

Received February 2, 2020, accepted February 6, 2020, date of current version February 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974406

Deep Learning Data-Intelligence Model Based on Adjusted Forecasting Window Scale: Application in Daily Streamflow Simulation

MINGLEI FU¹, TINGCHAO FAN¹, ZI'ANG DING¹, SINAN Q. SALIH²,
NADHIR AL-ANSARI³, AND ZAHER MUNDHER YASEEN⁴

¹College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

²Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

³Civil, Environmental, and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden

⁴Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Corresponding author: Zaher Mundher Yaseen (yaseen@tdtu.edu.vn)

ABSTRACT Streamflow forecasting is essential for hydrological engineering. In accordance with the advancement of computer aids in this field, various machine learning (ML) models have been explored to solve this highly non-stationary, stochastic, and nonlinear problem. In the current research, a newly explored version of an ML model called the long short-term memory (LSTM) was investigated for streamflow prediction using historical data for forecasting for a particular period. For a case study located in a tropical environment, the Kelantan river in the northeast region of the Malaysia Peninsula was selected. The modelling was performed according to several perspectives: (i) The feasibility of applying the developed LSTM model to streamflow prediction was verified, and the performance of the developed LSTM model was compared with the classic backpropagation neural network model; (ii) In the experimental process of applying the LSTM model to the prediction of streamflow, the influence of the training set size on the performance of the developed LSTM model was tested; (iii) The effect of the time interval between the training set and the testing set on the performance of the developed LSTM model was tested; (iv) The effect of the time span of the prediction data on the performance of the developed LSTM model was tested. The experimental data show that not only does the developed LSTM model have obvious advantages in processing steady streamflow data in the dry season but it also shows good ability to capture data features in the rapidly fluctuant streamflow data in the rainy season.

INDEX TERMS Deep learning model, streamflow forecasting, tropical environment, window scale forecasting, LSTM.

I. INTRODUCTION

Studies have consistently evidenced the complex nature of predicting streamflow due to the involvement of natural variabilities, such as the complex nature, non-linearity, randomness, and non-stationarity of river systems [1], [2]. Several research efforts have been reported in the field of hydrology pertaining to the improvement of the reliability and accuracy of hydrological variable prediction [3]–[7]. Until now, many hydro-meteorological studies have been reported, but there

The associate editor coordinating the review of this manuscript and approving it for publication was Kok Lim Alvin Yau¹.

has been no single approach that can be said to be efficient in modelling hydrological events under varying conditions, particularly as there are various catchment features that play a part. This is due to certain physical processes that characterize river flows, such as the periodicity of the methods and the current patterns in the model data [8]. It could be said that, currently, a model that outperforms other models in various hydrological conditions is non-existent. It may not be feasible to generate consistent prediction using several models due to the dynamic nature and non-stationarity of historical data; hence, studies are needed to develop more efficient models based on the available historical data [9].

The recent advancements in computational models can also be exploited to improve modelling accuracy of such models [7], [10], [11]. The feasibility of using novel data-intelligence techniques in developing efficient forecasting models should also be explored.

For the purposes of water resource management, it is important to understand the hydrological processes that control streamflow patterns. Several studies have been conducted in recent years on streamflow phenomena due to the interest in both global and regional changes in hydrologic patterns that result in drought and flooding [1], [12]–[14]. Evidence in the literature suggests that streamflow patterns can be modelled using either physically based models or artificial intelligence (AI) based models [15]. Meanwhile, there is a need for more studies on the hydrological parameters to develop the required initial and boundary criteria for simulating the elemental processes of specific watersheds with the aid of physically based models [16], [17].

From the reported studies on streamflow simulation, it is evident that classical regression tools are commonly used despite being associated with low accuracy levels [18]; this has motivated the development of AI methods, which offer more accuracy. Several reviews in the field of hydrology suggest that various AI models have been investigated for streamflow simulation; such models include support vector machine (SVM), artificial neural network (ANN), adaptive neuro-fuzzy inference system, complementary wavelet-AI, as well as hybrid evolutionary computing models [1], [12]–[14], [19]–[21]. However, several challenges are still associated with these models, especially regarding their implementation as expert systems for sustainable river engineering. Other challenges are their time-inefficiency and manual modelling processes. In various studies, various model optimisation methods such as input vector optimisation, prediction interval optimisation, integration, hybridisation, and data decomposition have been proposed [22]–[24]. Thus, it is necessary to come up with a universally applicable and automated AI model that is applicable across different local scales.

The non-linear relationships between the estimators and the simulated parameters can be understood using conceptually-based methods as they rely on historical input data and do not require previous flow information [25]. Such models could be beneficial as they demand few hydrological inputs [26]. The basic steps in the AI-based models are as follows: problem analysis, data collection & preprocessing, data-driven model selection, identification of the optimal model from the host of trained models, and evaluation of the selected model. The most important of these steps is the data-driven model identification as this is where the learning process is performed, and the features are extracted. The optimal model approximation is performed by reducing the training error between the actual and predicted matrix; the selection of the optimal model is made from the host of trained models as the model with the lowest mean squared error after a series of independent validation processes.

The approximation performance of AI models can also be influenced by factors that affect actual hydrological conditions, such as the determination of the model input, the configuration of the model, and the time-scale horizon.

As compared to the traditional ANN model, the long short-term memory (LSTM) network model adds the setting of the forget gate. The addition of thresholds allows the LSTM model to be reset internally at the appropriate times to avoid network crashes. LSTM has better time series capture capabilities and better long-term memory, and can solve complex artificial long-time-lag problems [27], [28]. During the past two years, the application of LSTM to time series forecasting has received increasing attention [29]. LSTM is widely used for forecasting various time series, including flood debris, groundwater levels, lake water levels, watershed runoff, and meteorological problems [30]–[38].

The advantages and feasibility of LSTM in capturing the long-term dependence of time series were highlighted by comparing LSTM with the autoregressive integrated moving average and the generalised regression neural network models [39]. In 2018, Zhang et al. applied LSTM to the analysis of countercurrent problems in wastewater treatment plants and compared it with the Elman and NARX (non-linear autoregressive network with exogenous inputs) networks [40]. Zhang et al. added a loss layer based on LSTM to predict the depth of the water table [41]. In the group's study, the new forecasting model was compared with the feed-forward neural network and the double LSTM model. The results show that the new model is more accurate for groundwater level forecasting. Later, LSTM was combined with the lion optimiser algorithm model of the ant colony optimisation (ALO) model to form the LSTM-ALO model, which was applied to flow forecasting for the Astor River basin [42]. Widiyari et al. applied an LSTM model to forecast river water levels in the lower reaches of the Semarang region. The results demonstrated the role of LSTM in flood control and disaster mitigation [43].

In an additional study, LSTM was applied to studying the water level fluctuations and reservoir operations in Dongting Lake [44]. The LSTM model was compared with the SVM model, and a comprehensive examination of the influence of the Three Gorges Dam on the water level of Dongting Lake was conducted. The accuracy of the LSTM model used in the study was much better than that of the SVM model, especially in forecasting high water-level values, where it showed an excellent performance [44]. In the research related to reservoir operation, multiple studies were conducted for different time scales and different flow regimes, and the forecasting results of three different models, LSTM, back-propagation (BP), and SVM, were compared. Compared to the BP and SVM models, the LSTM model predicted the operating modes under different conditions more quickly and accurately [45]. Tian et al. explored the potential of the LSTM model for runoff simulations of the Xiangjiang River and the Qujiang River basins. In this study, LSTM demonstrated excellent time series capture and better long-term memory

compared to SVM [46]. In 2019, Huang et al. published a study in which LSTM was combined with the El Niño-Southern Oscillation index. In comparison with the linear regression model, LSTM used the nonlinear evolution of the data better and demonstrated its statistical advantage in long leads [47].

As highlighted in the previous paragraphs, the LSTM model has been increasingly used for various aspects of hydrology research. In the literature survey, the application of LSTM in groundwater level forecasting, reservoir operation forecasting, runoff problems in the Xiangjiang River and Qujiang River basins, and El Niño-Southern Oscillation positively demonstrated the feasibility of using LSTM to predict water conservancy and climate-related issues [48]. The streamflow problem is an important issue that combines water conservancy and climate. The application of LSTM in streamflow forecasting research has important practical significance for flood control and disaster reduction.

As compared with previous studies, this innovative study was conducted from the following perspectives.

- i. The first objective was to examine the feasibility of applying the developed LSTM to streamflow forecasting located in a highly stochastic environment. The time series characteristics of streamflow periodicity and the reliable accuracy of the developed LSTM model for streamflow forecasting results illustrate the feasibility of using the LSTM in streamflow forecasting.
- ii. Based on the first objective, the effects of the size of the training set on the performance of the model were explored.
- iii. Based on the second objective, the effects of the time interval between the training set data and testing set data on the performance of the model were explored.
- iv. Based on the third objective, the effects of the time span of predicted data on the performance of the model were explored.

II. STUDY AREA DESCRIPTION AND DATA AVAILABILITY

In this study, the daily scale streamflow information used was collected from the Kelantan River located in the northeastern region of the Malaysia Peninsula. The location of the river is presented in Fig. 1. Permission to collect the historical data was obtained from the Malaysian Department of Irrigation and Drainage. The measurements of the streamflow were collected using telemetry methods, where a sensor placed at the targeted area of the river calculates the water level pressure. The case study site is predisposed to flooding because of heavy monsoon rainfall events. Hence, the adoption of such an intelligent model for streamflow prediction can be an important step in flood hydrology assessments. It can also be of significant economic, agricultural, and infrastructural importance. The total length of the Kelantan River is 248 km, and its drainage area is approximately 13,170 km². The daily flow data of the river were obtained for the observed



FIGURE 1. Kelantan river case study area located in the northeastern region of the Malaysia Peninsula.

period of 1964 – 2004. The data set includes data such as streamflow and daily precipitation. The typical characteristics of the streamflow of the Kelantan River are presented in Table 1.

The original data of the Kelantan River used in this study include two datasets: streamflow and rainfall. Fig. 2 (a) shows the streamflow. The time series includes a total of 14,976 data observations for the 50 years from January 1, 1964 to December 31, 2004. Fig. 2 (b) shows the rainfall. The time series includes a total of 14,976 data observations over the 50 years from January 1, 1964 to December 31, 2004. The changes in the streamflow of the Kelantan River have obvious periodic characteristics.

III. METHODOLOGY

In this study, the LSTM model was performed by TensorFlow to predict the streamflow. The complete forecasting process is shown in Fig. 3. The original streamflow data is standardised by z-score normalisation, and then a cyclic neural network is created. The LSTM model is well-trained by means of parameter adjustment, and the pre-processed data are input into the trained LSTM. Finally, the predicted data are output to complete the LSTM-based streamflow forecast.

A. LONG SHORT-TERM MEMORY MODEL

The LSTM model was proposed by Hochreiter et al. in 1997 [27], and Gers et al. conducted a detailed study and elaboration of LSTM's forget gate in 1999 [28]. As shown in Fig. 4, the input gate, output gate, and forget gate are adopted by the LSTM model cell, which makes the model choose the states that have more influence on the current state, compared with the recurrent neural network model.

TABLE 1. Kelantan streamflow characteristics from 1996 to 2005.

Year	February 15 th (m ³ /s)	Dry season (From April to June) (m ³ /s)	July 15 th (m ³ /s)	Rainy season (From August to October) (m ³ /s)	December 15 th (m ³ /s)
1996	27.12	0.56	200.89	1326	77.30
1997	20.13	5.69	254.19	1179	66.68
1998	24.75	2.18	129.30	1454	85.08
1999	26.88	2.49	134.5	1150	127.40
2000	36.88	6.59	183.30	1279	91.68
2001	28.54	0.53	164.80	1834	88.01
2002	25.01	0.44	220.60	1692	83.95
2003	24.12	0.61	445.29	1981	103.50
2004	36.95	5.72	365.50	983.70	86.30
2005	28.65	4.23	462.39	1772	80.26

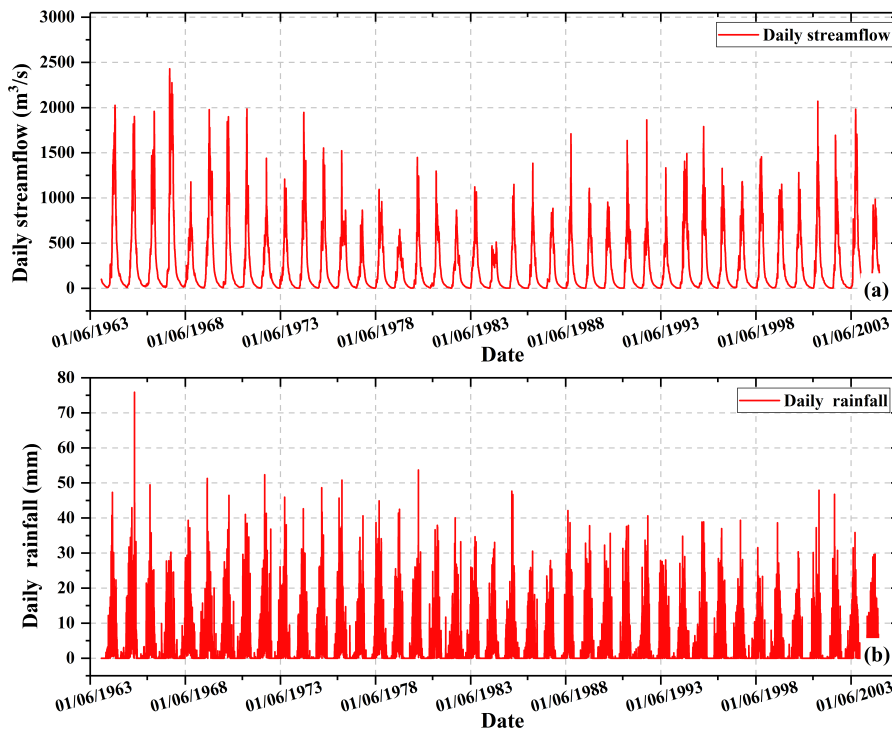


FIGURE 2. Schematic diagram of Kelantan River's original dataset, including (a) the streamflow, and (b) the rainfall.

B. MODEL STRUCTURE

1) FORECASTING PROCESS

The LSTM model adopted in this study is a three-layer neural network, which consists of an input layer, a hidden layer, and an output layer. The neuron number of the hidden layer is set according to Formula (1).

$$M = \sqrt{a + b} + c \tag{1}$$

where a is the neuron number of the input layer, b is the neuron number of the output layer, and c is a constant within the interval $[0, 10]$.

In the work described in this article, the input layer size is 2, and the output layer size is 1. Combining Formula (1)

and Table 2, we consider the number of hidden layer units as 30.

Three thresholds in the LSTM cell follow Formulas (2) to (4).

$$i^{(t)} = \sigma \left(W_i \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_i \right) \tag{2}$$

$$f^{(t)} = \sigma \left(W_f \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_f \right) \tag{3}$$

$$o^{(t)} = \sigma \left(W_o \cdot \left[h^{(t-1)}, x^{(t)} \right] + b_o \right) \tag{4}$$

where $i^{(t)}$ is the input threshold at time t , $f^{(t)}$ is the forget threshold at time t , $o^{(t)}$ is the output threshold at time t , $x^{(t)}$ is the input of the training sample at time t , $h^{(t)}$ is the

TABLE 2. Test results of various parameters of the developed LSTM model.

number	hidden layer units	learning rate	epochs	time (minutes)	RMSE	MAPE	NSE
1	30	0.00001	1000	16.50	99.07	0.341	0.872
2	40	0.00001	1000	17.30	41.36	0.113	0.978
3	30	0.00001	1000	16.37	42.83	0.114	0.976
4	30	0.00001	2000	31.25	40.12	0.095	0.979
5	30	0.00001	3000	47.03	39.72	0.092	0.979
6	20	0.00001	1000	14.85	43.05	0.120	0.976
7	20	0.00001	2000	29.25	42.71	0.115	0.976
8	10	0.00001	1000	13.65	50.05	0.149	0.967
9	30	0.0001	1000	15.77	39.81	0.094	0.979

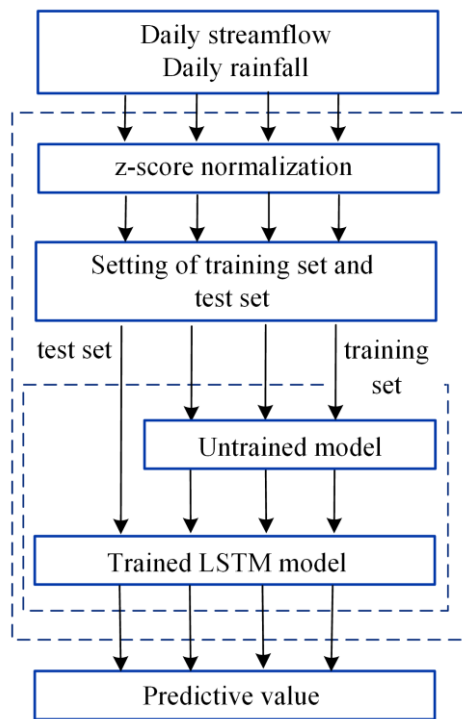


FIGURE 3. Schematic diagram of the modelling process of the streamflow prediction model.

output of the current unit at time t , W is the connection weight of the model, and b is the offset value of the model.

Thus, the state function and output function of the LSTM unit follow Formulas (5) to (7).

$$h^{(t)} = o^{(t)} * \tanh(C^{(t)}) \tag{5}$$

$$C^{(t)} = f^{(t)} * C^{(t-1)} + i^{(t)} * \tilde{C}^{(t)} \tag{6}$$

$$\tilde{C}^{(t)} = \tanh(W_C \cdot [h^{(t-1)}, x^{(t)}] + b_C) \tag{7}$$

where $h^{(t-1)}$ is the output of the current unit, $h^{(t)}$ is the output of the unit at the previous time, $\tilde{C}^{(t)}$ is the unit status at the previous time, and $C^{(t)}$ is the state of the unit at the current moment.

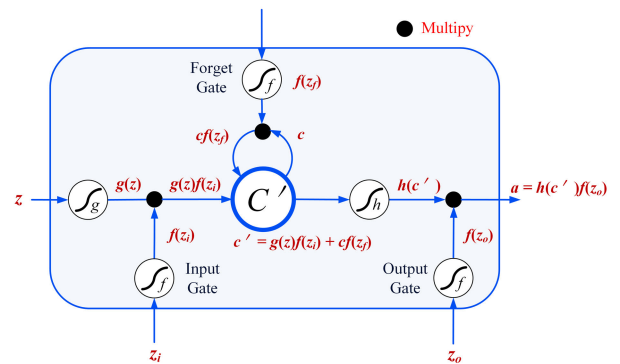


FIGURE 4. Schematic of a cell in the long short-term memory model.

The loss function of the LSTM model follows Formula (8).

$$L = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

where n is the number of the testing samples, y_i is the observed value at time i , and \hat{y}_i is the predicted value at time i .

During the training process of the model, the weights of the LSTM units are adjusted according to the error. The adjustment method follows Formulas (9) and (10).

$$x_h^{(t)} = \frac{\partial L}{\partial h^{(t)}} \tag{9}$$

$$x_C^{(t)} = \frac{\partial L}{\partial C^{(t)}} \tag{10}$$

where $x_h^{(t)}$ is the weight of the hidden layer, and $x_C^{(t)}$ is the weight between the hidden layer and the output layer.

During the training process, the error direction propagation of the input gate, output gate, and forget gate follow Formulas (11) to (13).

$$\delta_i^{(t)} = f'(h_i^{(t)}) \sum_{c=1}^C g(h^{(t)}) x_C^{(t)} \tag{11}$$

$$\delta_o^{(t)} = l'(h_o^{(t)}) \sum_{c=1}^C h(C^{(t)}) x_h^{(t)} \tag{12}$$

$$\delta_f^{(t)} = l'(h_f^{(t)}) \sum_{c=1}^C g(h^{(t)}) x_C^{(t)} \tag{13}$$

TABLE 3. Main functions involved in the long short-term memory model used in this study.

Define the neural network unit type	cell=tf.nn.rnn_cell.BasicLSTMCell(rnn_unit)
Define the loss function	loss=tf.reduce_mean(tf.square(tf.reshape(pred,[-1])-tf.reshape(Y, [-1])))
Define the gradient training algorithm	train_op=tf.train.AdamOptimiser(lr).minimize(loss)
Define the transfer method (input layer to hidden layer)	input_rnn=tf.matmul(input,w_in)+b_in
Define the transfer method (hidden layer to output layer)	pred=tf.matmul(output,w_out)+b_out

where l is the activation function of the control gate, g is the activation input function of the unit, and h is the output activation function of the unit.

If the error is less than the minimum value of the expected error, the model will converge, and if the maximum number of iterations is reached, the training process is complete. Then, the testing data are input into the neural network to forecast the streamflow.

2) LONG SHORT-TERM MEMORY MODEL PARAMETERS USED IN THE EXPERIMENTS

The operating system used in this study was Microsoft Windows 10. TensorFlow 1.7.0, Cuda 9.0, and Python 3.6 were selected as the development environment.

The learning rate is one of the key parameters of the developed LSTM model. The optimised value of the learning rate is set to 0.00001, as shown in Table 2.

The other optimised values of the parameters of the developed LSTM model are as follows: the number of hidden layer units is 30; the input layer size is 2; the output layer size is 1; the learning rate is 0.00001; and the cycle index is 2000.

The main functions involved in the LSTM model are shown in Table 3.

C. APPLICATION OF DATA AND EVALUATION PARAMETERS

1) DATA APPLICATION

In the experiments, 14976 sets of streamflow data and rainfall data in the period from 01/01/1964 to 31/12/2004 were selected as the original data set for streamflow prediction. Considering that the original data has a large variation range, the z-score-normalised preprocess method was performed, as shown by Formulas (14) to (16).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (14)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (15)$$

$$x'_i = \frac{x_i - \bar{x}}{s} \quad (16)$$

where \bar{x} is the average of the original time series data, s is the standard deviation of the original time series data, and x'_i is the new sequence data formed by z-score normalisation.

2) PERFORMANCE METRICS

Various statistical performance indicators were computed, including the mean absolute percentage error (*MAPE*), root mean square error (*RMSE*), root mean square relative error (*RMSRE*), mean absolute error (*MAE*), mean relative error (*MRE*), BIAS, and Nash-Sutcliffe efficiency (*NSE*) [49], [50].

MAPE is the average of absolute percentage errors. The closer the *MAPE* is to 0, the better the prediction result.

$$MAPE = 100 \frac{1}{n} \sum_{i=1}^n \frac{|Sf_o - Sf_f|}{\bar{Sf}_o} \quad (17)$$

where n is the number of the testing samples, Sf_o and Sf_f are the actual and forecasted streamflow values, respectively, and \bar{Sf}_o is the mean value of the actual streamflow.

RMSE is the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations. The *RMSE* reflects the precision of the measurement and is used to measure the deviation between the predicted value and the true value. The value range of *RMSE* is $[0, \infty]$. When *RMSE* is 0, the prediction result is the best.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Sf_o - Sf_f)^2}{n}} \quad (18)$$

RMSRE is the root of the mean squared relative error. The closer the *RMSRE* approaches 0, the better the prediction.

$$RMSRE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{Sf_o - Sf_f}{Sf_o} \right)^2} \quad (19)$$

MAE is the average of the absolute values of the deviations. It uses the absolute value of the deviation, which avoids the positive and negative offset of the deviation and reflects the actual situation of the predicted value error. The *MAE* value range is $[0, \infty]$. When *MAE* is 0, the prediction is the best.

$$MAE = \frac{\sum_{i=1}^n |Sf_o - Sf_f|}{n} \quad (20)$$

MRE is the average of the ratio of the deviation to the true value. The closer the *MRE* approaches 0, the better the prediction.

$$MRE = \frac{1}{n} \sum_{i=1}^n \left(\frac{Sf_o - Sf_f}{Sf_o} \right) \quad (21)$$

TABLE 4. Table of related information of experimental training set and test set of each group.

Serial number	forecast interval	training set			testing set		
		start	end	Size (days)	start	end	Size (days)
A		20/12/1995	06/03/2004	3000	07/03/2004	31/12/2004	300
B		15/09/1998		2000	07/03/2004	31/12/2004	300
		20/12/1995		3000			
		25/03/1993	06/03/2004	4000			
		29/06/1990		5000			
		03/10/1987		6000			
C	0	20/12/1995	06/03/2004	3000	07/03/2004	31/12/2004	300
	1000	10/06/2001	25/03/1993				
	2000	29/06/1990	14/09/1998				
	3000	03/10/1987	19/12/1995				
	6000	17/07/1979	02/10/1987				
	9000	30/04/1971	18/07/1979				
D		23/06/1993	08/09/2001	3000	09/09/2001	17/12/2001	100
						05/07/2002	300
						21/01/2003	500
						09/08/2003	700
						25/02/2004	900

BIAS reflects how much the predicted value deviates from the true value. The closer *BIAS* approaches 0, the better the prediction.

$$BIAS = \frac{\sum_{i=1}^n (Sf_o - Sf_f)}{\sum_{i=1}^n (Sf_o)} \quad (22)$$

NSE reflects how well the model predicts. The value range of *NSE* is $[-\infty, 1]$. The closer the *NSE* is to 1, the higher the reliability of the model. The closer the *NSE* is to 0, the closer the predicted value is to the average of the actual values. The overall model result is credible, but the process prediction error is large. If the *NSE* is much less than 0, the model is unreliable.

$$NSE = 1 - \frac{\sum_{i=1}^n (Sf_o - Sf_f)^2}{\sum_{i=1}^n (Sf_o - \bar{Sf}_o)^2} \quad (23)$$

IV. APPLICATION RESULTS AND ANALYSIS

The three aspects of research involved in the forecasting of streamflow are analysed in this section.

This section is divided into four parts: A, B, C, and D. These four parts discuss the performance stability of the LSTM model from four different perspectives: feasibility, differences in training set size, differences in the period of the dataset used in the training set, and predictions of different durations. As shown in Table 4, the four parts correspond to the relevant information of the training set and the test set in the experiment.

A. FEASIBILITY VERIFICATION OF THE STREAMFLOW PREDICTION MODEL BASED ON LSTM

In this section, the feasibility of applying the LSTM model for streamflow prediction is verified. The classic BP model is adopted as a reference model to be compared with the LSTM model. The streamflow data and rainfall data during the period from 20/12/1995 to 06/03/2004 were used as the training set. The streamflow data and rainfall data during the period from 07/03/2004 to 31/12/2004 were used as the testing set.

As shown in Fig. 5, the curve of LSTM prediction results fits the curve of actual observation data well. Whether it was a smooth stream flow in the dry season or a rapidly fluctuant stream flow in the rainy season, there were significant deviations for the prediction results by the BP model compared with the actual observation data.

As shown in Fig. 6, the residual distribution of prediction results by LSTM was better than the residual distribution of prediction results by BP. The residual of prediction results by LSTM was less than $175 \text{ m}^3/\text{s}$ in the worst case scenario where the streamflow changed rapidly at the points marked by the vertical dashed line in the rainy season. The LSTM model showed stronger robustness to residual distribution than the BP model. The corresponding residuals of prediction results by BP increased significantly in the worst case when the streamflow changed rapidly in the rainy season. Therefore, compared with BP, the residual fluctuation of LSTM has a clear advantage.

Fig. 7 shows the distribution of predicted and observed data by LSTM and BP. As shown in Fig. 7 (a), the predicted

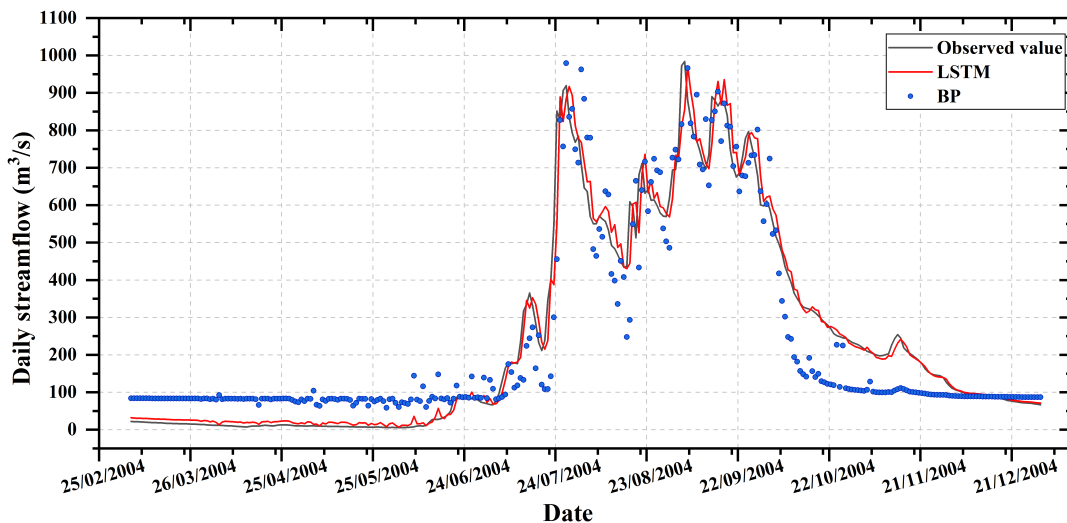


FIGURE 5. Schematic diagram of LSTM and BP prediction results.

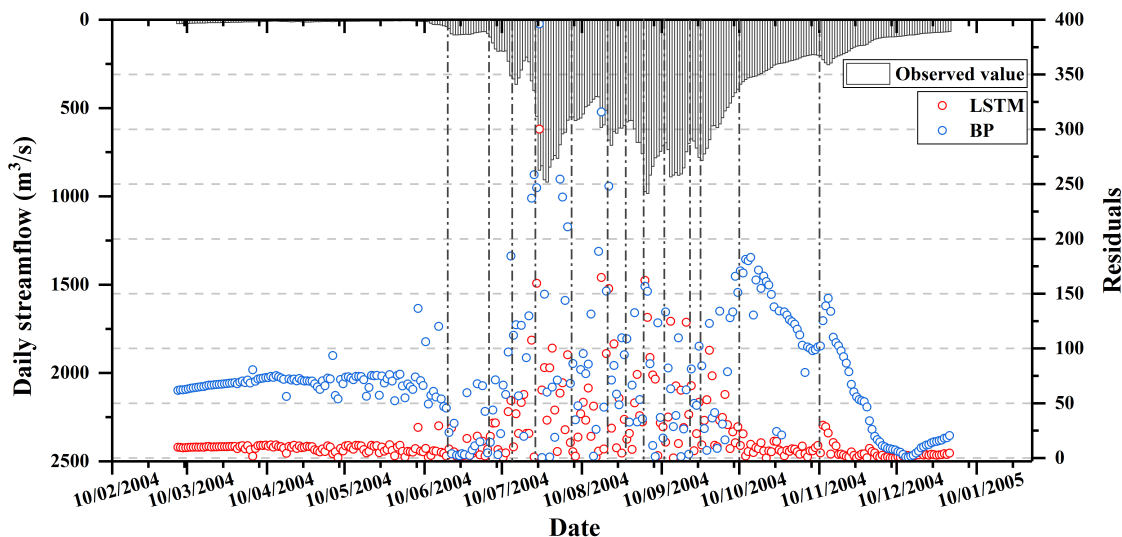


FIGURE 6. Residual distribution of LSTM and BP prediction results. The left vertical axis denotes the observed value of daily streamflow. Please note that the coordinate values are the opposite of the regular pattern. The right vertical axis denotes the residuals of the LSTM (red dot) and BP (blue dot) prediction results.

data of LSTM were close to the observed data except for a few numerical points with large errors (still in the acceptable range). As shown in Fig. 7 (b), the BP model was not sensitive to data changes in low stream flow segments, resulting in lower prediction accuracy. Although the sensitivity of the BP model to the data in the high streamflow segments is better, its prediction accuracy was still lower than that of LSTM.

Table 5 shows the evaluation index values for the prediction results of the LSTM and BP models. The LSTM model outperforms the BP model in all seven evaluation indices. The MAPE of LSTM is 0.08296, which is nearly 0.22 lower than that of BP. The RMSE of LSTM is 38.83468, which is 54.5 lower than that of BP. According to RMSRE, MAE, and MRE, there are substantial performance improvements for LSTM compared with BP. The absolute values of BIAS

TABLE 5. Evaluation indices of lstm and bp prediction results.

Index	Model	
	LSTM	BP
MAPE	0.08	0.30
RMSE	38.83	93.30
RMSRE	0.57	4.33
MAE	20.56	74.87
MRE	-0.29	-2.25
BIAS	-0.02	0.01
NSE	0.98	0.88

corresponding to the two models are similar, but BP’s BIAS is slightly better than LSTM’s. The NSE of LSTM is 0.98039, which is an improvement of nearly 0.1 compared to BP.

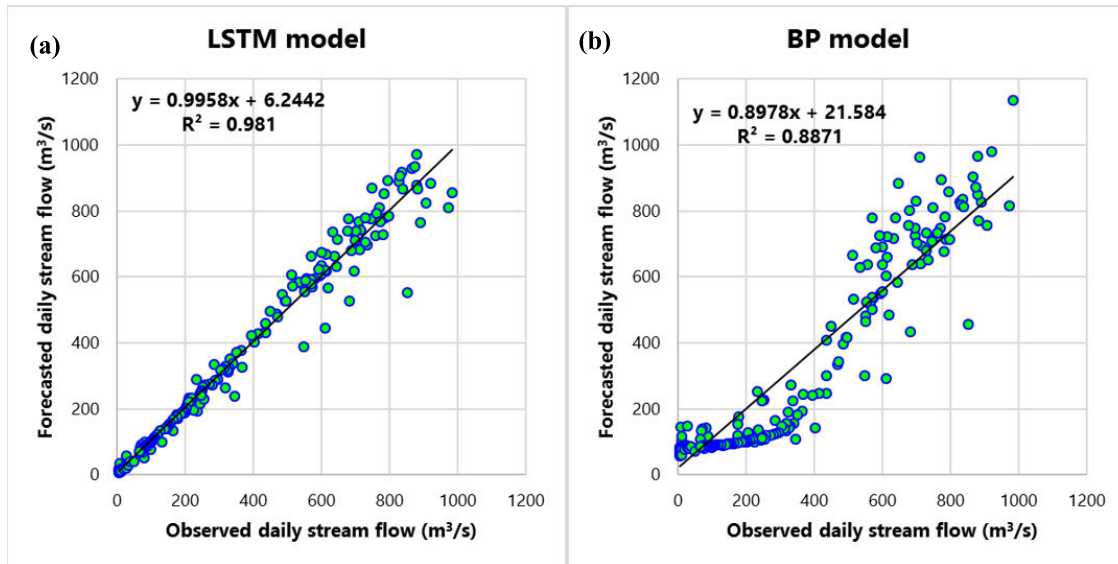


FIGURE 7. Distribution of predicted and observed data by LSTM and BP. (a) prediction result of the LSTM model, (b) prediction result of the BP model.

TABLE 6. Prediction results using different sized training sets.

Index	Training set size				
	2000	3000	4000	5000	6000
MAPE	0.09	0.08	0.08	0.09	0.09
RMSE	39.95	38.83	38.75	39.72	38.33
RMSRE	0.99	0.57	0.61	0.94	1.08
MAE	22.64	20.56	21.17	22.47	22.49
MRE	-0.54	-0.29	-0.25	-0.49	-0.55
BIAS	-0.03	-0.02	-0.02	-0.02	-0.03
NSE	0.97	0.98	0.98	0.97	0.98

The LSTM model is perfectly suitable for the task of streamflow prediction for the Kelantan River. Moreover, compared with the classic BP model, the LSTM model has obvious prediction accuracy advantages, irrespective of whether it was a smooth streamflow in the dry season or a rapidly fluctuant streamflow in the rainy season.

B. EVALUATION OF TRAINING SET SIZE ON THE PERFORMANCE OF THE STREAMFLOW PREDICTION MODEL BASED ON LSTM

In this section, the effect of training set size on the performance of the streamflow prediction model based on LSTM is investigated. In the experiment, original data in the period from 07/03/2004 to 31/12/2004 (almost 300 days) was used as the testing set. For the training set, the deadline was fixed at 06/03/2004. By adjusting the start date of the training set, five training sets with different sizes were created. The data size of the five training sets were: 2000, 3000, 4000, 5000, and 6000 days.

The experimental results of our investigation of the impact of training set size on the prediction performance of the

LSTM model are shown in Fig. 8 and Fig. 9. Fig. 8 is a data curve diagram of the prediction results of the LSTM model under different training set conditions. On the whole, the prediction result curves of the LSTM model have a good fit with the actual observations. For the streamflow prediction in the dry season, such as during the period from 5/2004 to 6/2004, as shown in the enlarged part (a), the curves have the best fit when 3000 and 4000 days of data are chosen as the training sets. However, for the streamflow prediction in the rainy season, such as during the period from 7/2004 to 8/2004 and from 8/2004 to 9/2004, as shown in the enlarged part (b) and (c), larger training sets make the prediction curve better fit the observed values.

According to Fig. 9, as the training set size increases, the distribution of the predicted values of the LSTM model has a tendency to approach the observed values gradually. Therefore, with an increase of the training set size, the prediction ability of the LSTM model improved.

Table 6 shows the evaluation indices of the prediction results by the LSTM model with the five training datasets. As the size of the training set increases, the MAPE, RMSRE,

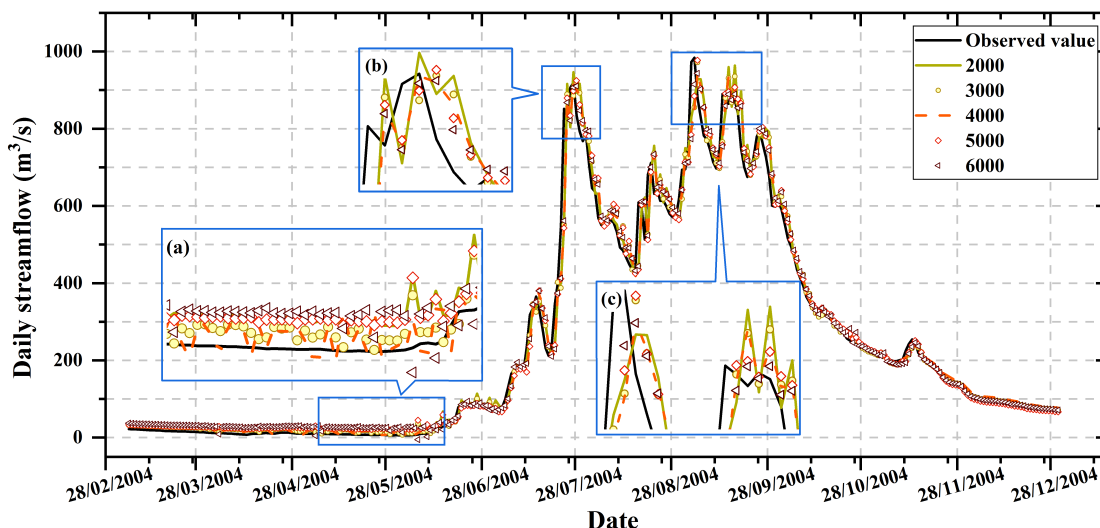


FIGURE 8. Data curves of prediction results of LSTM models under different training set sizes.

MRE, and BIAS values tend to decrease first and then increase. RMSE, MAE, and NSE are steady and are not sensitive to the training dataset size changes. From the point of view of all seven evaluation indices, the training sets with 3000 days data and 4000 days data have the optimal performance for streamflow prediction.

In Fig. 10, the distribution ranges of MAPE, RMSE, and NSE corresponding to the experimental results of each training set are shown. In general, the prediction results of each training set were good. As the training sets increase in size, the three indices showed certain regularity. Fig. 10 (a) shows the change trend of the MAPE. The MAPE of each training set was relatively stable. However, the MAPE floating ranges using 3000-day and 5000-day training sets were large, and the MAPE floating ranges using 2000-day and 6000-day training sets were small. The MAPE floating ranges had the optimal value when the training set size was 4000 days. Fig. 10 (b) shows the change trend of the RMSE and Fig. 10 (c) shows the change trend of the NSE. Both RMSE and NSE had similar variation regularity as MAPE. Hence, the training set with 4000 days of data was the best choice for optimal prediction performance of the developed model.

Fig. 11 shows the change curve of the loss value of the LSTM model under different training set conditions. The developed model converged faster as the size of the training set increased.

In summary, for the LSTM model trained with the historical streamflow data of the Kelantan River, as the size of training set data increases, the prediction performance of the model is optimised to a certain extent, and the model convergence speed is accelerated. However, there are certain limitations in optimising the prediction performance of the LSTM model simply by increasing the training set. The weights of parameters of the developed model will tend to occur overfitting phenomenon by an oversized training set.

C. EVALUATION OF THE TIME INTERVAL BETWEEN THE TRAINING SET AND TESTING SET ON THE PERFORMANCE OF THE STREAMFLOW PREDICTION MODEL BASED ON LSTM

In this section, the effect of the time interval between the training set and testing set on the prediction performance of the LSTM model is investigated. The purpose of the experiments was to test whether the selection of historical data from different time intervals as the training set will affect the prediction performance of the LSTM model. In the experiment, the size of the training set is set to 3000 days, and the size of the testing set is set to 300 days. Historical data in the period from 07/03/2004 to 31/12/2004 was selected as the testing set. Six groups of training sets with different time intervals to the testing set were selected, namely, no interval (that is, the data selected by the training set and the testing set are continuous), 1000 days, 2000 days, 3000 days, 6000 days, and 9000 days. The duration of the training set data was from 20/12/1995 to 06/03/2004. Hence, with a fixed size for the training set and different time intervals between the training set and testing set, experiments were performed to evaluate the performance of the streamflow prediction model based on LSTM.

To further investigate the effect of the time interval between the training set and testing set on the prediction performance of the LSTM model, more analysis has been done, as shown in Fig. 12 and Fig. 13. Fig. 12 is a data curve diagram of the prediction results by the LSTM model with different time intervals. The corresponding prediction result curves of each experimental group are close to the actual observation data, and the results of each group are satisfactory. According to the detailed local information of curves, the experimental data of each group accorded with the multi-peak fluctuations.

Fig. 13 shows the distribution of predicted and observed values for the LSTM model at different time intervals.

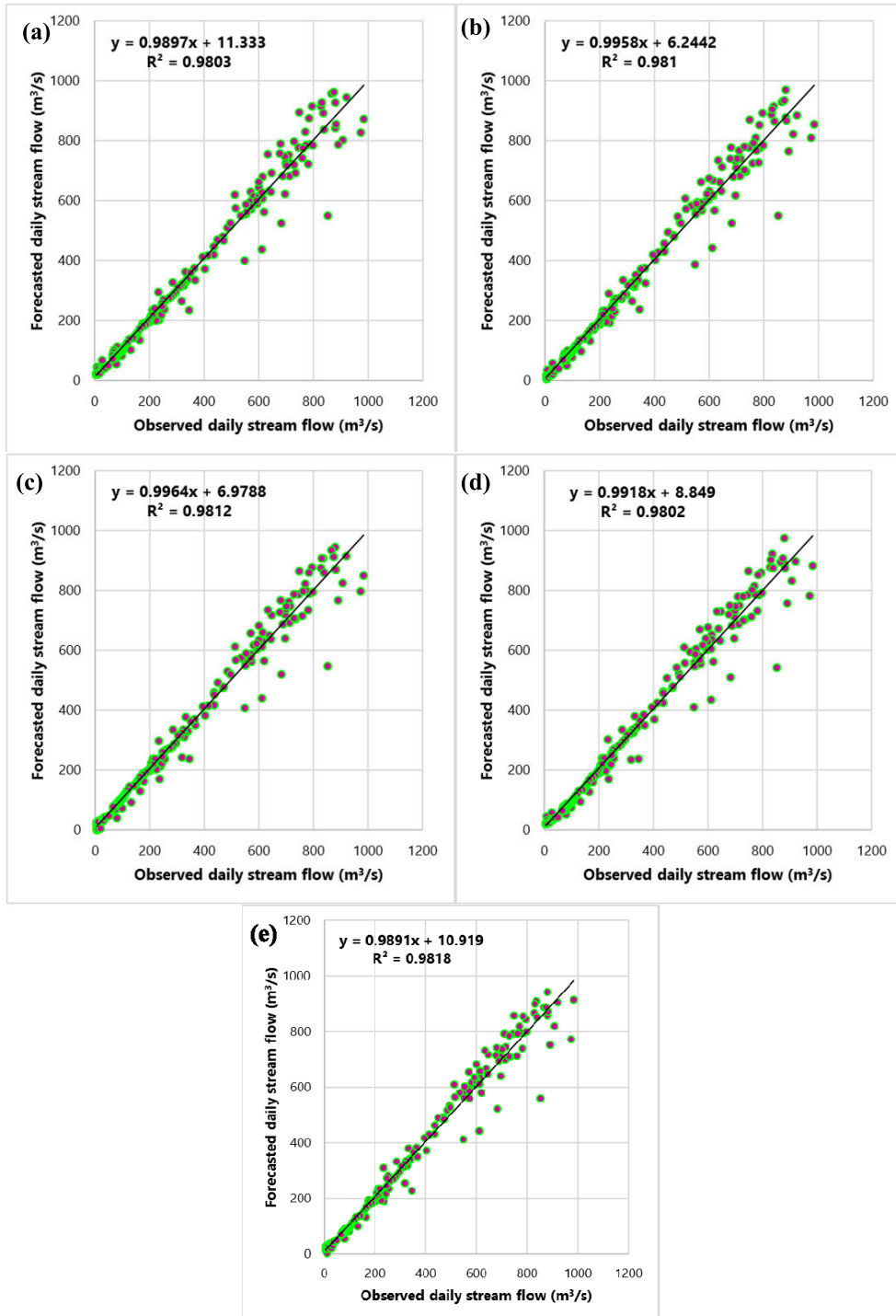


FIGURE 9. Distribution of predicted and observed values of LSTM models under different training set sizes. (a) - (e) correspond to the prediction results of the five training sets of 2000 - 6000 days, respectively.

As shown in the figures, there was no significant change in the overall distribution of the predicted values of each group, and only a few numerical points changed significantly.

Table 7 shows the predictive evaluation indices of the LSTM model with different training set and testing set intervals. MAPE, RMSE, RMSRE, and MAE achieve the

minimum value at the interval of 9000 days and the maximum value at the intervals of 1000 days and 3000 days. Both MRE and BIAS values are negative. The NSE values of each group are small and steady. Considering that the different data characteristics of different training sets, the seven evaluation indices of each group have a small range

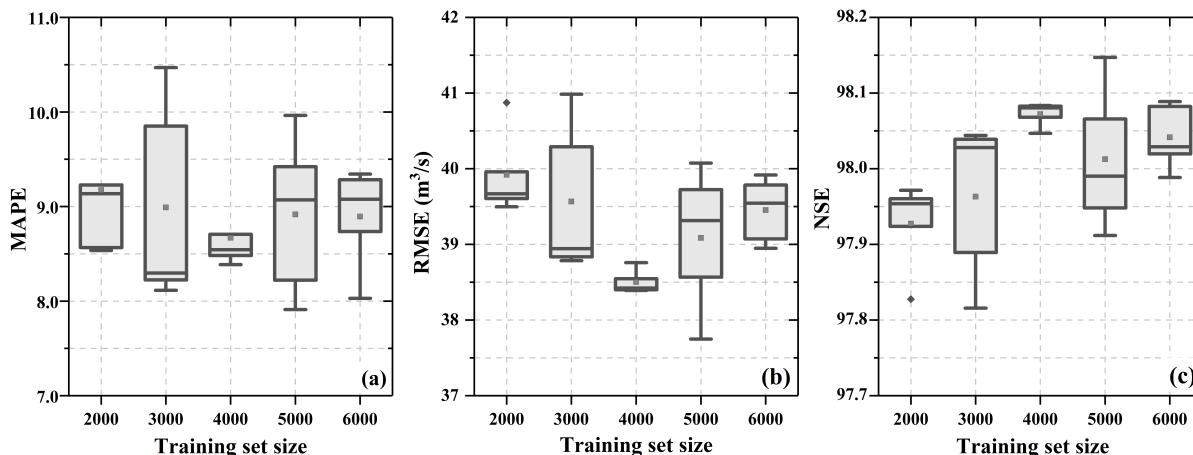


FIGURE 10. Schematic diagrams of the influence of training set size on prediction indicators (a) MAPE, (b) RMSE, (c) NSE.

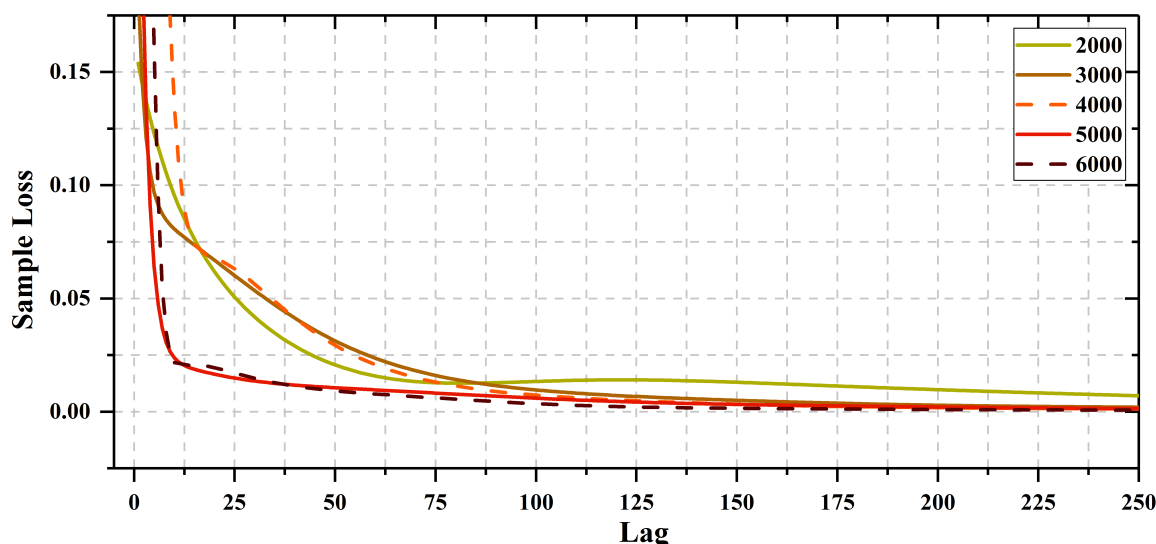


FIGURE 11. Loss value curve of LSTM model prediction process under the condition of different training set sizes.

of fluctuation. The interesting finding is that multiple predictive evaluation indices obtained the optimal value when the time interval between the training set and testing set was set to 9000 days.

Fig. 14 shows the MAPE, RMSE, and NSE of the predictions of the developed model by changing the time intervals between the training set and testing set. As the time interval increased, the value distribution of the three indices showed multi-peak fluctuations. When the interval was set to 1000 days, the three indices achieved poor results. When the interval was set to 6000 days, the three indices achieved better results. When the interval was set to 3000 days, the three indices were at their most stable stage, but the probability of the occurrence of extreme values increased.

The results shown in Table 7 are consistent with the trend of performance changes in Fig. 14. When the time

intervals between the training set and testing set data change, the prediction performance of the LSTM model shows a small multi-peak fluctuation. Extreme points appeared in experiments when the time intervals were set to 1000 days, 2000 days, and 3000 days.

As shown in Fig. 15, the time intervals between the training set and testing set had little impact on the convergence speeds of the model. When the time interval was set to 4000 days, the model convergence speed decreased slightly, and the other groups all have similar convergence speeds.

In summary, according to the experimental results shown in Table 7 and Fig. 12 to Fig. 15, the time interval between the training set and testing set has little impact on the prediction performance of the developed LSTM model trained with the historical streamflow data of the Kelantan River. It is a good feature for the model training process in the case that a small

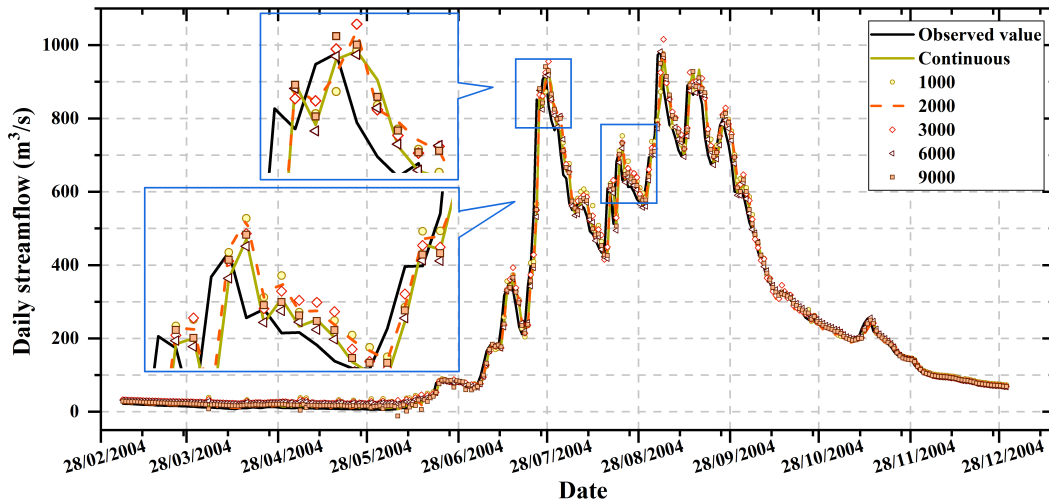


FIGURE 12. Prediction results of the LSTM model under different time intervals.

TABLE 7. Predictive performance evaluation indices of lstm models under different time intervals.

Index	Time Interval					
	Continuous	1000	2000	3000	6000	9000
MAPE	0.08	0.09	0.082	0.09	0.08	0.07
RMSE	38.83	40.19	39.70	40.01	39.54	38.59
RMSRE	0.57	1.01	0.46	1.01	0.90	0.50
MAE	20.56	23.62	20.36	23.00	21.17	19.48
MRE	-0.29	-0.51	-0.17	-0.53	-0.46	-0.21
BIAS	-0.02	-0.04	-0.02	-0.04	-0.02	-0.01
NSE	0.98	0.97	0.97	0.979	0.97	0.98

TABLE 8. Prediction results using different time spans of data.

Index	Time span of the prediction data (days)				
	100	300	500	700	900
MAPE	0.05	0.09	0.09	0.09	0.08
RMSE	42.38	43.72	60.08	55.79	62.70
RMSRE	0.08	5.82	3.88	2.10	2.57
MAE	24.31	15.08	23.79	20.71	24.92
MRE	-0.03	1.72	-0.97	-0.07	-0.24
BIAS	-0.01	0.06	0.00	0.01	0.00
NSE	0.98	0.97	0.96	0.96	0.97

amount of the historical data is missing. Missing historical data will not affect the prediction ability of the developed model.

D. EVALUATION OF THE TIME SPAN OF PREDICTION DATA ON THE PERFORMANCE OF THE STREAMFLOW PREDICTION MODEL BASED ON LSTM

In this section, the effect of the time span of prediction data on the prediction performance of the developed LSTM model is investigated. The experimental settings of this section are as follows: the training set size is set to 3000 days, and the

training set and the testing set have no time interval. The training set is selected from the data set in the period from 23/06/1993 to 08/09/2001. For the testing set, each group took 09/09/2001 as the initial date of the testing set. Then, testing sets with different time spans were selected. In the experiment, the time spans of the following five groups of testing data were set as 100 days, 300 days, 500 days, 700 days, and 1000 days.

Fig. 16 shows the curves of predicted data and the corresponding observed data using the five groups of testing data. As shown in Fig. 16 (a), the predicted data with a 100-day span can fit the observed data well. In Fig. 16 (b),

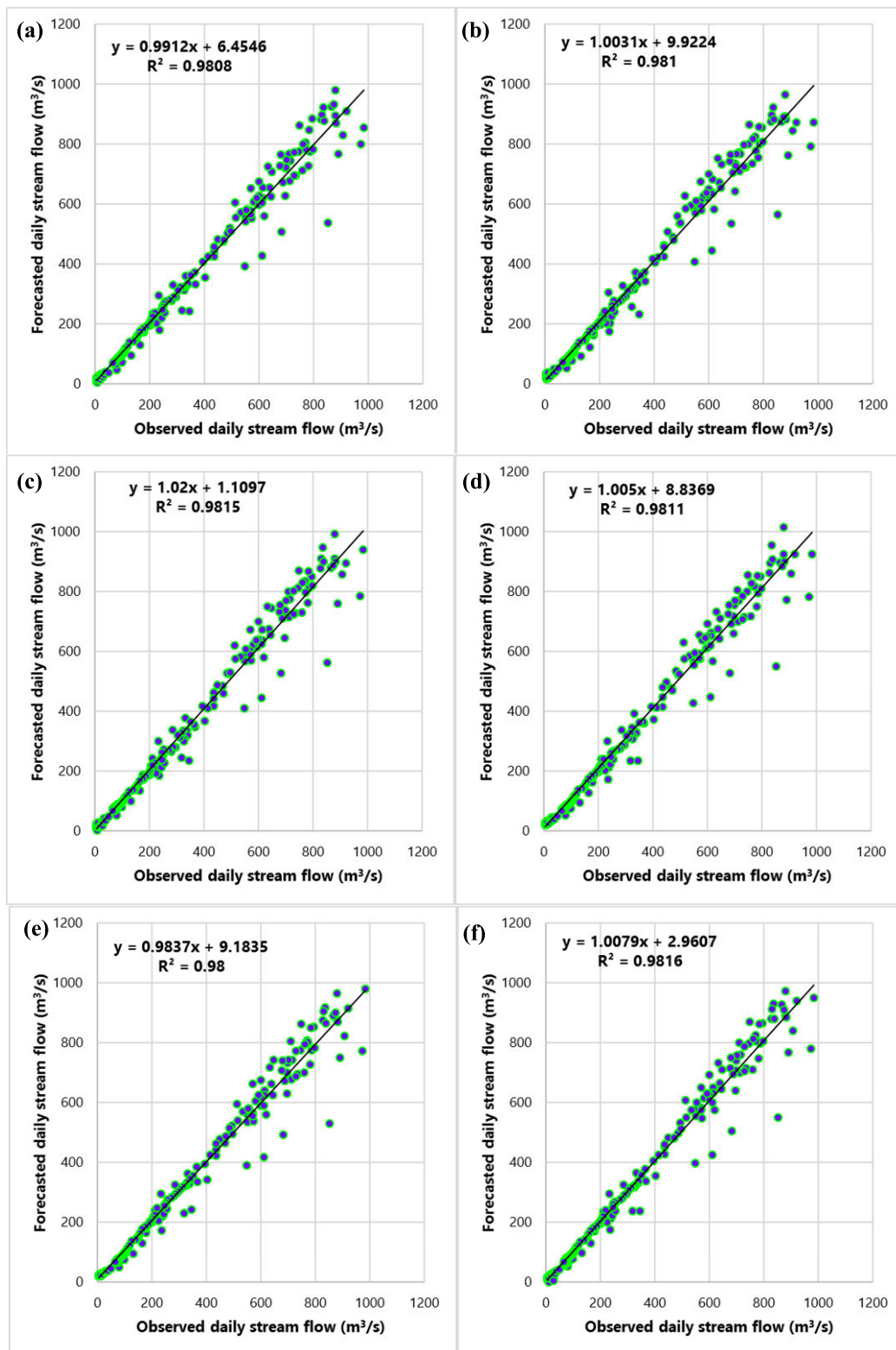


FIGURE 13. Distribution of predicted and observed values of the LSTM model at different time intervals. (a) – (f) correspond to time intervals of 0, 1000, 2000, 3000, 6000, and 9000 days, respectively.

the predicted data in the period from 01/09/2001 to 01/10/2001 has obvious errors. Hence, the prediction accuracy is lower for the 300-day span than for the 100-day

span. Similarly, the two curves cannot fit well in the period from 01/07/2002 to 01/08/2002 in Fig. 16 (c), in the period from 01/07/2002 to 01/08/2002 in Fig. 16 (d),

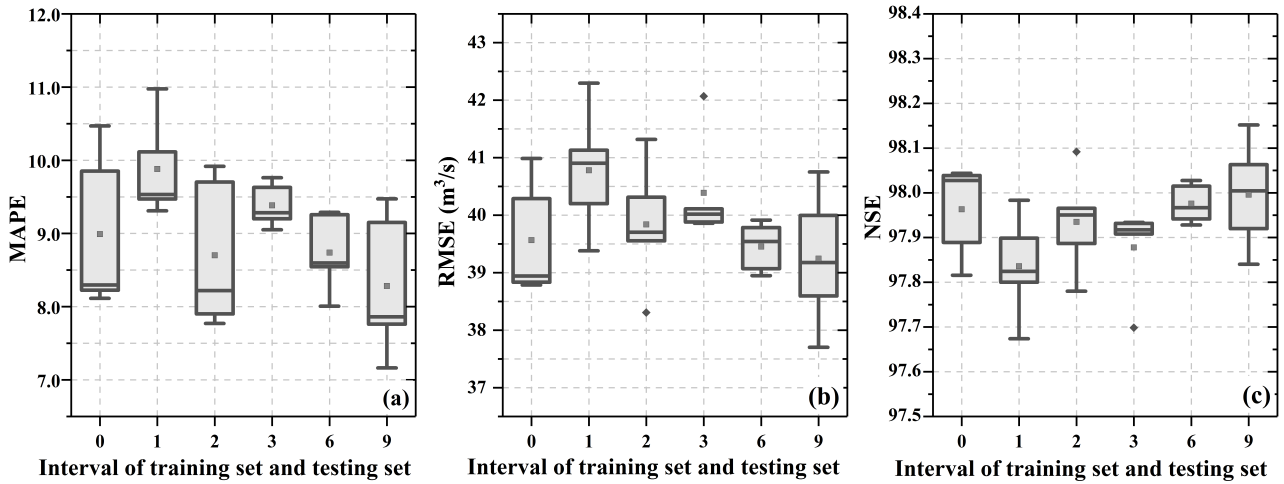


FIGURE 14. Effect of the time interval between the training set and testing set on the prediction index of the LSTM model. (a) MAPE, (b) RMSE, (c) NSE.

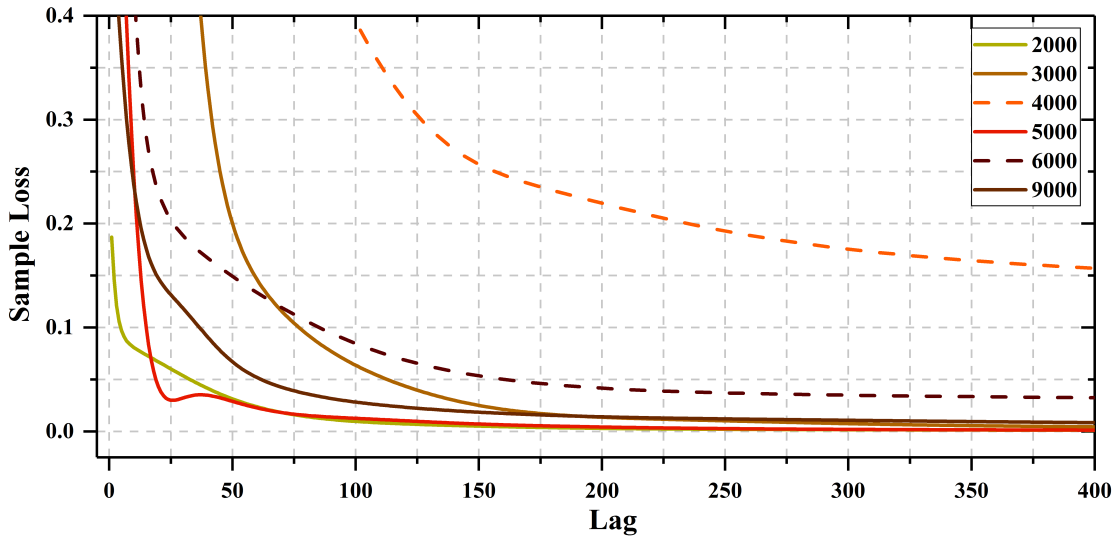


FIGURE 15. Loss value predicted by LSTM model under different time intervals.

and in the period from 01/09/2003 to 01/10/2003 in Fig. 16 (e).

In Fig. 17 (a) to Fig. 17 (e), the diagrams show the distribution of predicted data and observed data with the five time spans, showing that as the time span increased, the data distribution became worse.

Table 8 shows the evaluation index values of the LSTM model prediction results with different prediction data time spans. Prediction results with a 100-day span achieved the best performance in terms of the indices. The other four testing groups have similar MAPE and NSE values. Compared to the prediction results with 500 days, 700 days, and 900 days, the prediction results with 300 days have better RMSE and MAE, but worse RMSRE, MRE, and BIAS.

Fig. 18 is a schematic diagram of the stability of the prediction results of the LSTM model using different time spans for the prediction data. As shown in Fig. 18 (a), the MAPE using 100-day data shows a clear advantage, and the remaining four groups of prediction results are similar. If the time span is more than 500 days, the probability of extreme values will increase. As shown in Fig. 18 (b) and Fig. 18 (c), RMSE and NSE with 100-day data also achieved the best performance. Although the RMSE and NSE with 300-day data have good values, the predicted data have a large fluctuation range.

The results in Fig. 18 and Table 8 show that stability and accuracy of the prediction performance of the LSTM model slightly decreased as the time spans became larger. From the point of view of performance indices, each

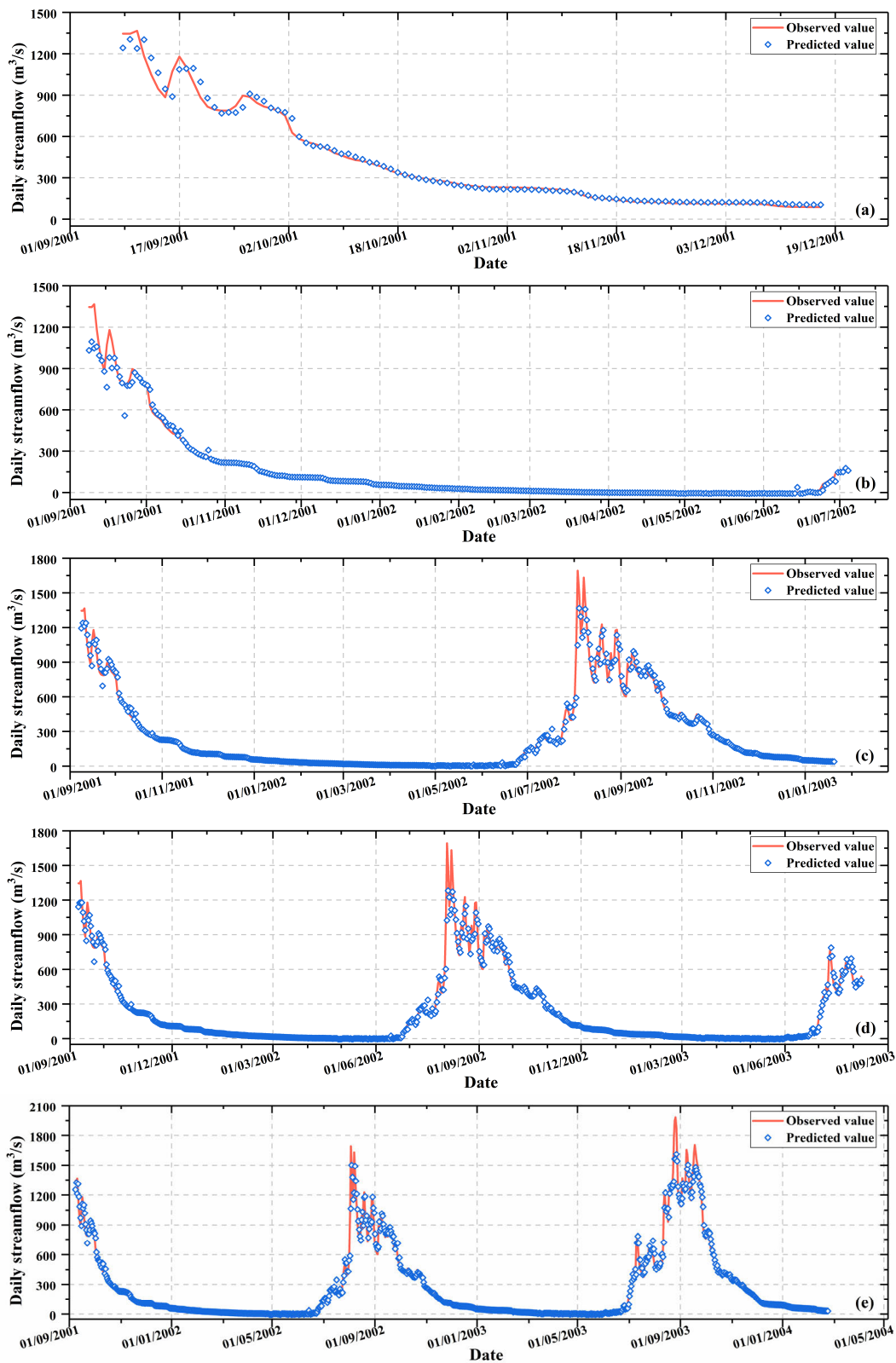


FIGURE 16. LSTM model prediction results using different time spans for the prediction data. (a) - (e) correspond to the prediction results of time spans of 100, 300, 500, 700, and 1000 days, respectively.

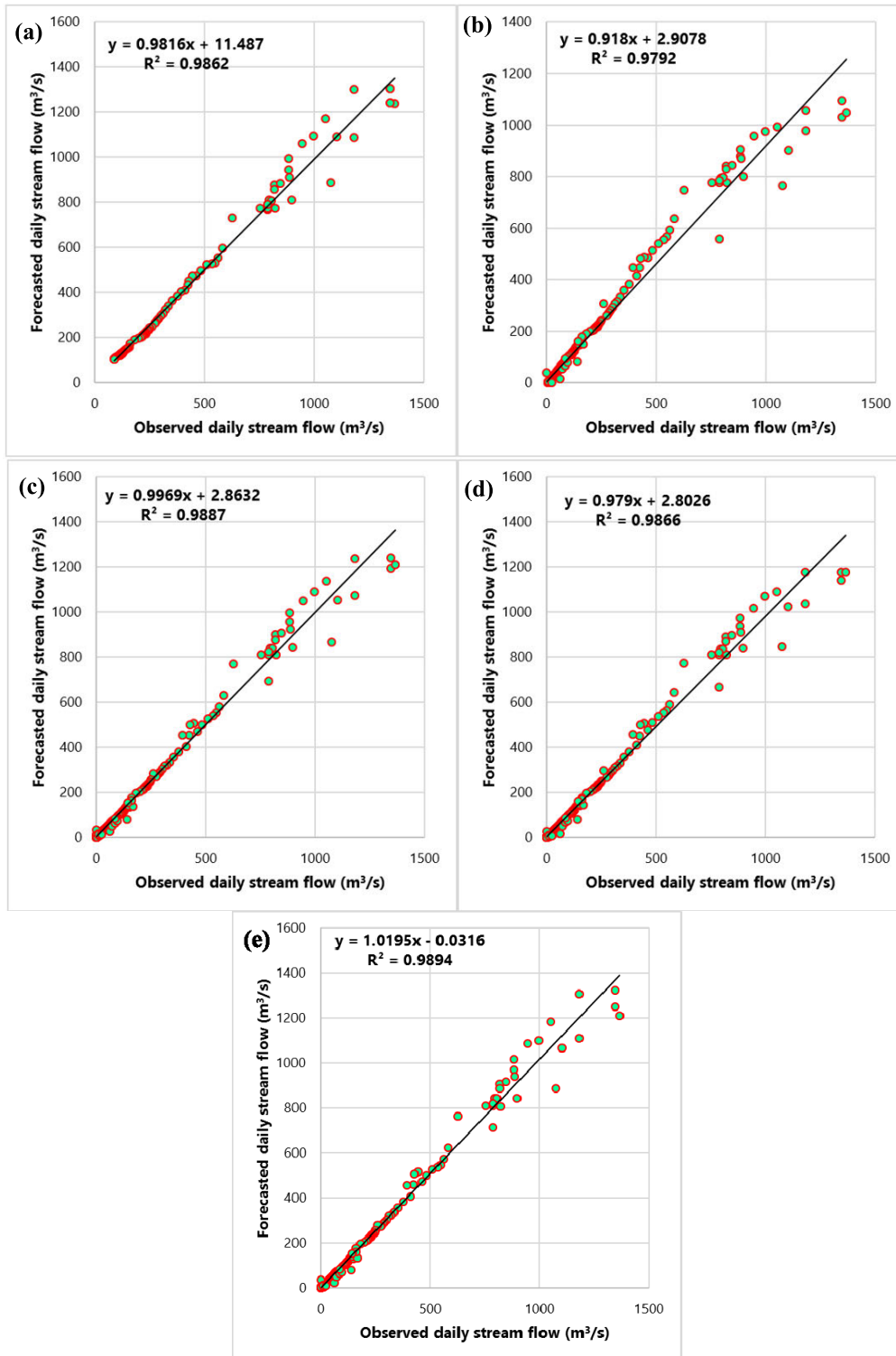


FIGURE 17. Distribution of predicted and observed values of the LSTM model using different time spans of predicted data. (a) - (e) correspond to time spans of 100, 300, 500, 700, and 1000 days, respectively.

experimental group also obtained good prediction results which still meet the requirements for high-accuracy prediction of streamflow.

In summary, for the LSTM model trained with the historical streamflow data of the Kelantan River, the results of this experiment show that the model can complete prediction

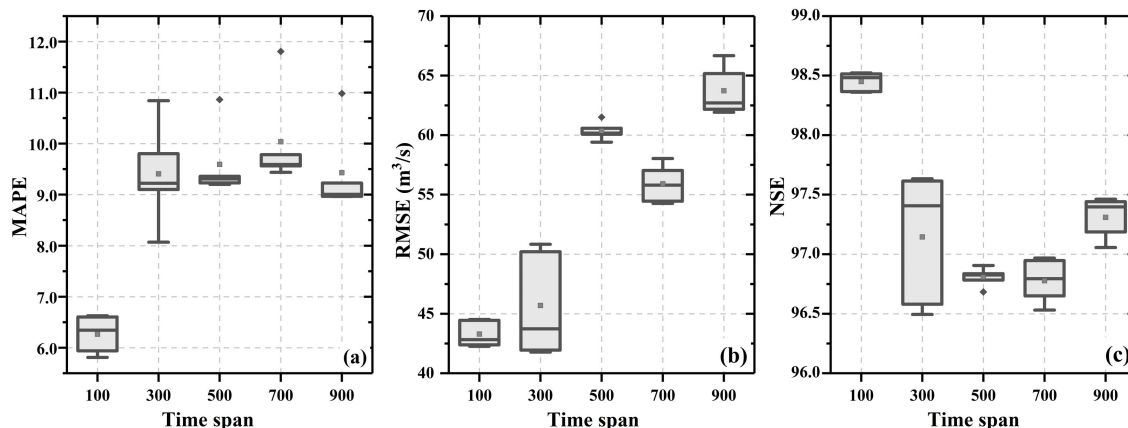


FIGURE 18. Effect of time span and the stability of the prediction results of the LSTM model. (a) MAPE, (b) RMSE, (c) NSE.

tasks with different prediction data time spans, and can ensure the accuracy and stability of prediction results. The LSTM model can be used for streamflow prediction in a variety of time spans.

V. CONCLUSION

In this study, a deep learning data-intelligence model based on LSTM was developed to forecast the streamflow of the Kelantan River. The following four aspects were studied: (i) the feasibility of applying the developed LSTM to streamflow forecasting, and the performance comparison between the developed LSTM model and the classic BP model; (ii) the impact of the amount of data in the training set on the prediction accuracy of the developed LSTM model; (iii) the impact of the time interval between the training set and testing set on the performance of the developed LSTM model; (iv) the impact of the time span of predicted data on the developed LSTM performance. The main conclusions from this study are as follows.

1. The developed LSTM model shows excellent accuracy in capturing the time series of streamflow. Compared with the classic BP model, the developed LSTM model has obvious advantages in prediction accuracy no matter whether it was smooth streamflow in the dry season or rapidly fluctuant streamflow in the rainy season.

2. As the size of the training set data increases, the prediction performance of the developed LSTM model is optimised to a certain extent, and the model convergence speed is accelerated. However, according to the performance indices, there are certain limitations in optimising the prediction performance of the developed LSTM model simply by increasing the training set. The reason is probably that the developed LSTM model is overfitting the data.

3. The time interval between the training set and testing set has little impact on the prediction performance of the developed LSTM model trained with historical streamflow data of the Kelantan River. It is a good feature of the model training process that small amounts of missing historical data will not affect the prediction ability of the developed model.

4. In the task of predicting different time spans, although the stability and accuracy of the prediction performance of the developed LSTM model slightly decreases as the time spans become larger, from the point of view of performance indices, each experimental group can also obtain good prediction results that still meet the requirement for high-accuracy streamflow prediction. Based on this strength, the developed LSTM model can be used for streamflow prediction in a variety of time spans.

With this method, there is still room for improvement. In subsequent research, in the pre-processing phase, a bionic intelligent algorithm can be introduced into the LSTM model. In the forecasting model phase, a combination of the LSTM model and other models is an important consideration for future research.

VI. CONFLICT OF INTEREST

The authors have no conflict of interest to any party.

ACKNOWLEDGMENT

The authors would like to thank the hydrological data provider: Department of Irrigation and Drainage (DID), Malaysia. They would also like to thank Editage for English language editing.

REFERENCES

- [1] Z. M. Yaseen, A. El-Shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, "Artificial intelligence based models for stream-flow forecasting: 2000–2015," *J. Hydrol.*, vol. 530, pp. 829–844, Nov. 2015.
- [2] K.-W. Chau, "Use of meta-heuristic techniques in rainfall-runoff modelling," *Water*, vol. 9, no. 3, p. 186, Mar. 2017.
- [3] K.-H. Ahn and V. Merwade, "Quantifying the relative impact of climate and human activities on streamflow," *J. Hydrol.*, vol. 515, pp. 257–266, Jul. 2014.
- [4] J. W. Kirchner, "Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward," *Water Resour. Res.*, vol. 45, no. 2, Feb. 2009.
- [5] H. Chien, P. J.-F. Yeh, and J. H. Knouft, "Modeling the potential impacts of climate change on streamflow in agricultural watersheds of the Midwestern United States," *J. Hydrol.*, vol. 491, pp. 73–88, May 2013.

- [6] M. Yadav, T. Wagener, and H. Gupta, "Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins," *Adv. Water Resour.*, vol. 30, no. 8, pp. 1756–1774, Aug. 2007.
- [7] H. R. Maier et al., "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions," *Environ. Model. Softw.*, vol. 62, pp. 271–299, Dec. 2014.
- [8] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.
- [9] M. A. Ghorbani, R. Khatibi, V. Karimi, Z. M. Yaseen, and M. Zounemat-Kermani, "Learning from multiple models using artificial intelligence to improve model prediction accuracies: Application to river flows," *Water Resour. Manage.*, vol. 32, no. 13, pp. 4201–4215, Oct. 2018.
- [10] H. R. Maier, Z. Kapelan, J. Kasprzyk, and L. S. Matott, "Thematic issue on evolutionary algorithms in water resources," *Environ. Model. Softw.*, vol. 69, pp. 222–225, Jul. 2015.
- [11] J. M. Szemis, H. R. Maier, and G. C. Dandy, "A framework for using ant colony optimization to schedule environmental flow management alternatives for rivers, wetlands, and floodplains," *Water Resour. Res.*, vol. 48, no. 8, Aug. 2012.
- [12] F. Fahimi, Z. M. Yaseen, and A. El-shafie, "Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review," *Theor. Appl. Climatol.*, vol. 128, nos. 3–4, pp. 875–903, May 2017.
- [13] V. Nourani, A. Hosseini Baghanam, J. Adamowski, and O. Kisi, "Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review," *J. Hydrol.*, vol. 514, pp. 358–377, 2014.
- [14] A. Danandeh Mehr, V. Nourani, E. Kahya, B. Hrnjica, A. M. A. Sattar, and Z. M. Yaseen, "Genetic programming in water resources engineering: A state-of-the-art review," *J. Hydrol.*, vol. 566, pp. 643–667, Nov. 2018.
- [15] Z. M. Yaseen, O. Kisi, and V. Demir, "Enhancing long-term streamflow forecasting and predicting using periodicity data component: Application of artificial intelligence," *Water Resour. Manage.*, vol. 30, no. 12, pp. 4125–4151, Sep. 2016.
- [16] R. Moazenzadeh, B. Mohammadi, S. Shamsirband, and K.-W. Chau, "Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 584–597, Jan. 2018.
- [17] S. Shamsirband, T. Rabczuk, and K.-W. Chau, "A survey of deep learning techniques: Application in wind and solar energy resources," *IEEE Access*, vol. 7, pp. 164650–164666, 2019.
- [18] Z. M. Yaseen, M. Fu, C. Wang, W. H. M. W. Mohtar, R. C. Deo, and A. El-shafie, "Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons," *Water Resour. Manage.*, vol. 32, no. 5, pp. 1883–1899, Mar. 2018.
- [19] G. J. Bowden, G. C. Dandy, and H. R. Maier, "Input determination for neural network models in water resources applications. Part 1—background and methodology," *J. Hydrol.*, vol. 301, nos. 1–4, pp. 75–92, Jan. 2005.
- [20] S. Raghavendra. N and P. C. Deka, "Support vector machine applications in the field of hydrology: A review," *Appl. Soft Comput.*, vol. 19, pp. 372–386, Jun. 2014.
- [21] Ö. Kişi, "Streamflow forecasting using different artificial neural network algorithms," *J. Hydrol. Eng.*, vol. 12, no. 5, pp. 532–539, Sep. 2007.
- [22] W.-C. Wang, K.-W. Chau, D.-M. Xu, L. Qiu, and C.-C. Liu, "The annual maximum flood peak discharge forecasting using hermite projection pursuit regression with SSO and LS method," *Water Resour. Manage.*, vol. 31, no. 1, pp. 461–477, Jan. 2017.
- [23] R. Taormina and K.-W. Chau, "ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 429–440, Oct. 2015.
- [24] K. Rasouli, W. W. Hsieh, and A. J. Cannon, "Daily streamflow forecasting by machine learning methods with weather and climate inputs," *J. Hydrol.*, vols. 414–415, pp. 284–293, Jan. 2012.
- [25] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, Oct. 2018.
- [26] H. Zamani Sabzi, J. P. King, and S. Abudu, "Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: A case study," *Expert Syst. Appl.*, vol. 83, pp. 145–163, Oct. 2017.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1–32, 1997.
- [28] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw.*, 1999, pp. 850–855.
- [29] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transport Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.
- [30] F. Fotovatikhah, M. Herrera, S. Shamsirband, K.-W. Chau, S. F. Ardabili, and M. J. Piran, "Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 411–437, Jan. 2018.
- [31] S. Gope, S. Sarkar, P. Mitra, and S. Ghosh, "Early prediction of extreme rainfall events: A deep learning approach," in *Proc. Ind. Conf. Data Mining*, 2016, pp. 154–167.
- [32] B. D. Bowes, J. M. Sadler, M. M. Morsy, M. Behl, and J. L. Goodall, "Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks," *Water*, vol. 11, no. 5, p. 1098, May 2019.
- [33] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall-runoff simulation," *Water*, vol. 10, no. 11, p. 1543, Oct. 2018.
- [34] F. Kratzert, D. Klotz, C. Brenner, and K. Schulz, "Rainfall-runoff modelling using long short-term memory (LSTM) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022, 2018.
- [35] S. H. I. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [36] D. Klotz, F. Kratzert, M. Herrnegger, S. Hochreiter, and G. Klambauer, "Towards the quantification of uncertainty for deep learning based rainfall-runoff models," *Geophys. Res. Abstr.*, vol. 21, Jan. 2019.
- [37] S. Aswin, P. Geetha, and R. Vinayakumar, "Deep learning models for the prediction of rainfall," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2018, pp. 657–661.
- [38] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018.
- [39] L. Yunpeng, H. Di, B. Junpeng, and Q. Yong, "Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network," in *Proc. 14th Web Inf. Syst. Appl. Conf. (WISA)*, Nov. 2017, pp. 305–310.
- [40] D. Zhang, N. Martinez, G. Lindholm, and H. Ratnaweera, "Manage sewer in-line storage control using hydraulic model and recurrent neural network," *Water Resour. Manage.*, vol. 32, no. 6, pp. 2079–2098, Apr. 2018.
- [41] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas," *J. Hydrol.*, vol. 561, pp. 918–929, Jun. 2018.
- [42] X. Yuan, C. Chen, X. Lei, Y. Yuan, and R. Muhammad Adnan, "Monthly runoff forecasting based on LSTM-ALO model," *Stoch Environ. Res. Risk Assess.*, vol. 32, no. 8, pp. 2199–2212, Aug. 2018.
- [43] I. R. Widiyarsi, L. E. Nugoho, and R. Efendi, "Context-based hydrology time series data for a flood prediction model using LSTM," in *Proc. 5th Int. Conf. Inf. Technol., Comput., Elect. Eng. (ICITACEE)*, Sep. 2018, pp. 385–390.
- [44] C. Liang, H. Li, M. Lei, and A. Q. Du, "Dongting lake water level forecast and its relationship with the three gorges dam based on a long short-term memory network," *Water*, vol. 10, no. 10, p. 1389, Oct. 2018.
- [45] D. Zhang, J. Lin, Q. Peng, D. Wang, T. Yang, S. Sorooshian, X. Liu, and J. Zhuang, "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm," *J. Hydrol.*, vol. 565, pp. 720–736, Oct. 2018.
- [46] Y. Tian, Y.-P. Xu, Z. Yang, G. Wang, and Q. Zhu, "Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting," *Water*, vol. 10, no. 11, p. 1655, Nov. 2018.
- [47] A. Huang, B. Vega-Westhoff, and R. L. Sriver, "Analyzing el Niño–Southern oscillation predictability using long-short-term-memory models," *Earth Space Sci.*, vol. 6, no. 2, pp. 212–221, Feb. 2019.
- [48] C. Shen, "A transdisciplinary review of deep learning research and its relevance for water resources scientists," *Water Resour. Res.*, vol. 54, no. 11, pp. 8558–8593, Nov. 2018.

- [49] H. Tao, L. Diop, A. Bodian, K. Djaman, P. M. Ndiaye, and Z. M. Yaseen, "Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso," *Agricult. Water Manage.*, vol. 208, pp. 140–151, Sep. 2018.
- [50] Z. M. Yaseen, S. M. Awadh, A. Sharafati, and S. Shahid, "Complementary data-intelligence model for river flow simulation," *J. Hydrol.*, vol. 567, pp. 180–190, Dec. 2018.



MINGLEI FU is currently an Associate Professor with the Zhejiang University of Technology. He has published over 50 articles and held 15 patents. He is also the coauthor of four monographs. His research interests include signal processing, machine learning, and image processing.



TINGCHAO FAN is currently pursuing the master's degree with the Zhejiang University of Technology. Her research interests include deep learning and data analysis. She has been participating in Baidu's artificial intelligence creative competition.



ZI'ANG DING is currently pursuing the master's degree with the Zhejiang University of Technology. He is also an Algorithmic Engineer working for an Image Processing Company. His research interests include deep learning and data analysis.



SINAN Q. SALIH received the B.Sc. degree in information systems from the University of Anbar, Anbar, Iraq, in 2010. He completed his M.Sc. degree in computer sciences from Universiti Tenaga Nasional (UNITEN), Malaysia, in 2012. He received a Ph.D. degree in soft modelling and intelligent systems from Universiti Malaysia Pahang (UMP). His current research interests include optimization algorithms, nature-inspired metaheuristics, machine learning, and feature selection problem for real world problems.



NADHIR AL-ANSARI received the B.Sc. and M.Sc. degrees from the University of Baghdad in 1968 and 1972 respectively, and the Ph.D. degree in water resources engineering from Dundee University, in 1976. From 1976 to 1995, he was with Baghdad University. From 1995 to 2007, he was with Al al-Bayt University, Jordan. He is currently a Professor with the Department of Civil, Environmental and Natural Resources Engineering, Lulea Technical University, Sweden.

He has served several academic administrative posts (the Dean, the Head of Department). His publications include more than 424 articles in international/national journals, chapters in books and 13 books. He executed more than 60 major research projects in Iraq, Jordan, and U.K. He has one patent on physical methods for the separation of iron oxides. He has supervised more than 66 postgraduate students at Iraq, Jordan, and U.K. Australia universities. His research interests are mainly in geology, water resources, and environment. He is also a member of several scientific societies, e.g., International Association of Hydrological Sciences, Chartered Institution of Water and Environment Management, Network of Iraqi Scientists Abroad, the Founder and the President of the Iraqi Scientific Society for Water Resources, and so on. He is also a member of the editorial board of ten international journals. He received several scientific and educational awards, among them are the British Council on its 70th Anniversary awarded him top five scientists in Cultural Relations.



ZAHER MUNDHER YASEEN received the master's and Ph.D. degrees from the National University of Malaysia (UKM), Malaysia, in 2017. He is currently a Senior Lecturer and a Senior Researcher in the field of civil engineering at Ton Duc Thang University. He is major in surface hydrology, water resources engineering, hydrological processes modeling, environmental engineering, and climate. In addition, he has an excellent expertise in machine learning and advanced data

analytics. He has published over 100 articles in international journals with a Google Scholar H-Index of 23, and a total of 1588 citations.

...