

A WaveNet-based convolutional neural network for river water level prediction

Jun Chen^{a,b,*}, Yanhua Huang^{a,c}, Teng Wu^d and Jing Yan^{a,b}

^a College of Water Conservancy and Hydropower Engineering, Hohai University, 1 Xikang Road, Nanjing 210098, China

^b State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, 1 Xikang Road, Nanjing 210098, China

^c Guangdong South China High-tech Co., Ltd., 105 Tianshou Road, Guangzhou 510610, China

^d College of Harbour, Coastal and Offshore Engineering, Hohai University, 1 Xikang Road, Nanjing 210098, China

*Corresponding author. E-mail: chenjunhhu@hhu.edu.cn

id JC, 0000-0002-2563-7564

ABSTRACT

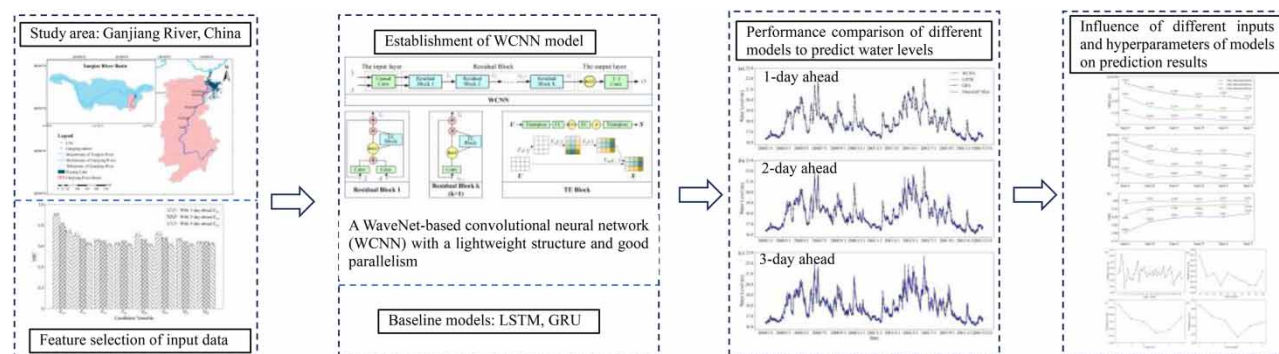
River water level prediction (WLP) plays an important role in flood control, navigation, and water supply. In this study, a WaveNet-based convolutional neural network (WCNN) with a lightweight structure and good parallelism was developed to improve the prediction accuracy and time effectiveness of WLP. It was applied to predict 1/2/3 days ahead the water levels at the Waizhou gauging station of the Ganjiang River (GR) in China, and it was compared with two recurrent neural networks (long short-term memory (LSTM) and gated recurrent unit (GRU)). The results showed that the WCNN model achieved the best prediction performance with the fewest training parameters and time. Compared with the LSTM and GRU models in the 1-day ahead prediction, the training parameters were reduced from 73,851 and 55,851 to 32,937, respectively. The root mean square error (RMSE) was reduced from 0.071 and 0.076 to 0.057, respectively. The mean absolute error (MAE) was reduced from 0.052 and 0.059 to 0.038, respectively. The Nash–Sutcliffe efficiency (NSE) and coefficient of determination (R^2) both increased to 0.998. This result indicated that the improved model was more efficient for WLP.

Key words: convolutional neural network, Ganjiang River, machine learning model, water level prediction, WaveNet

HIGHLIGHTS

- A WaveNet-based convolutional neural network was proposed for water level prediction.
- WCNN with a lightweight structure and good parallelism achieved better prediction performance.
- WCNN obtained higher accuracy results with the fewest parameters and training time than RNN.
- Influence of different inputs and hyperparameters of models on prediction results was revealed.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The water level is an important hydrological feature of rivers, and accurate water level prediction (WLP) is crucial to flood control, shipping and water supply planning and management (Deng *et al.* 2021). However, variations in the water level are highly nonlinear due to various factors, such as rainfall, runoff, topography, water conservancy projects, and human activities, which increase the difficulty of accurate prediction (Lai *et al.* 2013; Zhang *et al.* 2016).

The WLP models mainly include hydrologic models based on physical processes and machine learning (ML) models based on a data-driven approach (abbreviations: physical models and ML models). Physical models can obtain the meticulous simulation of water level processes by establishing basic equations to express the interaction mechanism between variables (Lai *et al.* 2013). However, the establishment of physical models requires a wealth of professional knowledge, a large number of basic data and physical parameters, the modelling process is complex, difficult and time-consuming. ML models do not need to understand the mechanism of physical systems, and they predict the water level by directly detecting correlations between variables and realizing the mapping of inputs to outputs. The data that ML models require are easy to obtain, and the modelling process is relatively simple (Yin *et al.* 2021).

Various ML models were used in hydrological prediction tasks. For example, in Ahmed *et al.* (2022), six different ML algorithms (i.e., linear regression (LR), interaction regression (IR), robust regression (RR), stepwise regression (SR), support vector regression (SVR), boosted trees ensemble regression (BOOSTER), bagged trees ensemble regression (BAGER), XGBoost, tree regression (TR), and Gaussian process regression (GPR)) were developed to predict the river's water level, on a daily basis based on collected data from 1990 to 2019 which were used to train and test the proposed models. They found that the GPR model was capable of predicting the water level of the river with high precision and less uncertainty. In Latif & Ahmed (2020), (2023) and Latif (2023), the LSTM model was used to predict the daily streamflow of the Kowmung River at Cedar Ford in Australia, daily reservoir inflow of the Dokan Dam in Iraq and daily pan evaporation at Sydney airport in Australia, respectively. The results showed the LSTM model outperformed other conventional ML models (e.g., random forest (RF), tree boost (TB), multilayer perceptron neural network (MLP-NN), and boosted regression tree (BRT)). Huang *et al.* (2021) built three ML models, namely, an artificial neural network (ANN), a nonlinear autoregressive model with exogenous input (NARX) and GRU to simulate the daily Poyang Lake level from 2003 to 2016. They found that ML models with historical memory (i.e., the GRU model) were more suitable for simulating the Poyang Lake level under the influence of the Three Gorges Dam. In Ho *et al.* (2022), the LSTM model was proposed to predict short-term water levels in tidal sluice gates from 6 to 48 h ahead in the Bac Hung Hai irrigation system in Vietnam. The findings of this study highlighted the performance of LSTM models in providing high-accuracy short-period water level forecasts for areas near estuaries. Kima *et al.* (2022) used ML models such as gradient boosting (GB), support vector model (SVM), and LSTM to predict the flood water level of the Heungcheon bridge station which was located in the downstream of Bokha bridge. Rainfall, water level, and discharge data of Bokha bridge station from 2005 to 2020 were collected, especially the collected rainfall data were classified into 53 rainfall events using Interevent Time Definition (IETD) analysis of rainfall data. The LSTM model showed the best predictive power and was selected as the optimal model for real-time floodwater level forecasting in this study. Zhang *et al.* (2018) built four different neural networks to predict the water level of the combined sewer overflow structure in Norway, through a comparison of other different neural networks (e.g., multilayer perceptron (MLP) and wavelet neural network (WNN)), the LSTM and GRU presented superior capabilities for multistep-ahead time series prediction. In Cai *et al.* (2021), a GRU model was built for groundwater level simulation in 78 catchments in the central-eastern continental United States, the results showed that the GRU model performed better in regions where hydrogeological properties could promote more effective responses of groundwater to external changes. The above shows that the LSTM and GRU models in recurrent neural networks (RNNs) can usually obtain better prediction performance. However, they still have some shortcomings that need to be improved. For example, their internal recurrent connections required processing inputs according to time sequence, resulting in models that could not be trained in parallelism and increased training time (Fan *et al.* 2021). Furthermore, they consumed additional memory to hold long-term information. For practical problems with very large datasets, the LSTM and GRU models are relatively low in terms of cost-efficiency.

A convolutional neural network (CNN) is a lightweight structure that has unique advantages in capturing spatial dependencies of input data (Collado-Villaverde *et al.* 2021). Nevertheless, its performance on time series regression tasks is poor (Yan *et al.* 2020). A new convolutional structure named WaveNet, proposed by DeepMind in 2016, has performed well in

sequential analysis problems (van den Oord *et al.* 2018). It combines the advantages of dilated causal convolutions, residual, and skip connections. It not only receives nonfixed-length inputs but also learns complex long-term dependencies of sequences.

Compared with RNN models, WaveNet has a smaller number of parameters, large receptive fields, and good parallelism. It has good application prospects in solving sequence analysis problems with large datasets (Borovykh *et al.* 2019; Rizvi *et al.* 2021). However, WaveNet was initially proposed for audio generation tasks, and it cannot directly address the time series regression problem. Many researchers have improved it and achieved better prediction performance than LSTM or GRU in the prediction tasks on the power load (Wang *et al.* 2021), traffic flow (Zhang *et al.* 2021a, 2021b), and air quality (Benhaddi & Ouarzazi 2021) datasets. However, to the best of our knowledge, WaveNet has not been improved for WLP tasks.

This study aims to develop a highly efficient WLP model of the convolutional structure. A more lightweight CNN, named WaveNet-based convolutional neural network (WCNN), was built by improving WaveNet on the basis of WLP characteristics. The GR in China was selected as the study area, the 1/2/3 days ahead water levels were predicted for the Waizhou gauging station (abbreviation: Waizhou station), and the long-term sequence data of gauging stations could be used for verification. To make full use of the ability of the WCNN to extract spatial information, the water level and discharge sequences of three upstream gauging stations were used as auxiliary input features. Two RNN models (LSTM and GRU) were established, and the performance of the WCNN model was evaluated by comparing their results on the test set.

2. STUDY AREA AND DATA DESCRIPTION

The GR is the largest river of the Poyang Lake water system in the Yangtze River Basin (Figure 1), and its annual mean runoff accounts for 46% of the Poyang Lake basin. The GR originates in Wuyi Mountain at the junction of Fujian and Jiangxi

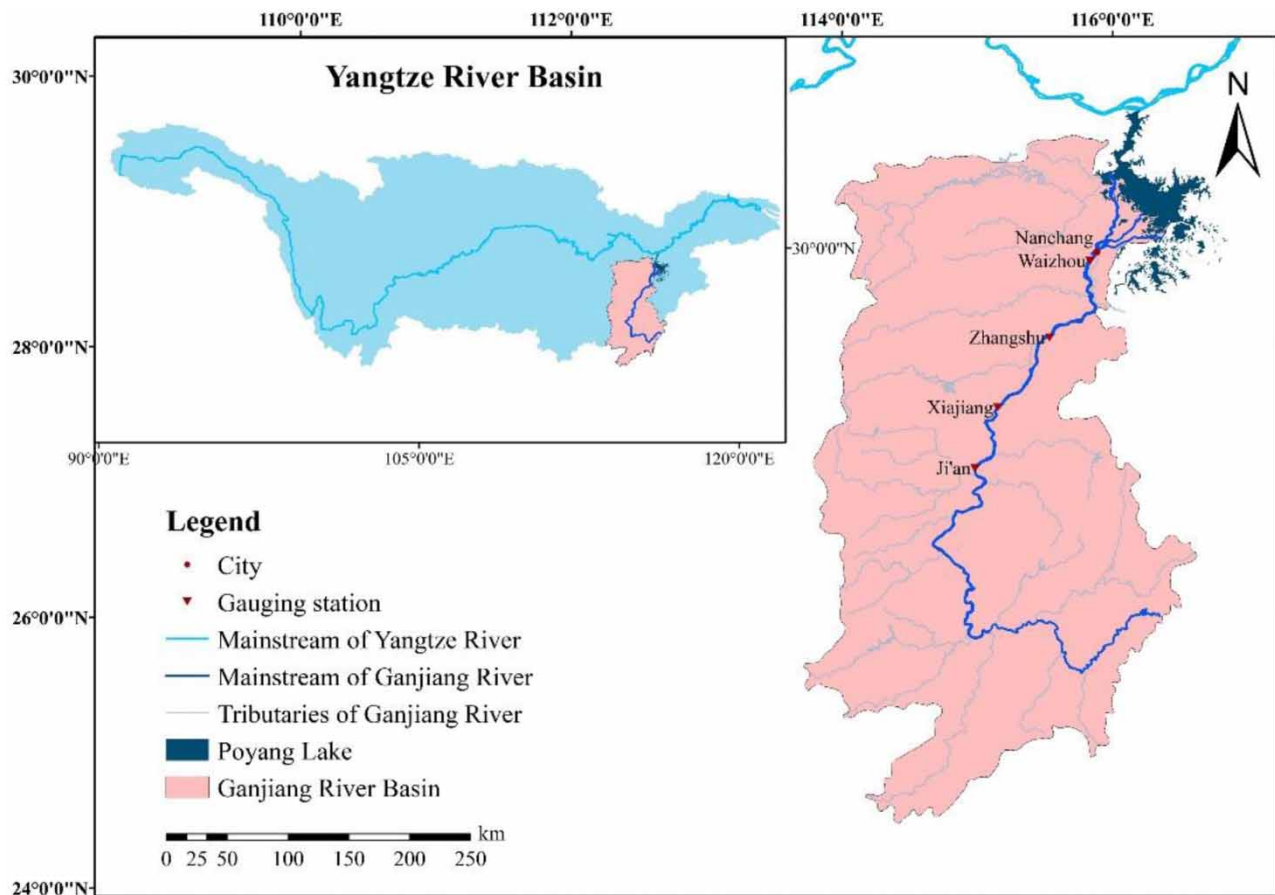


Figure 1 | Ganjiang River basin and distribution location of gauging stations.

Provinces. The length of the mainstream is 823 km, and the drainage area of the GR is 82,900 km² (Wen *et al.* 2019). The GR crosses through Jiangxi Province from south to north, and it is bifurcated near Nanchang City, which is the provincial capital. Then it flows into Poyang Lake. The Waizhou station, the catchment area which accounts for 97% of the GR basin, is a more important gauging station in the lower reach of the GR, and the station is located near Nanchang City. The WLP of this station is significant to the safety of flood control, shipping and water supply of the Nanchang reach, thus this station is selected as the prediction station.

Before the 21st century, the hydrological rule of the Nanchang reach was relatively stable. Since the early 2000s, affected by human activities such as large-scale sand mining in the Nanchang reach, the riverbed descended. From 2001 to 2019, the topographic elevation of the riverbed cross-section of Waizhou station descended more than 3 m (Figure 2), and the water level had obviously decreased, resulting in changes in the mapping relationships between the water level and influencing factors (Yao *et al.* 2019). In this study, the sequence data of the Waizhou station and upstream gauging stations from 1965 to 2001 (37 years in total) with relatively stable hydrological rules were used to set up the model. Considering the water transmission time and predicted time, Ji'an gauging station with a distance of 230 km was selected as the farthest station in the upstream of the Waizhou station. The locations of the gauging stations are shown in Figure 1. Specifically, the sequences of daily average water level and discharge data belonged to Waizhou, Zhangshu, Xiajiang and Ji'an stations (hereafter referred to as Z_{WZ} , Z_{ZS} , Z_{XJ} , Z_{JA} , Q_{WZ} , Q_{ZS} , Q_{XJ} , and Q_{JA} , respectively). These data were supplied by the Hydrological Bureau of Jiangxi Province. The statistical information from 1965 to 2001 is shown in Figure 3, which shows that the distributions of the water level and discharge at all stations are uneven, and there are many extreme values on the right side of the mean value.

3. METHODS

3.1. Procedure of WLP of models

The procedure of WLP by ML models in this study was as follows: (1) Feature selection was carried out on input data of models by calculating the maximal information coefficient (MIC), and variables with MIC values larger than 0.6 were selected to be input features. (2) The input data were divided into the training dataset and the test dataset. The data from 1965 to 1999 were used as the training dataset to train models, and the data from 2000 to 2001 were used as the test dataset to evaluate the generalization ability of models. Then the Min-Max normalization was performed on the input data. (3) The WCNN, LSTM and GRU models were established. Based on the training dataset, the grid search method was used to optimize the hyperparameters of models, and the optimal combination of hyperparameters was obtained. 80% of the training dataset was randomly selected as the training set for the model fitting training, and the remaining 20% were used as the validation set for calibration to improve the performance of models. (4) The data from 2000 to 2001 were used to test models, and the 1/2/3 days ahead predicted and measured water levels of the Waizhou station were compared. Evaluation metrics of MAE, RMSE and NSE were calculated to evaluate the performance of models. The flowchart is shown in Figure 4.

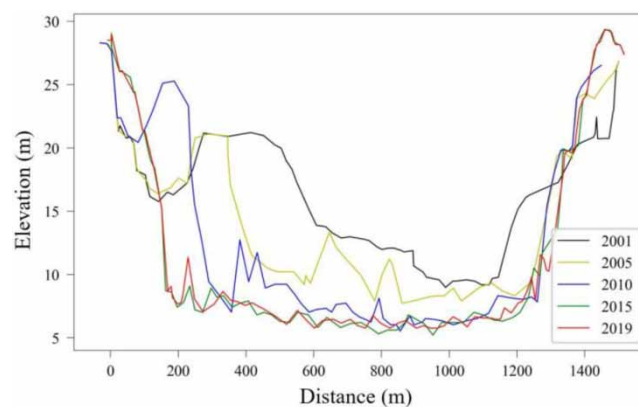


Figure 2 | Topographic changes of riverbed cross-section of the Waizhou station from 2001 to 2019.

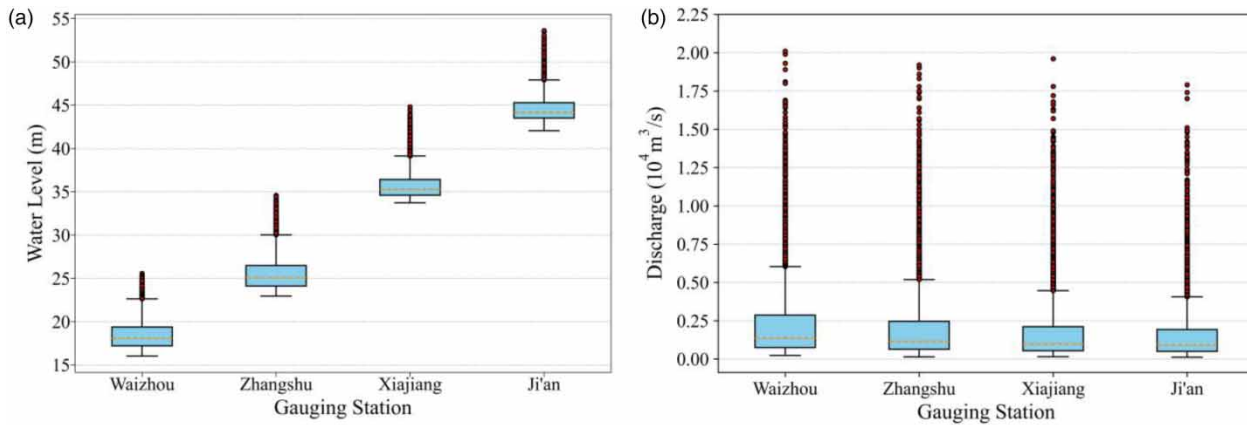


Figure 3 | Statistics of (a) water level and (b) discharge in the dataset from 1965 to 2001. The coloured box indicates the interquartile range (IQR), while the orange lines within the IQR represent the median line. Dots outside the lines at both ends (upper and lower edge lines) indicate extreme values.

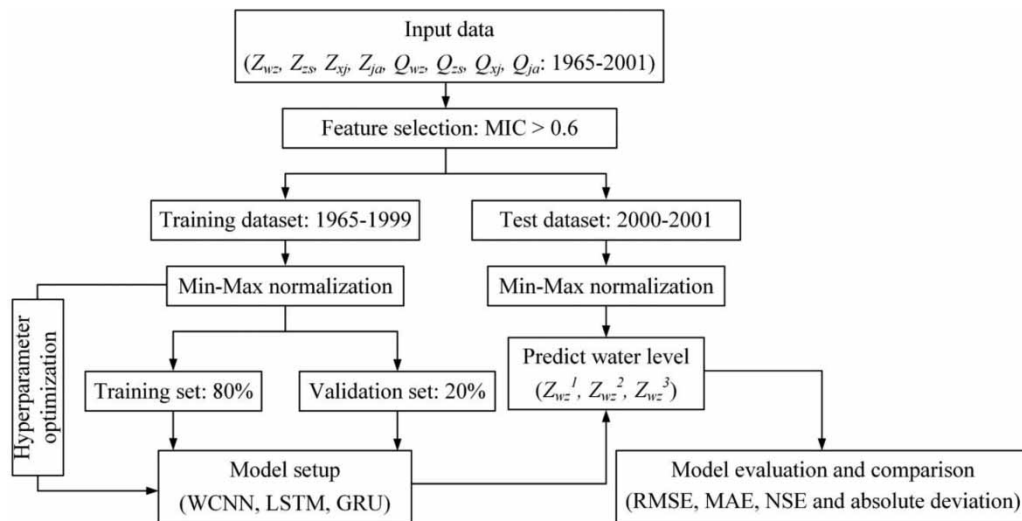


Figure 4 | Flowchart of water level prediction by ML models.

3.2. Feature selection

There are highly complex nonlinear correlations between predicted variables and influencing factors, and feature selection by calculating the MIC between predicted and candidate variables was conducted. The MIC was proposed by Reshef *et al.* (2011), which could express the nonlinear correlations between two variables and was widely used in feature selection in hydrological datasets (Lu *et al.* 2021; Zhang *et al.* 2021b). The calculation formula of MIC is shown in Formula (1), and the results between each variable (Z_{wz} , Z_{zs} , Z_{xj} , Z_{ja} , Q_{wz} , Q_{zs} , Q_{xj} , Q_{ja} , which are regarded as H) and predicted variables ($1/2/3$ days ahead water levels at the Waizhou station expressed as Z_{wz}^1 , Z_{wz}^2 , Z_{wz}^3 , respectively, which are regarded as L) from 1965 to 1999 are shown in Table 1. Variables with MIC values larger than 0.6 were selected in the original datasets to constitute input features. In the three forecast periods, all candidate variables met the requirements. Thus, the input features were Z_{wz} , Z_{zs} , Z_{xj} , Z_{ja} , Q_{wz} , Q_{zs} , Q_{xj} , and Q_{ja} . In addition, Min-Max normalization was used to scale features to $[0, 1]$.

$$\text{MIC}(D, H, L) = \max_{HL < B(|D|)} \frac{\max_G I(D|G)}{\log_2 \min\{H, L\}} \quad (1)$$

Table 1 | MICs between predicted and candidate variables

	Water levels of gauging stations				Discharges of gauging stations			
	Waizhou station	Zhangshu station	Xiajiang station	Ji'an station	Waizhou station	Zhangshu station	Xiajiang station	Ji'an station
1-day ahead water level Z_{wz}	0.89	0.70	0.64	0.61	0.70	0.71	0.65	0.62
2-day ahead water level Z_{wz}	0.80	0.65	0.63	0.63	0.64	0.66	0.62	0.62
3-day ahead water level Z_{wz}	0.73	0.61	0.60	0.60	0.60	0.60	0.60	0.61

where $H = \{h_1, h_2, \dots, h_q\}$ and $L = \{l_1, l_2, \dots, l_q\}$ are independent variables and dependent variables, respectively, and q is the length of H and L ; D is a set of ordered pairs, and $D = \{(h_i, l_i), i = 1, 2, \dots, q\}$; G is the cells of grid; D/G is the probability distribution caused by data D on G ; $I(D/G)$ denotes mutual information of D on G ; $B(q) = q^{0.6}$; and $|D|$ is the size of D .

3.3. The WCNN

WaveNet is a deep network model proposed by DeepMind for generating raw audio waveforms that have large and flexible receptive fields and good parallelism, and can capture long-term dependencies of sequences (Rethage *et al.* 2018; van den Oord *et al.* 2018). WaveNet can be extrapolated to datasets that are outside of audio, and many achievements have been made in research on time series prediction (Luo *et al.* 2021; Rueda *et al.* 2021; Nie *et al.* 2022).

In this study, a lightweight CNN framework was proposed based on WaveNet. The variables (X) and other auxiliary input features (A) before the t th timestep were used to predict x_t (t th timestep predictor), and the mapping formula is shown in Equation (2). Here, A is treated as the local condition LC , which is received using the ReLU function.

$$f(X|LC) = \prod_{t=1}^T f(x_t|x_1, \dots, x_{t-1}, LC) = \prod_{t=1}^T f(x_t|x_1, \dots, x_{t-1}, a_1^i, \dots, a_{t-1}^i) \quad (2)$$

where $X = \{x_1, x_2, \dots, x_{T-1}\}$, and T is the input length; $t = \text{range}(1, T)$; $A = \{a^1, a^2, \dots, a^p\}$, a^i denotes the i th auxiliary input features, $i = \text{range}(1, p)$, and p is the number of auxiliary input features; and a_t^i denotes the value of auxiliary input condition a^i at the t th timestep.

Figure 5 shows the network structure of the WCNN, which mainly includes the input layer, residual block and output layer, as described below.

- (1) The input layer received the sequences of predicted variables X and auxiliary input A , and the input shape was [timesteps, features]. As mentioned in Section 3.2, X was Z_{wz} , and A included Z_{zs} , Z_{xj} , Z_{ja} , Q_{wz} , Q_{zs} , Q_{xj} , and Q_{ja} in this study. To implement the residual operation, inputs were first transformed into the output shape of the residual blocks through a causal convolutional layer.
- (2) The main components of the residual block were the convolutional layer and ReLU function. Through these two structures, the order of data was ensured and the spatiotemporal nonlinear mapping of data was learned. Moreover, a temporal-excitation (TE) block based on a squeeze-and-excitation (SE) block was proposed to learn the long-term dependencies of data (Hu *et al.* 2019). The TE block obtained the global temporal information by explicitly modelling the relationships between the timesteps of the convolution channel U . The structure is shown in Figure 5. First, the transpose function $F_{tr}(\cdot)$ is used to swap the coordinate systems of the temporal features and channel features of the channel U . Then, the excitation operation $F_{ex}(\cdot)$ is used to capture the temporal dependence of the channel U , generating a set of modulation weights for each channel. Specifically, a fully connected (FC) layer with a dimensionality-reduction ratio r ($r = 2$) and ReLU function are used to parameterize the nonlinearity between the time steps, with an FC layer that restores the dimension and a sigmoid function that scales the weights, as shown in Figure 5. Finally, $F_{tr}(\cdot)$ is used to restore the coordinate system, and a multiplication operation $F_{mul}(\cdot, \cdot)$ is used to integrate the results into the backbone network.

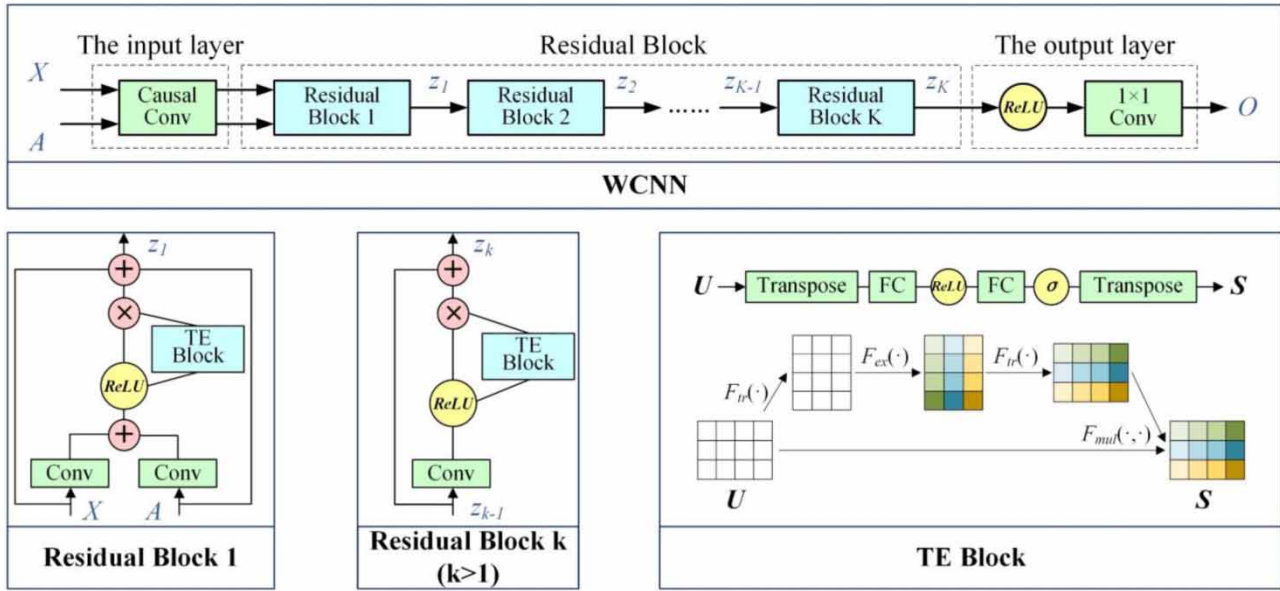


Figure 5 | WCNN network structure. Conv, 1×1 Conv and FC indicate the 1-D convolutional layer, 1×1 convolutional layer and FC layer, respectively, while σ is the sigmoid activation function.

The number and size of the convolution kernels for all residual blocks in the WCNN were the same, which ensured that all residual blocks outputted a uniform shape. As shown in Figure 5, in the first residual block, the predicted variables X and condition A are based on a convolution operation and ReLU function to obtain channel U containing temporal and spatial features. The calculation formula is shown in Equation (3). Subsequently, the TE block is utilized to learn the global temporal information of the channel U to recalibrate the convolutional channel features, and the calculation formula is shown in Equation (4). The output of the TE block is fed into the backbone network by multiplying it with U , and then, it is added to the inputs to obtain the final output of the residual block z_1 , as shown in Equations (5) and (6). The other residual blocks receive z_{k-1} and output z_k , and the calculation formulas are as follows in Equations (7)–(10).

For the first residual block ($k = 1$):

$$u_k = \delta \left(W_{f,k} \cdot X + b_{f,k} + \sum_{i=1}^p V_g^i \cdot a^i + b_g \right) \quad (3)$$

$$s_k = F_{tr}(F_{ex}(F_{tr}(u_k))) = (\sigma(W_{k,2} \cdot \delta(W_{k,1} \cdot u'_k + b_{k,1}) + b_{k,2}))' \quad (4)$$

$$e_k = u_k \odot s_k \quad (5)$$

$$z_k = e_k + X + \sum_{i=1}^p a^i \quad (6)$$

For other residual blocks ($k > 1$):

$$u_k = \delta(W_{f,k} \cdot z_{k-1} + b_{f,k}) \quad (7)$$

$$s_k = F_{tr}(F_{ex}(F_{tr}(u_k))) = (\sigma(W_{k,2} \cdot \delta(W_{k,1} \cdot u'_k + b_{k,1}) + b_{k,2}))' \quad (8)$$

$$e_k = u_k \odot s_k \quad (9)$$

$$z_k = e_k + z_{k-1} \quad (10)$$

where $k = \text{range}(1, K)$ denotes the k th residual blocks, and K is the number of residual blocks; $W_{f,k}$ and $b_{f,k}$ represent the weight and bias of the convolutional filters of X in the k th layer, respectively; $j = \text{range}(1, p)$ is the conditional index; V_g^i

represent the weight of the convolutional filters of a^i in the first layer, and b_g represent the bias of the convolutional filters in the first layer; δ is the ReLU function; u_k is the output of δ in the k th residual block. Additionally, $W_{k,1}$, $b_{k,1}$ and $W_{k,2}$, $b_{k,2}$ denote the weights and biases of the first and second FC layers in the TE block of the k th residual block; σ is the *sigmoid* function; u_k' is the transpose of u_k ; s_k is the output of the TE block in the k th residual block; e_k and z_k indicate the intermediate and final outputs of the k th residual block, respectively; \odot represents the multiplication of corresponding elements; and the remaining parameters are the same as in Formula (2).

- (3) The output layer is a 1×1 convolutional layer with linear activation. The output z_k of the last residual block first performs the calculation of the ReLU function and then enters the output layer and outputs $O = (o_{t-n+1}, o_{t-n+2}, \dots, o_t)$, n is the input length, and o_t denotes the t th timestep prediction, which is the required predicted value. The calculation formula is shown in Equation (11).

$$O = W_o \cdot \text{ReLU}(z_K) + b_o \quad (11)$$

where z_K is the output of the last residual block; W_o and b_o represent the weight and bias of the output layer, respectively; and O refers to the result of the output layer.

The WCNN inherits the advantages of WaveNet. It has a lightweight structure, good learning ability and parallelism, and can solve network degradation (vanishing/explosion gradient) problems. It differs from the original WaveNet in the following three points:

- (1) Replace the gated activation function in the residual structure with the ReLU function. It can learn nonlinear correlations of sequences effectively, making the WCNN more applicable to 1-D sequential regression problems. This replacement has been proven to be effective in nonstationary and noisy time series forecasting tasks (Borovykh *et al.* 2019). In addition, the application of the ReLU function can also reduce the complexity of the model and training time.
- (2) Replace the dilated convolutional structure with a TE block. A TE block based on an SE block is proposed to learn the long-term temporal dependencies of sequences. It obtains global temporal information by capturing the information between time steps and feeds back this information to the convolutional channel. Thus, the network can select key temporal features for mapping.
- (3) For application to 1-D sequential regression studies, the softmax distribution of the output layer is replaced with a linear function.

3.4. Baseline model

LSTM and GRU are widely used ML models, and have already been applied to various WLP tasks (Zhang *et al.* 2018; Ren *et al.* 2020; Noor *et al.* 2022). Thus, LSTM and GRU were chosen as the baseline models in this study. LSTM solves the short-term memory and vanishing gradient problems by setting three gate (e.g., forget, input, and output gate) structures and two hidden states (e.g., short-term and long-term state). The specific structure and calculation formulas of LSTM are shown in Gers *et al.* (1999). GRU is a simplified version of LSTM that combines the two state variables in the LSTM unit into one. It controls the forget gate and input gate by a single gate controller (Géron 2019).

3.5. Model parameter settings and evaluation metrics

The hyperparameters have a great influence on the prediction performance of the model. The basic network architecture of WCNN is different from that of LSTM and GRU models, thus, its hyperparameters are also different. The network hyperparameters of the WCNN model include the residual block number, kernel size and filter number of the convolutional layer. The network hyperparameters of LSTM and GRU models are the recurrent layer number and neurons number in each layer. The hyperparameters of the three models in the training stage are the same, which are optimizer, epochs, batch size and learning rate. For the above hyperparameters, the grid search method was used to determine the optimum value. This method selected one value from the value range of each hyperparameter of the model and combined the parameters to build the model, and then the training data were used to validate it. The optimal hyperparameter combination was determined by comparing the loss between the predicted value and the measured value on the validation set under different parameter combinations. The range of hyperparameters and the optimal parameter configuration of the three models are shown in Table 2. After optimization by the grid search method, the optimal parameter configuration of the three models is as follows: the WCNN model

Table 2 | The optimal parameter configuration of three models

Parameters	Optimization range	WCNN	LSTM	GRU
Layers	[1, 2, 3]	–	2	2
Neurons	[50, 100, 150]	–	[100, 50]	[100, 50]
Dropout	[0.1, 0.2, 0.3, 0.4, 0.5]	–	0.2	0.2
Residual blocks	[1, 2, 3, 4, 5, 6]	4	–	–
Kernel size	[1, 2, 3, 4, 5, 6, 7, 8]	5	–	–
Filter number	[20, 30, 40, 50, 60]	40	–	–
Timesteps or Input Length	[6, 10, 17, 30, 45]	10	10	10
Optimizer	[SGD,RMSprop,Adam,Nadam]	Adam	Adam	Adam
Epochs	[50,100,200,300,400,500]	500	500	500
Batch size	[32,64,100,128,256]	100	100	100
Initial learning rate	[0.01,0.005,0.001,0.0005,0.0001]	0.001	0.001	0.001

consists of four residual blocks. The filter number and kernel size of each residual block are 40 and 5, respectively. The LSTM and GRU models consist of two recurrent layers, the number of neurons in the first and second layers are 100 and 50, respectively. The dropout rate is 0.2. For the hyperparameters of models in the training stage, the optimal parameters of the three models are the same. Adam is selected to be the optimizer, the batch size is set to 100 and the timestep is 10. The initial learning rate is set to 0.001, which is reduced by 50% while the validation loss does not decrease in 20 epochs for scheduling.

The number of training parameters is an important measure of model training efficiency. The fewer the training parameters, the shorter the training time of the model. The number of training parameters of WCNN, LSTM and GRU under the same input conditions identified in Section 3.2 was calculated, which were 32,937, 73,851, and 55,851, respectively. The WCNN model had the fewest parameters, and the LSTM model had the most parameters.

MAE was chosen as the loss function of the LSTM, GRU, and WCNN models, as shown in Equations (12)–(14). To reduce the risk of overfitting, an early stopping technology was employed in three models. The random seed was set to 42 for all experiments to guarantee reproducibility of the model results. All coding in this study was implemented based on Python and the deep learning framework of TensorFlow-CPU 2.3.0, and the type of CPU was an Intel(R) Core (TM) i5-9600 K CPU @3.70 GHz.

$$\text{Loss}_{\text{LSTM}} = \frac{1}{m} \sum_{i=1}^m |y_i^{\hat{}} - y_i| \quad (12)$$

$$\text{Loss}_{\text{GRU}} = \frac{1}{m} \sum_{i=1}^m |y_i^{\hat{}} - y_i| \quad (13)$$

$$\text{Loss}_{\text{WCNN}} = \frac{1}{m} \sum_{i=1}^m |y_{i-1}^{\hat{}} - y_{i-1}| \quad (14)$$

where $\text{Loss}_{\text{LSTM}}$, Loss_{GRU} , and $\text{Loss}_{\text{WCNN}}$ represent the loss functions of the LSTM, GRU, and WCNN models, respectively; m is the batch size of the training set; $y_i^{\hat{}}$ and y_i represent the predicted and measured values of the i th input, respectively; $y_{i-1}^{\hat{}}$ and y_{i-1} represent the last predicted and measured values of the i th input.

Four criteria, the mean absolute error (MAE), root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), and coefficient of determination (R^2) were used to evaluate the performance of different models (Huang *et al.* 2021; Latif & Ahmed 2023). The closer the values of MAE and RMSE are to 0, and NSE and R^2 are to 1, the better the performance of the models.

The formulas for calculating MAE, RMSE, NSE and R^2 are as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (15)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (16)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (17)$$

$$R^2 = \left[\frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \right]^2 \quad (18)$$

where N is the test set size; \bar{y} and $\bar{\hat{y}}$ are the mean value of the measured and predicted values, respectively; and the remaining parameters are the same as in Equation (12).

4. RESULTS

4.1. Performance comparison of different models

The WCNN, LSTM and GRU models were built to simulate Z_{wz} for 1/2/3 days ahead, and the results are shown in Table 3. Compared with the LSTM and GRU models for 1-day ahead prediction of total water levels, the MAE and RMSE of the WCNN models are improved by 26.30 and 34.98% and by 19.97 and 24.93%, respectively. The NSE and R^2 are both improved to 0.998 on the complete test set.

Figures 6–8 show the hydrographs, deviations and scatter plots of the 1/2/3 days ahead measured and predicted water levels of three models. It can also be seen from Figure 6 that the WCNN model has the best prediction performance of

Table 3 | Evaluation metrics of the 1/2/3 days ahead predicted water level on the test set of three models

Range of water level	Metrics	Model & Forecast period								
		WCNN			LSTM			GRU		
		t + 1	t + 2	t + 3	t + 1	t + 2	t + 3	t + 1	t + 2	t + 3
All water levels	MAE	0.038	0.115	0.213	0.052	0.127	0.222	0.059	0.131	0.238
	RMSE	0.057	0.171	0.333	0.071	0.179	0.337	0.076	0.180	0.348
	NSE	0.998	0.979	0.920	0.996	0.977	0.918	0.996	0.977	0.912
	R^2	0.998	0.980	0.921	0.997	0.979	0.920	0.997	0.979	0.918
Water level < 17 m	MAE	0.021	0.047	0.077	0.036	0.077	0.087	0.071	0.092	0.104
	RMSE	0.031	0.067	0.102	0.044	0.089	0.112	0.080	0.106	0.130
	NSE	0.982	0.916	0.801	0.964	0.848	0.758	0.878	0.786	0.676
	R^2	0.982	0.922	0.831	0.977	0.914	0.819	0.973	0.904	0.805
17 m ≤ Water level ≤ 19 m	MAE	0.041	0.117	0.204	0.052	0.120	0.215	0.048	0.121	0.230
	RMSE	0.058	0.153	0.278	0.069	0.159	0.285	0.064	0.159	0.297
	NSE	0.990	0.930	0.768	0.986	0.924	0.757	0.988	0.924	0.736
	R^2	0.990	0.941	0.805	0.987	0.933	0.805	0.989	0.936	0.807
Water level > 19 m	MAE	0.059	0.219	0.466	0.078	0.230	0.469	0.071	0.229	0.484
	RMSE	0.082	0.297	0.619	0.106	0.306	0.621	0.101	0.301	0.633
	NSE	0.985	0.809	0.170	0.976	0.797	0.165	0.978	0.804	0.131
	R^2	0.986	0.832	0.451	0.979	0.842	0.442	0.978	0.836	0.427

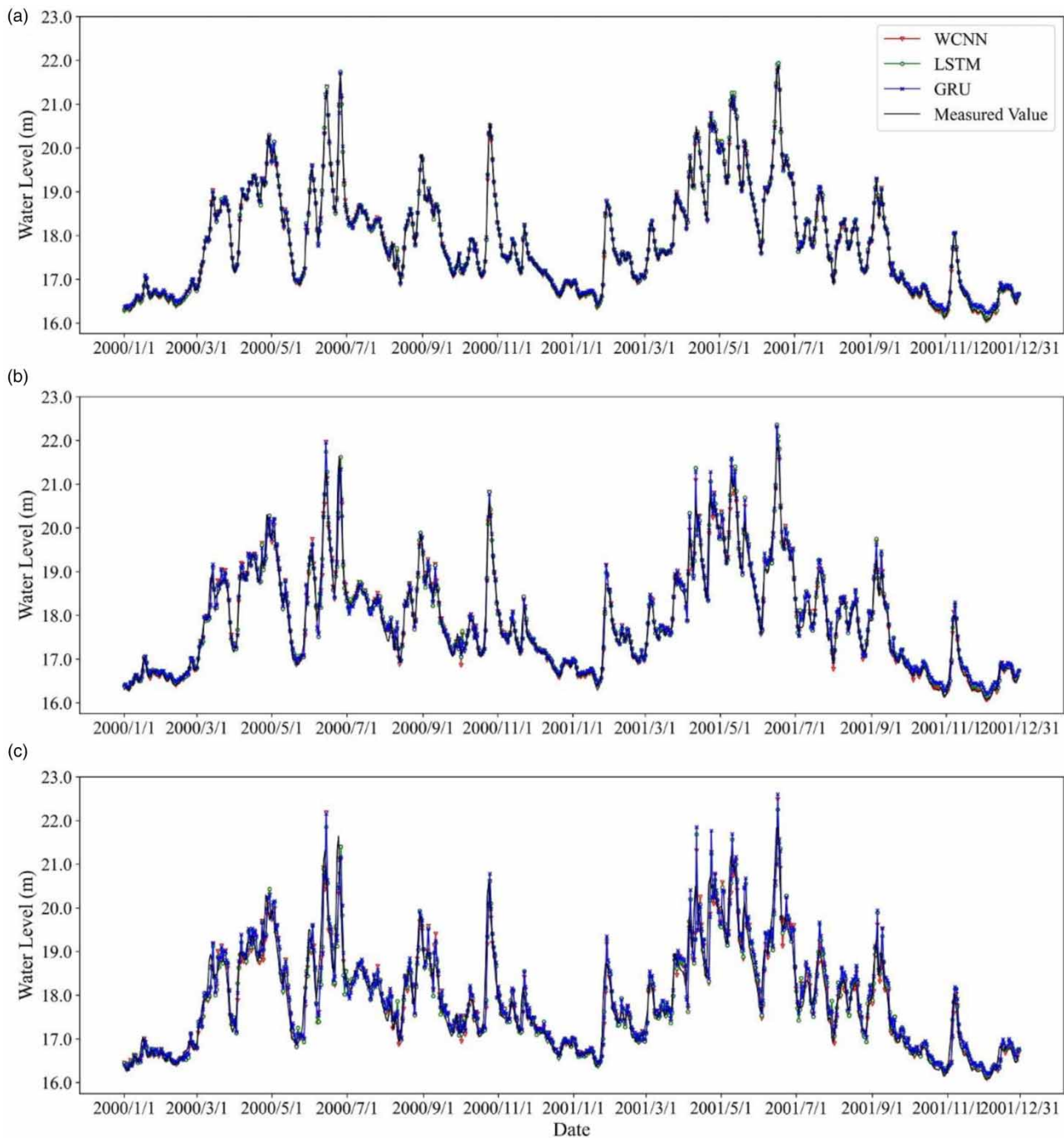


Figure 6 | The measured and (a) 1-day ahead, (b) 2-day ahead, and (c) 3-day ahead predicted water levels of three models on the test set 2000–2001.

water levels, and the fitting ability for extreme values is also better. The LSTM and GRU models tend to overestimate the maximum and underestimate minimum values, especially the GRU model. For 2 days ahead of WLP, the performance of all models decreased. The MAE of the WCNN, LSTM and GRU model on the complete test set (2000–2001) were increased from 0.038, 0.052 and 0.059 to 0.115, 0.127 and 0.131, respectively. When the forecast period reaches 3 days, the prediction error increases further. Overall, the WCNN model leads by a narrow margin. In addition, the predicted accuracy of low water levels (lower than 17 m) by different models is higher than that of middle water levels (between 17 and 19 m). The predicted

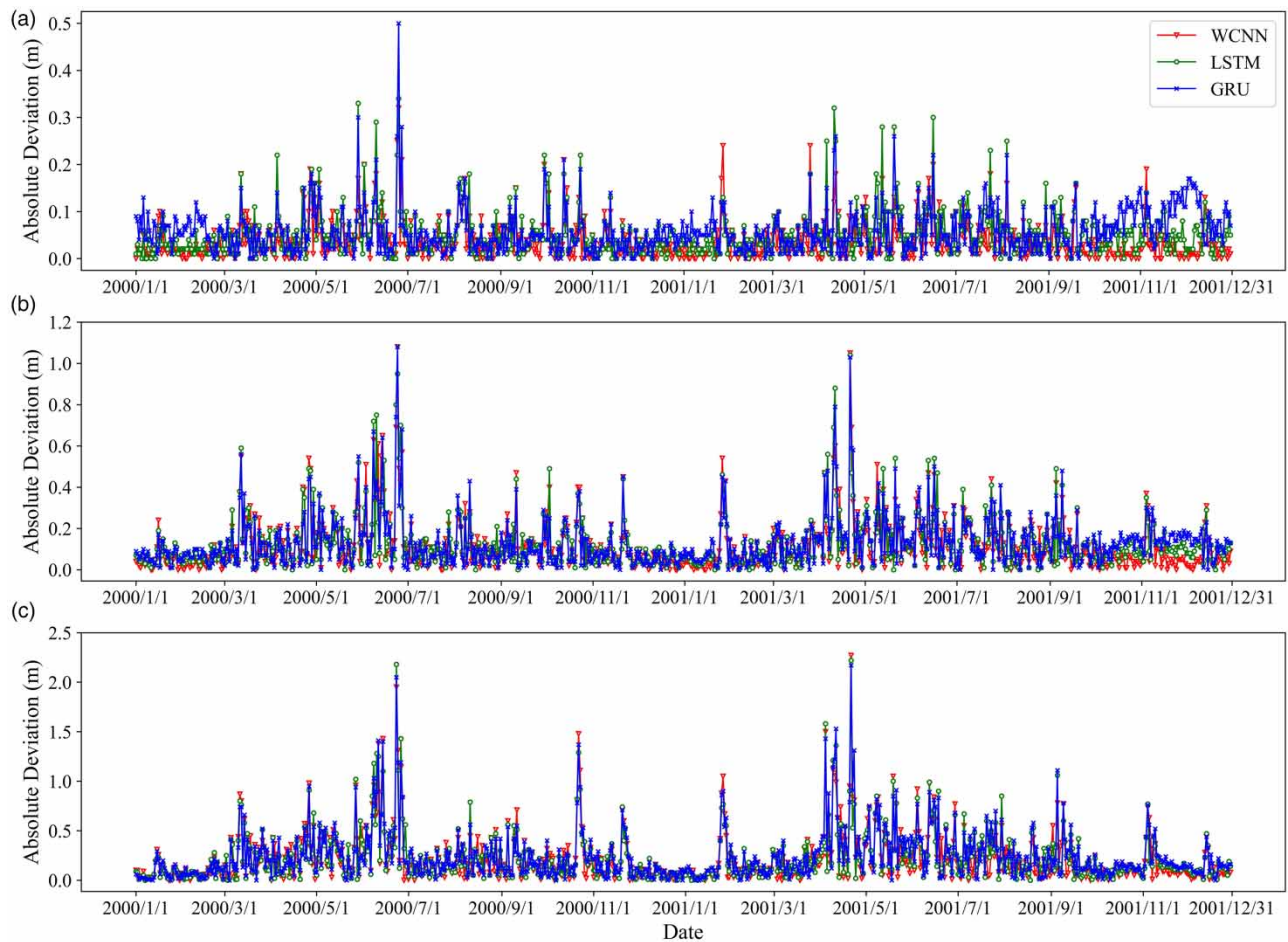


Figure 7 | Absolute deviations between the measured and (a) 1-day ahead, (b) 2-day ahead, and (c) 3-day ahead predicted water levels of three models on the test set 2000–2001.

accuracy of high water levels (higher than 19 m) is lower than that of middle water levels because there are more maximum values of water levels.

4.2. Influence of different inputs on results

To further explore the influence of the inputs on the predicted accuracy of the WCNN model, six types of input feature combinations were set for the experiments, as follows: (1) Input A: $[Z_{wz}, Q_{wz}]$; (2) Input B: $[Z_{wz}, Q_{wz}, Z_{zs}, Q_{zs}]$; (3) Input C: $[Z_{wz}, Q_{wz}, Z_{xj}, Q_{xj}]$; (4) Input D: $[Z_{wz}, Q_{wz}, Z_{ja}, Q_{ja}]$; (5) Input E: $[Z_{wz}, Q_{wz}, Z_{zs}, Q_{zs}, Z_{xj}, Q_{xj}]$; (6) Input F (recommended model input): $[Z_{wz}, Q_{wz}, Z_{zs}, Q_{zs}, Z_{xj}, Q_{xj}, Z_{ja}, Q_{ja}]$. The parameters of the models were the same in the experiments. Evaluation metrics of the 1/2/3 days ahead predicted water levels on the test set (2000–2001) of the WCNN model under different inputs are plotted in Figure 9. On the basis of the input data of the Waizhou station (input A), the input data of an upstream station is added separately (e.g., inputs B, C, D), and the prediction accuracy of the models is all significantly improved. Especially, when adding the data of Zhangshu, Xiajiang and Ji'an gauging stations to the Waizhou station separately, which are 92, 170, and 230 km away from the Waizhou station, to predict 1/2/3 days of water levels of the Waizhou station, the accuracy of the results are improved the most. The result shows that when the forecast period increases, adding the data of gauging stations farther away from the Waizhou station to the inputs, the prediction accuracy is better. The result also shows the influence of the water propagation characteristics from upstream to downstream stations on the WLP. On the basis of input data of two gauging stations (e.g., inputs B, C, D), adding the data of one or two gauging stations to the inputs respectively, and the prediction accuracy of the models are further improved (input B compares with inputs E, F; input C compares

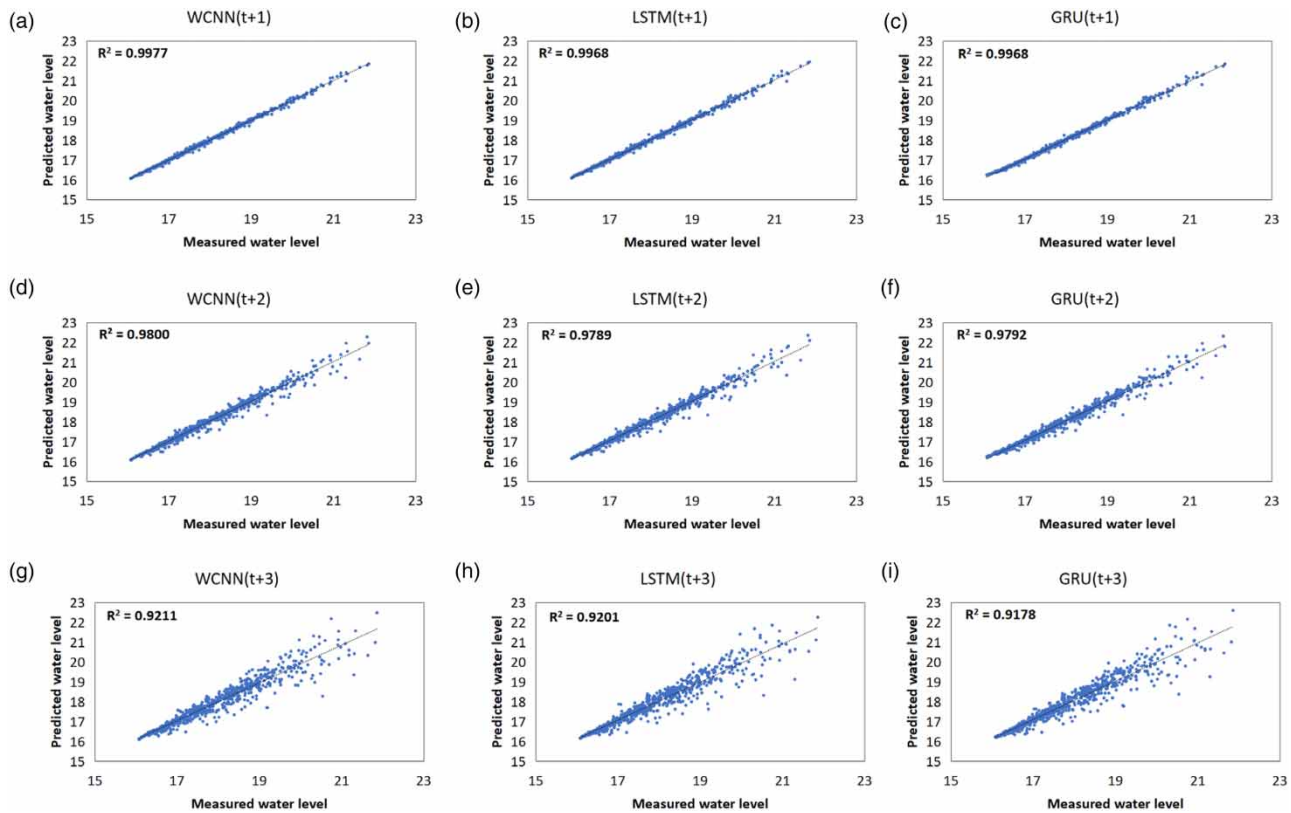


Figure 8 | Scatter plots of water level prediction results for (a) WCNN($t+1$); (b) LSTM($t+1$); (c) GRU($t+1$); (d) WCNN($t+2$); (e) LSTM($t+2$); (f) GRU($t+2$); (g) WCNN($t+3$); (h) LSTM($t+3$); and (i) GRU($t+3$) on the test set 2000–2001.

with inputs E, F; input D compares with input F). But with the increase of input features, the increase rate of the prediction accuracy gradually slows down.

4.3. Influence of hyperparameters

The effect of hyperparameters in the WCNN model was discussed, including input length, layer number, kernel size and filter number. In each experiment, the other parameters of the models were kept the same. Figure 10 shows the influence of the above four hyperparameters on the 1-day ahead WLP. It can be seen that there are multiple extreme values in the variation curve of the validation loss of the input length and filter number, which means that the model is easily trapped in local optima. Thus, the optimization range of the two parameters should be determined reasonably. The input length and filter number of the WCNN model were set as 10 and 40, respectively. The validation loss curves for the layer number and the kernel size are concave, which indicates that there is an optimal choice to optimize the results of the model. Therefore, the parameter corresponding to the lowest loss is directly selected. Notably, when the input length, layer number, kernel size or filter number are too small, the performance of the model is limited due to insufficient learning capacity. As hyperparameters increase, the training parameters and time of the model will increase. Therefore, under the premise of similar accuracy, it is of higher priority to choose smaller parameters.

5. DISCUSSION

The proposed WCNN model was compared with LSTM and GRU models in the WLP of 1/2/3 days ahead of the Waizhou station in GR, China. The result shows that the best prediction with the fewest training parameters was achieved in the WCNN model, which was consistent with the findings of variant models of WaveNet proposed by researchers (Benhaddi & Ouarzazi 2021; Zhang *et al.* 2021a) in air quality prediction and other sequence modelling tasks, respectively. The WCNN model can be generalized to other rivers with different climate conditions. Because the inputs of the model

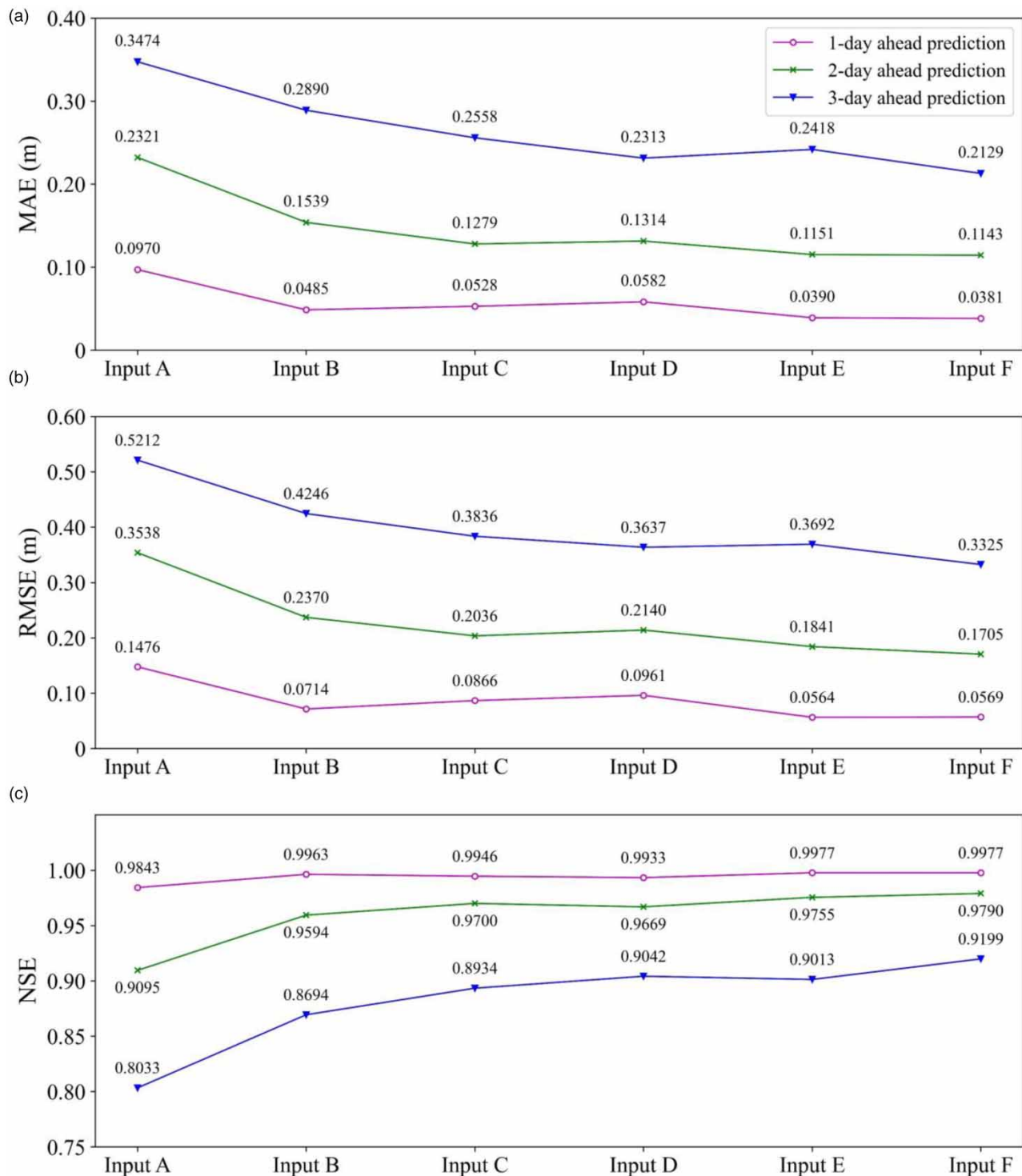


Figure 9 | Evaluation metrics of (a) MAE, (b) RMSE, and (c) NSE of the 1/2/3 days ahead predicted water levels on the test set of the WCNN model under different inputs.

(i.e., water level and discharge data of the gauging stations) reflect the characteristics of the weather conditions, the data of the predicted station and its upstream stations are easy to obtain.

On the other hand, according to the propagating physical characteristics of rivers, the water level and discharge sequences of gauging stations were selected to be the model inputs. The reasonable input features were selected by calculating the MIC

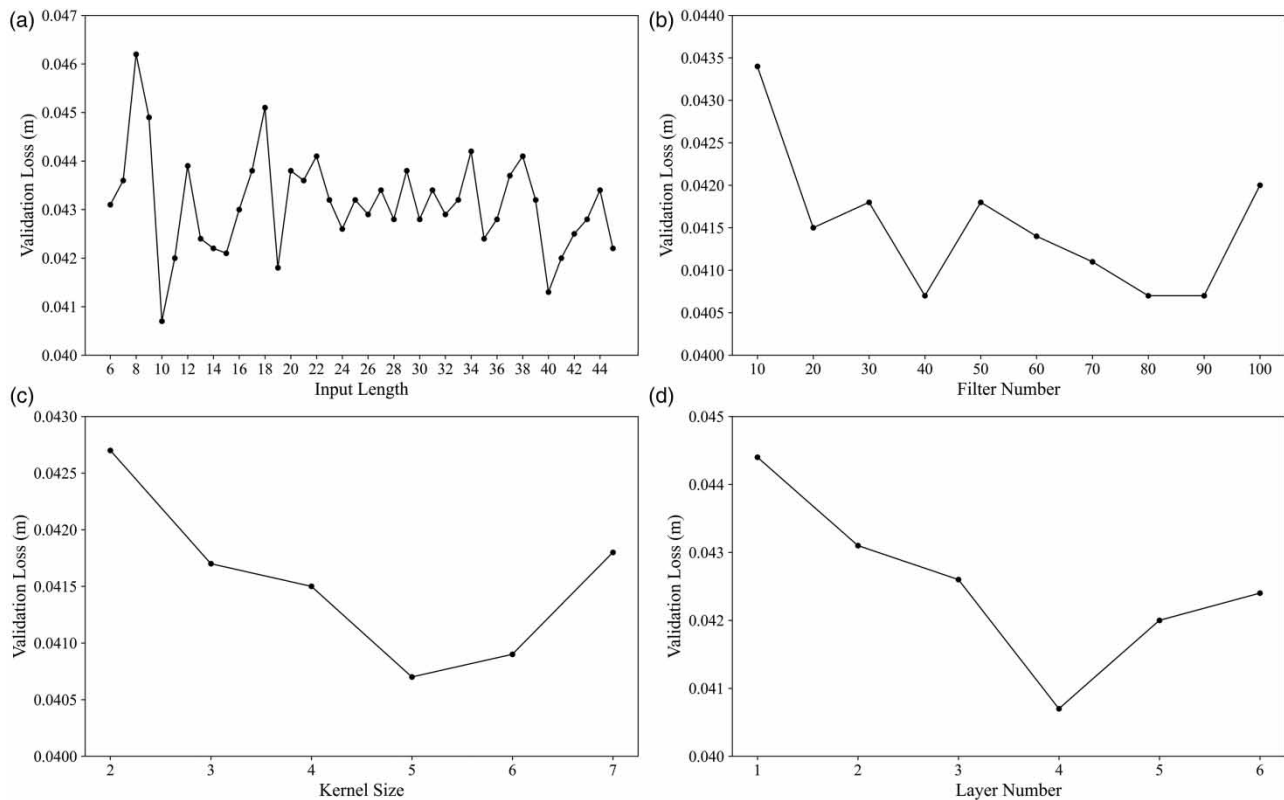


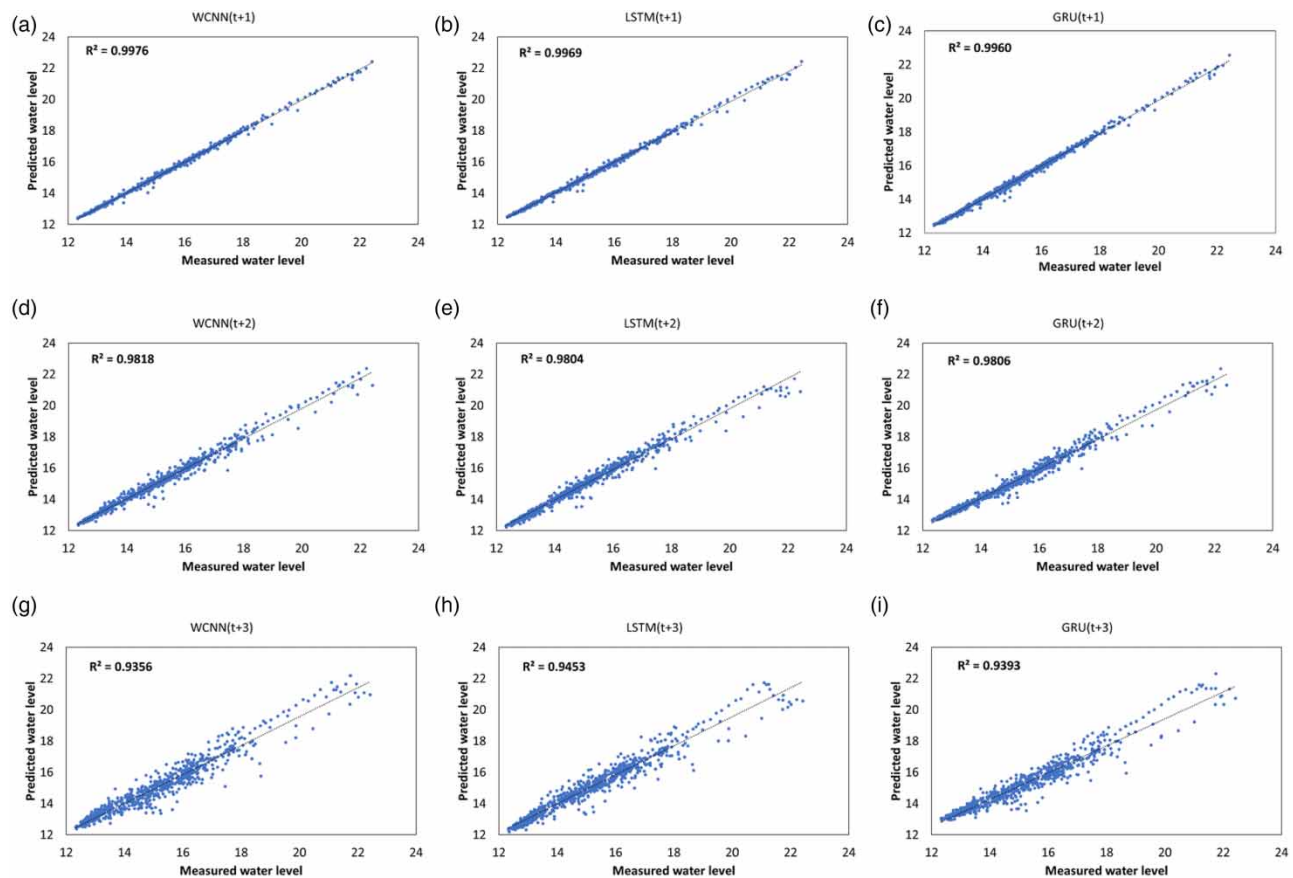
Figure 10 | Influence of (a) input length, (b) filter number, (c) kernel size, and (d) layer number on the 1-day ahead prediction performance of the WCNN model.

of variables, referring to the studies of [Lu et al. \(2021\)](#) and [Zhang et al. \(2021b\)](#). The influence of input features on the prediction results of models was quantified. The results showed that adding variables of high correlation with predicted variables to model inputs could improve prediction performance. But with the increase of input features, the increase rate of prediction accuracy gradually slowed down, and the training time of models also increased. Therefore, when selecting input features, it was necessary to comprehensively consider the relationship between the time cost caused by increasing input features and the degree of performance improvement. When setting the model parameters, the grid search method was used to optimize the parameters and determine the optimal values. The influence of the main hyperparameters of the WCNN model, such as input length, layer number, kernel size, and filter number, on the prediction results was analyzed and discussed in section 4.3. At present, there are various methods for hyperparameter optimization, such as Random Search, Bayesian Optimization, and Genetic Algorithm. These methods will be selected and adopted for our future research.

In order to compare the performance of the proposed WCNN model with the LSTM and GRU models in training time, parallelism and simulation accuracy, the hydrological datasets of 1956–2001 that were not affected by riverbed topographic changes were selected for training and testing models, which could reduce the interference of external factors. For testing the prediction performance of three models under the condition of riverbed topographic changes in the last 20 years of this century, the average daily water level and discharge of Waizhou, Zhangshu, Xiajiang and Ji'an stations from 2002 to 2016 were selected as the training set, and the 2017–2018 data were set as the test dataset. The parameter configuration used by the three models is according to [Table 2](#), and the predicted results are shown in [Table 4](#) and [Figure 11](#). It can be seen that the three models also have good accuracy for 1-day ahead WLP. The RMSE of three models are 0.098, 0.123, and 0.130, respectively. The R^2 are 0.998, 0.997, and 0.996, respectively. However, compared with the WLP results of 2001–2002, the accuracy was reduced due to the influence of riverbed topographic changes. In further research, the impact of riverbed topographic changes, tributary inlet, rainfall and water conservancy projects should be considered to improve the prediction accuracy of models ([Zhang et al. 2018](#); [Deng et al. 2021](#)). Furthermore, the multistep ahead WLP could not be significantly improved by the WCNN model, while [Borovykh et al. \(2019\)](#), [Benhaddi & Ouarzazi \(2021\)](#) only studied the variant models of WaveNet

Table 4 | Evaluation metrics of the 1/2/3 days ahead predicted water level on the test set 2017–2018 of three models

Range of water level	Metrics	Model & Forecast period								
		WCNN			LSTM			GRU		
		t + 1	t + 2	t + 3	t + 1	t + 2	t + 3	t + 1	t + 2	t + 3
All water levels	MAE	0.063	0.180	0.370	0.088	0.181	0.316	0.097	0.212	0.408
	RMSE	0.098	0.267	0.503	0.123	0.281	0.470	0.130	0.284	0.517
	NSE	0.997	0.981	0.934	0.996	0.979	0.943	0.996	0.979	0.930
	R ²	0.998	0.982	0.936	0.997	0.980	0.945	0.996	0.981	0.939
Water level < 14 m	MAE	0.047	0.127	0.245	0.099	0.107	0.183	0.079	0.208	0.452
	RMSE	0.069	0.164	0.298	0.113	0.147	0.248	0.097	0.235	0.492
	NSE	0.970	0.837	0.454	0.921	0.869	0.623	0.941	0.665	0.486
	R ²	0.975	0.894	0.670	0.972	0.903	0.744	0.957	0.897	0.696
14 ≤ Water level ≤ 16 m	MAE	0.060	0.172	0.390	0.070	0.179	0.298	0.107	0.173	0.312
	RMSE	0.097	0.258	0.516	0.111	0.267	0.411	0.146	0.252	0.418
	NSE	0.971	0.792	0.171	0.962	0.778	0.473	0.933	0.802	0.454
	R ²	0.971	0.809	0.500	0.962	0.815	0.601	0.934	0.824	0.560
Water level > 16 m	MAE	0.082	0.245	0.479	0.096	0.261	0.479	0.103	0.261	0.474
	RMSE	0.122	0.353	0.641	0.145	0.386	0.670	0.141	0.357	0.633
	NSE	0.994	0.946	0.823	0.991	0.936	0.806	0.991	0.945	0.827
	R ²	0.994	0.949	0.853	0.992	0.942	0.836	0.993	0.948	0.854

**Figure 11** | Scatter plots of water level prediction results for (a) WCNN(t + 1); (b) LSTM(t + 1); (c) GRU(t + 1); (d) WCNN(t + 2); (e) LSTM(t + 2); (f) GRU(t + 2); (g) WCNN(t + 3); (h) LSTM(t + 3); and (i) GRU(t + 3) on the test set 2017–2018.

to predict 1-day variables. Therefore, follow-up work should be carried out to extend the effective forecast periods of the model. Considering the poor effect of the sequence-to-value structure, a sequence-to-sequence structure using an encoder-decoder network (Rueda *et al.* 2021) should be adopted.

6. CONCLUSIONS

To improve the effectiveness of WLP models under the conditions of large datasets, a new CNN model named WCNN was proposed by adapting the WaveNet architecture in the audio generation domain. It was applied to predict the water level at the Waizhou station of the GR in China. The main conclusions are as follows:

- (1) For 1/2/3 days ahead WLP of the Waizhou station, the WCNN, LSTM, and GRU models all had good predicted accuracy. Moreover, the WCNN showed better prediction performance than the baseline models of LSTM and GRU. Compared with the other two models, the model structure of WCNN was lighter, the training parameters were minimal, and the model had good parallelism, which could significantly reduce the training time of the model. It is a more efficient method for WLP.
- (2) It is recommended to select the water level and discharge of upstream stations that are highly correlated with the prediction station as input features to improve the prediction accuracy. However, when the input features increase to a certain amount, the increase rate of the prediction accuracy gradually slows down, and the training time of models also increases. The time and data costs associated with the addition of input features should be considered in combination with the incremental performance improvement.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Programme of China (2021YFD1700802).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Ahmed, A. N., Yafouz, A., Birima, A. H., Kisi, O., Huang, Y. F., Sherif, M., Sefelnasr, A. & El-Shafie, A. 2022 [Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia](#). *Eng. Appl. Comput. Fluid Mech.* **16** (1), 422–440. <https://doi.org/10.1080/19942060.2021.2019128>.
- Benhaddi, M. & Ouarzazi, J. 2021 [Multivariate time series forecasting with dilated residual convolutional neural networks for urban air quality prediction](#). *Arabian J. Sci. Eng.* **46** (4), 3423–3442. <https://doi.org/10.1007/s13369-020-05109-x>.
- Borovykh, A., Bohte, S. & Oosterlee, C. W. 2019 [Dilated convolutional neural networks for time series forecasting](#). *J. Comput. Finance* **22** (4), 73–101. <https://doi.org/10.21314/JCF.2018.358>.
- Cai, H. J., Shi, H. Y., Liu, S. N. & Babovic, V. 2021 [Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: the case of central eastern continental United States](#). *J. Hydrol. Reg. Stud.* **37**, 100930. <https://doi.org/10.1016/j.ejrh.2021.100930>.
- Collado-Villaverde, A., Munoz, P. & Cid, C. 2021 [Deep neural networks with convolutional and LSTM layers for SYM-H and ASY-H forecasting](#). *Space Weather* **19** (6), e2021SW002748. <https://doi.org/10.1029/2021SW002748>.
- Deng, B., Lai, S. H., Jiang, C. B., Kumar, P., El-Shafie, A. & Chin, R. J. 2021 [Advanced water level prediction for a large-scale river-lake system using hybrid soft computing approach: a case study in Dongting Lake, China](#). *Earth Sci. Inform.* **14** (4), 1987–2001. <https://doi.org/10.1007/s12145-021-00665-8>.
- Fan, J., Zhang, K., Huang, Y. P., Zhu, Y. F. & Chen, B. P. 2021 [Parallel spatio-temporal attention-based TCN for multivariate time series prediction](#). *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-021-05958-z>.
- Géron, A. 2019 *Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd edn. O'Reilly Media, Inc., Canada.

- Gers, F. A., Schmidhuber, J. & Cummins, F. 1999 Learning to forget: continual prediction with LSTM. *Neural Comput.* **12** (10), 2451–2471. <https://doi.org/10.1162/089976600300015015>.
- Ho, H. V., Nguyen, D. H., Le, X. H. & Lee, G. 2022 Multi-step-ahead water level forecasting for operating sluice gates in Hai Duong, Vietnam. *Environ. Monit. Assess.* **194**, 441–456. <https://doi.org/10.1007/s10661-022-10115-7>.
- Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. H. 2019 Squeeze-and-Excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42** (8), 2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>.
- Huang, S., Xia, J., Zeng, S. D., Wang, Y. L. & She, D. X. 2021 Effect of Three Gorges Dam on Poyang Lake water level at daily scale based on machine learning. *J. Geogr. Sci.* **31** (11), 1598–1614. <https://doi.org/10.1007/s11442-021-1913-1>.
- Kima, D., Leeb, J., Kimc, J., Leec, M., Wanga, W. & Kima, H. S. 2022 Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea. *J. Hydrol.* **606**, 127415. <https://doi.org/10.1016/j.jhydrol.2021.127415>.
- Lai, X. J., Jiang, J. H., Liang, Q. H. & Huang, Q. 2013 Large-scale hydrodynamic modelling of the middle Yangtze River Basin with complex river-lake interactions. *J. Hydrol.* **492**, 228–243. <https://doi.org/10.1016/j.jhydrol.2013.03.049>.
- Latif, S. D. 2023 Evaluating deep learning and machine learning algorithms for forecasting daily pan evaporation during COVID-19 pandemic. *Environ. Dev. Sustain.* <https://doi.org/10.1007/s10668-023-03469-6>.
- Latif, S. D. & Ahmed, A. N. 2020 Application of deep learning method for daily streamflow time-series prediction: a case study of the Kowmung River at Cedar Ford, Australia. *Int. J. Sustainable Dev. Plann.* **16** (3), 497–501. <https://doi.org/10.18280/ijssdp.160310>.
- Latif, S. D. & Ahmed, A. N. 2023 Streamflow prediction utilizing deep learning and machine learning algorithms for sustainable water supply management. *Water Resour. Manage.* **37**, 3227–3241. <https://doi.org/10.1007/s11269-023-03499-9>.
- Lu, P. Y., Lin, K. R., Xu, C. Y., Lan, T., Liu, Z. Y. & He, Y. H. 2021 An integrated framework of input determination for ensemble forecasts of monthly estuarine saltwater intrusion. *J. Hydrol.* **598**, 126225. <https://doi.org/10.1016/j.jhydrol.2021.126225>.
- Luo, H. F., Dou, X., Sun, R. & Wu, S. J. 2021 A multi-step prediction method for wind power based on improved TCN to correct cumulative error. *Front. Energy Res.* **9**, 723319. <https://doi.org/10.3389/fenrg.2021.723319>.
- Nie, Q. Q., Wan, D. S., Zhu, Y. L., Li, Z. J. & Yao, C. 2022 Hydrological model based on temporal convolutional neural network. *J. Comput. Appl.* <https://doi.org/10.11772/j.issn.1001-9081.2021061366>.
- Noor, F., Haq, S., Rakib, M., Ahmed, T., Jamal, Z., Siam, Z. S., Hasan, R. T., Adnan, M. S. G., Dewan, A. & Rahman, R. M. 2022 Water level forecasting using spatiotemporal attention-based long short-term memory network. *Water* **14** (4), 612. <https://doi.org/10.3390/w14040612>.
- Ren, T., Liu, X. F., Niu, J. W., Lei, X. H. & Zhang, Z. 2020 Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *J. Hydrol.* **585**, 124783. <https://doi.org/10.1016/j.jhydrol.2020.124783>.
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. & Sabeti, P. C. 2011 Detecting novel associations in large data sets. *Science* **334** (6062), 1518–1524. <https://doi.org/10.1126/science.1205438>.
- Rethage, D., Pons, J. & Serra, X. 2018 A Wavenet for Speech Denoising. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada. pp. 5069–5073. <https://doi.org/10.1109/ICASSP.2018.8462417>.
- Rizvi, S. M. H., Syed, T. & Qureshi, J. 2021 Real-time forecasting of petrol retail using dilated causal CNNs. *J. Ambient Intell. Hum. Comput.* **13** (2), 989–1000. <https://doi.org/10.1007/s12652-021-02941-3>.
- Rueda, F. D., Suarez, J. D. & Torres, A. D. 2021 Short-term load forecasting using encoder-decoder WaveNet: application to the French grid. *Energy* **14** (9), 2524. <https://doi.org/10.3390/en14092524>.
- Van den Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., van den Driessche, G., Lockhart, E., Cobo, L. C., Stimberg, F., Casagrande, N., Grewe, D., Noury, S., Dieleman, S., Elsen, E., Kalchbrenner, N., Zen, H., Graves, A., King, H., Walters, T., Belov, D. & Hassabis, D. 2018 Parallel WaveNet: Fast high-fidelity speech synthesis. In *Proceedings of the 35th International Conference on Machine Learning, PMLR*, Stockholm, Sweden, pp. 3918–3926.
- Wang, Y. Y., Chen, J., Chen, X. Q., Zeng, X. J., Kong, Y., Sun, S. F., Guo, Y. S. & Liu, Y. 2021 Short-term load forecasting for industrial customers based on TCN-LightGBM. *IEEE Trans. Power Syst.* **36** (3), 1984–1997. <https://doi.org/10.1109/TPWRS.2020.3028133>.
- Wen, T. F., Xiong, L. H., Jiang, C., Hu, J. M. & Liu, Z. J. 2019 Effects of climate variability and human activities on suspended sediment load in the Ganjiang River basin, China. *J. Hydrol. Eng.* **24** (11), 05019029. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001859](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001859).
- Yan, J. N., Mu, L., Wang, L. Z., Ranjan, R. & Zomaya, A. Y. 2020 Temporal convolutional networks for the advance prediction of ENSO. *Sci. Rep.* **10** (1), 8055. <https://doi.org/10.1038/s41598-020-65070-5>.
- Yao, J., Zhang, D., Li, Y. L., Zhang, Q. & Gao, J. F. 2019 Quantifying the hydrodynamic impacts of cumulative sand mining on a large river-connected floodplain lake: Poyang Lake. *J. Hydrol.* **579**, 124156. <https://doi.org/10.1016/j.jhydrol.2019.124156>.
- Yin, W. J., Fan, Z. W., Tangdamrongsub, N., Hu, L. T. & Zhang, M. L. 2021 Comparison of physical and data-driven models to forecast groundwater level changes with the inclusion of grace – a case study over the state of Victoria, Australia. *J. Hydrol.* **602**, 126735. <https://doi.org/10.1016/j.jhydrol.2021.126735>.
- Zhang, Q., Liu, J. Y., Singh, V. P., Gu, X. H. & Chen, X. H. 2016 Evaluation of impacts of climate change and human activities on streamflow in the Poyang Lake basin, China. *Hydrol. Process.* **30** (14), 2562–2576. <https://doi.org/10.1002/hyp.10814>.

- Zhang, D., Lindholm, G. & Ratnaweera, H. 2018 [Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring](#). *J. Hydrol.* **556**, 409–418. <https://doi.org/10.1016/j.jhydrol.2017.11.018>.
- Zhang, R. J., Sun, F., Song, Z. W., Wang, X. L., Du, Y. C. & Dong, S. L. 2021a [Short-term traffic flow forecasting model based on GA-TCN](#). *J. Adv. Transp.* **2021**, 1338607. <https://doi.org/10.1155/2021/1338607>.
- Zhang, Z. D., Qin, H., Yao, H., Liu, Y. Q., Jiang, Z. Q., Feng, Z. K., Ouyang, S., Pei, S. Q. & Zhou, J. Z. 2021b [Downstream water level prediction of reservoir based on convolutional neural network and long short-term memory network](#). *J. Water Resour. Plan. Manage.* **147** (9), 04021060. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001432](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001432).

First received 18 July 2023; accepted in revised form 26 August 2023. Available online 6 September 2023