Accepted version Click here to view linked References

https://doi.org/10.1016/j.jenvman.2023.118232

Wordcount: 7,249 words

# Exploring a multi-output Temporal Convolutional Network driven Encoder-Decoder framework for ammonia nitrogen forecasting

Sheng Sheng[1], Kangling Lin[1], Yanlai Zhou[1], Hua Chen[1], Yuxuan Luo[1], Shenglian Guo[1], Chong-Yu Xu[2]

[1.] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China.
[2.] Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, N-0316 Oslo, Norway.

**Correspondence to*: Yanlai Zhou (yanlai.zhou@whu.edu.cn); Hua Chen (chua@whu.edu.cn).

**Abstract**

Artificial neural networks exhibit significant advantages in terms of learning capability and generalizability, and have been increasingly applied in water quality prediction. Through learning a compressed representation of the input data, the Encoder-Decoder (ED) structure not only could remove noise and redundancies, but also could efficiently capture the complex nonlinear relationships of meteorological and water quality factors. The novelty of this study lies in proposing a multi-output Temporal Convolutional Network based ED model (TCN-ED) to make ammonia nitrogen forecasts for the first time. The contribution of our study is indebted to systematically assessing the significance of combining the ED structure with advanced neural networks for making accurate and reliable water quality forecasts. The water quality gauge station located at Haihong village of Hengsha island in Shanghai City of China constituted the case study. The model input contained one hourly water quality factor and hourly meteorological factors of 32 observed stations, where each factor was traced back to the previous 24 hours and each meteorological factor of 32 gauge stations was aggregated into one areal average factor. A total of 13,128 hourly water quality and meteorological data were divided into two datasets corresponding to model training and testing stages. The Long Short-Term Memory based ED (LSTM-ED), LSTM and TCN models were constructed for comparison purposes. The results demonstrated that the developed TCN-ED model can succeed in mimicking the complex dependence between ammonia nitrogen and water quality and meteorological factors, and provide more accurate ammonia nitrogen forecasts (1- up to 6-hour-ahead) than the LSTM-ED, LSTM and TCN models. The TCN-ED model, in general, achieved higher accuracy, stability and reliability compared with the other models. Consequently, the improvement can facilitate river water quality forecasting and early warning, as well as benefit water pollution prevention in the interest of river environmental restoration and sustainability.

**Keywords**: Water quality forecast; Artificial Neural Networks (ANNs); Sequence-to-Sequence; Encoder-Decoder structure; Deep learning

## 1. Introduction

Due to fast industrialization and urbanization, riverine water qualities in many regions worldwide have gradually deteriorated (Ming et al., 2022). Water pollution directly threatens human life safety and becomes a principal obstacle to sustainable development (Mu et al., 2023; Walsh et al., 2022). Ammonia nitrogen is a common pollutant in aquatic environments, and its concentration is usually used as an important indicator of river water quality (Menon et al., 2021; Yan et al., 2023). Accurate and reliable water quality forecasts can provide valuable support and guidance for water pollution prevention and control, and have received broad attention from plenty of researchers (Mohammed et al., 2021; Wiering et al., 2023).

Water quality forecast models, in general, can be separated into two types: physically-based models and data-driven models (Noori et al., 2020). Physically-based models describe specific water chemical processes with physically meaningful equations and parameters (Uddin et al., 2023b), and are broadly used to mimic water quality forecasts (Aloui et al., 2023, Pyo et al., 2021; Quevedo-Castro et al., 2022). However, physically-based models have some limitations, including their idealistic assumptions (Wan et al., 2021), the need for prior knowledge of water physics and chemistry (Wan et al., 2022), and high computation costs (Ahmed et al., 2019), which reduce their efficiency in real-time or short-term water quality forecasts. In pace with data mining techniques developing and the monitoring data increasing, more and more data-driven models are gradually applied in modeling river water quality forecasts (Bertone et al., 2023). Rather than attempting to explain the physical characteristics and chemical processes of water quality factors, data-driven models usually identify the complex nonlinear correlation of meteorological factors and water quality factors (Zhang and Li, 2021; Zheng et al., 2023). As known, ANNs show significant advantages in learning capability, noise immunity, and generalizability, and have been successfully applied to water quality forecasting (Deng et al., 2021; Guo and Cui, 2022). Among ANNs, the Long Short-

Term Memory (LSTM) cell-based ANNs stand out for their ability to selectively memorize long-term features (Chen et al., 2021; Wang et al., 2023; Yang et al., 2021).

Recently, Temporal Convolutional Network (TCN) block-based ANNs have received increasing interest from researchers due to their superiority over the LSTM in modeling temporal predictions (Bai et al., 2018). Gopali et al. (2021) compared the performance and training time of the TCN and the LSTM and found that the TCN models have good performance and require less computation time to achieve model convergence. Hewage et al. (2020) utilized the TCN to make 9-hour-ahead weather forecasts and the results show that the TCN could produce better forecast accuracy compared with the LSTM. Similar conclusions have been drawn in the application of the TCN to water quality forecasting (Zhang et al., 2019; Fu et al., 2021). A multi-output TCN model was proposed by Zhang and Li (2023) to predict water quality, and the results verified its superiority over the LSTM and other commonly used machine learning models. Furthermore, the tensor flow-based machine learning has become a popular approach to improving the prediction performance of ANNs (Kao et al., 2020; Laubscher, 2019; Park et al., 2019). The Encoder-Decoder (ED) structure is a powerful neural network framework that can efficiently improve the performance and flexibility of ANNs. The aim of the structure is to translate the input sequence into a context value in the encoder part and parse the context value back to the output sequence in the decoder part. Through learning a compressed representation of the input data, the ED structure can effectively reduce the dimensionality of the input data and extract features from them, which helps capture the complex relationships of sequences and improve the accuracy and performance of ANNs. Compared with ANNs, the ANN-based ED models have exhibited superior performance in various fields including power load prediction (Dorado Rueda et al., 2021), medical image segmentation (Mahmud et al., 2021), and language translation (Abbaszade et al., 2021), and others. Due to the superiority of the ED in learning the patterns of time series (Bian et al., 2019;

Zhang et al., 2021), the water quality prediction problem can be modeled as a sequence-to-sequence problem using the ED structure. It is interesting to combine a more advanced and computationally efficient time series model (i.e., TCN) with the ED structure to enhance model accuracy as the effectiveness and robustness of the LSTM with the ED structure (LSTM-ED) have already been verified (Han et al., 2021). From the perspective of water quality forecasts, to date, no studies have fully analyzed the effect of the ED framework on multi-output ANNs (e.g., TCN) by considering sequence-to-sequence learning processes.

The novelty of our study lies in proposing a multi-output Temporal Convolutional Network based Encoder-Decoder (TCN-ED) framework to make accurate and reliable water quality forecasts for the first time. The developed TCN-ED model is utilized to provide technical support for water quality early warning and water pollution prevention. Meanwhile, we comprehensively compare and evaluate the predictive performance of the LSTM, LSTM-ED, TCN and TCN-ED models for water quality forecasting. Firstly, the ED framework is used to construct a multi-output structure in a sequence-to-sequence learning way. Secondly, the two TCN units and the learning way are incorporated into the ED framework for establishing a deep learning-based multi-output prediction model (i.e. TCN-ED). Lastly, to validate the applicability of the developed TCN-ED model in water quality forecasting, this study adopts an ammonia nitrogen time series of a water quality station located at Hengsha island of Shanghai City in China as a case study.

## 2. Study area and materials

Hengsha island is located at the estuary of the Yangtze River and covers an area of 52 km$^2$ (Zhou, 2020). The island lies in a subtropical monsoon climate zone with a mean annual temperature of 15.4℃, and experiences extremely high temperatures of 33-36℃ and extremely low temperatures of minus 2-5℃. The total annual rainfall is about 1,100 mm, and the tide level ranges between -0.27 and 5.9 meters. The water quality monitoring station in Haihong village is located in the southwest of Hengsha island. The map of Hengsha island, along with the water quality and meteorological gauge stations in Haihong village, is presented in Fig.1.

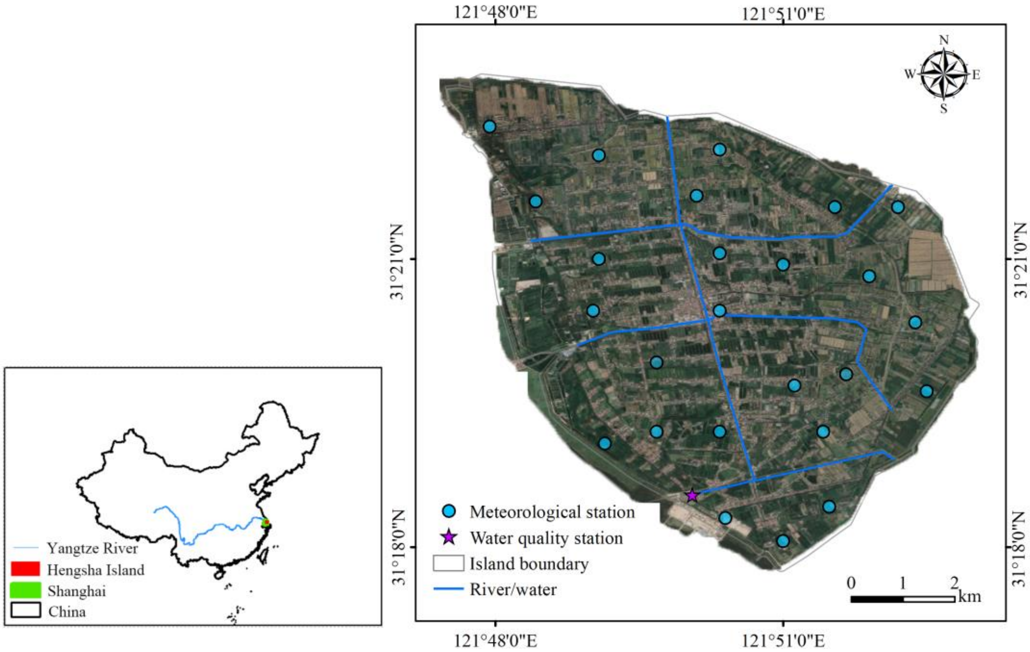



**Fig.1** Spatial distribution of meteorological stations and river water quality observed station in Hengsha island of Shanghai City in China

The study collected continuous hourly monitoring data from the water quality station in Haihong village and observation data from 25 meteorological stations (Fig.1). After data cleaning and correlation analysis, hourly water quality data, including Water Temperature (WT), Conductivity (COND), turbidity (TURB), and Ammonia Nitrogen ($NH_3$-N), as well as hourly meteorological data, including areal precipitation (P), areal wind speed (WS), and areal relative humidity (RH) from February 2019 to July 2020 were selected to constitute the case study, where the areal meteorological data were aggregated using the observed data of 25 meteorological stations. To reduce the negative influence of data scales on models' stability, normal standardization is employed to preprocess the input data. Table 1 presents the Pearson correlation coefficients between $NH_3$-N and water quality and meteorological factors used in this study.

**Table 1** Pearson correlation coefficient among $NH_3$-N and other water quality and meteorological factors.

| | Water quality factors | | | Meteorological factors | | |
|---|---|---|---|---|---|---|
| | WT | COND | TURB | P | WS | RH |
| $NH_3$-N | -0.32 | 0.78 | 0.69 | -0.27 | -0.19 | 0.15 |

## 3. Methodology

Fig. 2 illustrates the fusion of the LSTM cell (Fig. 2(a)) or the Temporal block (Fig. 2(b)) to the ED framework to construct the forecast models (LSTM-ED & TCN-ED, Fig. 2(c)) in this study. The $NH_3$-N concentration forecast (1- up to 6-hour-ahead) is considered as the function of three meteorological factors and four water quality factors (traced back to the previous 24 h). The TCN (LSTM) is fused into the ED framework to construct the TCN-ED (LSTM-ED) prediction model. For comparison purposes, the TCN and LSTM models are established in this study. The related methods are briefly described below.

### *3.1 Encoder-Decoder framework*

The Encoder-Decoder framework developed by Cho et al. (2014) takes two ANN layers as the encoder part and the decoder part, respectively. The Encoder part can translate the information of model input into a context value, while the Decoder part can decode the context value into the targeted value. The goal of the ED framework is to characterize several different information resided in the input data as a fixed-length vector (Xu et al., 2019).

The purpose of sequence-to-sequence learning process is to mimic the most likely next sequence of $\tilde{Y}_{t+1}, \dots, \tilde{Y}_{t+k}$ according to the previous observation $Y_{t-j+1}, \dots, Y_t$, which can be described below (Shi et al., 2015).

$$\tilde{Y}_{t+1}, \dots, \tilde{Y}_{t+k} = \underset{\hat{Y}_{t+1},\dots,\hat{Y}_{t+k}}{argmax}\, p\left(\hat{Y}_{t+1}, \dots, \hat{Y}_{t+k} | Y_{t-j+1}, Y_{t-j+2}, \dots, Y_t\right) \quad (4)$$

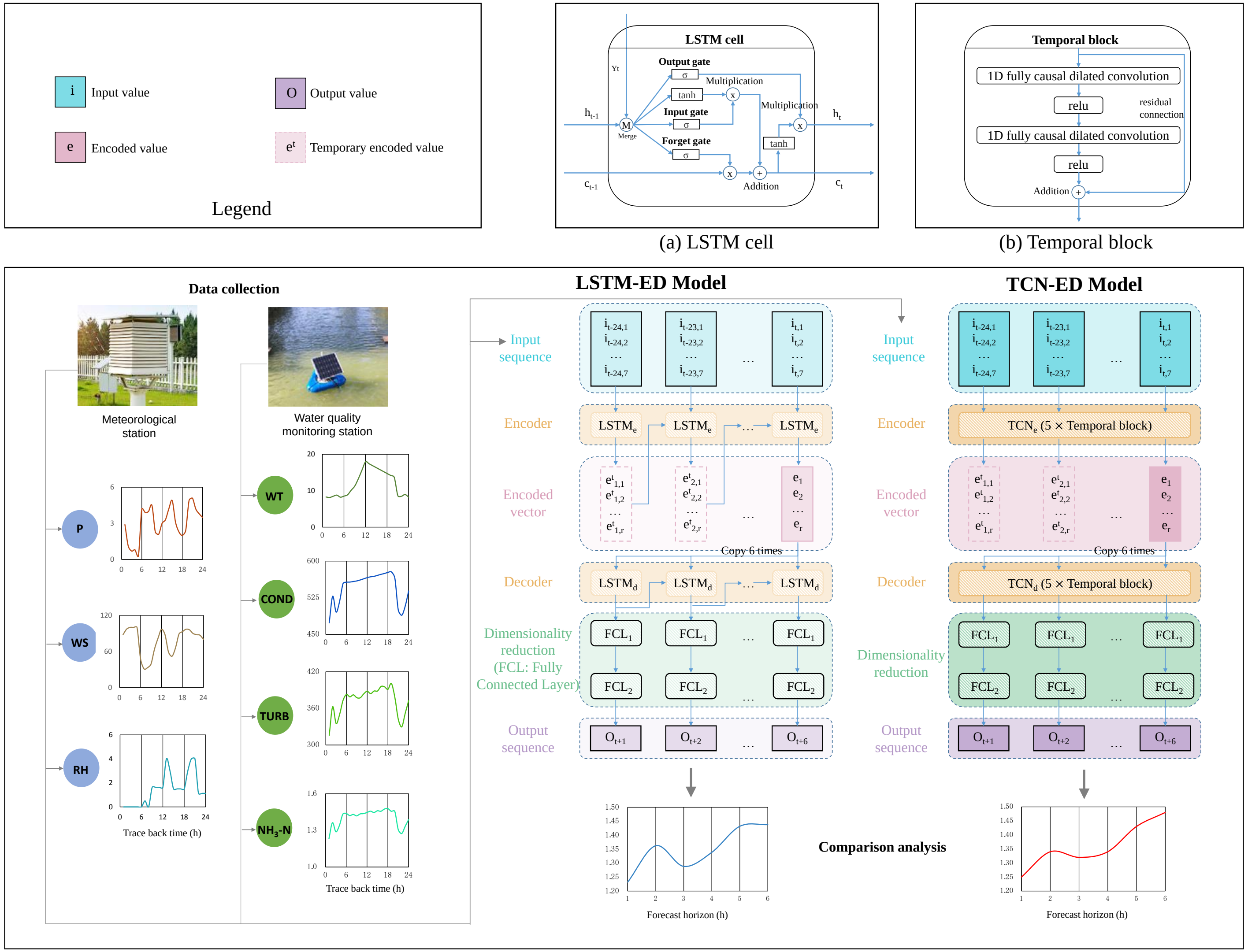
where $Y_t$ is the observed data at the time step *t*. *k* and *j* are the lengths of the predicted sequence and the observed sequence, respectively.

The Encoder part refines the various information related to the input sequence, and temporarily stored it in the context value:

$$\tilde{Y}_{t+1}, \dots, \tilde{Y}_{t+k} \approx \underset{\hat{Y}_{t+1},\dots,\hat{Y}_{t+k}}{argmax}\, p\left(\hat{Y}_{t+1}, \dots, \hat{Y}_{t+k} \middle| f_{encoder}\left(Y_{t-j+1}, Y_{t-j+2}, \dots, Y_t\right)\right) \tag{5}$$

Then the Decoder part decodes the refined context value into the expected value:

$$\tilde{Y}_{t+1}, \dots, \tilde{Y}_{t+k} \approx g_{decoder}\left(f_{encoder}\left(Y_{t-j+1}, Y_{t-j+2}, \dots, Y_t\right)\right) \tag{6}$$




(c) Forecast Models

**Fig.2** Research methodology framework. **a**. LSTM cell. **b**. Temporal block. **c**. Forecast models.

### 3.2 LSTM model and LSTM-ED model

The LSTM cell (Fig. 2(a)) consists of the input, forget and output gates. When a new input at each time step $X_t$ is received, the input gate evaluates the importance of current information and decides whether to retain it. The cell status in the previous time step $c_{t-1}$ will be discarded if the forget gate is off. The aim of the output gate is to determine which parts of the cell state $c_t$ need to be propagated to the output state $h_t$. The sigmoid activation function $\sigma$ is used in all three gates, with a result range of [0, 1] to allow the gates to be half-open (Zhao et al., 2019). $h_t$ can be specified whether to be sent to the next layer, and then $h_t$ and $c_t$ will be sent the LSTM cell when being used again.

The LSTM model is constructed according to the above LSTM unit structure. The LSTM model directly stacks two LSTM layers. The time steps in the two LSTM layers are equal to those of the input data. The output of the LSTM cell at each time step in the first LSTM layer is directly used as the input of the second LSTM layer. The second LSTM layer links its outputs at the last time step with the two fully connected layers ($FCL_1$ and $FCL_2$) to reduce computation dimensions, and then produces the $NH_3$-N sequence with horizons ($t$+1 up to $t$+6).

Through fusing the LSTM into the ED framework, we construct the LSTM-ED model (Fig. 2(c)). The LSTM cell is used several times for translating the input sequence into the output sequence continuously. The number of times to use the LSTM cell in the encoder (decoder) part depends on the input (output) length (Kao et al., 2020). In the encoding phase, each sample contains 24-time step inputs, so the cell of the LSTM encoder ($LSTM_e$) is used 24 times. The output state at the last time step $e_1, e_2, \ldots, e_r$ is selected as the context value, which contains the $NH_3$-N information refined by the LSTM gating structure using the entire sample. Since the output length is 6, the context value in the encoder part is copied 6 times correspondingly and sent to the decoder part. In the decoding phase, the output state of each time step of the LSTM decoder ($LSTM_d$) contains high-dimensional $NH_3$-N forecasts. Therefore, $FCL_1$ and $FCL_2$ are employed to reduce the dimensionality at each time step. In the final output layer, the LSTM-ED model outputs the $NH_3$-N sequence with forecast horizons of 6 hours.

### *3.3 TCN model and TCN-ED model*

The TCN is a generic architecture used for convolutional sequence prediction (Bai et al., 2018; Deng et al., 2019; He and Zhao, 2019; Liu et al., 2019). The fundamental building block of the TCN is the temporal block (Fig. 2(c)), which contains a residual connection and two 1-dimensional (1D) fully causal dilated convolution layers activated by the rectified linear unit (relu). The purpose of the 1D fully convolution is to ensure that the lengths of the input and output sequences in the convolution processes remain the same. The purpose of the causal convolution is to ensure that future (i.e. forecast) information is not utilized in the convolution processes. The dilated convolution expands the receptive field of the convolution processes with a small increase in computational overhead; and residual connections alleviate the degradation problem of the convolution processes. The TCN integrates these convolution techniques to achieve efficient and high-precision time series prediction.

The TCN model is also constructed by stacking two TCN layers. The numbers of time steps of the two TCN layers keep the same as those of the sample, and both layers consist of five temporal blocks. The second TCN layer also links its output at the last time step with $FCL_1$ and $FCL_2$ to reduce computation dimension, and then produces the expected sequence.

Through fusing the TCN units into the ED framework, we construct the TCN-ED model (Fig. 2(c)). The number of temporal blocks in the Encoder part of the TCN is 5. The context value that is copied 6 times to keep consistent with the output sequence length will be sent to the TCN decoder ($TCN_d$). The number of temporal blocks in the $TCN_d$ is also 5. Through the ED framework, the $TCN_d$ output also executes dimension reduction using $FCL_1$ and $FCL_2$, and finally produces the prediction data.

Table 2 summarizes the hyperparameters and the input/output sequence configured for the four models, where the parameters were determined by using a trial-and-error procedure. To adequately assess the predictability of the LSTM-ED and TCN-ED models, the hyperparameters (units/filters) of the Encoder and Decoder parts are set to 256 and 128, respectively. Correspondingly, the hyperparameters of the first and second layers of the LSTM and TCN models are also 256 and 128, respectively. Each TCN unit (layer/encoder/decoder) contains 5 temporal blocks. The output dimensions of $FCL_1$ of the four models are all 64, but due to the different tensor transmission methods, the $FCL_2$ output dimensions of LSTM and TCN are both 6, and the $FCL_2$ output dimensions of LSTM-ED and TCN-ED are both 1. The batch size, learning rate and epoch of the four models are 32, 0.01 and 100, respectively. In each model, the amount of input variables is 168 (=(4 (water quality factors)+3 (meteorological factors)) × 24 (time-lags)) as well as the amount of output variables is 6 (=1 (water quality factor) × 6 (forecast horizons)).

**Table 2** Hyperparameters and the input/output sequence configured for the four models.

| Item | Model | | Item | Model | |
|---|---|---|---|---|---|
| | LSTM | TCN | | LSTM-ED | TCN-ED |
| **First layer unit/filter** | 256 | 256 | **Encoder unit/filter** | 256 | 256 |
| **First layer temporal block** | / | 5 | **Encoder temporal block** | / | 5 |
| **Second layer unit/filter** | 128 | 128 | **Decoder unit/filter** | 128 | 128 |
| **Second layer temporal block** | / | 5 | **Decoder temporal block** | / | 5 |
| **$FCL_1$ output dimension** | 64 | 64 | **$FCL_1$ output dimension** | 64 | 64 |
| **$FCL_2$ output dimension** | 6 | 6 | **$FCL_2$ output dimension** | 1 | 1 |
| Batch size | 32 | 32 | Batch size | 32 | 32 |
| Learning rate | 0.01 | 0.01 | Learning rate | 0.01 | 0.01 |
| Epoch | 100 | 100 | Epoch | 100 | 100 |
| Input sequence | 7 (factors)*24 (time-lags) | 7 (factors)*24 (time-lags) | Input sequence | 7 (factors)*24 (time-lags) | 7 (factors)*24 (time-lags) |
| Output sequence | 1 (factor)*6 (forecast horizons) | 1 (factor)*6 (forecast horizons) | Output sequence | 1 (factor)*6 (forecast horizons) | 1 (factor)*6 (forecast horizons) |

All models were calculated with 20 rounds to reduce the influence of initial weight parameters on the accuracy of prediction models. Consider that Shanghai City has a subtropical monsoon climate and has four seasons consisting of spring (March to May), summer (June to August), autumn (September to November), and winter (December to February). In this study, the training set is composed of the data from February to October 2019, covering late winter, spring, summer, and autumn, and the testing set is composed of the data from November 2019 to July 2020, covering late autumn, winter, spring, and summer. Therefore, the collected continuous hourly data were divided into two parts for model training (February 1st 2019 – October 31st 2019) and testing (November 1st 2019 – July 31st 2020), both of which cover the four seasons in the study area to mitigate the impact of the large time interval on prediction model accuracy. The Adam optimizer was used to optimize the model parameters, and the Mean Square Error (MSE) (Erdélyi et al., 2023) indicator was used as the objective function to evaluate the efficiency of model training. Since the Encoder-Decoder structure refines the tensor transmission, the TCN-ED (LSTM-ED) model has higher computational efficiency than the TCN (LSTM) model. Specifically, the average calculation times per round for LSTM, LSTM-ED, TCN, and TCN-ED are about 110 s, 80 s, 250 s, and 230 s, respectively (computer specifications: i7-12700H, GeForce RTX 3060, 16GB Memory).

### *3.4 Evaluation indicators*

In this study, three evaluation indicators including the Nash-Sutcliffe Efficiency (NSE), the Pearson Correlation Coefficient (CC), and the Root Mean Square Error (RMSE), were adopted to assess the forecast accuracy of the constructed models. The NSE indicator is a broadly used criterion for assessing the prediction models' accuracy, and its value lies in the interval $(-\infty, 1]$ (Nash and Sutcliffe, 1970). An NSE value of 1 indicates perfect prediction, while a negative NSE value suggests the average observed value is a better estimate than the model prediction (Jiang et al., 2018). The RMSE indicator measures the difference between forecasted and observed values, and its value lies in the interval [0, +∞) (Jamro et al., 2023). The CC indicator reveals the goodness of fit between the forecasted and observed time series, and its value ranges from 0 to 1, reflecting a low to high correlation, respectively (Pawan and Dhiman, 2023). The calculation equations of these indicators are described as follows.

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(\hat{Q}_i - Q_i\right)^2}{\sum_{i=1}^{n}\left(\bar{Q} - Q_i\right)^2} \tag{1}$$

$$CC = \frac{\mathrm{COV}(\hat{Q}, Q)}{\sqrt{D(\hat{Q})}\sqrt{D(\hat{Q})}} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Q}_i - Q_i)^2} \tag{3}$$

where $\hat{Q}$ is the forecasting $NH_3$-N concentration. $Q$ is the observed $NH_3$-N concentration and $\bar{Q}$ is the average value of the observed $NH_3$-N concentration. *n* is the number of observed data.

## 4. Results and discussion

The study aimed at evaluating the application of the TCN-ED (LSTM-ED) model in enhancing the accuracy, stability and reliability of 1- up to 6-hour-ahead $NH_3$-N forecasts. The results, findings, and discussion are presented as follows: assessment on model performance over 20 rounds (Section 4.1); and assessment on model reliability (Section 4.2).

### *4.1 Assessment on model performance*

We conducted 20 rounds of experiments for each model to ensure statistical significance. To evaluate the LSTM, LSTM-ED, TCN, and TCN-ED models' performance, the mean and best values of NSE, RMSE, and CC of the models from horizon *t*+1 up to *t*+6 were summarized in Table 3.

**Table 3** The mean and best values of evaluation indicators over six forecast horizons of the models corresponding to 20 rounds.

| Stage | Indicator | LSTM | | LSTM-ED | | TCN | | TCN-ED | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Best | Mean | Best | Mean | Best | Mean | Best |
| Training | NSE | 0.98 | **0.99** | 0.98 | 0.98 | 0.96 | 0.98 | 0.97 | 0.98 |
| | RSME | 0.12 | **0.09** | 0.12 | 0.11 | 0.15 | 0.12 | 0.14 | 0.12 |
| | CC | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| Testing | NSE | 0.80 | 0.84 | 0.83 | 0.89 | 0.81 | 0.90 | 0.91 | **0.93** |
| | RSME | 0.23 | 0.21 | 0.21 | 0.18 | 0.23 | 0.17 | 0.16 | **0.14** |
| | CC | 0.91 | 0.95 | 0.92 | 0.95 | 0.95 | 0.96 | 0.96 | **0.97** |

In the training stage, the NSE values of each model range from 0.96 to 0.99, with a small difference between the average and best values. Among the four models, the LSTM model displays a slight advantage in NSE, achieving the highest accuracy in both average and best values. Conversely, the TCN model exhibits the lowest training accuracy, with an average NSE value of 0.96. Moreover, the RSME values differ by 0.01-0.03 between the average and best values of each model, with the LSTM and LSTM-ED models outperforming the TCN and TCN-ED models. All models exhibit a CC value of 0.99, indicating good training accuracy.

In the testing stage, the average and best values of NSE and RSME differ for each model, with the TCN-ED and TCN models exhibiting the smallest and the largest differences, respectively. The average and best values of CC are different for the LSTM and LSTM-ED models but are similar for the TCN and TCN-ED models.

In general, the TCN-ED model displays the higher prediction accuracy, with average values superior to the best values of the other models. In terms of overall performance in the testing stage, the prediction models are ranked from higher (TCN-ED & TCN) to lower (LSTM-ED & LSTM) accuracy.

The Gaussian kernel density estimation (GKDE) curves of the three evaluation indicators of the prediction models (run for 20 rounds) for all forecast horizons are presented in Fig. 3. The sharpness of the curve suggests the concentration of the values of the evaluation indicator. A sharp GKDE curve indicates high stability for the model.

The GKDE curves of the NSE and CC indicators for the LSTM and LSTM-ED models are sharp and lean towards the right, where most NSE values concentrate between 0.95 and 1, and most CC values concentrate between 0.97 and 1. This indicates high and stable accuracy during the training stage. Although the stability of the TCN-ED and TCN models is weaker than that of the LSTM-ED and LSTM models in training phases, the GKDE curves of the TCN-ED and TCN models are sharper and tend more towards the right in testing phases, indicating the TCN-ED and TCN models provide better stability for the forecasted results. Regarding the GKDE curves of the RSME indicator, the sharpness of the GKDE curves corresponding to the LSTM-ED and LSTM models largely decreases in testing stages, resulting in shorter and fatter curves, compared with those in training stages. In contrast, the sharpness of the GKDE curves corresponding to the TCN-ED and TCN models displays a smaller difference between training and testing phases. That is to say, the TCN-ED and TCN models have high stability and low uncertainty.

Furthermore, the ED framework gives rise to higher density peaks in training phases, indicating better stability resided in the TCN-ED (LSTM-ED) model. In testing phases, the GKDE curves of the three evaluation indicators created by the TCN-ED (LSTM-ED) model are sharper than those of the TCN (LSTM) model, and the density peaks of the GKDE curves move towards better values, showing superior accuracy and stability. The ED framework improves the accuracy of the TCN unit more than the LSTM unit, particularly in testing phases. The TCN-ED model exhibits a large improvement over the LSTM-ED model in terms of GKDE curve sharpness, density peak, and occurrence position.

Fig. 4 explicitly presents the boxplots of the values of the three evaluation indicators corresponding to four prediction models from horizon $t$+1 up to $t$+6 in training and testing phases. In general, the results corresponding to 20 rounds support that the constructed TCN-ED (LSTM-ED) model is much more stable with better consistency than the TCN (LSTM) model according to the values of the interquartile ranges and min-max ranges. The boxplot results also reveal that the TCN-ED (LSTM-ED) model creates better NSE and CC values, as well as inferior RMSE values than the TCN (LSTM) model, especially significant at long horizons > 3h in training and testing phases. These results demonstrate that, from the perspective of model performance, the accuracy and stability of the TCN-ED (LSTM-ED) model is better than that of the TCN (LSTM) model. Despite that the LSTM-ED model outperforms the TCN-ED model in terms of accuracy in the training phase, the TCN-ED model displays better accuracy and stability in the testing phase, particularly for the long forecast horizons (t+3 – t+6). This suggests that the LSTM-ED model tends to produce unstable forecasts, while the TCN-ED model has strong generalization capability and stability. The widening gap between the two models with the increasing forecast horizons in the testing phase is mainly attributed to the LSTM-ED model's excessive reliance on the update of the cell state in the LSTM at the encoder's last time step. On the one hand, hydrological information would gradually be neglected by the forget gate as redundant information in multiple iterations, making it challenging for the encoder to fully capture temporal dependencies meanwhile resulting in inadequate information in the semantic vectors assigned by the decoder. On the other hand, the TCN encoder with dilated convolution has a receptive field that can encompass 24-time steps of the input sequence at the last time step, allowing for better characterization of hydrological patterns in the long forecast horizons. Therefore, the TCN-ED model achieves higher accuracy than the LSTM-ED model.

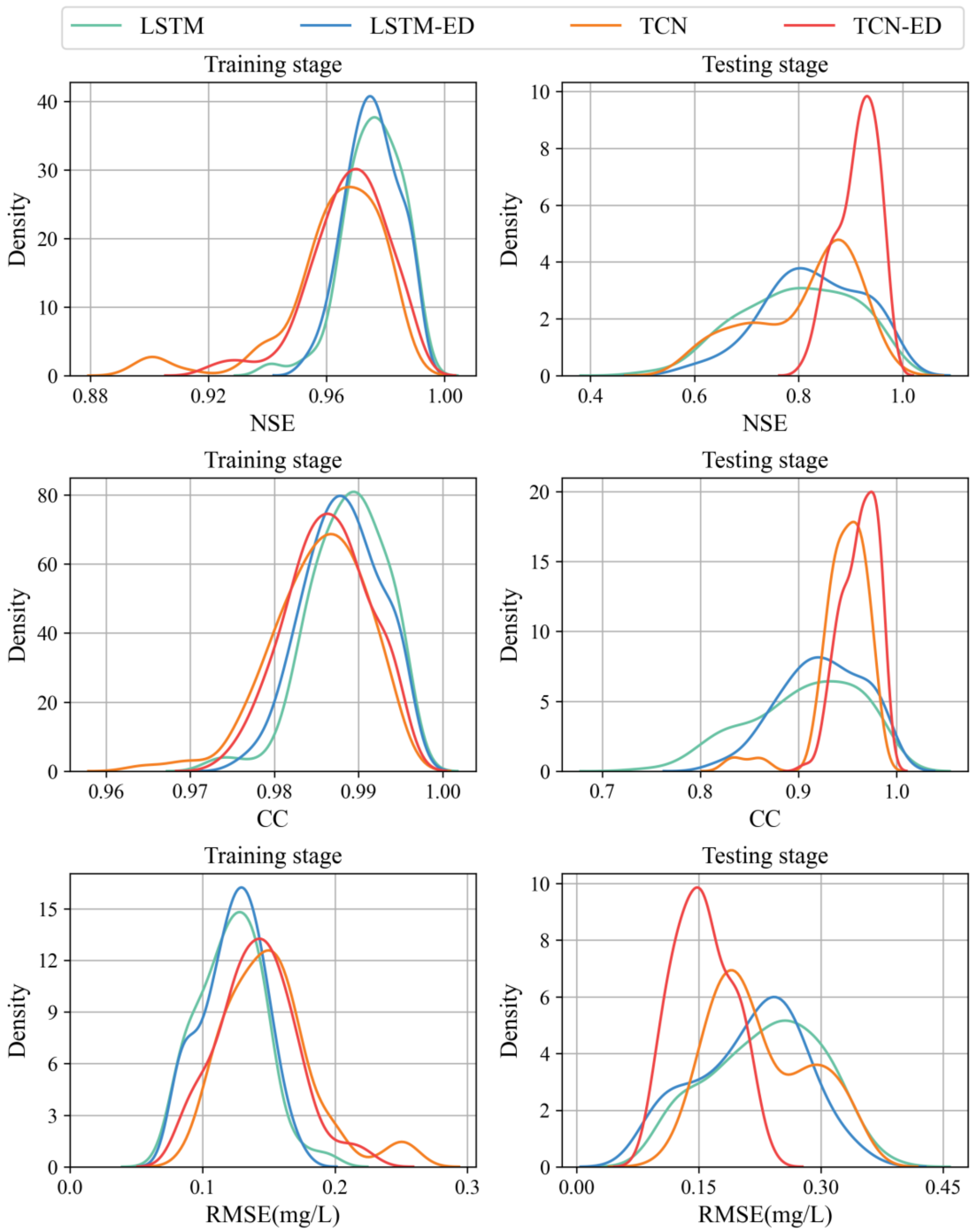



**Fig.3** Gaussian kernel density estimation curves (GKDEs) of the three evaluation indicators of prediction models corresponding to training and testing phases.

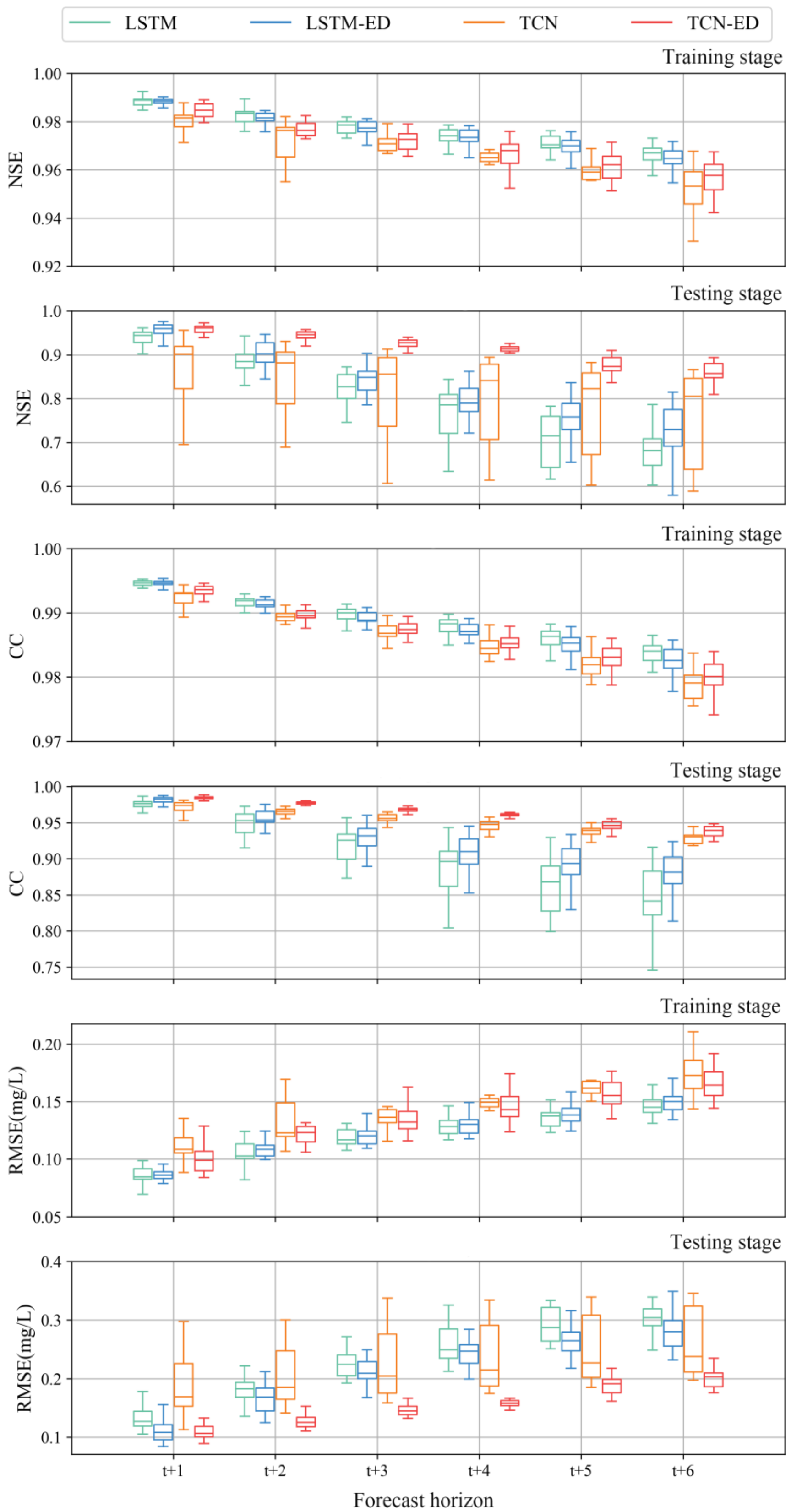



**Fig.4** Boxplots of the evaluation indicator values corresponding to four prediction models (each model runs 20 rounds) from horizon t+1 up to t+6 with respect to training and testing phases.

### *4.2 Assessment on model reliability*

The Taylor diagram graphically summarizes the goodness of fit between the observed and predicted data by comprehensively considering the CC, the Centered RMSE (CRMSE) and the Standard Deviation (STD) (Taylor, 2001), and has been widely used to evaluate the accuracy of hydrological models (Pal et al., 2022; Uddin et al., 2023a). Fig. 5 displays the normalized Taylor diagrams (Molina et al., 2021) of four prediction models with the highest accuracies over 20 rounds at three horizons ($t$+1, $t$+3 and $t$+6) in training and testing phases. It is noted that the CRMSE and the standard deviations of each time series are normalized by the standard deviation of the corresponding observed field, so both CRMSE and STD are unitless.

In training phases (Fig. 5(a)), all models have good prediction performance, although each model performance decreases slightly with the growing forecast horizons. The difference in the values of evaluation indicators of the prediction models is similar across the three forecast horizons. The STD values of the data simulated by the TCN model are obviously smaller than those of the observed data, indicating that the variability of the produced data is smaller than that of the observed data. The STD values of the data simulated by the TCN-ED model are slightly larger than 1, while those of the data simulated by the other three models are less than 1. In other words, the simulation data created by the TCN-ED model are closer to the observed data. Regarding the CC and CRMSE values, the effect of the ED framework on the TCN neural network is significantly larger than that of the ED framework on the LSTM neural network. Furthermore, from the standpoint of three evaluation indicators, the TCN-ED outperforms the other three prediction models, despite the superiority decreasing with the growing forecast horizons.

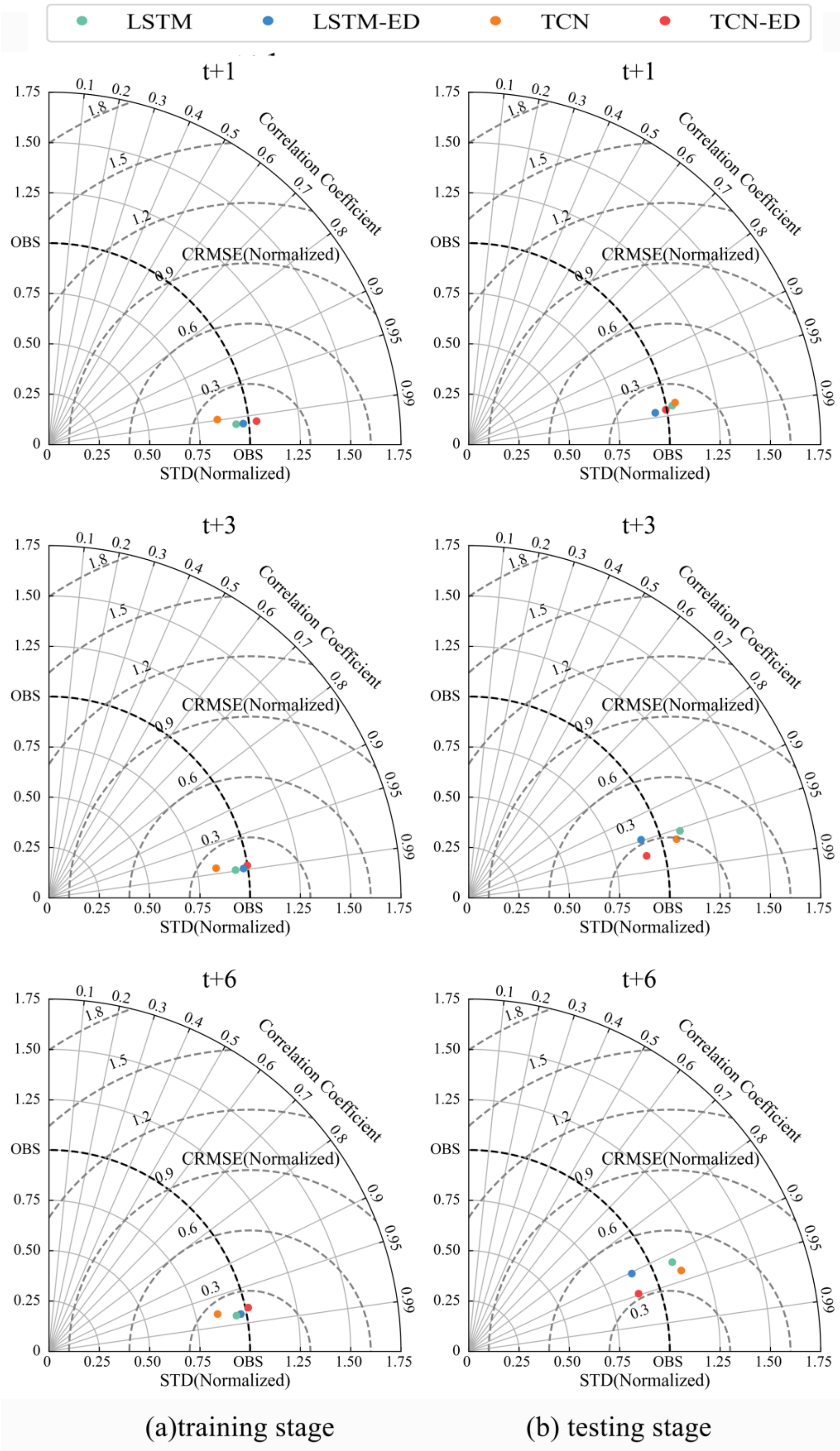



**Fig.5** Normalized Taylor diagrams of four prediction models with the highest accuracies over 20 rounds at the horizons *t*+1, *t*+3 and *t*+6 with respect to training and testing phases.

In testing phases (Fig.5(b)), the TCN-ED (TCN) achieves better reliability than the LSTM-ED (LSTM) in terms of the CC and CRMSE indicators at horizons t+3 and t+6. The variation degree (represented by the STD) of the data forecasted by the TCN and LSTM models without using the ED framework is larger than that of the observed data. The variation degree of the data forecasted by the TCN-ED and LSTM-ED models is smaller, and the goodness of fit between forecasted and observed values decreases with the growth of horizons. That is to say, the ED framework can improve the performance of the TCN and LSTM neural networks by reducing the CRMSE values and increasing the CC values. Besides, for the forecast horizon *t*+1, all models have good predictability. The TCN has slightly larger CRMSE values, lower CC values, and more significant differences in the STD values from 1, while the values of the CC indicator related to the LSTM are close to 0.98, and the STD values of the TCN-ED model are the best (close to 1), in comparison to other models. For the forecast horizons *t*+3 and *t*+6, the LSTM model performs the worst while the TCN-ED model performs the best, according to the CC and CRMSE values. This again demonstrates that the TCN-ED model could well capture the change in ammonia nitrogen concentrations and provide more reliable and accurate forecasts compared with the other models, especially true at long forecast horizons (> 3h).

Fig. 6 shows the forecast results of the maximum (the highest peak of $NH_3$-N concentrations exceeds 3.0 mg/L) ammonia nitrogen outbreak event (2019-11-27 up to 2019-12-2) in testing phases. The concentration of ammonia nitrogen reaches 3.78 mg/L, which causes toxicity to aquatic organisms, affecting the growth and reproduction of fish, crustaceans, and other aquatic organisms (Xu et al., 2021).

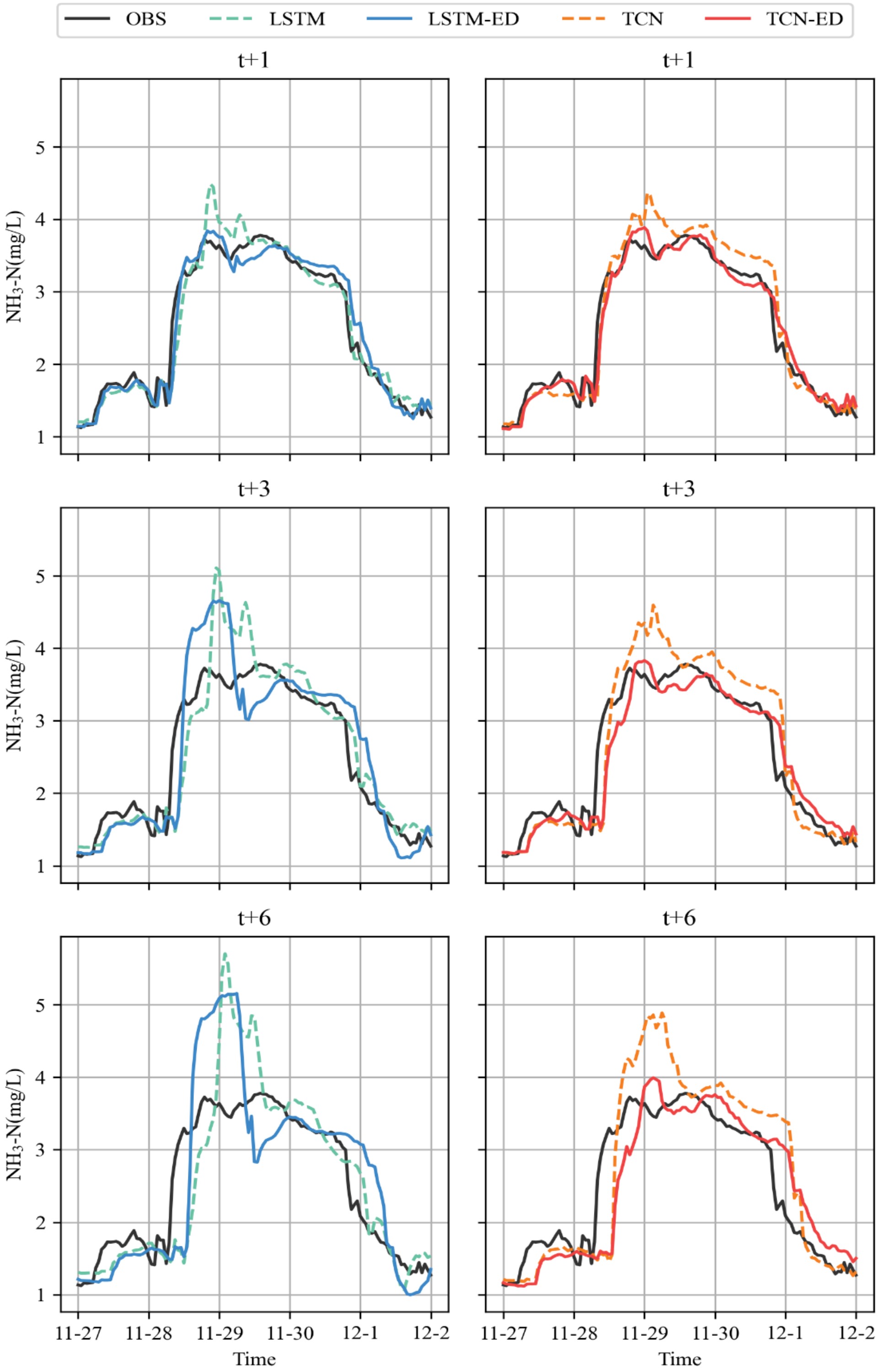



**Fig.6** Forecast results of the maximum (the highest peak of $NH_3$-N concentrations exceeds 3.0 mg/L) ammonia nitrogen concentration event in testing phases with respect to horizons *t*+1, *t*+3 and *t*+6.

It is obviously found that all models perform well and have good predictability at forecast horizon $t$+1. The forecasting processes of $NH_3$-N concentrations produced by the TCN (LSTM) model show an obvious oscillation around the highest peak of $NH_3$-N concentrations. However, such phenomena would be mitigated after fusing the TCN (LSTM) neural network into the ED framework. The forecasted $NH_3$-N concentrations of the TCN-ED could trace the observed $NH_3$-N concentrations, while the forecasting processes of the LSTM model could not fit the observed processes. At the forecast horizon $t$+3, the TCN-ED (TCN) has fewer oscillations in the forecasting processes nearby the highest peak of $NH_3$-N concentrations, compared with those of the LSTM-ED (LSTM). Thus, the LSTM-ED (LSTM) cannot capture the dynamic processes of ammonia nitrogen concentrations. Despite that the forecasting processes of the TCN model do not show obvious oscillations, the problem of time-delay effect exists, and the processes of high ammonia nitrogen concentrations on November 28th - 29th do not fit well. After fusing the TCN into the ED framework, the forecasting processes of the TCN-ED not only could well fit the highest peak of ammonia nitrogen concentrations, but also could trace the subsequent receding processes of ammonia nitrogen concentrations. At the forecast horizon $t$+6, the forecasting processes of the LSTM-ED (LSTM) oscillate more obviously. The highest peak of ammonia nitrogen concentrations is about 3.5 mg/L on November 29th, but the data forecasted by the LSTM-ED (LSTM) is large than 5 mg/L, and the forecasting processes have more intensive oscillations. In contrast, the TCN-ED (TCN) model shows a small oscillation after November 29th.

In summary, the TCN-ED can produce more reliable, stable and accurate ammonia nitrogen forecasts as well as effectively reduce the time-delay effect on the highest peak of ammonia nitrogen concentrations, compared with other three prediction models.

## 5. Conclusion

The ED structure has shown its superiority in capturing the complex and nonlinear relationships between the predictive and dependent variables, thereby enhancing the generalizability of ANNs. The ANN-based ED framework has become a valuable tool in data-driven water quality prediction models. This study developed a TCN-based ED framework (TCN-ED) to forecast ammonia nitrogen processes of a water quality gauge station in Hengsha island of Shanghai City in China. The TCN-ED model could better accomplish the multi-step-ahead ammonia nitrogen forecasts based on three meteorological factors and four water quality factors than the LSTM-ED, TCN and LSTM models. The findings are summarized as follows:

(1) The TCN and LSTM models can generate small errors in the training phases but undergo overfitting and instability. The TCN-ED and LSTM-ED models can mitigate the error propagation associated with multi-horizon forecasts and efficiently overcome the over-fitting bottleneck to achieve better performance with less computation time for model training.

(2) In the ammonia nitrogen forecasting of the highest peak test event, the TCN-ED model shows the shortest time-delay phenomenon among the four models at the long forecast horizon (6h). The TCN-ED model could provide more accurate, stable and reliable forecasts and trace the dynamic processes of the ammonia nitrogen event, even outperforming the LSTM-ED model.

(3) The TCN-ED adequately combines predictive variables with dependent variables patterns in the Encoder and accurately produces the sequence of ammonia nitrogen concentrations in a systematic way in the Decoder. The ED framework largely improves the accuracy, stability and reliability of the ANNs.

The study systematically evaluated the influence of the ED framework on ANNs for water quality forecasting. There is a significant need for multi-task learning in water environmental management to capture the complex nonlinear correlation between multi-input and multi-output factors. Although this study developed an ANN-based ED model to forecast ammonia nitrogen concentration and applied it to a local case study, the developed TCN-ED can be easily extended to predict more water quality factors and model time series in other fields. Furthermore, some studies could be conducted to explore the developed models for forecasting and early warning water pollution events (e.g., algal bloom outbreak and heavy metal pollution). In future research, the confidence interval of the forecast model will be considered. Probabilistic forecasts that take into account uncertainty in the input data and model parameters will also be carried out as probabilistic forecasts can provide more comprehensive and informative predictions compared with deterministic forecasts.


## Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2021YFC3200304), the National Natural Science Foundation of China (U20A20317), and the Research Council of Norway (FRINATEK Project 274310, KLIMAFORSK Project 302457). The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that greatly contributed to improving the manuscript.


## 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.

## Data availability

The data that has been used is confidential.

## References


Abbaszade, M., Salari, V., Mousavi, S.S., Zomorodi, M., Zhou, X., 2021. Application of Quantum Natural Language Processing for Language Translation. IEEE Access 9, 130434-130448. https://doi.org/10.1109/ACCESS.2021.3108768.

Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578, 124084. https://doi.org/10.1016/j.jhydrol.2019.124084.

Aloui, S., Mazzoni, A., Elomri, A., Aouissi, J., Boufekane, A., Zghibi, A., 2023. A review of Soil and Water Assessment Tool (SWAT) studies of Mediterranean catchments: Applications, feasibility, and future directions. J. Environ. Manag. 326, 116799. https://doi.org/10.1016/j.jenvman.2022.116799.

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv, 1803.01271. https://doi.org/10.48550/arXiv.1803.01271

Bertone, E., Rousso, B.Z., Kufeji, D., 2023. A probabilistic decision support tool for prediction and management of rainfall-related poor water quality events for a drinking water treatment plant. J. Environ. Manag. 332, 117209. https://doi.org/10.1016/j.jenvman.2022.117209.

Bian, C., He, H., Yang, S., Huang, T., 2019. State-of-charge sequence estimation of lithium-ion battery based on bidirectional long short-term memory encoder-decoder architecture. J. Power Sources. 449, 227558. https://doi.org/10.1016/j.jpowsour.2019.227558.

Chen, Z., Xu, H., Jiang, P., Yu, S., Lin, G., Bychkov, I., Hmelnov, A., Ruzhnikov, G., Zhu, N., Liu, Z., 2021. A transfer Learning-Based LSTM strategy for imputing Large-Scale consecutive missing data and its application in a water quality prediction system. J. Hydrol. 602, 126573. https://doi.org/10.1016/j.jhydrol.2021.126573.

Cho, K., Van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv Prepr., arXiv1406.1078. https://doi.org/10.48550/arXiv.1406.1078

Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J., Chen, H., 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. Companion Proceedings of The 2019 World Wide Web Conference, 678-685. https://doi.org/10.1145/3308560.3317701.

Deng, T., Chau, K.-W., Duan, H.-F., 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. J. Environ. Manag. 284, 112051. https://doi.org/10.1016/j.jenvman.2021.112051.

Dorado Rueda, F., Durán Suárez, J., del Real Torres, A., 2021. Short-Term Load Forecasting Using Encoder-Decoder WaveNet: Application to the French Grid. Energies 14, 2524. https://doi.org/10.3390/en14092524.

Erdélyi, D., Hatvani, I.G., Jeon, H., Jones, M., Tyler, J., Kern, Z., 2023. Predicting spatial distribution of stable isotopes in precipitation by classical geostatistical- and machine learning methods. J Hydrol 617, 129129. https://doi.org/10.1016/j.jhydrol.2023.129129.

Fu, Y., Hu, Z., Zhao, Y., Huang, M., 2021. A Long-Term Water Quality Prediction Method Based on the Temporal Convolutional Network in Smart Mariculture. Water 13, 2907.

Gopali, S., Abri, F., Siami-Namini, S., Namin, A.S., Year. A Comparison of TCN and LSTM Models in Detecting Anomalies in Time Series Data. 2021 IEEE International Conference on Big Data (Big Data), 2415-2420. https://doi.org/10.1109/BigData52589.2021.9671488.

Guo, C., Cui, Y., 2022. Machine learning exhibited excellent advantages in the performance simulation and prediction of free water surface constructed wetlands. J. Environ. Manag. 309, 114694. https://doi.org/10.1016/j.jenvman.2022.114694.

Han, J., Xu, L., Cao, K., Li, T., Tan, X., Tang, Z., Shi, T., Liao, G., 2021. Online estimation of the heat flux during turning using long short-term memory based encoder-decoder. Case Stud Therm Eng 26, 101002. https://doi.org/10.1016/j.csite.2021.101002.

He, Y., Zhao, J., 2019. Temporal convolutional networks for anomaly detection in time series. Journal of Physics: Conference Series 1213, 042050. https://doi.org/10.1088/1742-6596/1213/4/042050.

Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., Liu, Y., 2020. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Comput 24, 16453-16482. https://doi.org/10.1007/s00500-020-04954-0.

Jamro, I.A., Raheem, A., Khoso, S., Baloch, H.A., Kumar, A., Chen, G., Bhagat, W.A., Wenga, T., Ma, W., 2023. Investigation of enhanced H2 production from municipal solid waste gasification via artificial neural network with data on tar compounds. J. Environ. Manag. 328, 117014. https://doi.org/10.1016/j.jenvman.2022.117014.

Jiang, S., Zheng, Y., Babovic, V., Tian, Y., Han, F., 2018. A computer vision-based approach to fusing spatiotemporal data for hydrological modeling. J. Hydrol. 567, 25-40. https://doi.org/10.1016/j.jhydrol.2018.09.064.

Kao, I.F., Zhou, Y., Chang, L.-C., Chang, F.-J., 2020. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583, 12. https://doi.org/10.1016/j.jhydrol.2020.124631.

Laubscher, R., 2019. Time-series forecasting of coal-fired power plant reheater metal temperatures using encoder-decoder recurrent neural networks. Energy 189, 116187. https://doi.org/10.1016/j.energy.2019.116187.

Liu, J., Zhu, H., Liu, Y., Wu, H., Lan, Y., Zhang, X., 2019. Anomaly detection for time series using temporal convolutional networks and Gaussian mixture model. Journal of Physics: Conference Series 1187, 042111. https://doi.org/10.1088/1742-6596/1187/4/042111.

Loyola, P., Liu, C., Hirate, Y., Year. Modeling user session and intent with an attention-based encoder-decoder architecture. Proceedings of the Eleventh ACM Conference on Recommender Systems, 147-151. https://doi.org/10.1145/3109859.3109917.

Mahmud, T., Paul, B., Fattah, S.A., 2021. PolypSegNet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images. Comput Biol Med 128, 104119. https://doi.org/10.1016/j.compbiomed.2020.104119.

Menon, P., Anantha Singh, T.S., Pani, N., Nidheesh, P.V., 2021. Electro-Fenton assisted sonication for removal of ammoniacal nitrogen and organic matter from dye intermediate industrial wastewater. Chemosphere 269, 128739. https://doi.org/10.1016/j.chemosphere.2020.128739.

Ming, H., Yan, G., Zhang, X., Pei, X., Fu, L., Zhou, D., 2022. Harsh temperature induces Microcystis aeruginosa growth enhancement and water deterioration during vernalization. Water Res. 223, 118956. https://doi.org/10.1016/j.watres.2022.118956.

Mohammed, H., Tornyeviadzi, H.M., Seidu, R., 2021. Modelling the impact of weather parameters on the microbial quality of water in distribution systems. J. Environ. Manag. 284, 111997. https://doi.org/10.1016/j.jenvman.2021.111997.

Molina, M.O., Gutiérrez, C., Sánchez, E., 2021. Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. Int J Climatol 41, 4864-4878. https://doi.org/10.1002/joc.7103.

Mu, L., Mou, M., Tang, H., Gao, S., 2023. Exploring preference and willingness for rural water pollution control: A choice experiment approach incorporating extended theory of planned behaviour. J. Environ. Manag. 332, 117408. https://doi.org/10.1016/j.jenvman.2023.117408.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 0-290. https://doi.org/10.1016/0022-1694(70)90255-6.

Noori, N., Kalin, L., Isik, S., 2020. Water quality prediction using SWAT-ANN coupled approach. J Hydrol 590, 125220. https://doi.org/10.1016/j.jhydrol.2020.125220.

Pal, S.C., Ruidas, D., Saha, A., Islam, A.R.M.T., Chowdhuri, I., 2022. Application of novel data-mining technique based nitrate concentration susceptibility prediction approach for coastal aquifers in India. J Clean Prod 346, 131205. https://doi.org/10.1016/j.jclepro.2022.131205.

Park, C., Kim, D., Yu, H., 2019. An encoder–decoder switch network for purchase prediction. Knowl. Based. Syst. 185, 104932. https://doi.org/10.1016/j.knosys.2019.104932.

Pawan, Dhiman, R., 2023. Electroencephalogram channel selection based on pearson correlation coefficient for motor imagery-brain-computer interface. Measurement: Sensors 25, 100616. https://doi.org/10.1016/j.measen.2022.100616.

Pyo, J., Kwon, Y.S., Min, J.-H., Nam, G., Song, Y.-S., Ahn, J.M., Park, S., Lee, J., Cho, K.H., Park, Y., 2021. Effect of hyperspectral image-based initial conditions on improving short-term algal simulation of hydrodynamic and water quality models. J. Environ. Manag. 294, 112988. https://doi.org/10.1016/j.jenvman.2021.112988.

Quevedo-Castro, A., Bustos-Terrones, Y.A., Bandala, E.R., Loaiza, J.G., Rangel-Peraza, J.G., 2022. Modeling the effect of climate change scenarios on water quality for tropical reservoirs. J. Environ. Manag. 322, 116137. https://doi.org/10.1016/j.jenvman.2022.116137.

Shi, X.J., Chen, Z.R., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst 28, 802-810.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106, 7183-7192. https://doi.org/10.1029/2000JD900719.

Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I., 2023a. A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. Water Res. 229, 119422. https://doi.org/10.1016/j.watres.2022.119422.

Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I., 2023b. A sophisticated model for rating water quality. Sci. Total Environ. 868, 161614. https://doi.org/10.1016/j.scitotenv.2023.161614.

Walsh, C.J., Imberger, M., Burns, M.J., Bos, D.G., Fletcher, T.D., 2022. Dispersed Urban-Stormwater Control Improved Stream Water Quality in a Catchment-Scale Experiment. Water Resour Res 58, e2022WR032041. https://doi.org/10.1029/2022WR032041.

Wan, H., Mao, Y., Cai, Y., Li, R., Feng, J., Yang, H.J.E.w.C., 2021. An SPH-based mass transfer model for simulating hydraulic characteristics and mass transfer process of dammed rivers. Engineering with Computers 38, 3169–3184. https://doi.org/10.1007/s00366-021-01354-2.

Wan, H., Xu, R., Zhang, M., Cai, Y., Li, J., Shen, X., 2022. A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. J. Hydrol. 612, 128081. https://doi.org/10.1016/j.jhydrol.2022.128081.

Wang, L., Dong, H., Cao, Y., Hou, D., Zhang, G., 2023. Real-time water quality detection based on fluctuation feature analysis with the LSTM model. J. Hydroinf. 25, 140–149. https://doi.org/10.2166/hydro.2023.127.

Wiering, M., Kirschke, S., Akif, N.U., 2023. Addressing diffuse water pollution from agriculture: Do governance structures matter for the nature of measures taken? J. Environ. Manag. 332, 117329. https://doi.org/10.1016/j.jenvman.2023.117329.

Xu, J., Luo, W., Huang, Y., 2019. Dadu river runoff forecasting via Seq2Seq. Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science, 494-498. https://doi.org/10.1145/3349341.3349457.

Xu, Z., Cao, J., Qin, X., Qiu, W., Mei, J., Xie, J., 2021. Toxic Effects on Bioaccumulation, Hematological Parameters, Oxidative Stress, Immune Responses and Tissue Structure in Fish Exposed to Ammonia Nitrogen: A Review. Animals 11, 3304. https://doi.org/10.3390/ani11113304.

Yan, K., Li, C., Zhao, R., Zhang, Y., Duan, H., Wang, W., 2023. Predicting the ammonia nitrogen of wastewater treatment plant influent via integrated model based on rolling decomposition method and deep learning algorithm. Sustain Cities Soc 94, 104541. https://doi.org/10.1016/j.scs.2023.104541.

Yang, Y., Xiong, Q., Wu, C., Zou, Q., Yu, Y., Yi, H., Gao, M., 2021. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. Environ. Sci. Pollut. Res. 28, 55129-55139. https://doi.org/10.1007/s11356-021-14687-8.

Zhang, L., Wang, W., Gao, Q., Yang, M., Ji, Y., Geng, S., 2021. Study on the Rapid Prediction Model of Water Quality for Emergency Water Pollution. 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), 211-215. https://doi.org/10.1109/ICCEAI52939.2021.00041.

Zhang, Q., Li, Z., 2021. Data-driven interval credibility constrained quadratic programming model for water quality management under uncertainty. J. Environ. Manag. 293, 112791. https://doi.org/10.1016/j.jenvman.2021.112791.

Zhang, X., Li, D., 2023. Multi-input multi-output temporal convolutional network for predicting the long-term water quality of ocean ranches. Environ. Sci. Pollut. Res. 30, 7914-7929. https://doi.org/10.1007/s11356-022-22588-7.

Zhang, Y.-F., Thorburn, P.J., Fitch, P., 2019. Multi-task Temporal Convolutional Network for Predicting Water Quality Sensor Data. Neural Information Processing, 122-130. https://doi.org/10.1007/978-3-030-36808-1_14.

Zhao, J.C., Deng, F., Cai, Y.Y., Chen, J., 2019. Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere 220, 486-492. https://doi.org/10.1016/j.chemosphere.2018.12.128.

Zheng, H., Liu, Y., Wan, W., Zhao, J., Xie, G., 2023. Large-scale prediction of stream water quality using an interpretable deep learning approach. J. Environ. Manag. 331, 117309. https://doi.org/10.1016/j.jenvman.2023.117309.

Zhou, Y., 2020. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. J. Hydrol. 589, 125164. https://doi.org/10.1016/j.jhydrol.2020.125164.