodland (14.70%), residential land (18.93%), and other land types (9.65%). The urban land is concentrated in the lower reaches of the watershed, including the urban area of Ziyang, part of the suburb, and seven towns, while village land is scattered throughout the watershed. The watershed experiences a subtropical humid monsoon continental climate, with abundant rainfall and mild climate conditions. The rainy season extends from April to October, and the dry season spans from November to March. Local pollution often occurs because of the complex point source and non-point source pollution conditions in the watershed. Notably, the river has exhibited concentrations of ammonia nitrogen and total phosphorus surpassing the Chinese water quality standard (GB 3838-2002) by factors of 3.5 and 2.0, respectively.

The urban area of Ziyang is located in the lower reaches of the Jiuqu River, with a total area of 41.04 km^2 and a population of 2.3 million people. The urban sewer network uses a combined drainage system with 195.46 km of pipelines and 23.18 km of drainage channels. The river section flowing through the urban area is an artificial channel where the banks and part of the riverbed have been solidified. Additionally, dredging was carried out in the Jiuqu River between December 2017 and February 2018. Therefore, the influence of sediment re-suspension on river water quality was ignored in this research. For domestic sewage treatment, a WWTP is located in an urban area with a daily capacity of 50,000 tons. In terms of data collection, a weather station located in the urban area downstream and an automatic water flow and quality monitoring station located near the confluence point, where the Jiuqu River meets a larger river, provided the main data support in this study.

This research mainly obtained the required data from local government departments and stations, as listed in Table 1. Data from December 2018 to December 2019 were used for model construction, and the remaining data were used as a reference to ensure data rationality. Geographical information, including the river length, cross-sectional profiles, location coordinates, digital elevation model (DEM) (30 m × 30 m), and remote sensing images, originated from the database of the local water bureau, and the data was collected in December 2018. Watershed meteorological information originated from the local weather station, including the temperature, rainfall, wind speed, and solar radiation intensity observations. The meteorological data ranged from January 2010 to December 2020, with measurements recorded every 3 h. River hydrological and quality information, including river flow, water depth, and pollutant concentration (ammonia nitrogen and total phosphorus) observations, originated from the automatic water quality monitoring station, with a frequency of once a week. River pollution information, including pollution point sources such as sewage treatment plants, separate system outlets (sewage and rainwater), combined system outlets (overflow), and non-point source pollution statistics, such as planting and breeding industries, were retrieved from local environmental protection bureaux. Monitoring data of sewage treatment plants and drainage outlets were collected once a day, and non-point source pollution data were estimated once a month (calculated based on the planting area and breeding quantity) during the data collection period from January 2018 to December 2020 [39]. The estimation coefficient for non-point sources was derived from the calculation of Sichuan Province in 2017 [40], and the results are close to the estimates of non-point sources in the Tuojiang River watershed (the larger river into which the Jiuqu River flows) in the recently published article [41].

2. 2.2. Integrated sewer network and urban river model

In this study, an integrated sewer network and urban river model using the SWMM and WASP was constructed to predict the water quality of the Jiuqu River. The complete structure of the integrated model is depicted in Fig. 1.

The SWMM is a sewer network model widely used in 200 cities worldwide, with functions for the simulation of rainfall confluence, rainwater and sewage collection, and sewage mixing and transportation inside sewers [42]. In regard to urban rainwater and sewage simulation, an SWMM of the sewer network in the Ziyang urban area was built based on the User Manual of the Storm Water Management Model (SWMM) version 5.1 of the US Environmental Protection Agency (EPA) and the work of Baek et al. and Li et al. [43,44]. The required sewer.shp files were updated according to the latest sewer network status to establish a drainage system model of Ziyang with 39.5 km of confluence pipes, 169.1 km of sewage pipes, and 204.2 km of rainwater pipes. The model covered an area of 129.4 km^2, which was divided into 145 subcatchments. Meteorological and rainfall data were provided by the meteorological monitoring station, and sewage production data were estimated according to the population distribution in each region [45]. The model was run with a simulation period of one year (January 01,

Table 1
Data sources for SWMM and WASP setup.

Boundary condition group	Boundary condition name	Data source	Time period	Data frequency
Meteorological data	Temperature Rainfall Wind speed and direction Sunlight intensity Cloudiness	Ziyang weather station	January 01, 2019, to December 31, 2019	3 h
River hydrological and quality information	River flow Water depth Ammonia nitrogen concentration Total phosphorus concentration	Jiuqu River water quality monitoring station	January 01, 2019, to December 31, 2019	1 week
	Tributary flow and quality	Ziyang water affairs bureau		1 day
Pollution information	Sewage treatment plant discharge Planting-related pollutants Breeding-related pollutants	Ziyang environmental protection bureau	January 01, 2019, to December 31, 2019	1 day 1 month

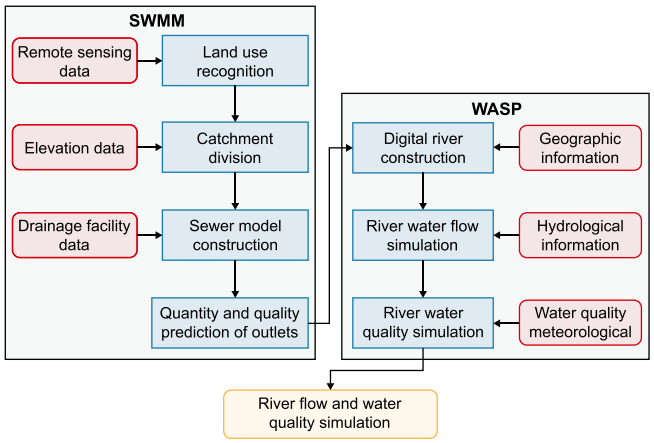

Fig. 1. Structure of the integrated sewer network and urban river model.

2019, to December 31, 2019).

The WASP is a widely used hydraulic and quality model for river water developed by the EPA, especially in nitrogen and phosphorus simulations [46–48]. First, we selected the mainstream section of the Jiuqu River as the model domain. Cross-sections of the river were created in the WASP according to the obtained elevation information for spatial model establishment. Subsequently, boundary conditions, such as tributaries, agricultural non-point sources, WWTP point sources, and drainage outlets, were added to the model based on the spatial position. Specifically, the model encompassed four tributaries, one WWTP point source, and 134 drainage outlets. Water flow and pollutant concentration (such as nitrogen, phosphorus) time series created with the data collected in Section 2.1 were added to each boundary condition. The non-point source pollution data for the upper and middle reaches were available monthly, while the sewage discharge (from the SWMM model) was updated daily. The model parameters were assigned based on expert experience, as listed in Table 3. Finally, we set the simulation time from January to December 2019, and the simulated flow, ammonia nitrogen, and total phosphorus concentration results at the outlet of the catchment were compared to the measurements at the station to evaluate the model performance. The SWMM and WASP were coupled using drainage network outlets as a link to create an integrated sewer network and urban river model. Discharge and pollutant loading outputs at all drainage outlets, retrieved from the SWMM, were transferred to the corresponding boundary conditions in the WASP, with a temporal resolution of once a day. Hence, the river model could respond to inshore water quality changes promptly. This automatic process was realized using a script written in Python.

2.3. Surrogate model using the LSTM algorithm

The LSTM is a neural network algorithm with memory characteristics that can save long- and short-term memory resources. Long-term memory is stored in the hidden layer, whereas short-term memory is saved using a recurrent neural network [49]. The LSTM usually includes forward and backward propagation steps. The forward calculation requires three inputs (current network input, previous output, and previous unit states) and two outputs. The forget, input, and output gates are crucial in determining the extent of element retention, input storage, and output utilization, respectively. The forward propagation process can be formulated as follows:

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

$$\left[W_f\right]\begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} = \begin{bmatrix} W_{fh} & W_{fx} \end{bmatrix}\begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\overline{c}_t = \mathrm{ReLU}(W_c[h_{t-1}, x_t] + b_c) c_t = f_t c_{t-1} + i_t \overline{c}_t \tag{3}$$

$$o_t = \sigma[W_o[h_{t-1}, x_t] + b_o] h_t = o_t \mathrm{ReLU}(c_t) \tag{4}$$

where f_t is the forget gate; i_t is the input gate; $\overline{c}_t$ is the current input; c_t is the current element; o_t is the output gate; and h_t is the LSTM output. W_f, W_{fh}, W_{fx}, W_i, W_c, and W_o are the weight matrices of f_t, h_{t-1}, X_t, i_t, $\overline{c}_t$, and o_t, respectively. $[h_{t-1}, x_t]$ connects two vectors into a longer vector; b_f is the bias term of f_t; σ is the activation function; b_i is the bias entry of i_t; b_c is the offset item of the current input unit $\overline{c}_t$; b_o is the output gate offset; and ReLU() is the activation function.

In contrast to conventional scenarios, our study incorporates LSTM to capture the impact of model parameters on the results. In order to emulate the process-based model, we utilized both the boundary conditions and parameters of the integrated model as LSTM input samples, while the simulated river water quantity and quality were employed as output samples. The inputs of the LSTM were time-based sequence data composed of dynamic inputs and steady parameters of the process-based model, as shown in Table S2, which made LSTM play an equivalent role to the process-based model during a single event. Limited by the high computational cost of the integrated model, it was crucial to cover a wide range of rainfall scenarios in the training data within the shortest time series. Therefore, 30 consecutive days in April and May covering seven sequential rainfall events were selected for model training. The training samples comprised 20 boundary conditions, ten model parameters, and three model outputs. The model boundary conditions, as listed in Table 2, were obtained from the variable parts of the integrated model. Ten model parameters (Table 3), considered the most influential variables (five sewer network-related and five river-related parameters), were selected based on other relevant studies to constitute various parameter combinations of training data [50–52]. During training database sampling, only the ten considered parameters were varied, and all the other parameters remained fixed based on expert experience. The model outputs included the river flow, ammonia nitrogen concentration, and total phosphorus concentration at a location of interest. Model boundary conditions and corresponding model outputs over 30 days were extracted from the process-based model. Parameter sets were generated via weighted random sampling. The value ranges of the parameters were divided into 12 or 6 ranges on average according to the relative sensitivity. The Monte Carlo simulation method was adopted within these ranges to randomly sample the ten parameters, generating 473 parameter sets based on the training data requirements and the total time needed to obtain model results [53]. By combining these parameter sets with the model boundary conditions spanning 30 days, we generated a substantial dataset consisting of 14,190 training samples, which included corresponding model outputs.

According to the above design, the training period of the LSTM was set from April 5 to May 5, 2019, including seven representative rainfalls with intensities of 1.1, 3.7, 5.5, 5.9, 11.9, 13.8, and 30.3 mm d^{-1}. The LSTM model employed a sample size of 14,190. The LSTM model was designed with a four-layer neural network

Table 2
Model boundary conditions of the LSTM training samples.

Boundary condition group	Boundary condition name	Unit
Rainfall intensity		mm d^{-1}
Temperature		°C
Non-point source pollution	Water flow	m^3 s^{-1}
	Ammonia nitrogen concentration	mg L^{-1}
	Total phosphorus concentration	mg L^{-1}
Tributaries	Water flow	m^3 s^{-1}
	Ammonia nitrogen concentration	mg L^{-1}
	Total phosphorus concentration	mg L^{-1}
WWTP	Water flow	m^3 s^{-1}
	Ammonia nitrogen concentration	mg L^{-1}
	Total phosphorus concentration	mg L^{-1}

Abbreviations: WWTP, wastewater treatment plant.

Table 3
Optimized parameters of the integrated model.

Parameter type	Parameter	Unit
Sewer-network-related	Manning impermeability coefficient	–
	Manning permeability coefficient	–
	Houghton maximum permeability coefficient	–
	Impervious rainfall Lost	–
	Previous rainfall lost	–
River-related	Depth exponent	–
	Dissolved organic phosphorus mineralization rate constant (20 °C)	d^{-1}
	Nitrification rate constant (20 °C)	d^{-1}
	Nitrification temperature coefficient	–
	Half-saturation constant for the nitrification oxygen limit	mg O_2 L^{-1}

structure, including one fully connected, one LSTM, and two linear layers. The dimension of the LSTM output layer was 30, while the linear layers had dimensions of 90 each. The rectified linear unit (ReLU) function, mean-square error function (MSEloss), and stochastic gradient descent were used as the activation function, loss function, and optimizer, respectively. The next seven days were used as the test period, and the LSTM performance was verified by the river flow and water quantity results of the process-based model. Model settings for the LSTM emulation model are listed in Table 4.

2.4. Ant colony optimization of the parameters

Parameter optimization was realized based on the ACO method. The ACO algorithm is an iterative calculation method that simulates the foraging process of ants in the real world. The optimization algorithm was executed in ten algorithmic repetitions, with the results averaged. Each repetition entailed 500 iterations with 100 ants. To ensure that all the edges attained the same probability of being selected, all edge pheromones were set to the same maximum level ($\tau_{max}^{0} = 100{,}000$) before the repeated calculation steps. The Nash–Sutcliffe efficiency (NSE) coefficient value of the model prediction results (river flow, ammonia nitrogen concentration, and total phosphorus concentration) was used as the objective function to obtain the global algorithm search optimum. This objective function can be expressed as follows:

$$NSE = 1 - \left(\frac{\sum_{t=1}^{N} \left(Q_{\mathrm{sim},t} - Q_{\mathrm{obs},t}\right)^2}{\sum_{t=1}^{N} \left(Q_{\mathrm{obs},t} - \overline{Q}\right)^2} \right) \quad (5)$$

$$NSE_{\mathrm{overall}} = 0.5NSE_{\mathrm{flow}} + 0.25NSE_{\mathrm{N}} + 0.25NSE_{\mathrm{P}} \quad (6)$$

where N is the sample size, which is the number of model prediction results and corresponding observations ; $Q_{\mathrm{sim},t}$ denotes the model prediction results in group t; $Q_{\mathrm{obs},t}$ denotes the corresponding observations; and $\overline{Q}$ denotes the average observation data. Water quality prediction is based on hydraulic simulation for the current process-based model. High-precision water quality results with low-precision hydraulic results indicate that the model may be overfitting, which is detrimental to optimization. To address this concern, we assign a higher weight to NSE_{flow} than NSE_{N} and NSE_{P}.

Before optimization, the parameters were discretized as decision nodes of the ACO algorithm. The implementation process involved the following steps:

(1) Definition of parameter optimization range

The optimization range for each parameter was determined according to the model reference value (Table 6).

Table 4
Model settings for the LSTM emulation model.

Layers	Dimension of the output layer	Dimension of the linear layer	Activation function	Loss function	Optimizer
4	30	90	ReLU()	MSEloss	Stochastic gradient descent

(2) Discretization of parameter ranges

The discretization method for each variable can be expressed as follows:

$$h_j = \frac{x_{j,\text{upper}} - x_{j,\text{lower}}}{\text{N}} (j = 1, 2, \cdots, n) \quad (7)$$

where $x_{j,\text{lower}}$ is the lower limit of the jth parameter; $x_{j,\text{upper}}$ is the upper limit; and N is the discreteness of the parameter, which is set to ten.

The optimization method was constructed following the state transition rule, which can be defined as follows:

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha} \times [\eta_{ij}(t)]^{\beta}}{\sum_{s \in \text{allowed}_k} [\tau_{is}(t)]^{\alpha} \times [\eta_{is}(t)]^{\beta}}, \text{if } j \in \text{allowed}_k \\ 0, \text{otherwise} \end{cases} \quad (8)$$

where $P_{ij}^{k}(t)$ is the probability that ant k moves from positions i to j; $\tau_{ij}(t)$ is the pheromone trail; n_{ij} is the heuristic information; and β is the weight of the heuristic information. n_{ij} is an auxiliary method that helps guide the search for ants. The edges along which the ant has already traveled in one iteration can no longer be selected and this memory function is realized via the tabu list. $\text{Allowed}_k = \{0, 1, \ldots, n-1\} - \text{tabu}_k$ is applied, and $\text{tabu}_k(\text{k} = 1, 2, \ldots, m)$ is the tabu list used to record the locations where the ants have traveled. This list constantly changes as the algorithm is operated.

The pheromone update rules of the ACO algorithm include local and global dynamic updating. The optimal path can be continuously developed by updating the pheromone trail along the current optimal path. According to Moriasi et al.'s research on the evaluation criteria of water environment models, the model results are considered "good" when 0.65 < NSE ≤0.75 [54]. In this study, we hope that the optimized model can meet the "good" standard while retaining good generalization ability, so we set the threshold of the NSE metric to 0.7. Using the ACO algorithm's positive feedback mechanism, the pheromone trail along the current optimal path can be retained for the next iteration until the NSE value reaches 0.7. A flow diagram of the parallel system is depicted in Fig. 2.

2.5. Model performance evaluation

The developed LSTM surrogate method for parameter estimation consisted of two essential components: LSTM emulation and parameter optimization processes. In order to assess the performance of the model, a two-step analytical approach was employed. First, the emulation effect of the LSTM model was analyzed by comparing the prediction results of the LSTM and integrated models in terms of the river flow, ammonia nitrogen concentration, and total phosphorus concentration during the multiple rainfall events. R^2 was calculated to determine whether the LSTM model could accurately emulate the results of the process-based model, including model errors and variations due to parameter changes, which constituted the basis for the model surrogate feasibility.

The prediction effect of the optimized model was then analyzed by comparing the original model, optimized model, and monitoring results from January 1, 2019, to December 31, 2019, using the percent bias (PBIAS), correlation coefficient, root mean square error (RMSE), standard deviation, and relative percentage difference (RPD). These metrics served as crucial indicators to gauge the accuracy enhancement achieved by the optimized model.

3. Results and discussion

3.1. Evaluation of the LSTM surrogate model

The trained LSTM model employed the same model parameters as the process-based model to simulate the river water quality during eight consecutive rainfall events from April 5 to May 12, 2019 (30 days of training and seven days of testing), and the results are shown in Fig. 3. The changes in the water flow and pollutant concentration are shown in Fig. 3a–c for the LSTM and mechanistic models during the rainfall events. The R^2 value was determined by plotting the river water flow and pollutant concentration predicted with the LSTM model against the process-based model results to assess the emulation accuracy. Notably, the deviation of the LSTM results from the process-based model results was negligible. During the training and test periods, the R^2 values of the water flow, ammonia nitrogen concentration, and total phosphorus concentration were 0.9993, 0.9999, and 0.9998, respectively, indicating that the LSTM model described more than 99.9% of the total variability in the water quantity and quality, which was desirable. The Pearson correlation coefficients of river flow, ammonia nitrogen, and total phosphorus concentration of the LSTM model and the process-based model are 0.9996, 0.9999, and 0.9998, as shown in Fig. 3 and Table S1, respectively, indicating a significant positive linear correlation between the outputs of the LSTM and the process-based model. A large number of points are concentrated near the coordinate origin indicating that during this period, the main weather conditions are low-intensity rainfall or sunny days (Fig. 3d), while the distribution of points is scattered (Fig. 3e,f), indicating that the concentration of nitrogen and phosphorus in river water fluctuates significantly during rainfall events. The LSTM model provided suitable results under both the light and heavy rain scenarios, highlighting its feasibility as a surrogate model for parameter optimization without interference by the boundary conditions.

The LSTM model achieved high accuracy in the 37-day emulation of the water quantity and quality relative to the process-based model. Regarding the ammonia nitrogen and total phosphorus concentrations in river water during the eight rainfall events, the emulation results were highly similar to the integrated model results without any apparent over- or underestimation tendency, which may be closely related to the similar specific concentration

Table 5
Performance of the model under the empirical and optimized parameters.

Model result name	Model result type	PBIAS	Correlation coefficient	RMSE	Standard deviation	RPD
River flow (m s^{-1})	Unoptimized	462.8%	0.285	8.54	7.27	0.9
	Optimized	4.1%	0.999	0.13	1.91	14.7
Ammonia nitrogen (mg L^{-1})	Unoptimized	52.0%	0.503	17.99	19.46	1.1
	Optimized	8.1%	0.996	1.51	9.99	6.6
Total phosphorus (m L^{-1})	Unoptimized	−196.5%	0.493	1.86	2.04	1.1
	Optimized	9.5%	0.995	0.33	1.78	5.4

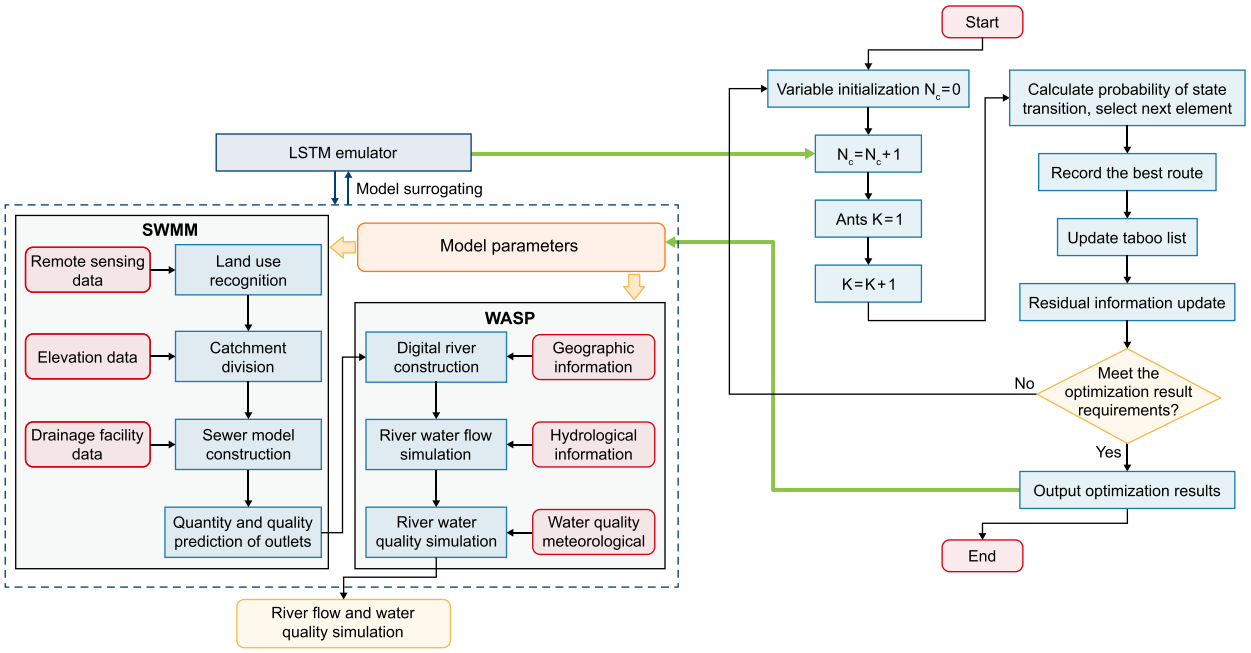

Fig. 2. Flow diagram of the parallel system.

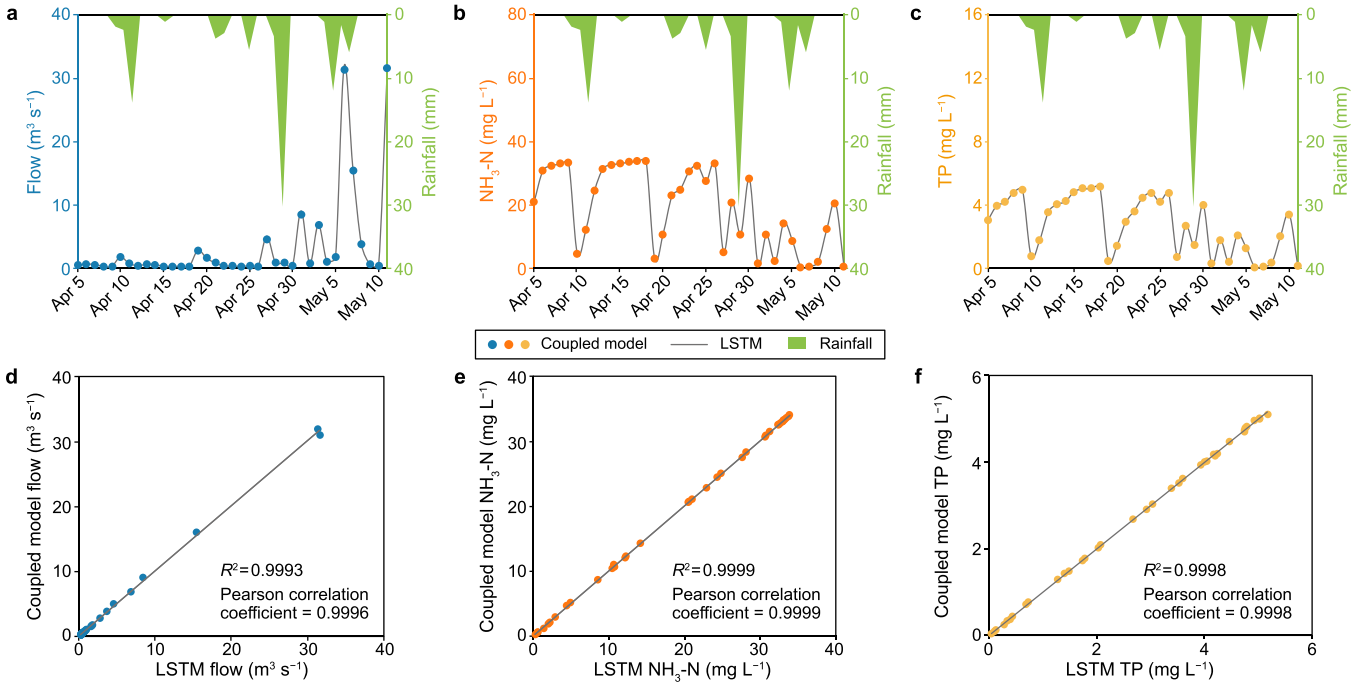

Fig. 3. Result comparison between the LSTM and process-based models under the same boundary conditions and parameters: **a**, water flow results; **b**, ammonia nitrogen concentration results; **c**, total phosphorus concentration results; **d**, water flow correlation; **d**, ammonia nitrogen concentration correlation; **e**, total phosphorus concentration correlation.

profiles. Regarding water flow, however, small deviations occurred in the emulation of peaks and valleys due to the rapid change, large fluctuations, and relatively low regularity. Due to the notable temporal correlation between the water quantity and quality, the LSTM model provided an excellent emulation effect for the process-based model, which notably depended on the long- and short-term memory capacities of the LSTM model. Other researches also prove the feasibility of using LSTM as a surrogate model for the integrated sewer network and urban river model in this study. Similar research by Kratzert et al. [55] indicates that LSTM is more suitable for predicting streamflow than traditional recurrent neural networks (RNN), which works ineffectively when the length of sequences is over ten [56]. Additionally, Li et al. [57] have found that when used as a surrogate model for groundwater contamination

source identification, LSTM exhibits the highest accuracy compared to other methods such as RBF, Kriging, and KELM.

Notably, although the emulation R^2 value of the LSTM model for the process-based model exceeded 99.9%, this does not suggest that the LSTM model alone can provide precise predictions for water quality. In contrast, the errors of the process-based model were also contained in the LSTM model. The process-based model still plays an irreplaceable role in sudden pollution and rainfall events due to its heightened sensitivity to variations in boundary conditions. The accurate emulation of the process-based model by the LSTM model can potentially reduce the prediction error.

The ACO algorithm in this study is designed for ten parallel optimizations, with each optimization comprising 500 iterations. In each iteration, the integrated sewer network and urban river model need to be invoked once to obtain output, resulting in 5000 calls. If the process-based model is called directly in each iteration, the time required for a single invocation is approximately 1.5–2 h (the simulation period is from January 1, 2019, to December 31, 2019, a total of 365 days). Consequently, the total ACO operation time exceeds 7500 h, which is not implementable. When using the LSTM surrogate model for iteration, a single invocation takes less than 1 min, reducing the overall ACO operation time to about 80 h.

3.2. Performance of the optimized, integrated model

Two simulations were conducted in 2019 using the integrated sewer network and urban river model to assess the water quantity and quality of the Jiuqu River. In one model, empirical values of the parameters were determined according to the local geographical environment. While in the other model, parameters were optimized via the ACO algorithm based on LSTM emulation, as summarized in Table 6. Daily simulation results of the river flow, ammonia nitrogen concentration, and total phosphorus concentration at the river's end were collected for comparison to online station monitoring data, as shown in Fig. 4.

In 2019, the river flow remained relatively stable, with notable fluctuations occurring only during heavy rain events or storms. The ammonia nitrogen concentration in river water ranged from 0.1 to 33.9 mg L^{-1}, while the total phosphorus concentration ranged from 0.03 to 6.77 mg L^{-1}. Pollutant concentration fluctuations occurred more frequently during the dry season than the rainy season, which may be related to the different sources of the main pollutants during the different periods. As the main pollution source, point source significantly affected the concentration of river water pollutants during the dry season. Therefore, point sources with large differences in daily sewage discharge and pollutants concentration, such as factory and merchant sewage discharge, may cause stronger fluctuations in water quality. During the rainy season, under the dilution effect of rainfall, non-point source pollution became the main influencing factor of river water quality. Compared with the point source, the pollutant concentration in rainwater tended to be relatively low and stable.

The optimized model achieved a better performance in water quality prediction than the model with empirical parameters, especially during the dry season (October to the following February). The predicted ammonia nitrogen concentration after optimization (average: 18.9 mg L^{-1}) was approximately 42.2% of that before optimization (average: 44.8 mg L^{-1}). Similarly, the predicted total phosphorus concentration after optimization (average: 3.66 mg L^{-1}) was approximately 82.1% of that before optimization (average: 4.46 mg L^{-1}). Notably, although not apparent in Fig. 4, the predicted river flow was also reduced by 47.0% (from 1.59 to 0.84 m^3 s^{-1} on average), suggesting a corresponding decrease of 77.7% and 56.5% in ammonia nitrogen and total phosphorus pollutant amounts after optimization, respectively. As listed in Table 5, PBIAS of the optimized model results reached 4.1% (river flow), 8.1% (ammonia nitrogen), and 9.5% (total phosphorus), indicating excellent model performance in river quantity and quality prediction according to the research by Moriasi et al. [54]. The model performance under the empirical parameters during the rainy season was relatively favorable. However, during the dry season, a large deviation was observed. In contrast, the optimized model yielded accurate prediction results during both seasons. The higher adaptability to different scenarios of the optimized model resulted in a quick response to changes in boundary conditions, such as the river flow or pollutant concentration, which is vital in engineering applications such as emergency pollution control and facility operation.

The Taylor diagram (Fig. 5a) shows that the optimized model achieved a high correlation and small RMSE value in water flow quantity and quality simulation. The Taylor diagram was drawn based on a polar coordinate system, with the radial coordinate ρ of point P (ρ, θ) representing the standard deviation. The projection curve of point P on the circle of $\rho = 0.3$ indicates the correlation coefficient, while the position of point P in the green dashed curve coordinate system represents RMSE. Taking the river flow of dry season (square marker) in Fig. 5a as an example, it could be seen that the optimized model demonstrates a standard deviation of 0.02, a correlation coefficient over 0.99, and an RMSE of 0.1 when compared to the observed data. The standard deviation of the models indicated that, after parameter optimization, the integrated urban drainage model produced results close to the observed values. The pollutant concentration simulation results during the dry and rainy seasons were similar. The standard deviation of the water flow during the dry season was smaller (0.02), while the RMSE was larger (0.10) than the rainy season, indicating that the prediction results for the dry season were better in terms of discrete values, while the prediction results for the rainy season deviated less from the monitored values overall. The integrated urban drainage model produced more effective simulations during the rainy season than during the dry season. During the rainy season, the model obtained a similarly high correlation (>0.99) but a smaller RMSE value (<0.10), indicating a close alignment of the model's results with the observed data in terms of data distribution and discretization. This may be attributed to the relatively stable soil infiltration during the rainy season, which slightly affected the river water quantity.

Violin plots were generated to analyze the data distribution of the river flow and pollutant concentration simulation results for 2019. The violin plots (Fig. 5b–d) indicated a satisfactory performance of the optimized model, which exhibited a favorable fit with the observations. The outlier cutoffs (the thin lines in the violin plots) and the 25th and 75th percentiles (the thick lines in the plots) of the optimized model also indicated a suitable fit for the observations. The frequency of extreme rainfall events was low, and the river discharge number was mainly less than ten during the rainy season and less than three during the dry season. Additionally, the violin plot shows the probability density function for the water flow quantity and quality dataset. As indicated by the data distribution, there was a certain similarity between the ammonia nitrogen and total phosphorus concentrations in river water. The pollutant concentration in river water during the rainy season was generally lower than in the dry season, but the peak ammonia nitrogen concentration during the rainy season was higher than during the dry season. The strong fluctuation of pollutant concentration in the river during the rainy season can be attributed to the dilution effect caused by the rainfalls of different intensities, while the sustained high concentration in the dry season is primarily attributed to continuous pollution of point sources. However, point sources are not the only reason. The reduced river water level

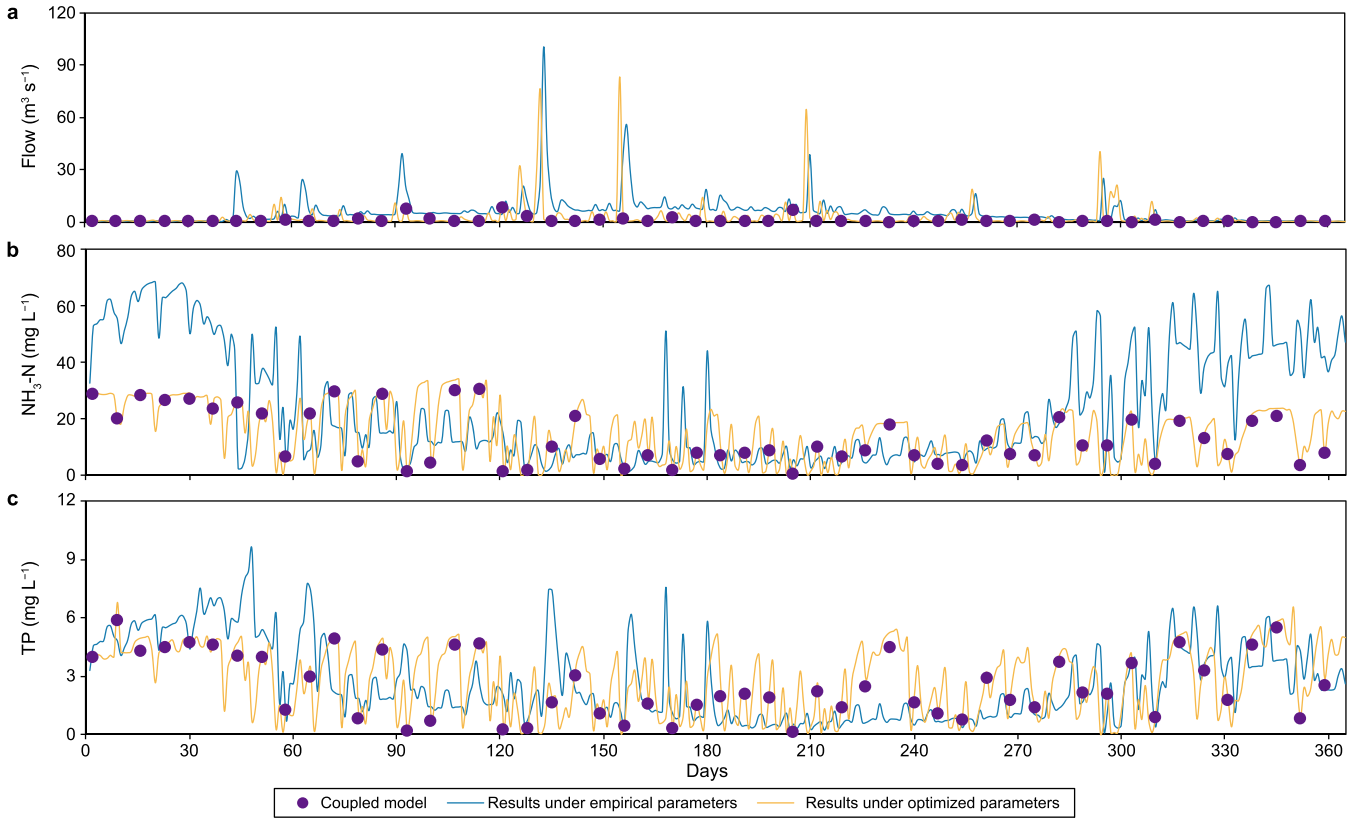

Fig. 4. Simulation of the river water flow (**a**), ammonia nitrogen concentration (**b**) and total phosphorus concentration (**c**) in 2019 via the integrated model.

Table 6
Changes in the model parameters before and after optimization.

Parameters	Unit	Ranges	Empirical	Optimized
Manning impermeability coefficient	–	1.20–2.40	1.78	1.79
Manning permeability coefficient	–	2.45–7.35	5.95	5.68
Houghton maximum permeability coefficient	–	25.40–127.00	46.50	52.45
Depth exponent	–	0.30–0.60	0.43	0.43
Dissolved organic phosphorus mineralization rate constant (20 °C)	d^{-1}	0–0.22	0.01	0.03
Impervious rainfall lost	–	1.27–2.54	1.91	1.91
Previous rainfall lost	–	3–10	5	5
Nitrification rate constant (20 °C)	d^{-1}	0–0.40	0.01	0.10
Nitrification temperature coefficient	–	1.04–1.10	1.04	1.04
Half-saturation constant for the nitrification oxygen limit	mg O_2 L^{-1}	0–5.0	2.0	1.9

during dry days hampers the biodegradation of pollutants, diminishes the self-purification ability, and limits environmental capacity, thereby exacerbating pollutant accumulation. Overall, the optimized model notably simulates the urban water environment during both dry and rainy seasons.

3. 3.3. *Possible rationale for accuracy improvement*

Compared to the integrated sewer network and urban river model using empirical parameters, the optimized model based on the coupled LSTM + ACO approach significantly improved river water quality prediction, especially during the dry season. The changes in each parameter after model optimization were analyzed, as listed in Table 6, to explore the possible reason for model accuracy improvement due to parameter optimization and the different efficiencies during the dry and rainy seasons. Among the ten sensitive model parameters, the nitrification rate constant at 20 °C changed the most, increasing from 0.01 to 0.10 within the range of 0–0.40. Additionally, the dissolved organic phosphorus mineralization rate constant at 20 °C experienced a change from 0.01 to 0.03 within the range of 0–0.22. The Houghton maximum permeability coefficient changed from 46.50 to 52.45 within the range of 25.40–127.00. The remaining parameters, namely Manning impermeability coefficient, Manning permeability coefficient, depth exponent, impervious rainfall lost, previous rainfall lost, nitrification temperature coefficient, and half-saturation constant for the nitrification oxygen limit, changed slightly.

Due to the absence of emergent plants and the artificial hardening of the urban section of the Jiuqu River, the uptake process of microorganisms and a few plants is an important factor affecting the concentration of nitrogen and phosphorus pollutants in the river water [58]. These may cause different efficiencies during the dry and rainy seasons after parameter optimization. During the dry season, the uptake process predominantly determined the

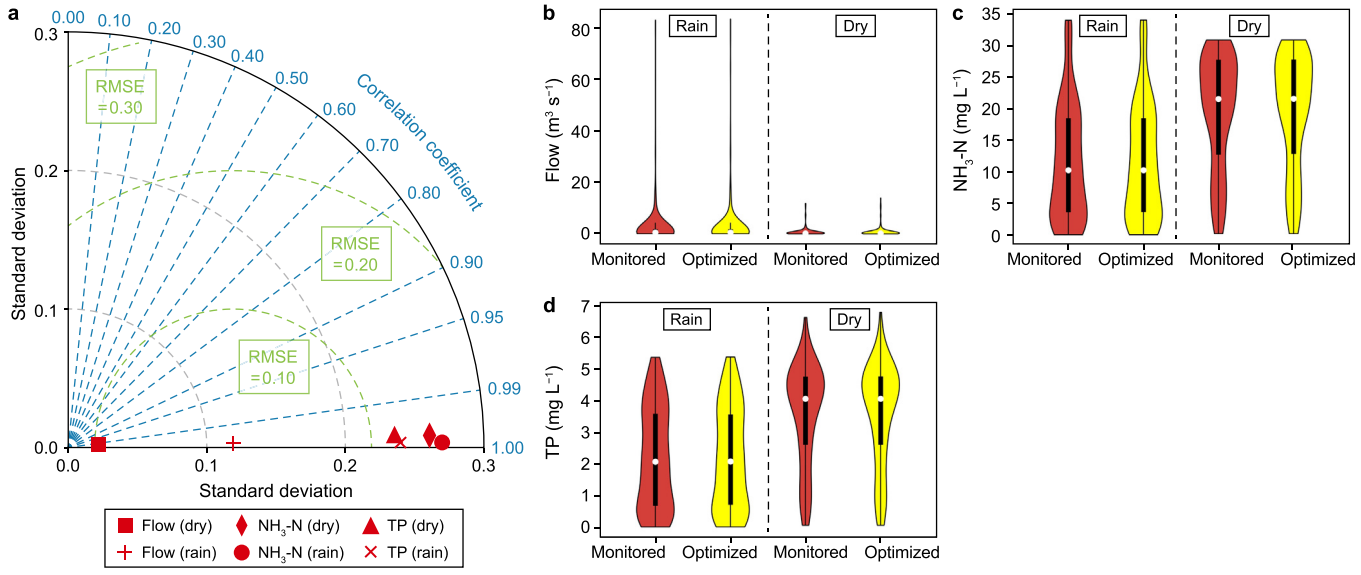

Fig. 5. Correlation and comparative analysis of the optimized model prediction results with the monitored values during the dry and rainy seasons: **a**, standard deviation, RMSE and correlation coefficient of water flow, ammonia nitrogen concentration and total phosphorus concentration; **b**, water flow distribution; **c**, ammonia nitrogen concentration distribution; **d**, total phosphorus concentration distribution.

pollutant concentration in river water due to the river's low velocity and limited flow. Parameter optimization could compensate for the deficiency in the empirical model in this part, which could significantly improve the prediction accuracy of the integrated model during the dry season. During the rainy season, river water mixing and dilution dominated due to the high velocity and notable flow [59], resulting in a relatively insignificant optimization effect. At the same time, significant differences exist in soil permeability between the dry and rainy seasons. In contrast to the high-frequency rainfall during the rainy season leading to a reduction in the infiltration capacity of the watershed, a large proportion of runoff directly enters the soil through infiltration during the dry season [60], corresponding to the reduced river flow predicted with the optimized model. The optimization of these parameters enables the model to accurately estimate the total amount of pollutants entering the river. Specifically, a slight decrease in the Manning permeability coefficient leads to an approximate 5% increase in rainwater runoff velocity in the optimized model, resulting in earlier peak water flow in the river during each rainfall event, without considering the influence of the other factors, as shown in Fig. 4. The increase in the Houghton maximum permeability coefficient indicates enhanced soil permeability in the saturated state, leading to a significant reduction in the peak water level in the river during the rainy season. However, this reduction is not observed during the dry season, as the soil experiences limited saturation during this period. The small increase in the dissolved organic phosphorus mineralization rate constant and the large increase in the nitrification rate constant indicate that the degradation rates of dissolved phosphorus and ammonia nitrogen in river water could increase, leading to a decrease in the nitrogen and phosphorus concentrations, respectively. This phenomenon is more likely to be observed during the dry season than during the rainy season due to the slower river flow during the dry season. In summary, optimizing the pollutant degradation and water infiltration processes during the dry and rainy seasons contributes significantly to the improved prediction accuracy of the integrated sewer network and urban river model during the dry season in the context of the LSTM + ACO optimization method.

3.4. Potentials for urban water pollution control strategies

Water pollution control technology based on the sewer–WWTP–river system constitutes a hotspot to solve the problem of urban surface water pollution. In contrast to previous pollution discharge control or degradation enhancement strategies, the sewer–WWTP–river system-based approach for water pollution control offers the potential to mitigate the adverse impacts of urban pollution emissions on watershed water quality by addressing pollution generation, accumulation, transmission, discharge, and degradation. The high resource and energy consumption levels of end-of-pipe treatments could be effectively prevented by maximizing the total pollutant amount reduction at each stage. This strategy suits cities with complex and multiple pollution sources or rapid population and area growth. Water quantity and quality predictions provided by the integrated sewer network and urban river model are vital for water pollution control technology based on the sewer–WWTP–river system. To improve the prediction accuracy of the integrated sewer network and urban river model, a parameter estimation method based on the LSTM model surrogates has been developed. This method overcomes the model accuracy limitation in converting this technology from theoretical research to practical application. Furthermore, runoff simulation with the integrated model may be combined with urban low-impact development techniques to explore the effect of runoff pollution reduction on the urban water environment. Water transport simulation within the sewer network may be combined with separate sewer system construction and sewage circulation to investigate the effect of various sewage treatment and reuse methods. Pollution discharge simulation may be combined with pollution control technology involving planting and breeding to examine the effect of non-point source pollution reduction strategies. Given the wide applicability of the LSTM, the model optimization system based on LSTM + ACO built in this research holds the potential for generalization to diverse process-based models. Since LSTM is the key to correlating ACO with process-based models, the accuracy of LSTM emulation determines the successful construction and application feasibility of the LSTM + ACO optimization system. Recent studies indicate that the excellent emulation effect

of LSTM for time series process-based models extends beyond river or water quality models, including groundwater, lake, reservoir, drinking water, and even air temperature, $PM_{2.5}$. Consequently, the LSTM + ACO optimization system is also suitable for the above process-based models [61–66]. The LSTM model and ACO algorithm in this research are objective techniques and ready for automatic operation. In the future, the LSTM + ACO method might be integrated with online environmental sensors to realize digital environment management with the help of emerging technologies such as cloud computation.

4. Conclusions

In this study, we proposed a machine learning-based parameter search parallel system for integrated process-based models. Our approach involved employing the LSTM model, renowned for its exceptional data series prediction capabilities, as a simulator for the coupled mechanism model. By combing the favorable operational characteristics of the ACO algorithm, we optimized the parameters of the LSTM simulator and incorporate the optimized parameters into the integrated process-based model. The parallel system could offer a solution to address the limitations of traditional integrated process-based models, such as insufficient monitoring data, long modeling time, high computational cost, and overdependence on the parameters. By achieving efficiency and performance enhancements, the original model preserves its complexity, accuracy, and stability.

The feasibility of the parallel system with optimized parameters was verified via the SWMM-WASP integrated process-based model in the Jiuqu River basin. The performance assessment of the LSTM simulation model revealed a remarkable resemblance between the LSTM fitting accuracy and that of the coupled mechanism model, particularly for the river flow, ammonia nitrogen concentration, and total phosphorus concentration, yielding a high coefficient of determination values (R^2 = 0.999, 0.996, and 0.995, respectively). Furthermore, model prediction performance analysis indicated that the integrated process-based model with optimized parameters was highly accurate regarding the annual water flow, ammonia nitrogen concentration, and total phosphorus concentration. The analysis of the model parameters demonstrated that the parameters related to degradation exhibited the most significant changes, followed by those related to infiltration. The optimization of pollutant degradation and the dry season water infiltration process were pivotal factors in the significant improvement in the model prediction accuracy during the dry season.

The integrated process-based model with optimized parameters via the LSTM-ACO approach could accurately simulate the river water quality under the influence of the sewer–WWTP–river system within a short period. This model provides valuable support for pollution control across various stages, including pollution generation, accumulation, transmission, discharge, and degradation. Moreover, it offers potential solutions to urban water challenges. In the future, this technique may be combined with online water sensors, cloud computing, and other emerging technologies to play a significant role in developing digital cities, making it feasible to efficiently and systematically tackle complex multimedia water environment issues.

CRediT author contribution statement

Yundong Li: Conceptualization, Writing - Original Draft, Methodology, Validation, Formal Analysis, Visualization. **Lina Ma:** Conceptualization, Software, Resources, Visualization. **Jingshui Huang:** Writing - Review & Editing. **Markus Disse:** Resources. **Wei Zhan:** Formal Analysis, Visualization. **Lipin Li:** Project Administration, Funding Acquisition. **Tianqi Zhang:** Formal Analysis, Investigation. **Huihang Sun:** Formal Analysis, Investigation. **Yu Tian:** Formal Analysis, Supervision, Funding Acquisition.

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 study was supported by the National Key R&D Program of China (2019YFD1100300) and the Fellowship of China Postdoctoral Science Foundation (2020M681105). The authors also acknowledge the State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology (No. 2021TS23).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2023.100320.

References

[1] D. Yin, et al., Integrated 1D and 2D model for better assessing runoff quantity control of low impact development facilities on community scale, Sci. Total Environ. 720 (Jun. 2020) 137630, https://doi.org/10.1016/J.SCITOTENV.2020.137630.
[2] L. Leng, et al., Performance assessment of coupled green-grey-blue systems for Sponge City construction, Sci. Total Environ. 728 (Aug. 2020) 138608, https://doi.org/10.1016/J.SCITOTENV.2020.138608.
[3] M. Mair, et al., The application of a Web-geographic information system for improving urban water cycle modelling, Water Sci. Technol. 70 (11) (2014) 1838–1846, https://doi.org/10.2166/WST.2014.327.
[4] W. Rauch, et al., Modelling transitions in urban water systems, Water Res. 126 (Dec. 2017) 501–514, https://doi.org/10.1016/J.WATRES.2017.09.039.
[5] C. Urich, W. Rauch, Modelling the urban water cycle as an integrated part of the city: a review, Water Sci. Technol. 70 (11) (2014) 1857–1872, https://doi.org/10.2166/WST.2014.363.
[6] A. Casal-Campos, G. Fu, D. Butler, A. Moore, An integrated environmental assessment of green and gray infrastructure strategies for robust decision making, Environ. Sci. Technol. 49 (14) (Jul. 2015) 8307–8314, https://doi.org/10.1021/ES506144F.
[7] C. Sun, B. Parellada, J. Feng, V. Puig, G. Cembrano, Factors influencing the stormwater quality model of sewer networks and a case study of Louis Fargue urban catchment in Bordeaux, France, Water Sci. Technol. 81 (10) (May 2020) 2232–2243, https://doi.org/10.2166/WST.2020.280.
[8] S. Gato, N. Jayasuriya, P. Roberts, Temperature and rainfall thresholds for base use urban water demand modelling, J. Hydrol. (Amst.) 337 (3–4) (Apr. 2007) 364–376.
[9] N. Mostafavi, H.R. Shojaei, A. Beheshtian, S. Hoque, Residential water consumption modeling in the integrated urban metabolism analysis tool (IUMAT), Resour. Conserv. Recycl. 131 (Apr. 2018) 64–74, https://doi.org/10.1016/J.RESCONREC.2017.12.019.
[10] F. Tscheikner-Gratl, M. Lepot, A.M. Moreno-Rodenas, A.N.A. Schellart, 'QUICS D.6.7 - A Framework for the Application of Uncertainty Analysis', Oct. 2017, https://doi.org/10.5281/ZENODO.1240926.
[11] L. Benedetti, et al., Modelling and monitoring of integrated urban wastewater systems: review on status and perspectives, Water Sci. Technol. 68 (6) (2013) 1203–1215, https://doi.org/10.2166/WST.2013.397.
[12] A. Deletic, et al., Assessing uncertainties in urban drainage models, Phys. Chem. Earth, Parts A/B/C 42–44 (Jan. 2012) 3–10, https://doi.org/10.1016/J.PCE.2011.04.007.
[13] H.R. Maier, et al., Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions, Environ. Model. Software 62 (Dec. 2014) 271–299, https://doi.org/10.1016/J.ENVSOFT.2014.09.013.
[14] S. Razavi, B.A. Tolson, D.H. Burn, Review of surrogate modeling in water resources, Water Resour. Res. 48 (7) (2012), https://doi.org/10.1029/2011WR011527.
[15] J.P.C. Kleijnen, Kriging metamodeling in simulation: a review, Eur. J. Oper. Res. 192 (3) (Feb. 2009) 707–716, https://doi.org/10.1016/J.EJOR.2007.10.013.
[16] K. Behzadian, Z. Kapelan, D. Savic, A. Ardeshir, Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks, Environ. Model. Software 24 (4) (Apr. 2009) 530–541, https://doi.org/10.1016/J.ENVSOFT.2008.09.013.

[17] W. Gong, et al., Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models, Water Resour. Res. 52 (3) (Mar. 2016) 1984–2008, https://doi.org/10.1002/2015WR018230.
[18] G. Kourakos, A. Mantoglou, Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models, Adv. Water Resour. 32 (4) (Apr. 2009) 507–521, https://doi.org/10.1016/J.ADVWATRES.2009.01.001.
[19] S. Yan, B. Minsker, Applying dynamic surrogate models in noisy genetic algorithms to optimize groundwater remediation designs, J. Water Resour. Plann. Manag. 137 (3) (May 2011) 284–292, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000106.
[20] J. Yazdi, A. Hokmabadi, M.R. JaliliGhazizadeh, Optimal size and placement of water hammer protective devices in water conveyance pipelines, Water Resour. Manag. 33 (2) (Jan. 2019) 569–590, https://doi.org/10.1007/S11269-018-2120-4/TABLES/7.
[21] A.M. Moreno-Rodenas, J.G. Langeveld, F.H.L.R. Clemens, Parametric emulation and inference in computationally expensive integrated urban water quality simulators, Environ. Sci. Pollut. Control Ser. 27 (13) (May 2020) 14237–14258, https://doi.org/10.1007/S11356-019-05620-1/FIGURES/18.
[22] J. Yin, F.T.C. Tsai, Bayesian set pair analysis and machine learning based ensemble surrogates for optimal multi-aquifer system remediation design, J. Hydrol. (Amst.) 580 (Jan. 2020) 124280, https://doi.org/10.1016/J.JHYDROL.2019.124280.
[23] K. Manley, C. Nyelele, B.N. Egoh, A review of machine learning and big data applications in addressing ecosystem service research gaps, Ecosyst. Serv. 57 (Oct. 2022) 101478, https://doi.org/10.1016/J.ECOSER.2022.101478.
[24] P. Arriagada, B. Karelovic, O. Link, Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm, J. Hydrol. (Amst.) 598 (Jul. 2021) 126454, https://doi.org/10.1016/J.JHYDROL.2021.126454.
[25] J.H. Cho, S.R. Ha, Parameter optimization of the QUAL2K model for a multiple-reach river using an influence coefficient algorithm, Sci. Total Environ. 408 (8) (Mar. 2010) 1985–1991, https://doi.org/10.1016/J.SCITOTENV.2010.01.025.
[26] L. Liu, et al., Comprehensive evaluation of parameter importance and optimization based on the integrated sensitivity analysis system: a case study of the BTOP model in the upper Min River Basin, China, J. Hydrol. (Amst.) 610 (Jul. 2022) 127819, https://doi.org/10.1016/J.JHYDROL.2022.127819.
[27] X. Lu, et al., Development and application of a hybrid long-short term memory – three dimensional variational technique for the improvement of PM2.5 forecasting, Sci. Total Environ. 770 (May 2021) 144221, https://doi.org/10.1016/J.SCITOTENV.2020.144221.
[28] M.T. Vu, A. Jardani, N. Massei, M. Fournier, Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network, J. Hydrol. (Amst.) 597 (Jun. 2021) 125776, https://doi.org/10.1016/J.JHYDROL.2020.125776.
[29] Z. Chen, et al., A transfer Learning-Based LSTM strategy for imputing Large-Scale consecutive missing data and its application in a water quality prediction system, J. Hydrol. (Amst.) 602 (Nov. 2021) 126573, https://doi.org/10.1016/J.JHYDROL.2021.126573.
[30] Y.W. Kim, et al., Forecasting abrupt depletion of dissolved oxygen in urban streams using discontinuously measured hourly time-series data, Water Resour. Res. 57 (4) (Apr. 2021) e2020WR029188, https://doi.org/10.1029/2020WR029188.
[31] P. Liu, J. Wang, A.K. Sangaiah, Y. Xie, X. Yin, Analysis and prediction of water quality using LSTM deep neural networks in IoT environment, Sustainability 11 (7) (Apr. 2019) 2058, https://doi.org/10.3390/SU11072058, 2019, Vol. 11, Page 2058.
[32] Y. Deng, et al., New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloketones in tap water, Sci. Total Environ. 772 (Jun. 2021) 145534, https://doi.org/10.1016/J.SCITOTENV.2021.145534.
[33] N. Noori, L. Kalin, S. Isik, Water quality prediction using SWAT-ANN coupled approach, J. Hydrol. (Amst.) 590 (Nov. 2020) 125220, https://doi.org/10.1016/J.JHYDROL.2020.125220.
[34] H. Wan, et al., Incorporating fish tolerance to supersaturated total dissolved gas for generating flood pulse discharge patterns based on a simulation-optimization approach, Water Resour. Res. 57 (9) (Sep. 2021) e2021WR030167, https://doi.org/10.1029/2021WR030167.
[35] J.L. Bertrand-Krajewski, G. Chebbo, A. Saget, Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon, Water Res. 32 (8) (1998) 2341–2356.
[36] M. Verdaguer, N. Clara, O. Gutiérrez, M. Poch, Application of Ant-Colony-Optimization algorithm for improved management of first flush effects in urban wastewater systems, Sci. Total Environ. 485–486 (1) (Jul. 2014) 143–152, https://doi.org/10.1016/J.SCITOTENV.2014.02.140.
[37] M.H. Afshar, A parameter free Continuous Ant Colony Optimization Algorithm for the optimal design of storm sewer networks: constrained and unconstrained approach, Adv. Eng. Software 41 (2) (Feb. 2010) 188–195, https://doi.org/10.1016/J.ADVENGSOFT.2009.09.009.
[38] X.L. Zhang, X.F. Chen, Z.J. He, An ACO-based algorithm for parameter optimization of support vector machines, Expert Syst. Appl. 37 (9) (Sep. 2010) 6618–6628, https://doi.org/10.1016/J.ESWA.2010.03.067.
[39] L. Hou, Z. Zhou, R. Wang, J. Li, F. Dong, J. Liu, Research on the non-point source pollution characteristics of important drinking water sources, Water 14 (2) (Jan. 2022) 211, https://doi.org/10.3390/W14020211, 2022, Vol. 14, Page 211.
[40] 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 (Jun. 2020) 110400, https://doi.org/10.1016/J.JENVMAN.2020.110400.
[41] J. Yao, M. Fan, Y. Xiao, X. Liang, C. Cai, Y. Wang, Spatial–temporal characteristics of corrected total phosphorus pollution loads from agricultural non-point sources in Tuojiang River watershed, Sichuan Province of south-western China, Environ. Sci. Pollut. Control Ser. (Jan. 2023) 1–22, https://doi.org/10.1007/S11356-023-25244-W/FIGURES/7.
[42] M.O. Arjenaki, · Hamed, R.Z. Sanayei, Heisam Heidarzadeh, Niloofar, A. Mahabadi, Modeling and investigating the effect of the LID methods on collection network of urban runoff using the SWMM model (case study: shahrekord City), Model. Earth Syst. Environ. 7 (1) (Jun. 2020) 1–16, https://doi.org/10.1007/S40808-020-00870-2, 2020 7:1.
[43] S.S. Baek, et al., Optimizing low impact development (LID) for stormwater runoff treatment in urban area, Korea: experimental and modeling approach, Water Res. 86 (Dec. 2015) 122–131, https://doi.org/10.1016/J.WATRES.2015.08.038.
[44] Q. Li, F. Wang, Y. Yu, Z. Huang, M. Li, Y. Guan, Comprehensive performance evaluation of LID practices for the sponge city construction: a case study in Guangxi, China, J. Environ. Manag. 231 (Feb. 2019) 10–20, https://doi.org/10.1016/J.JENVMAN.2018.10.024.
[45] A. Katsouli, A.S. Stasinakis, Production of municipal solid waste and sewage in European refugees' camps: the case of Lesvos, Greece, Waste Manag. 86 (Mar. 2019) 49–53, https://doi.org/10.1016/J.WASMAN.2019.01.036.
[46] P.F. Wang, J. Martin, G. Morrison, Water quality and eutrophication in tampa bay, Florida, Estuar. Coast Shelf Sci. 49 (1) (1999) 1–20, https://doi.org/10.1006/ECSS.1999.0490.
[47] T.A. Wool, S.R. Davie, H.N. Rodriguez, Development of three-dimensional hydrodynamic and water quality models to support total maximum daily load decision process for the neuse river estuary, North Carolina, J. Water Resour. Plann. Manag. 129 (4) (Jul. 2003) 295–306, https://doi.org/10.1061/(ASCE)0733-9496(2003)129:4(295).
[48] T. Wool, R.B. Ambrose, J.L. Martin, A. Comer, Wasp 8: the next generation in the 50-year evolution of USEPA's water quality model, Water 12 (5) (May 2020) 1398, https://doi.org/10.3390/W12051398, 2020, Vol. 12, Page 1398.
[49] Z. Liang, R. Zou, X. Chen, T. Ren, H. Su, Y. Liu, Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach, J. Hydrol. (Amst.) 581 (Feb. 2020) 124432, https://doi.org/10.1016/J.JHYDROL.2019.124432.
[50] X. Chen, N. Zang, F. Wu, B. He, Stormwater management model (SWMM): principles, parameters and applications, China Water & Wastewater 29 (2013) 4–7, 04.
[51] Zang, Z., Wang, X., Li, M., 2104. Uncertainty analysis of WASP based on global sensitivity analysis method. Environ. Sci. Resour. Util. 34(05), 1336-1346..
[52] F. Wang, J. Yang, Y. Li, X. Yang, X. Zhong, Modification of WASP model based on release of sediment phosphorus, Environ. Sci. Resour. Util. 33 (12) (2013) 3301–3308, https://doi.org/10.13671/j.hjkxxb.2013.12.021.
[53] M. Nayeb Yazdi, M. Ketabchy, D.J. Sample, D. Scott, H. Liao, An evaluation of HSPF and SWMM for simulating streamflow regimes in an urban watershed, Environ. Model. Software 118 (Aug. 2019) 211–225, https://doi.org/10.1016/J.ENVSOFT.2019.05.008.
[54] D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, T.L. Veith, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 50 (3) (2007) 885–900, https://doi.org/10.13031/2013.23153.
[55] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, M. Herrnegger, Rainfall-runoff modelling using long short-term memory (LSTM) networks, Hydrol. Earth Syst. Sci. 22 (11) (2018), https://doi.org/10.5194/hess-22-6005-2018.
[56] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Network. 5 (2) (1994), https://doi.org/10.1109/72.279181.
[57] J. Li, W. Lu, J. Luo, Groundwater contamination sources identification based on the Long-Short Term Memory network, J. Hydrol. (Amst.) 601 (2021), https://doi.org/10.1016/j.jhydrol.2021.126670.
[58] E.B. Mackay, et al., Dissolved organic nutrient uptake by riverine phytoplankton varies along a gradient of nutrient enrichment, Sci. Total Environ. 722 (Jun. 2020) 137837, https://doi.org/10.1016/J.SCITOTENV.2020.137837.
[59] S. Begum, M. Adnan, C.J. McClean, M.S. Cresser, A critical re-evaluation of controls on spatial and seasonal variations in nitrate concentrations in river waters throughout the River Derwent catchment in North Yorkshire, UK, Environ. Monit. Assess. 188 (5) (May 2016) 1–11, https://doi.org/10.1007/S10661-016-5305-4/FIGURES/6.
[60] H. Salo, L. Warsta, M. Turunen, M. Paasonen-Kivekäs, J. Nurminen, H. Koivusalo, Development and application of a solute transport model to describe field-scale nitrogen processes during autumn rains 65 (Mar. 2015) 30–43, https://doi.org/10.1080/09064710.2014.971861.
[61] C. Liang, H. Li, M. Lei, Q. Du, Dongting Lake water level forecast and its relationship with the three gorges dam based on a long short-term memory network, Water 10 (10) (Oct. 2018) 1389, https://doi.org/10.3390/W10101389, 2018, Vol. 10, Page 1389.
[62] D. Valadkhan, R. Moghaddasi, A. Mohammadinejad, Groundwater quality prediction based on LSTM RNN: an Iranian experience, Int. J. Environ. Sci. Technol. 19 (11) (Nov. 2022) 11397–11408, https://doi.org/10.1007/S13762-022-04356-9/TABLES/4.

[63] Q. Zhang, C. Wei, Y. Wang, S. Du, Y. Zhou, H. Song, Potential for prediction of water saturation distribution in reservoirs utilizing machine learning methods, Energies 12 (19) (Sep. 2019) 3597, https://doi.org/10.3390/EN12193597, 2019, Vol. 12, Page 3597.

[64] H. Mohammed, H.M. Tornyeviadzi, R. Seidu, Modelling the impact of weather parameters on the microbial quality of water in distribution systems, J. Environ. Manag. 284 (Apr. 2021) 111997, https://doi.org/10.1016/J.JENVMAN.2021.111997.

[65] J. Hou, Y. Wang, B. Hou, J. Zhou, Q. Tian, Spatial Simulation and Prediction of Air Temperature Based on CNN-LSTM 37 (1) (Dec. 2023), https://doi.org/10.1080/08839514.2023.2166235.

[66] T.H. Choe, C.S. Ho, An improvement of PM2.5 concentration prediction using optimized deep LSTM, Int. J. Environ. Pollut. 69 (3–4) (2022) 249–260, https://doi.org/10.1504/IJEP.2021.126976.