The exploration of a Temporal Convolutional Network combined with Encoder-Decoder framework for runoff forecasting

Kangling Lin, Sheng Sheng, Yanlai Zhou, Feng Liu, Zhiyu Li, Hua Chen, Chong-Yu Xu **IWA**, Jie Chen and Shenglian Guo

ABSTRACT

The Temporal Convolutional Network (TCN) and TCN combined with the Encoder-Decoder architecture (TCN-ED) are proposed to forecast runoff in this study. Both models are trained and tested using the hourly data in the Jianxi basin, China. The results indicate that the forecast horizon has a great impact on the forecast ability, and the concentration time of the basin is a critical threshold to the effective forecast horizon for both models. Both models perform poorly in the low flow and well in the medium and high flow at most forecast horizons, while it is subject to the forecast horizon in forecasting peak flow. TCN-ED has better performance than TCN in runoff forecasting, with higher accuracy, better stability, and insensitivity to fluctuations in the rainfall process. Therefore, TCN-ED is an effective deep learning solution in runoff forecasting within an appropriate forecast horizon.

Key words | artificial neural network, deep learning, Encoder-Decoder architecture, runoff forecasting, Temporal Convolutional Network

Kangling Lin
Sheng Sheng
Hua Chen (corresponding author)
Jie Chen
Shenglian Guo
State Key Laboratory of Water Resources and
    Hydropower Engineering Science,
Wuhan University,
Wuhan 430072,
China
and
Hubei Provincial Key Lab of Water System Science
    for Sponge City Construction,
Wuhan University,
Wuhan 430072,
China
E-mail: chua@whu.edu.cn

Yanlai Zhou
Chong-Yu Xu **IWA**
Department of Geosciences,
University of Oslo,
P O Box 1047, Blindern, N-0316, Oslo,
Norway

Feng Liu
School of Computer Science,
Wuhan University,
Wuhan 430072,
China

Zhiyu Li
Hua Chen
Department of Geography and Geographic
    Information Science,
University of Illinois at Urbana-Champaign,
Champaign, IL 61801,
USA

HIGHLIGHTS

- For the first time, TCN and TCN-ED models are proposed to forecast runoff.
- TCN-ED has better performance than TCN in runoff forecast in this study.
- The concentration time is a critical threshold to the effective forecast horizon.
- Both models perform better in median and high flow than in low flow.
- It is subject to the forecast horizon for both models to forecast peak flow.

INTRODUCTION

Runoff forecasting is of considerable significance to water resources management. Accurate runoff forecasting can guide the hydraulic engineering construction, reservoir operation, flood control, drought relief, and navigation. According to the extent of physical principles, models for runoff forecasting can be divided into two categories: process-driven models and data-driven models (Yuan *et al.* 2018). Process-driven models represent a specific physical process employing experimental formulas before inputting data. Because of the valuable interpretability of

process-driven models, they have been widely used by hydrologists (Beven *et al.* 1984; Ren-Jun 1992; Douglas-Mankin *et al.* 2010; Wang *et al.* 2012). However, given the uncertainty of the hydrological process and the limitations of artificially constructed process-driven models, the parameters and simulation process are challenging to represent hydrological phenomena fully. Due to process-driven models' physical meaning, researchers without a professional background cannot improve the models, resulting in slow model iteration and difficulty in introducing new technologies. With the improvement of data availability and quality, data-driven models can substitute or supplement to process-driven models for runoff forecasting (Yuan *et al.* 2018). Data-driven models focus on the optimal mathematical relationships between a forecast object and a predictor, without considering the physical mechanism (Adamowski & Sun 2010). Therefore, data-driven models are highly transferable.

The artificial neural network (ANN), a popular data-driven model, was applied in runoff modeling at the end of the last century already (Hsu *et al.* 1995; Smith & Eli 1995; Shamseldin 1997; Dawson & Wilby 1998). However, due to the limitation of theoretical research and computing power at that time, the structure of ANNs was relatively simple. Therefore, ANNs were tough to obtain high accuracy results and dwarfed by another representative data-driven model, support vector machine (SVM). Sivapragasam *et al.* (2001) used the prediction technique based on singular spectrum analysis coupled with SVM to predict the Tryggevælde catchment runoff data. Furthermore, they used the flood data at Dhaka to demonstrate the forecast ability of SVM superior to that of the ANN-based model, particularly at longer lead days (Liong & Sivapragasam 2002).

With the rise of deep learning, hydrologists have turned to study the application of advanced deep ANNs in hydrological modeling and forecasting (Kisi 2008; Awchi & Srivastava 2009; Trajkovic 2010; Abdellatif *et al.* 2013; Baba *et al.* 2013; Shiri *et al.* 2015). For instance, Chang *et al.* (2002) presented an ANN termed real-time recurrent learning for streamflow forecasting in the Da-Chia River. Hu *et al.* (2018) applied a long short-term memory (LSTM) network to the Fen River basin and obtained good prediction results. Yuan *et al.* (2018) used the hybrid LSTM network

and ant lion optimizer model in the prediction of monthly runoff. These studies prove that the ANNs, especially the recurrent neural network (RNN), can capture the non-linear and non-stationary dynamics of river flow effectively and stand out among multiple data-driven models.

Nevertheless, the ANN research on runoff forecasting mainly focused on the RNN (Chang *et al.* 2002), including the LSTM (Hu *et al.* 2018) and further improvement based on the LSTM (Yuan *et al.* 2018). The reason is conjectured to be that the RNN, particularly the LSTM, has a promising ability to process time-series data. However, the development of ANNs is changing soon with each passing day, and there are oncoming ANNs worthy of applying for hydrological modeling and runoff forecasting. The Temporal Convolutional Network (TCN) is one of the novel ANNs designed for sequence modeling and prediction (Bai *et al.* 2018). TCN is essentially a combination of the one dimension fully convolutional network (1D FCN) and causal convolutions (Bai *et al.* 2018). The 1D FCN guarantees the length of the input and output of TCN can be kept the same. The causal convolutions ensure that the future information will not be used during convolutions. The dilated convolutions and residual modules, which dramatically increase the receptive field of the convolutional neural network and make it easier to train, respectively, are also used in TCN. With the empirical study demonstrated by Bai *et al.* (2018), TCN has been proved superior for sequence modeling tasks to LSTM. Recently, more and more studies have confirmed the ability of TCN in time-series data processing such as the stock trend prediction, anomaly detection, and recognition of sepsis (Deng *et al.* 2019; He & Zhao 2019; Moor *et al.* 2019). Therefore, TCN is introduced for the first time in hydrology by this research to provide more options for runoff forecasting based on ANNs.

Besides, most studies mentioned above considered all hydrological information contained in input as equal factors for forecasting the runoff sequences (global relationship) without further refining influence factors, therefore missing the interactions of the dependent variables (local relationship) (Park *et al.* 2019). Recently, Kao *et al.* (2020) introduced Encoder-Decoder architecture in flood forecasting by applying the LSTM based Encoder-Decoder model, whose output sequence of models stepped into 1- up to 6-h-ahead. This is the first journal article on the application

of the Encoder-Decoder architecture in hydrology. The Encoder-Decoder architecture is helpful to overcome some obstacles in the application of ANNs on runoff forecasting, which has been successfully applied to many other similar themes such as the Shihmen Reservoir flood forecast (Kao *et al.* 2020), Dadu River runoff forecast (Xu *et al.* 2019), and Yangtze River streamflow prediction (Liu *et al.* 2020). The Encoder-Decoder architecture is comprised of two sub-models that work in a symbiotic manner (Loyola *et al.* 2017). An encoder captures the local relationship of influence factors by encoding the sequence of inputs. Then, a decoder extracts the encoded vectors for the runoff forecasting by jointly modeling the global and local relationships. Therefore, sequential regularities contained in a time series can be learned in a more comprehensive, fine-grained way than using a single network (Bian *et al.* 2019).

The objective of this study is to explore the ability and stability of TCN and the integration of TCN with Encoder-Decoder architecture (TCN-ED) for runoff forecasting with multi-step ahead times. To achieve this goal, the remainder of this study is organized as follows. The study area, hydrological data, and the data preprocessing method are given first. TCN, the Encoder-Decoder architecture, the proposed data-driven runoff forecasting model, the comparison model, and metrics for the evaluation are then introduced. Next, the experimental results and thorough discussion are presented. Finally, the conclusions of this study are summarized.

STUDY AREA AND DATA

The Jianxi River is the primary up-stream tributary of the Minjiang River in southeast China. The Qilijie station (118°18′16″E, 27°01′21″N) is located in the Jianxi River. The Qilijie station with a drainage area of 14,787 km^2 is selected for this study (Jie *et al.* 2018). The primary soils in this area are red, yellow, and paddy soils. The regional climate is dominated by southeast Pacific Ocean and southwest Indian Ocean subtropical monsoons and partly influenced by regional landforms (Tang *et al.* 2013). The catchment is moist and rainy, with the mean annual rainfall from 1,800 to 2,200 mm, most of which occurs from March to September. The map of the study catchment is shown in Figure 1.

As can be seen from Figure 1, there are 16 precipitation stations, three evaporation stations, and seven runoff stations in the study area. The hourly data of these monitoring sites from 2009 to 2015 are obtained from the local Hydrology Bureau. The hourly runoff of the Qilijie station, the arithmetic mean of the hourly precipitation, and hourly evaporation among all corresponding stations in the basin, abbreviated hereafter as the 'runoff', 'precipitation', and 'evaporation', respectively, are used in this study. The runoff concentration time is about 12 h in this basin (Jie *et al.* 2018).

In order to ease the negative effect of the different scales of data on models' learning ability, the normal standardization is applied to data preprocessing, which is defined as follows:

$$X_*(t) = \frac{X(t) - \bar{X}}{\sigma} \tag{1}$$

where $X^*(t)$ is the normal standardization for input data in the tth time. $X(t)$ is the input sequence, and $\bar{X}$ and σ are the mean and standard deviation of the input sequence, respectively.

METHODS

Temporal Convolutional Network

The TCN, based on a 1D convolutional network, is a generic network structure for sequence modeling (Liu *et al.* 2019). With the empirical study demonstrated by Bai *et al.* (2018), TCN has been proved superior performance for sequence modeling tasks to LSTM.

In order to meet the requirements of sequence modeling tasks, TCN utilizes causal convolutions. Therefore, outputs are only influenced by present and past inputs in each layer (Moor *et al.* 2019). In addition, TCN uses a 1D FCN structure, whose convolution layers have the same size as RNNs by adding zero paddings (He & Zhao 2019).

In order to increase the receptive field with less computation cost, dilated convolutions have been introduced to TCN. The filter of a dilated convolution is applied over a region larger than its size by skipping input values with a given step. The dilated factor commonly increases
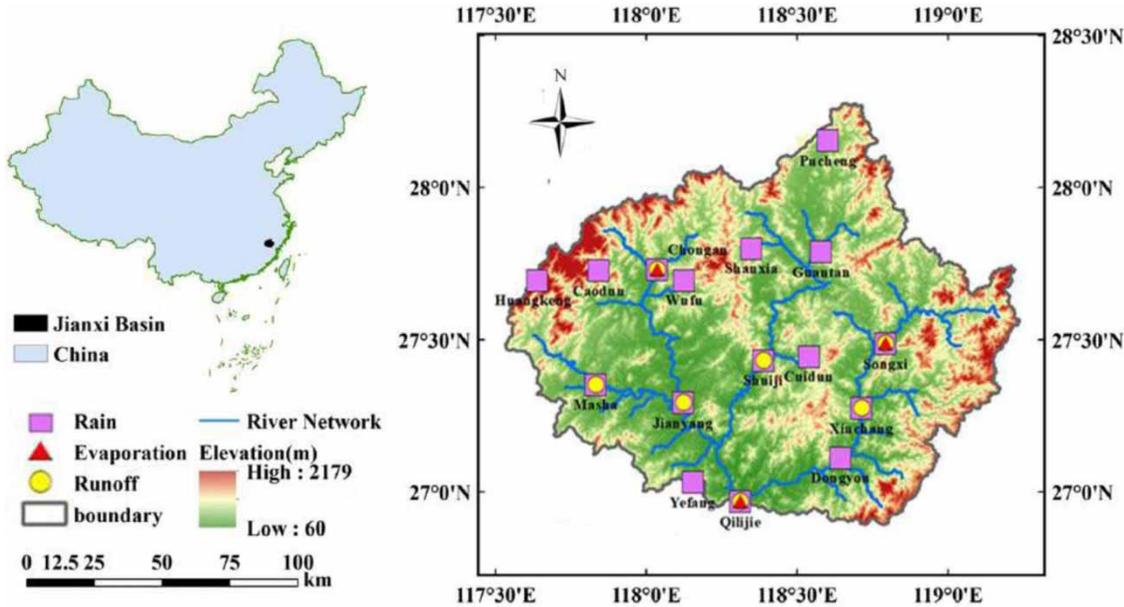
Figure 1 │ Map of the study catchment.

exponentially with the depth of the network, which ensures the receptive field covering each input in the history (He & Zhao 2019). Moreover, TCN applies residual connections that combine the previous input and the result of the convolution with an addition to ease deterioration, which has been proved very beneficial for deep networks (He *et al.* 2016).

Encoder-Decoder architecture

The Encoder-Decoder architecture was proposed by Cho *et al.* (2014) who used RNNs both as the encoder and decoder. This architecture is comprised of two ANNs: the first one is to take the information as input and encode it into a context value. The second one is used for decoding the context value into the expected output sequence. The purpose of this architecture is to compress various information contained in the whole input sequences into a fixed-length vector to make tensor flow more stable (Xu *et al.* 2019).

If the observation at the time step t is represented by X_t, the sequence to sequence forecasting problem is to predict the next sequence $\hat{X}_{t+1}, \ldots, \hat{X}_{t+k}$ whose length is k, according to the previous j observation $X_{t-j+1}, \ldots, X_t$ (Shi *et al.* 2015). Among them, the most likely next sequence

$\tilde{X}_{t+1}, \ldots, \tilde{X}_{t+k}$ is the target.

$$\tilde{X}_{t+1}, \ldots, \tilde{X}_{t+k} = \underset{\hat{X}_{t+1},\ldots,\hat{X}_{t+k}}{\arg\max}$$
$$p(\hat{X}_{t+1}, \ldots, \hat{X}_{t+k}|X_{t-j+1}, X_{t-j+2}, \ldots, X_t) \tag{2}$$

The encoder is to refine the information contained in the input and fix it into a context value.

$$\tilde{X}_{t+1}, \ldots, \tilde{X}_{t+k} \approx \underset{\hat{X}_{t+1},\ldots,\hat{X}_{t+k}}{\arg\max}$$
$$p(\hat{X}_{t+1}, \ldots, \hat{X}_{t+k}|f_{\text{encoder}}(X_{t-j+1}, X_{t-j+2}, \ldots, X_t)) \tag{3}$$

Then, the decoder decodes the context value and outputs the final prediction.

$$\tilde{X}_{t+1}, \ldots, \tilde{X}_{t+k} \approx g_{\text{decoder}}(f_{\text{encoder}}(X_{t-j+1}, X_{t-j+2}, \ldots, X_t)) \tag{4}$$

TCN combined with Encoder-Decoder model

TCN-ED is proposed in this study, which is compared with the TCN model to verify their ability in runoff forecasting. The precipitation, evaporation, and runoff for the first 48 h are used as input to both models to forecast the runoff for

the next 24 h. The framework of the TCN-ED model (Figure 2) is explained below.

As shown in Figure 2, the sample enters the temporal block through 1×1 convolution. Each temporal block contains two layers of the causal convolutions. The zero paddings are added to make the output length between the layers the same, and the rectified linear unit is used to activate the output. The input of each temporal block is added to the output after 1×1 convolution with reference to Bai *et al.* (2018), to increase the nonlinearity of the residual link when transferring information across layers. The dilation factor generally increases exponentially with the depth of the network. Regarding the settings of Moor *et al.* (2019), for the first TCN (TCN encoder), dilation factors correspond to temporal blocks are 1, 2, 4, 8, and 16, respectively, to ensure that the receptive field can cover the input sequence. In order to meet the requirements of the output sequence length, the

context value is copied 24 times, and then enters the second TCN (TCN decoder). Since the time step of the second TCN's input is half of the first TCN, the number of temporal blocks is correspondingly reduced by one. In this way, overfitting is eased on the premise that the receptive field of the TCN decoder can still cover the context value sequence. The output of the TCN decoder at each time step performs dimension reduction through two fully connected layers, and finally outputs the forecast runoff sequence.

In this study, the output dimensions at each time step of the TCN encoder and the TCN decoder are 256 and 128, respectively. The output dimensions at each time step of the first fully connected layer and the second fully connected layer are 64 and 1, respectively. With such a setting, the tensor flow among the network layers is hoped to be smoother under the premise of giving the models enough parameter space.
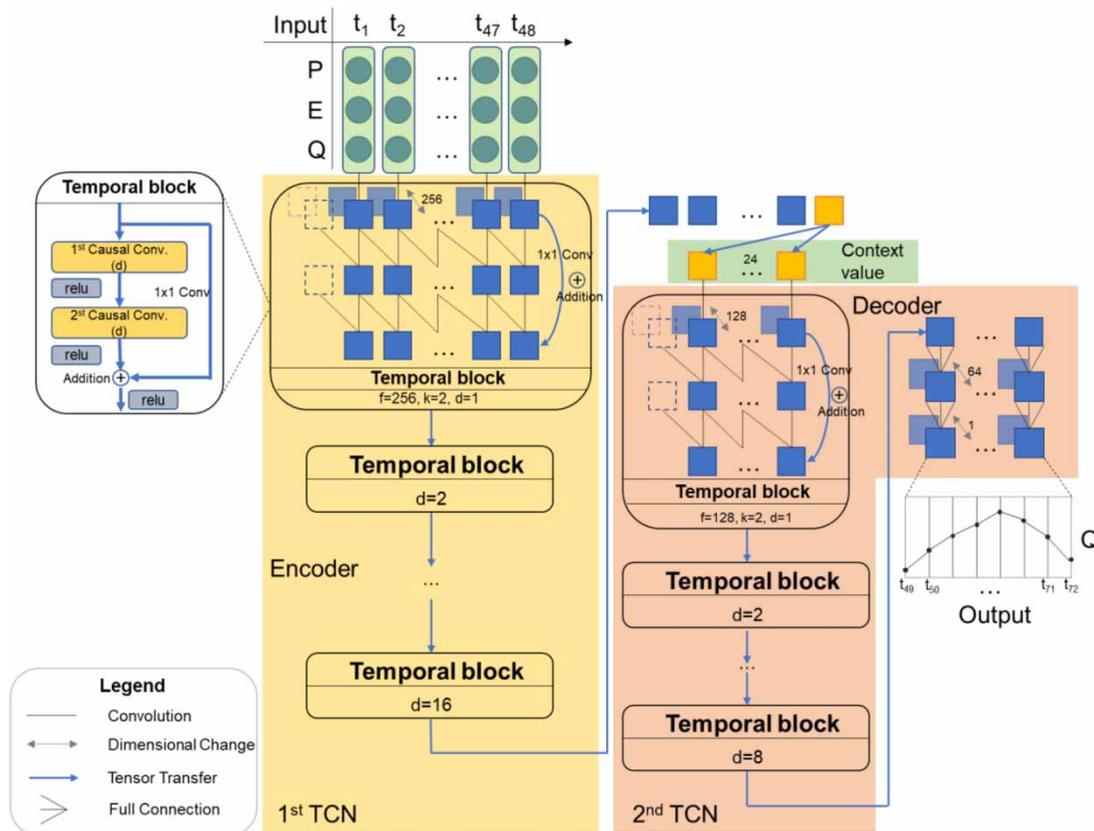
Figure 2 │ The framework of the proposed TCN-ED. P, E, and Q represent the precipitation, evaporation, and runoff, respectively. *f*, *k*, and *d* represent the number of filters, filter size, and dilation factor, respectively.

Flood forecasting model based on the proposed method

To build a dataset to train and test the model, the preprocessed data from 2009 to 2015 are sorted in chronological order, which are further divided into training samples with a 5-year period and testing samples with a 2-year period. Therefore, the training and testing samples are provided with temporal continuity. Although the runoff concentration time is about 12 h in the study basin, the forecast horizons are set to range from 1 to 24 h to detect the forecast ability of TCN and TCN-ED during different forecast horizons. Each model continually trains and optimizes the performance by reducing the errors in the training set. The Adam optimizer (Kingma & Ba 2014) and mean square error (MSE) objective function (Allen 1971) are used in the training stage, and the number of epochs is set to 100. The test set is used to evaluate the performance of the models. The numerical calculations in this paper are accomplished on the supercomputing system in the Supercomputing Center of Wuhan University. Calculating the flow structure of each model is as follows: First, the data are preprocessed and divided into training data and testing data. Second, the training data are further divided into continuous 48 time step samples and 24 time step targets, input into the constructed TCN and TCN-ED, with MSE as the objective function, and Adam optimizer for model training in 100 epochs. Finally, evaluate the model at each forecast horizon using the testing data.

Evaluation of the model performance

In order to objectively illustrate the improvement of the Encoder-Decoder models compared with the ANN models without the Encoder-Decoder architecture, TCN is set as the comparison model for TCN-ED with the same input and output. In order to make TCN as similar as possible to TCN-ED, the comparison model stacks two TCN layers as well. The time steps of the first and second TCN layer are the same as the sample's time steps, so the number of temporal blocks for both TCN layers is 5. The second TCN connects its output at the last time step to the two fully connected layers for dimension reduction, and finally outputs the runoff sequence with forecast horizons from $t + 1$ to $t + 24$. The output dimensions at each time step of the first TCN layer and the second TCN layer are 256 and 128, respectively. The output dimensions of the first fully connected layer are 64. In order to output the runoff sequence with 24 forecast horizons, the output dimensions of the last fully connected layer are 24.

In this study, Nash–Sutcliffe efficiency (NSE), volumetric efficiency (VE), and coefficient of determination (R^2) (Dodge 2006) are employed as hydrological metrics. NSE is a popular indicator in runoff forecasting defined by Nash & Sutcliffe (1970), but it is oversensitive to the extremal flood (Legates & McCabe 1999; Jiang *et al.* 2018). VE is an alternative indicator focusing on the measurement of overall performance that addresses certain limitations of NSE (Criss & Winston 2008). The formulae of NSE and VE are expressed as follows:

$$\text{NSE} = 1 - \frac{\sum (\hat{Q} - Q)^2}{\sum (\bar{Q} - Q)^2} \tag{5}$$

$$\text{VE} = 1 - \frac{\sum |\hat{Q} - Q|}{\sum Q} \tag{6}$$

where $\hat{Q}$ is the predicted runoff, Q is the observed runoff, and $\bar{Q}$ is the average of the observed runoff. Both indicators range from negative infinity to 1, where 1 represents a perfect fit (Jiang *et al.* 2018). Four classes are chosen for each criterion: 'very good' when >0.66, 'good' when 0.33–0.66, 'average' when 0–0.33, and 'poor' when <0 (Ecrepont *et al.* 2019).

Since ANNs have strong randomness, all models are recalculated 40 rounds (with different random initial weights) (Kao *et al.* 2020). With the repeated calculation results, the accuracy, stability, and robustness of each model are analyzed. The best model is used to forecast the largest flood event in the testing stage, which is determined as the model that produces the highest NSE value averaging over 24 forecast horizons in the testing stages.

RESULTS AND DISCUSSION

Evaluation and comparison of models performances

In order to compare the performance of TCN and TCN-ED in the learning and forecasting phase, the minimum, mean, and maximum NSE and VE values averaging over 24

forecast horizons of each model in the training and testing stages with 40 rounds are shown in Table 1.

As can be seen from Table 1, TCN and TCN-ED show good performance, while TCN-ED maintains higher NSE and VE values than TCN in the training and testing stages. Especially for the minimum accuracy, TCN-ED has a 4.48% improvement rate for NSE and a 4.65% improvement rate for VE compared with TCN in the testing stages,

indicating that the Encoder-Decoder architecture can improve the stability with higher minimum accuracy.

For evaluating the performance of each model intuitively and holistically, Figure 3 shows Gaussian kernel density estimation (GKDE) for all NSE and VE values of each model in the training and testing stages, respectively. For the GKDE curve, the accuracy when the density peak appears is the mode accuracy of the model, the sharpness of the curve

Table 1 │ The minimum, mean, and maximum NSE and VE values averaging over 24 forecast horizons of each model in the training and testing stages with 40 rounds

	NSE (%)						VE (%)					
	TCN			TCN-ED			TCN			TCN-ED		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Training	87.47	90.12	93.70	89.37	91.64	94.42	71.05	75.68	80.31	74.79	78.27	81.76
Testing	82.67	87.40	89.15	87.15	88.84	90.22	71.41	77.14	79.47	76.06	79.07	81.14
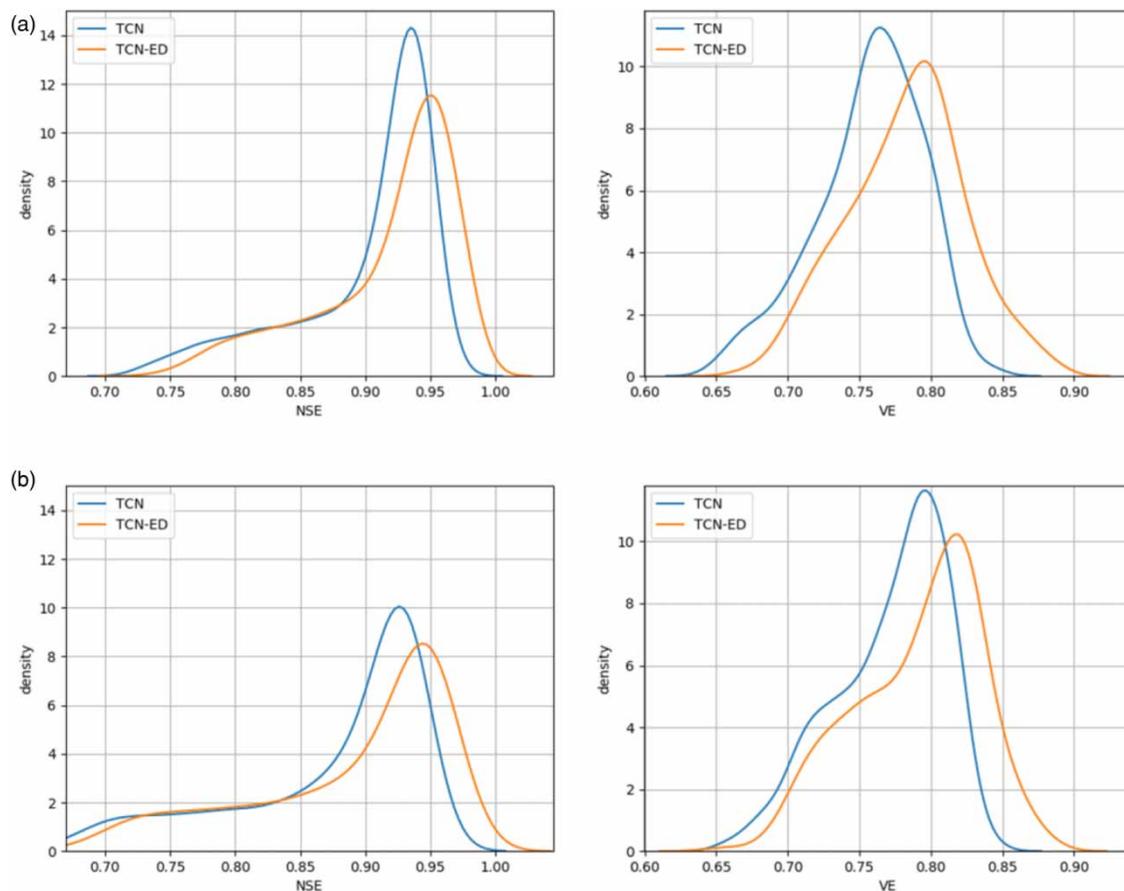
Figure 3 │ GKDE for NSE and VE values of TCN and TCN-ED in the (a) training and (b) testing stages.

represents the concentration of the accuracy. If the GKDE curve is on the right of the *X*-axis and the shape is sharp, it means that the model has stable and high-precision results.

As shown in Figure 3, distributions of NSE and VE values of the two models are both left-skewed, i.e., the left tails of the GKDE curves are longer, the mass of distributions are concentrated on the right, showing both models have a few relatively low values. For each subgraph, the GKDE curve's density peak of TCN is higher than that of TCN-ED. However, the value at the GKDE curve's density peak of TCN-ED is higher than that of TCN-ED, and the GKDE curve of TCN-ED has a longer tail in the high end than that of TCN, causing the GKDE curve of TCN-ED is more rightward than that of TCN. Therefore, TCN-ED has a higher mode accuracy and more high-precision results than TCN. Both models perform better under the VE indicator from the training stage to the testing stage. However, the NSE GKDE curve's density peak and the NSE value at the density peak of TCN drop more than those of TCN-ED from the training stage to the testing stage, respectively.

The above phenomenon shows that TCN may overfit in the training stage, resulting in weaker transferability than TCN-ED whose GKDE curves change less between the training stage and the testing stage.

Models performance with multi-step ahead times

The effective forecast horizons of ANNs have always been the bottleneck and difficulty of the artificial intelligence model research (Zhu *et al.* 2008; Cardenas-Barrera *et al.* 2013; Claveria *et al.* 2016). In order to compare the accuracy of TCN and TCN-ED models during different forecast horizons, boxplots for NSE and VE values in the testing stage are shown in Figure 4.

It can be seen from Figure 4 that both models perform very well as their NSE and VE values are very high and stable within $t + 12$ forecast horizons, while exceeding the $t + 12$ forecast horizon, the forecast accuracy of both models decreases gradually with the forecast horizon increasing. This shows that the forecast ability of the two
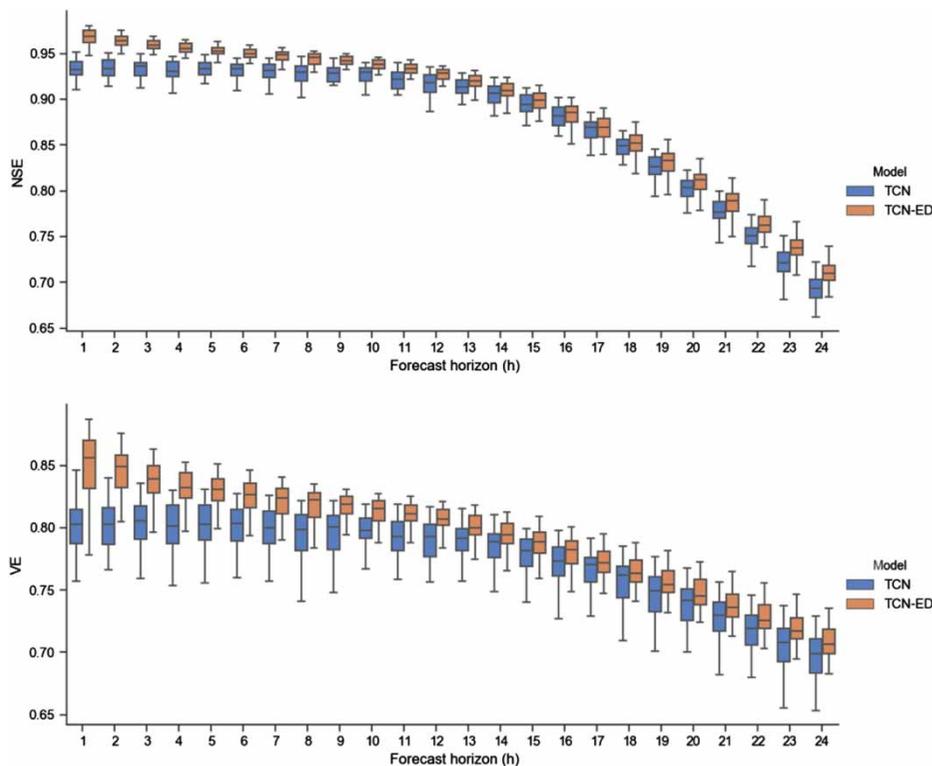

Figure 4 | Boxplots for NSE and VE values of TCN and TCN-ED during forecast horizons from $t + 1$ to $t + 24$ in the testing stage.

models during different forecast horizons is closely related to the concentration time of the basin, which needs to be further verified by more cases in more basins. It is evident that TCN-ED's forecast accuracy and stability are higher than those of TCN at almost each forecast horizon, respectively, especially for short forecast horizons up to $t + 12$ and long horizons close to $t + 24$. In the process of the Encoder-Decoder, the encoder simulates the process of reading and preprocessing in the brain. The context value with a specific length symbolizes the formed memory. The decoder represents the phase when combining known memory and new information to react by the brain. Compared with ANNs whose network layers are directly stacked, the Encoder-Decoder ANNs are not only more conducive to our understanding of the learning process but also make the tensor transmission between network layers more efficient and stable. Therefore, TCN-ED has higher forecast accuracy and stability than TCN at almost every forecast horizon. For TCN and TCN-ED, in order to keep the output length consistent, the zero paddings are added, resulting in the redundant information which interferes with causal convolutions more

seriously during the short forecast horizons. In comparison to TCN, TCN-ED's context value which has been refined already ease the problem caused by redundant information and make TCN-ED more advantageous during the short forecast horizons. In addition, the context value is the output of the last time step of the encoder, so the learning memory is time-sensitive, and TCN-ED is superior obviously to TCN during long horizons close to $t + 24$.

In flood control forecasting, more attention is paid to the model's ability to forecast large-scale floods. Therefore, the best model results of TCN and TCN-ED are used to evaluate the largest flood event, whose peak is maximum among 19 flood events in the testing stage at forecast horizons $t + 6$, $t + 12$, $t + 18$, and $t + 24$, as shown in Figure 5.

The largest flood event, which is considered moderately hazardous, has a maximal flow peak reaching 7,043 m^3/s in the study period, and the accumulated precipitation in the basin during the flood rising period achieves 134 mm. It can be found from Figure 5 that because there are multiple rain peaks, the forecast of rising limb and flood peaks by both models are unstable, causing the hydrograph jagged.
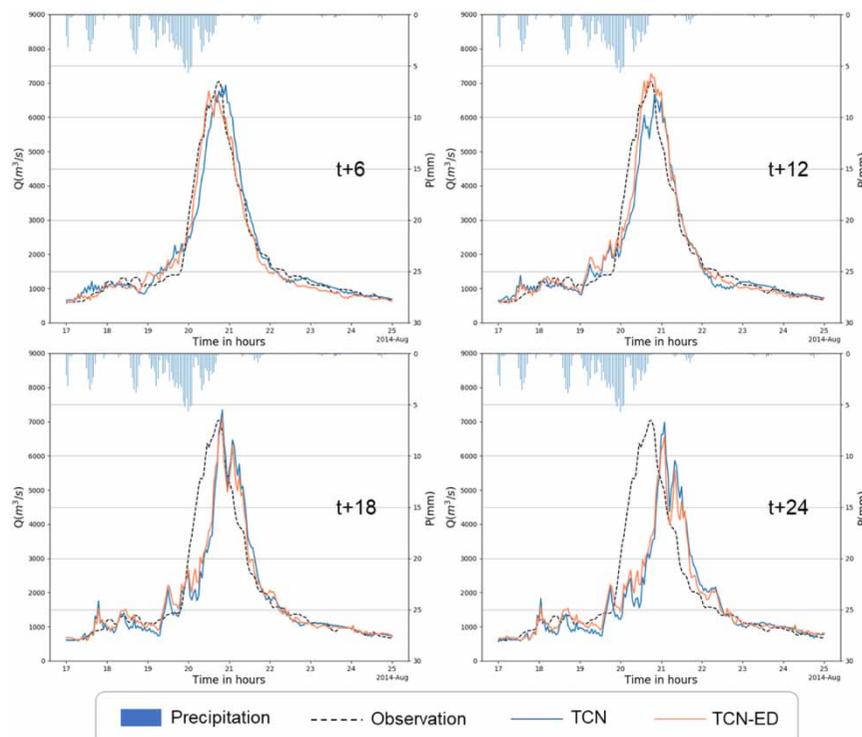
Figure 5 | The observed and predicted floods with the maximum peak in the testing stage using the best model results of TCN and TCN-ED during horizons $t + 6$, $t + 12$, $t + 18$, and $t + 24$.

This situation becomes more and more evident as the forecast horizon increases. In terms of peak time, within $t+12$ forecast horizons, the forecast peak time of each model is basically the same as the observed one, but at the $t+24$ forecast horizon, the models' forecast peaks appear significantly later.

At the $t+6$ forecast horizon, the forecast runoff curves of TCN and TCN-ED both fit the observed runoff curve well, which reflects the excellent forecast ability of TCN and TCN-ED at the short forecast horizon. However, TCN's forecast peak flow is later than the observed peak flow, which will put more pressure on flood control. In contrast, the forecast results of TCN-ED are more practical. At the $t+12$ forecast horizon, it is obvious that TCN-ED has higher accuracy in the peak flow forecast than TCN, as TCN underestimates the peak flow compared to the observed. The forecast runoff of TCN near the peak flow has a quick drop, which may be due to TCN being sensitive to the fluctuation of precipitation. The context value used by TCN-ED at each time step of decoding is uniformly the output of the last time step of the encoder, so the drastic changes in the rain sequence will not be immediately reflected in the output sequence, which is more in line with the smooth change process of the runoff. At the forecast horizons $t+18$ and $t+24$, due to the lack of hydrological information, the forecast runoff hydrograph lags far behind the observed with increasing ahead hours. This indicates that both TCN and TCN-ED models may gradually lose its forecast ability when the forecast horizons are beyond the concentration time of the basin.

Robustness of the models based on regression analysis

In order to explore the robustness of the models, the relationships between the observation and predictions of the 40 rounds are drawn at the $t+6$, $t+12$, $t+18$, and $t+24$ forecast horizons in the testing stage, which are displayed as scatter plots with the KDE curves in Figure 6. The runoff range is divided into the low flow section (178 m^3/s), medium section (178–550 m^3/s), high section (550–1,374 m^3/s), peak flow section (1,374–2,885 m^3/s), and extreme peak flow section (>2,885 m^3/s) by the 25th, 75th, 95th, and 99th percentiles of the observations (Jiang *et al.* 2018). The five sections contain 25, 50, 20, 4, and 1%

of the observations, respectively, which are drawn on the top of each subplot. The percentages of the predictions and their mean R^2 values over the 40 rounds in the five sections are shown on the right side of each subplot. R^2 values of the whole observed and 40 predicted series in the testing stage are calculated, and the maximum, median, and minimum R^2 values among 40 rounds are shown in each subplot.

As Figure 6 shows, both TCN and TCN-ED have high R^2 values (>0.9) for the whole series at the $t+6$ and $t+12$ forecast horizons in the testing stage, while R^2 values decrease a great deal at the $t+18$ and $t+24$ forecast horizons, especially at the $t+24$ horizon. The scatters are closer to the 45-degree line at the $t+6$ and $t+12$ forecast horizons, while they are farther away from the 45-degree line at the $t+18$ and $t+24$ forecast horizons. These indicate that the observation and predication scatters have good fitness at the $t+6$ and $t+12$ forecast horizons, and poor fitness at the $t+18$ and $t+24$ forecast horizons, which is consistent with the results above.

The percentage distribution and R^2 value of each flow section can further reflect the forecast ability of the models for different magnitude flow. It can be seen from each subplot in Figure 6 that the difference of percentage between the two models in different sections is relatively small, and the percentage of TCN-ED is closer to that of the observation than TCN, especially in the low and peak flow sections. It can be seen from Figure 6 that R^2 values of both models in the low flow section are less than 0.3 at four forecast horizons, reflecting poor fitting between the observation and prediction in the low flow section, while in the medium and high flow sections, R^2 values are greater than 0.5, indicating that both models have better forecast ability for median and high flow. It also can be found from Figure 6 that forecast horizons have a great influence on the prediction of peak and extreme peak flow. In Figure 6(a) and 6(b), R^2 values of both models in the peak flow and extreme peak flow sections are greater than 0.5, indicating that both models have good forecast ability for peak and extreme peak flow at the $t+6$ and $t+12$ forecast horizons. However, R^2 values drop to about 0.3 in Figure 6(c), and drop below 0.1 in Figure 6(d), indicating that both models gradually lose the forecast ability to peak and extreme peak flow with forecast horizons increasing. In terms of the R^2 value of each section in each subplot in Figure 6, TCN-ED performs better than TCN as its higher R^2 value.
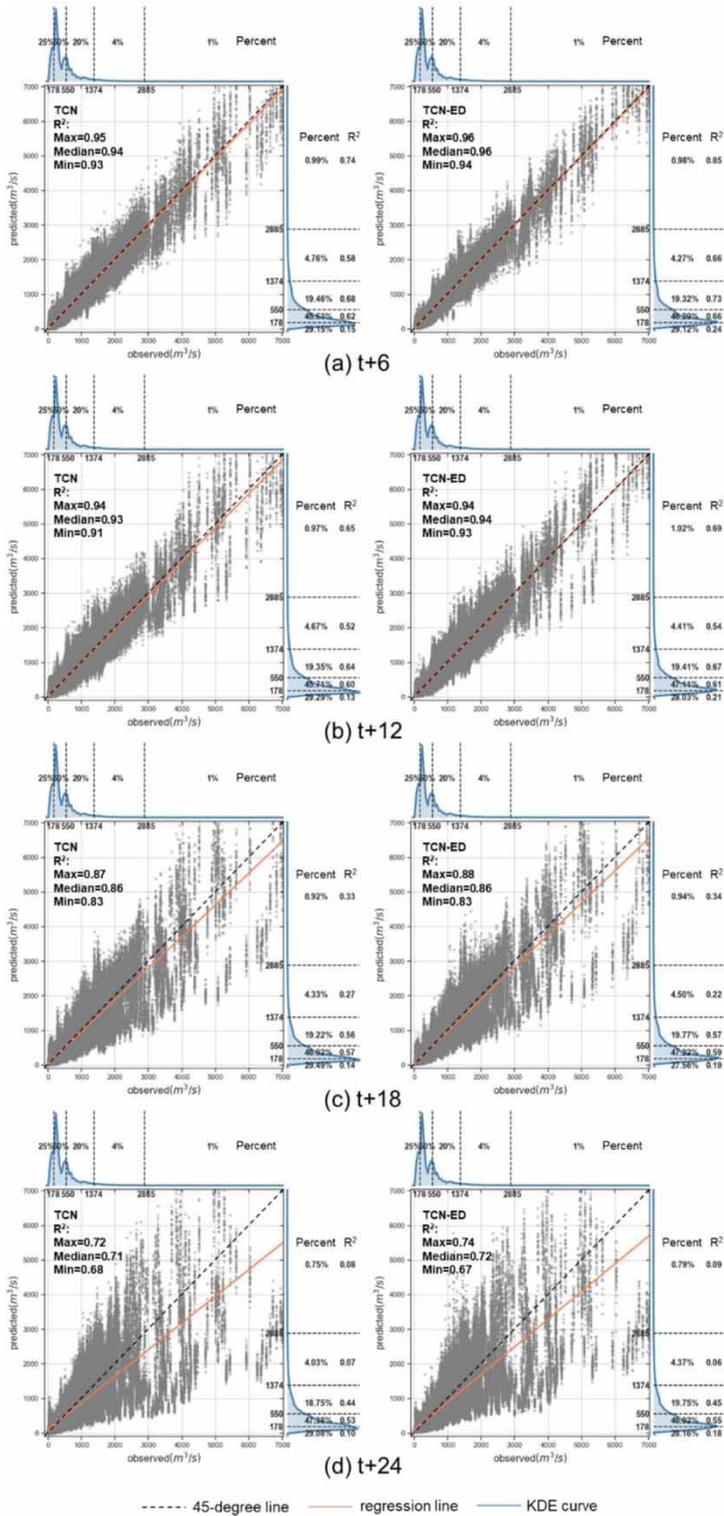
Figure 6 │ Relationships between the observed and predicted runoff in the testing stage using the 40 rounds results of TCN and TCN-ED at horizons (a) $t+6$, (b) $t+12$, (c) $t+18$, and (d) $t+24$.

According to the above analysis, within an appropriate forecast horizon, both TCN and TCN-ED have good forecast ability. For example, the appropriate forecast horizon in this study is the concentration time (12 h) of the basin. Both models perform poorly in the low flow section, and good in the medium and high flow sections, while the forecasting for peak flow and extreme peak flow is greatly affected by the variation of the forecast horizons.

CONCLUSIONS

This study constructs TCN and TCN-ED and demonstrates their applicability in the Jianxi basin, China. In addition, with the comparison of TCN and TCN-ED, the ability of the Encoder-Decoder architecture is further illustrated for the runoff forecasting. First, with the repeated calculation results of each model in the training and testing stages, the accuracy and transferability of each model are analyzed. Second, the performance of each model during various forecast horizons in the testing stage is comprehensively compared. Third, the best model results at the forecast horizons $t+6$, $t+12$, $t+18$, and $t+24$ are used to analyze the model's forecast ability for the maximum floods. Finally, the robustness of the model is explored by the relationships between the observation and predictions of the 40 rounds at the $t+6$, $t+12$, $t+18$, and $t+24$ forecast horizons. The major findings of this study are summarized as follows.

1. The forecast horizon has a significant impact on the forecast ability of TCN and TCN-ED. In this study, NSE and VE values of both models are high and stable within $t+12$ forecast horizons. As the forecast horizon increases after the $t+12$ forecast horizon, NSE and VE values decrease rapidly, indicating that the forecast ability of the models becomes poor. Since the concentration time of the study basin is about 12 h, it can be inferred that the concentration time is a critical threshold to the effective forecast horizon for both models, which needs to be further demonstrated in more basins.
2. Both models perform poorly in the low flow section, and good in the medium and high flow sections during most forecast horizons, while it is conditional to the forecast horizon in forecasting peak flow for both models. Whether this rule is universal or not also needs to be

further verified in more basins, which will also help to improve the forecast ability of ANNs in runoff forecasting.

3. In general, TCN-ED has better performance than TCN in runoff forecasting in this study. TCN-ED shows better transferability from the training stage to the testing stage. NSE and VE values of TCN-ED for most forecast horizons are higher, and R^2 values of TCN-ED for most flow sections are higher. In addition, with the unique context value which maintains the continuity of hydrological information, TCN-ED shows better stability and is insensitive to fluctuations in the rainfall process. Liu *et al.* (2020) proved that the LSTM combined with Encoder-Decoder architecture has better performance in streamflow forecast than the LSTM. More follow-up studies are needed to verify the advantages of Encoder-Decoder architecture integrated with more ANN models.

DATA AVAILABILITY STATEMENT

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

REFERENCES

Abdellatif, M., Atherton, W. & Alkhaddar, R. 2013 A hybrid generalised linear and Levenberg-Marquardt artificial neural network approach for downscaling future rainfall in North Western England. *Hydrology Research* **44** (6), 1084–1101.

Adamowski, J. & Sun, K. R. 2010 Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology* **390** (1–2), 85–91.

Allen, D. M. 1971 Mean square error of prediction as a criterion for selecting variables. *Technometrics* **13** (3), 469–475.

Awchi, T. A. & Srivastava, D. K. 2009 Analysis of drought and storage for Mula project using ANN and stochastic generation models. *Hydrology Research* **40** (1), 79–91.

Baba, A. P. A., Shiri, J., Kisi, O., Fard, A. F., Kim, S. & Amini, R. 2013 Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). *Hydrology Research* **44** (1), 131–146.

Bai, S., Kolter, J. Z. & Koltun, V. 2018 An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*:1803.01271.

Beven, K., Kirkby, M., Schofield, N. & Tagg, A. 1984 Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. *Journal of Hydrology* **69** (1–4), 119–143.

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. *Journal of Power Sources* **449**, 227558.

Cardenas-Barrera, J. L., Meng, J. L., Castillo-Guerra, E. & Chang, L. C. 2013 A neural network approach to multi-step-ahead, short-term wind speed forecasting. In: *12th International Conference on Machine Learning and Applications*, Miami. Vol. 2, pp. 243–248.

Chang, F.-J., Chang, L.-C. & Huang, H.-L. 2002 Real-time recurrent learning neural network for stream-flow forecasting. *Hydrological Processes* **16** (13), 2577–2588.

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*:1406.1078.

Claveria, O., Torra, S. & Monte, E. 2016 Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection. *Revista De Economia Aplicada* **24** (72), 109–132.

Criss, R. E. & Winston, W. E. 2008 Do Nash values have value? Discussion and alternate proposals. *Hydrological Processes* **22** (14), 2723–2725.

Dawson, C. W. & Wilby, R. 1998 An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal* **43** (1), 47–66.

Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. & Chen, H. 2019 Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In: *Companion Proceedings of the 2019 World Wide Web Conference*, San Franscisco, CA, pp. 678–685.

Dodge, Y. 2006 Coefficient of determination. *Alphascript Publishing* **31** (1), 63–64.

Douglas-Mankin, K., Srinivasan, R. & Arnold, J. 2010 Soil and water assessment tool (SWAT) model: current developments and applications. *Transactions of the ASABE* **53** (5), 1423–1431.

Ecrepont, S., Cudennec, C., Anctil, F. & Jaffrézic, A. 2019 PUB in Québec: a robust geomorphology-based deconvolution-reconvolution framework for the spatial transposition of hydrographs. *Journal of Hydrology* **570**, 378–392.

He, Y. & Zhao, J. 2019 Temporal convolutional networks for anomaly detection in time series. *Journal of Physics: Conference Series* **1213** (4), 042050.

He, K., Zhang, X., Ren, S. & Jian, S. 2016 Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, CA, pp. 770–778.

Hsu, K. l., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research* **31** (10), 2517–2530.

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

Jiang, S., Zheng, Y., Babovic, V., Tian, Y. & Han, F. 2018 A computer vision-based approach to fusing spatiotemporal data for hydrological modeling. *Journal of Hydrology* **567**, 25–40.

Jie, M.-X., Chen, H., Xu, C.-Y., Zeng, Q., Chen, J., Kim, J.-S., Guo, S.-l. & Guo, F.-Q. 2018 Transferability of conceptual hydrological models across temporal resolutions: approach and application. *Water Resources Management* **32** (4), 1367–1381.

Kao, I. F., Zhou, Y. L., Chang, L. C. & Chang, F. J. 2020 Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. *Journal of Hydrology* **583**, 12.

Kingma, D. P. & Ba, J. 2014 Adam: a method for stochastic optimization. *arXiv*:1412.6980.

Kisi, O. 2008 River flow forecasting and estimation using different artificial neural network techniques. *Hydrology Research* **39** (1), 27–40.

Legates, D. R. & McCabe Jr, G. J. 1999 Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35** (1), 233–241.

Liong, S. Y. & Sivapragasam, C. 2002 Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association* **38** (1), 173–186.

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.

Liu, D., Jiang, W., Mu, L. & Wang, S. 2020 Streamflow prediction using deep learning neural network: case study of Yangtze River. *IEEE Access* **8**, 90069–90086.

Loyola, P., Liu, C. & Hirate, Y. 2017 Modeling user session and intent with an attention-based encoder-decoder architecture. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*, Como, Italy, pp. 147–151.

Moor, M., Horn, M., Rieck, B., Roqueiro, D. & Borgwardt, K. 2019 Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping. *Machine Learning for Healthcare* Ann Arbor, USA, **106**, 2–26.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology* **10** (3), 0–290.

Park, C., Kim, D. & Yu, H. 2019 An encoder–decoder switch network for purchase prediction. *Knowledge-Based Systems* **185**, 104932.

Ren-Jun, Z. 1992 The Xinanjiang model applied in China. *Journal of Hydrology* **135** (1–4), 371–381.

Shamseldin, A. Y. 1997 Application of a neural network technique to rainfall-runoff modelling. *Journal of Hydrology* **199** (3–4), 272–294.

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. *Advances in Neural Information Processing Systems* **28**, 802–810.

Shiri, J., Marti, P., Nazemi, A. H., Sadraddini, A. A., Kisi, O., Landeras, G. & Fard, A. F. 2015 Local vs. external training of neuro-fuzzy and neural networks models for estimating reference evapotranspiration assessed through k-fold testing. *Hydrology Research* **46** (1), 72–88.

Sivapragasam, C., Liong, S.-Y. & Pasha, M. 2001 Rainfall and runoff forecasting with SSA–SVM approach. *Journal of Hydroinformatics* **3** (3), 141–152.

Smith, J. & Eli, R. N. 1995 Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning Management* **121** (6), 499–508.

Tang, Q., He, X., Bao, Y., Zhang, X., Guo, F., Zhu, H., Tang, Q., He, X., Bao, Y. & Zhang, X. 2013 Determining the relative contributions of climate change and multiple human activities to variations of sediment regime in the Minjiang River, China. *Hydrological Processes* **27** (25), 3547–3559.

Trajkovic, S. 2010 Testing hourly reference evapotranspiration approaches using lysimeter measurements in a semiarid climate. *Hydrology Research* **41** (1), 38–49.

Wang, G., Zhang, J., Jin, J., Pagano, T., Calow, R., Bao, Z., Liu, C., Liu, Y. & Yan, X. 2012 Assessing water resources in China using PRECIS projections and a VIC model. *Hydrology and Earth System Sciences* **16** (1), 231–240.

Xu, J., Luo, W. & Huang, Y. 2019 Dadu river runoff forecasting via Seq2Seq. In: *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, Hangzhou, China, pp. 494–498.

Yuan, X. H., Chen, C., Lei, X. H., Yuan, Y. B. & Adnan, R. M. 2018 Monthly runoff forecasting based on LSTM-ALO model. *Stochastic Environmental Research and Risk Assessment* **32** (8), 2199–2212.

Zhu, X. T., Wang, H., Xu, L. & Li, H. Z. 2008 Predicting stock index increments by neural networks: the role of trading volume under different horizons. *Expert Systems with Applications* **34** (4), 3043–3054.