Daily Streamflow Prediction and Uncertainty Using a Long Short-Term Memory (LSTM) Network Coupled with Bootstrap

Zhuoqi Wang[1] · **Yuan Si**[2] · **Haibo Chu**[1]

Abstract

Long short-term memory (LSTM) models with excellent data mining ability have great potential in streamflow prediction. The parameters and structure of the LSTM model, which should be completely determined in an explanatory manner based on the observed datasets, have a significant impact on the model performance. Due to the limitations and uncertainty in the observed datasets, the uncertainty in daily streamflow prediction needs to be quantitatively assessed. In this work, LSTM models are used to predict daily streamflow for two stations in the Mississippi River basin in Iowa, USA, and the performance of LSTM models with different parameters and inputs is investigated to demonstrate the process of determining the optimal parameters. The results show that the LSTM model with optimized parameters and an optimized structure performs the best among the four data-driven models, and the model with selected predictors (inputs) performs better than that without selected predictors. Moreover, the bootstrap method is employed to generate different realizations of the observed datasets that are used for developing LSTM models; thus, the prediction streamflow values from different LSTM models are finally used for uncertainty analysis in daily streamflow prediction. LSTM can be a promising tool for daily streamflow prediction. When LSTM is combined with Bootstrap method, reliable uncertainty quantification of streamflow prediction is also provided.

Keywords Streamflow prediction · Long short-term memory network · Bootstrap · Uncertainty

1 Introduction

Accurate daily streamflow prediction is an important condition for ensuring reasonable water resource planning and management (Eum and Kim 2010; Alemu et al. 2011). Streamflow prediction is necessary for hydropower generation, flood prediction and water

✉ Haibo Chu
  chuhaibo@bjut.edu.cn

[1] College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China

[2] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

supply management (Ni et al. 2020; Sammen et al. 2021). However, it is difficult to fore-cast daily streamflow because streamflow data are nonstationary and nonlinear and display great temporal and spatial variability (Nourani and Komasi 2013; Chu et al. 2018).

Many prediction models have been developed, and they can be divided into two cat-egories: process-based hydrological models (PHMs) and data-driven machine learn-ing models (DMLs) (Kan et al. 2015; Kim et al. 2021). PHMs were developed based on the physical knowledge of the interrelationships between various hydrological processes in a basin; however, high-quality spatial–temporal data and a large amount of running time are required when using these methods, which may limit their application. Different from PHMs, DMLs can be utilized to simulate streamflow generation and forecast daily streamflow by extracting the evolution characteristics of the streamflow generation process from historical observation data. These methods have great advantages in many aspects, including efficiency, accuracy and flexibility (Pandey and Srinivas 2015). From shallow learning to deep learning, DMLs, such as support vector regression (SVR) (Parisouj et al. 2020), general regression neural networks (GRNNs), and long short-term memory (LSTM) (Sahoo et al. 2019; Cho and Kim 2022), have attracted considerable attention in terms of streamflow prediction in hydrological applications.

A long short-term memory (LSTM) network, which is modified from the RNN model, can be utilized to process hydrological data with long-term dependence well. It has been applied to many fields and shows great potential in streamflow prediction. For instance, Kratzert et al. (2018) explored the ability of using an LSTM network for rainfall-streamflow simulation in a large number of basin experiments, and the experiments showed that LSTM has advantages in processing long-time series data. Hu et al. (2018) compared the ability of LSTM and ANN in rainfall-streamflow prediction, and the results showed that the perfor-mance of the LSTM model was better than that of the ANN model. Zhang et al. (2019) used an LSTM network to forecast sewage flow, and the results showed that the LSTM model has important application value in predicting sewage flow. However, the influence of the struc-ture and parameters of the LSTM model on the performance of the model still needs to be studied.

Uncertainty affects the reliability of streamflow prediction to a certain extent and risk may be introduced in some applications, such as real-time reservoir operation and flood defence (Chen et al. 2016; Xu et al. 2021). Input data uncertainty is one of the most sig-nificant uncertainty sources, and it also has an impact on the model structure and param-eters (Engeland et al. 2016); therefore, input data uncertainty needs to be studied further. Dehghani et al. (2014) investigated uncertainties in discharge and drought indices using a Monte Carlo simulation approach. Kasiviswanathan et al. (2016) coupled an artificial neural network (ANN) and a bootstrap method for streamflow prediction and uncertainty in Canada. The bootstrap method, which is simple and practical, can be used to reduce data uncertainty and evaluate uncertainty (Zhang et al.2018). Therefore, the combination of an LSTM network with a bootstrap method needs to be explored for the assessment of predic-tion uncertainty.

The objectives of this study are as follows: (1) investigate the potential of LSTM models for daily streamflow prediction and compare the performance of these models with other mod-els, (2) analyse the effect of different parameters and predictors on the model performance, and (3) evaluate the prediction uncertainty using LSTM coupled with a bootstrap method. In this paper, two stations in the Mississippi River basin in Iowa, USA, were used as case stud-ies. We first explored the applicability of LSTM models for daily streamflow prediction at these two stations, discussed the influence of the parameters on the model performance, and compared the performance with other models, including the multiple linear regression (MLR),

GRNN and SVR models. Then, four different input combinations were used as the inputs of the LSTM models to investigate the influence of different inputs on the model performance. Finally, we combined the LSTM model with the bootstrap method to evaluate the prediction uncertainty.

2 Method

2.1 Long Short-Term Memory (LSTM)

LSTM was originally proposed as a special recurrent neural network (RNN) (Xiang et al. 2020), and its long-short memory structure was designed to overcome the gradient disappearance and gradient explosion problems in RNNs (Rahimzad et al. 2021). LSTM has more complex memory units and can retain long-term time sequence information. Therefore, the LSTM model has outstanding performance in the prediction of time series and has been a research hotspot in the field of machine learning in recent years.

The LSTM cell is controlled and protected by three gates: the input gate, forget gate and output gate (Cheng et al. 2021). The information flow in LSTM units can be described by the following three steps: first, the information to be discarded from the cell state is decided. This decision is made through the forget gate. The gate will read h_{t-1} and x_t and output a value between 0 and 1 for each cell state c_{t-1}, where 1 means "completely retain", and 0 means "completely discard".

$$F_t = \sigma(W_f g[h_{t-1}, x_t + b_{if}]) \tag{1}$$

The next step is to determine how much new information should be added to the cell state. This requires two steps: first, a sigmoid layer called the "input gate layer" determines which information should be retained; a tanh layer generates a vector, which is the optional content to retain, $\widetilde{C}_t$. In the next step, the two parts are linked to retain the state of the cell.

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \widetilde{C}_t \tag{4}$$

Finally, it is necessary to resolve the value to output. This output will be based on the cell state, but it is also filtered. A sigmoid layer determines which part of the cell state will be explored. Then, the cell state is processed by tanh (to obtain a value between -1 and 1) and multiplied by the output of the sigmoid gate, and finally, only the output that we determined is output.

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = O_t * tanh(C_t) \tag{6}$$

where F_t represents the forget gate; I_t represents the input gate; $\widetilde{C}_t$ is another candidate gate created through the tanh function to compute the cell state of the current input; C_t represents the updated state of the new cell; O_t represents the output gate; h_t represents the final

output calculated by the tanh function; σ represents the sigmoid function; h_{t-1} represents the output of the previous cell; x_t represents the input of the current cell; tanh represents the hyperbolic cosine function; W_f, W_i, W_c, and W_o represent the weight parameter matrix between the hidden layer forget gate, input gate, candidate part and output gate and the previous layer of neurons at the current time step, respectively; and b_f, b_i, b_c, and b_o represent the biases of the forget gate, input gate, candidate part and output gate, respectively.

2.2 Bootstrap Method

The bootstrap method can be used to evaluate the uncertainty by using resampling technology (Saraiva et al. 2021). The bootstrap method uses computer simulations to replace complex and imprecise approximations of biases, variances, and other statistics (Zhang et al. 2014). When using this method, artificial assumptions about the unknown distribution are not required as the unknown distribution is obtained by resampling the original data (Chu et al. 2021). Therefore, the bootstrap method is a statistical inference method for medium-sized independent samples with equal distributions, and it can be utilized to improve the inference under the condition of insufficient statistical information (Gopala et al. 2019). More detailed information about the bootstrap method can be found in Belayneh et al. (2016).

2.3 Performance Measures

The coefficient of determination (R^2), root mean square error (RMSE), probability of detection (POD) and false acceptance rate (FAR) are used to qualitatively evaluate the performance of the models. The specific formulas for the indicators are as follows:

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n} (\widehat{y}_i - y_i)^2}{\sum\limits_{i=1}^{n} (\widehat{y}_i - \overline{y}_i)^2} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum\limits_{i=1}^{n} (\widehat{y}_i - y_i)^2} \tag{8}$$

$$POD = \frac{A}{A + C} \tag{9}$$

$$FAR = \frac{B}{B + D} \tag{10}$$

where y_i is the observed value of the daily streamflow; $\widehat{y}_i$ is the predicted value of the daily streamflow; $\overline{y}_i$ is the mean value of the daily streamflow; and A represents the number of days that the observed and predicted streamflows are both considered high-flow (i.e., streamflow values that are greater than the 75th percentile of the streamflow values in a station are considered high-flow). B represents the number of days in which the observed streamflow is in the high-flow zone and the predicted streamflow is not in the high-flow

zone. C represents the number of days in which the observed streamflow is not in the high-flow zone and the predicted streamflow is in the high-flow zone. D represents the number of days that neither the observed streamflow nor the predicted streamflow are in the high-flow zone. The range of R^2, POD and FAR is between 0 and 1. The closer to 1 the R^2 and POD are and the closer to 0 the RMSE and FAR are, the better the performance of the model.

3 Case Study

The Mississippi River is the fourth longest river in the world and is located in south central North America. The main stream originates from Lake Itasca, which is a very small lake that is 501 m above sea level in northwestern Minnesota, west of Lake Superior, and flows southward through the central plains to the Gulf of Mexico. The average discharge into the Gulf of Mexico is approximately 17,000 m³/s. The Mississippi River is 3,950 kms long. Two stations in the Mississippi River basin in Iowa, as shown in Fig. 1, were selected to explore the applicability of LSTM for streamflow prediction. Iowa has a temperate continental climate. The average temperature in January in this region is –9 °C in the northwest and –4 °C in the southeast. In the case of a strong storm, the temperature can be reduced to –34 °C. The average daytime temperature in July is 34 °C, which is very hot. The average annual precipitation in the northwest is 711 mm and 864 mm in the south. Most of this precipitation occurs during summer. There is less snow in winter than in the eastern and northern states. Station 1 (ID: 05458000) is located at the Little Cedar River near Ionia, IA, and data was collected from 1954/10 to 2021/7. The Cedar River Basin has a watershed area of approximately 20,280 km², 87 percent of which is located in Iowa. Station 2 (ID: 05412500) is located at the Turkey River near Garber, IA, and data was collected from

Fig. 1 Maps of the study area and hydrological stations
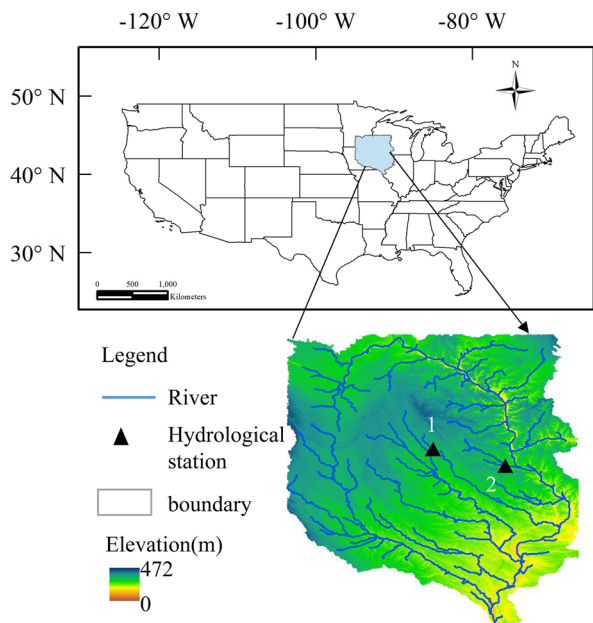
1977/7 to 2021/7. The Turkey River is 246 km long and occupies a catchment of 4384 km². The average precipitation in the Turkey River watershed is 915 mm, of which the precipitation in spring and summer accounts for approximately 70%. For streamflow data, the maximum and minimum values at Station 1 were 605.98 m³/s and 0.08 m³/s, respectively, with a difference of 605.9 m³/s. The maximum and minimum values at Station 2 were 1478.14 m³/s and 1.59 m³/s, respectively, with a difference of 1476.55 m³/s. The difference in the maximum value and variance between the two stations were 872.15 m³/s and 43.73 m³/s, respectively. For precipitation data, the maximum and minimum values at Station 1 were 180.30 m³/s and 0 m³/s, respectively, with a difference of 180.30 m³/s. The maximum and minimum values at Station 2 were 158.20 m³/s and 0 m³/s, with a difference of 158.20 m³/s. The maximum value of the two stations differs by only 22.1 m³/s, the minimum value is the same, and the variance differs by 0.40 m³/s.

4 Results and Discussion

4.1 Comparison of Different Models

In this study, the data were delimited as a training set for model calibration and a validation set for performance evaluation. The LSTM models were trained using data from 1954/10–2015/6 and validated using data from 2015/7–2021/7 for Station 1. The LSTM models were also trained using data from 1977/7–2015/6 and validated using data from 2015/7–2021/7 for Station 2.

4.1.1 Influence of the LSTM Parameters on the Model Performance

The LSTM parameters, including the number of neurons and the period (i.e., epochs, which are a single training iteration of all batches propagating forwards and backwards), were selected to analyse their influence on model performance. As shown in Fig. 2, at Station 1, when the period is 20, a gradual increase in the performance measure R^2 is observed as the number of neurons increases, and then, it tends to be flat. When the number of neurons is 20, there is a small fluctuation in the R^2 of the other periods, except when the period equals 20. The closer to 1 the R^2 is, the better the performance of the LSTM model; then, the corresponding parameters can be utilized as the final parameters of the LSTM model. The optimization parameters for Station 1 are 80 for the number of neurons and 60 for the
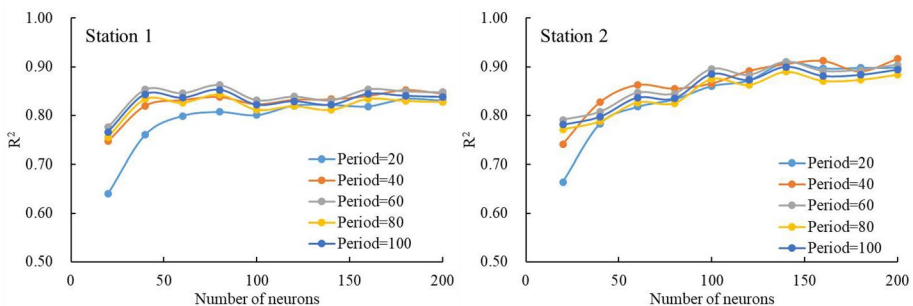
Fig. 2 Sensitivity analysis of different LSTM model parameters on the forecasting performance

period, and the corresponding R^2 is 0.85. The parameters with the highest R^2 of 0.92 for Station 2 are 200 for the number of neurons and 40 for the period. It can be clearly seen that the number of neurons has a great influence on the accuracy of the LSTM model, whereas the influence of the period is relatively small.

### 4.1.2 Influence of Different Models on the Model Performance

In this study, the MLR, GRNN, SVR and LSTM models were used to compare the model performance in terms of RMSE and R^2, and two measures, POD and FAR, were used to assess the high-flow performance. Table 1 shows the performance of four models at the two stations during the calibration and validation periods. As shown in Table 1, for the calibration period at Station 1 using LSTM, values of 10.07 and 0.87 were obtained for the RMSE and R^2, which were the minimum and maximum values among the four models, respectively. For the validation period, values of 9.36 and 0.85 were obtained for the RMSE and R^2, respectively. Compared with the other three models, when using LSTM, the RMSE decreased by 3.82, 2.97, 2.11, respectively, and the R^2 increased by 0.23, 0.13, and 0.1, respectively. At Station 2, for the validation period using LSTM, values of 28.05 and 0.92 were obtained for the RMSE and R^2, respectively, which were also the minimum and maximum values among the four models, and compared with the other three models, the R^2 increased by 0.31, 0.14 and 0.1 or 33.69%, 15.21% and 10.86%, respectively, compared with the other models.

As shown in Fig. 3, the fitting performance of the predicted value and observed value of LSTM at Station 1 is better than that of the other models during the calibration period. The overall performance of GRNN at Station 2 is similar to that of LSTM, but the performance of LSTM is obviously better than GRNN at low flows. A similar result can be found in Fig. 4. In general, LSTM has the best model performance among the four models.

In Table 2, for the calibration period at Station 1, a value 0.02 was achieved for the FAR using both MLR and LSTM. However, when using LSTM, a value of 0.98 was obtained for the POD, which is 0.24 higher than that of MLR, and is the largest value for this metric among the four models. For the validation period, although the FAR of MLR was the lowest among the four models, the POD value of 0.78 was also the lowest among the four models. The POD value achieved when using LSTM was 0.99, which is close to 1, while the FAR value achieved when using LSTM is only 0.04 higher than that MLR. At the same time, the same POD value was achieved using both SVR and LSTM. However, when using LSTM, a FAR value of 0.07 smaller than that of SVR was achieved. For Station 2, the POD value was 0.95 for LSTM, followed by 0.91 for GRNN, 0.64 for SVR and 0.60 for MLR for the calibration period, and a FAR value of 0.1 was not achieved using any of the four models. The same performance measure trend were also obtained for the validation
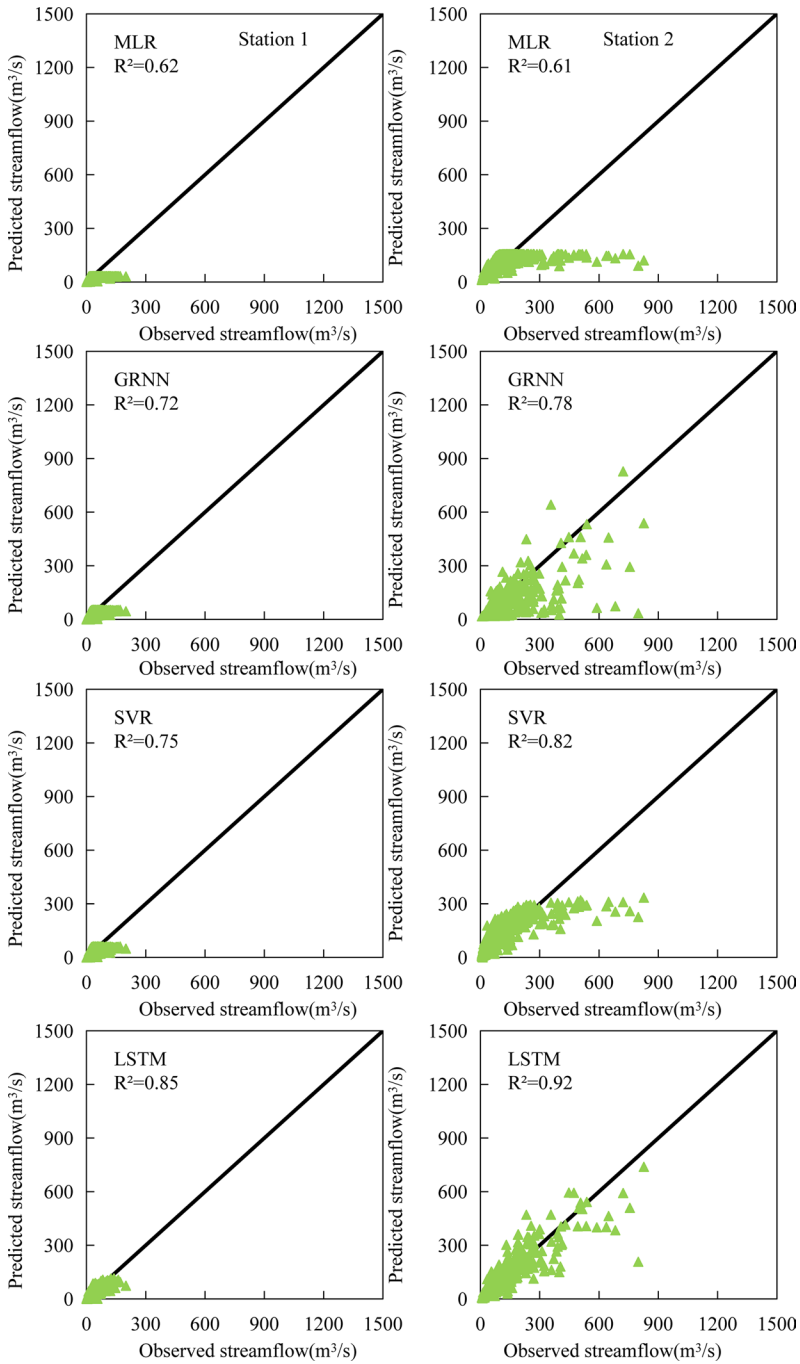
Table 1 The performance of MLR, GRNN, SVR and LSTM for different stations during calibration and validation period

Model	Station 1				Station 2			
	calibration		validation		calibration		validation	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
MLR	12.51	0.64	13.18	0.62	43.66	0.68	57.76	0.61
GRNN	11.55	0.74	12.33	0.72	28.69	0.87	47.05	0.78
SVR	11.45	0.75	11.47	0.75	34.70	0.82	41.80	0.82
LSTM	10.07	0.87	9.36	0.85	22.26	0.92	28.05	0.92
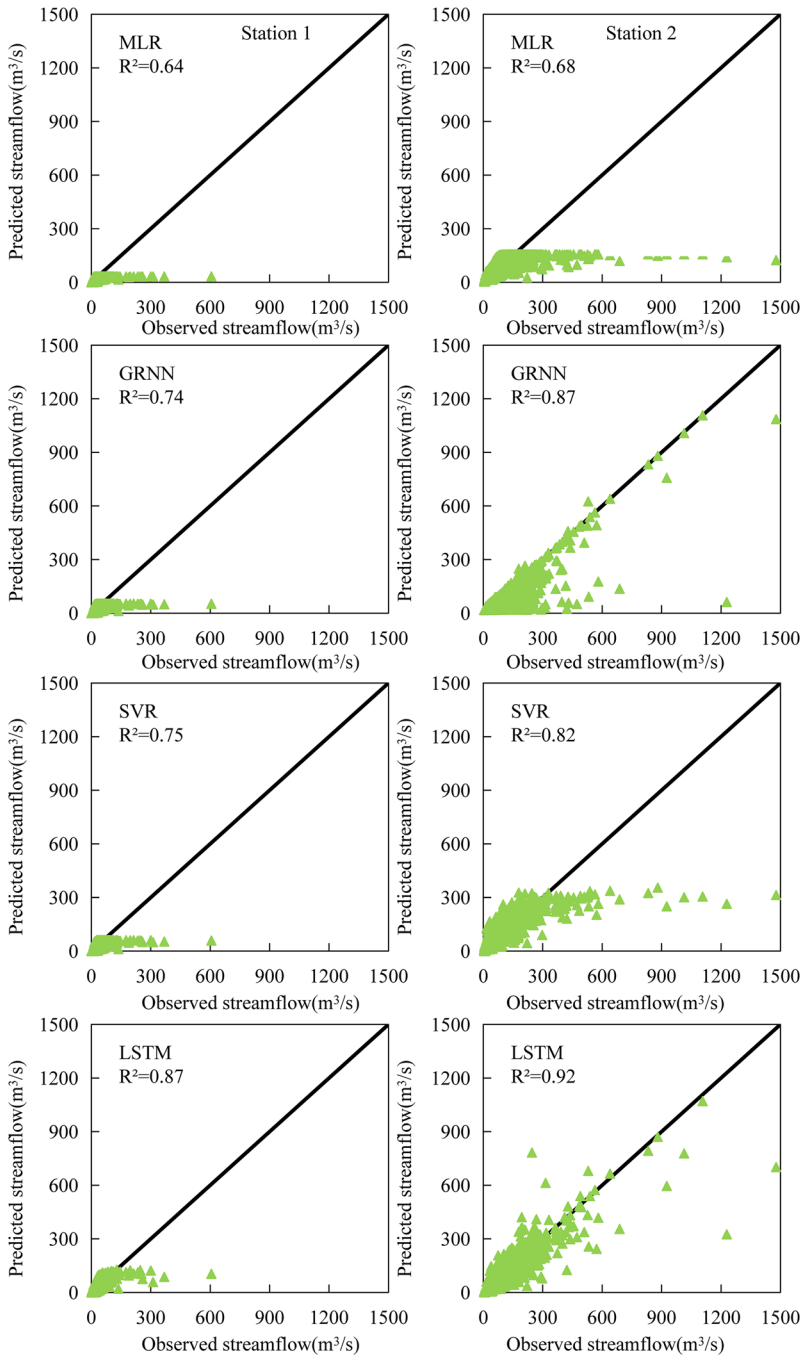
Fig. 3 Scatter plot of the observed vs. predicted streamflows for the four models during the calibration period (the left column shows the results at Station 1, and the right column shows the results at Station 2)

Fig. 4 Scatter plot of the observed vs. predicted streamflows for the four models during the validation period

Table 2 The performance of MLR, GRNN, SVR and LSTM for high-flow during calibration and validation period

Model	Station 1				Station 2			
	calibration		validation		calibration		validation	
	POD	FAR	POD	FAR	POD	FAR	POD	FAR
MLR	0.74	0.02	0.78	0.02	0.60	0.07	0.67	0.11
GRNN	0.86	0.11	0.98	0.11	0.91	0.01	0.91	0.04
SVR	0.90	0.06	0.99	0.13	0.64	0.01	0.71	0.01
LSTM	0.98	0.02	0.99	0.06	0.95	0.06	0.92	0.02

period. These results indicate that LSTM can better capture the characteristics of high-flow events in comparison to three other models.

4.2 Influence of Different Inputs on the Model Performance

Ten teleconnection candidates were selected in this study, including the Antarctic Oscillation (AAO), Southern Oscillation Index (SOI), Pacific North American Index (PNA), North Atlantic Oscillation (NAO), sunspots, East Central Tropical Pacific SST (Niño 3.4), Extreme Eastern Tropical Pacific SST (Niño 1+2), Central Tropical Pacific SST (Niño 4), Eastern Tropical Pacific SST (Niño 3), antecedent precipitation P and antecedent streamflow (S). Partial mutual information (PMI) was used to select the significant input variables, which can effectively select the variables that are linearly and nonlinearly related to streamflow without selecting redundant variables. Four input combinations were selected: (1) antecedent precipitation (P), (2) P and antecedent streamflow (S), (3) P, S and teleconnection factors (T), and (4) predictors selected by the PMI (SP). According to the PMI results, the streamflow at Station 1 showed a significant correlation with PNA, Niño 1+2, antecedent precipitation P and antecedent streamflow (S). In contrast, the streamflow at Station 2 showed a significant correlation with AAO, NAO, antecedent precipitation P and antecedent streamflow (S).

As shown in Table 3, at Station 1, in the calibration and validation periods, the RMSE and R^2 of the model with the predictors selected by PMI are the minimum and maximum values among the four combinations, which are 10.07 and 0.87 and 9.36 and 0.85, respectively. At Station 2, the RMSE and R^2 in the calibration and validation periods are 22.26 and 0.92 and 28.05 and 0.92, respectively. As seen from Fig. 5, it is obvious that the fitting

Table 3 The performance of LSTM with different inputs for different stations during calibration and validation period

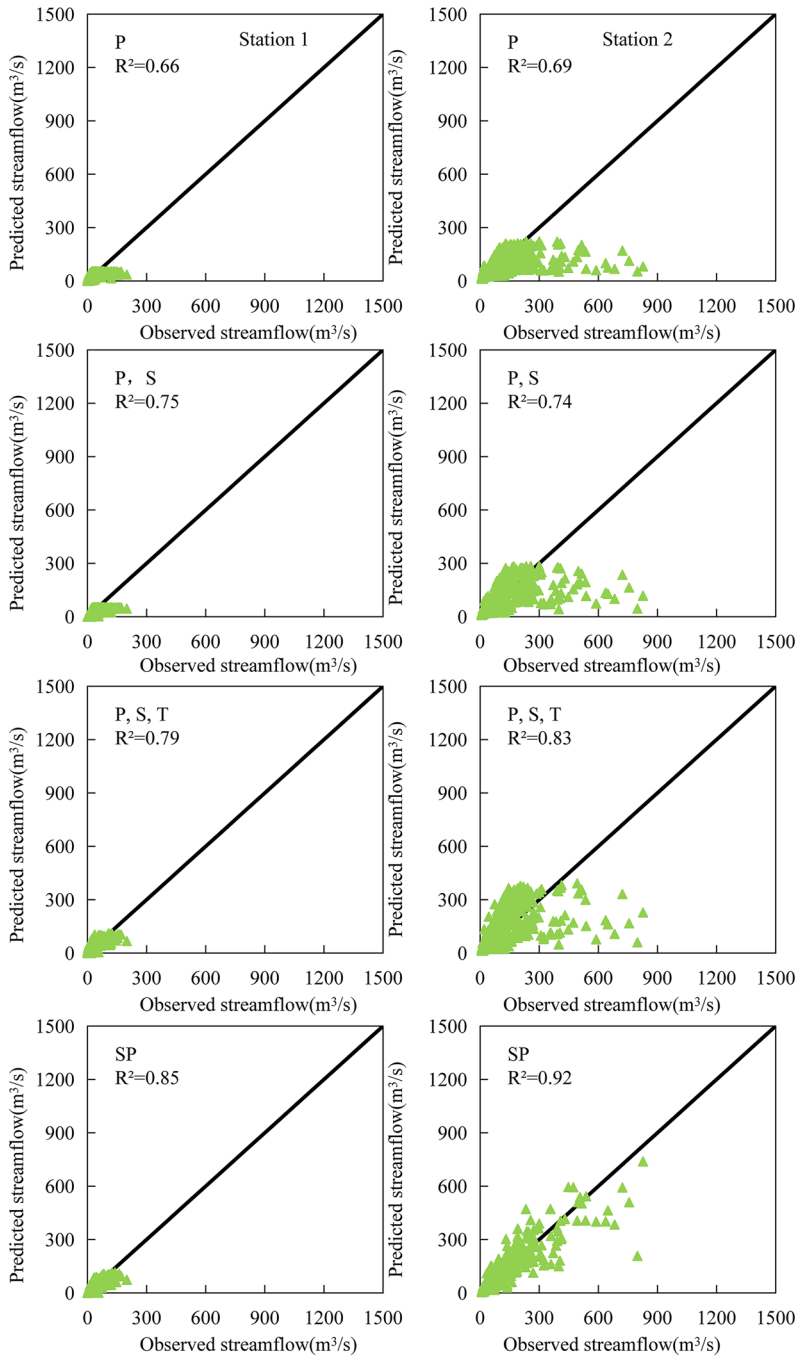| Inputs	Station 1				Station 2			
	calibration		validation		calibration		validation	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Antecedent precipitation (P)	12.19	0.67	13.74	0.66	41.36	0.72	52.36	0.69
P and Antecedent streamflow (S)	11.55	0.76	11.47	0.75	39.53	0.76	49.36	0.74
P, S and teleconnection factors (T)	10.55	0.80	13.24	0.79	36.08	0.83	48.25	0.83
Selected predictors by PMI (SP)	10.07	0.87	9.36	0.85	22.26	0.92	28.05	0.92

Fig. 5 Scatter plot of the observed vs. predicted streamflows for different inputs during the validation period

performance of the predictors selected by the PMI (SP) at Stations 1 and 2 is the closest to the 1:1 line. Therefore, the performance of the predictors selected by the PMI (SP) is the best. The selection process will be helpful not only for extracting the main characteristic relationship between the streamflow and predictors but also for reducing the noise impact of other factors.

4.3 Forecasting Uncertainty

The LSTM-bootstrap approach not only provides the estimated daily streamflow values but also assesses the confidence interval. In this study, three uncertainty performance measures were adopted based on different aspects to evaluate the uncertainty, namely, the coverage rate (CR), relative width (RB), and relative offset degree (RD) (Andrew et al. 2018). Their formulas are as follows:

$$CR = \frac{n}{N} \tag{11}$$

$$RB = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{(q_i^u - q_i^l)}{Q_{sim}^i} \tag{12}$$

$$RD = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\left| \frac{1}{2}(q_i^u + q_i^l) - Q_{obs}^i \right| / Q_{sim}^i \right) \tag{13}$$

where Q_{obs}^i and Q_{sim}^i are the observed and predicted values at moment i, respectively; q_i^u and q_i^l are the upper and lower limits of the corresponding uncertainty interval at moment i, respectively; n is the number of observed values within the uncertainty interval; and N is the total number of observed values.

The minimum value of CR is 0 and the maximum value is 1. The larger the value is, the higher the coverage rate of the interval is, where 1 means that the confidence intervals contain all observed streamflow values, and 0 means that the confidence intervals are unreliably with respect to whether they contain any observed streamflow values. RB is a measure of the average ratio of the uncertainty width to predicted values. The closer to 1 the CR value is, the more reliable the model. RD is a measure of the deviation of the centerline of the predicted interval from the observed flow hydrograph.

The CR values at Station 1 for the calibration and validation periods were both 0.99, indicating that the prediction results of the LSTM model are reliable. The CR values at Station 2 for the calibration and validation periods were 0.92 and 0.74, respectively. During the validation period, approximately 99% of the observed values at Station 1 were within the confidence interval, and only 1% of the values were not within the confidence interval. Approximately 74% of the observed values at Station 2 were within the confidence interval, and 26% of the observed values were lower or higher than the values within the confidence interval. These result indicate that the LSTM model is reliable for streamflow prediction at these two stations. The RB and RD values during the validation period (24.40 and 2.57 at Station 1 and 4.90 and 1.62 at Station 2, respectively)

at the two stations were higher than those during the calibration period (9.40 and 2.84 for Station 1 and 1.14 and 0.77 for Station 2, respectively), which is consistent with the change in the generalization ability. The values of RB and RD at Station 1 were higher than those at Station 2, which may be because the average, maximum and variance of the streamflows at Station 2 were greater than those at Station 1. The streamflows at Station 2 have the characteristics of larger fluctuation.

Figure 6 shows the streamflow prediction and confidence interval compared to the observations at the two stations during the validation period. To facilitate the presentation of the results, the confidence intervals at the two stations from 2021/5/1 to 2021/7/31 were separately enlarged in the upper right corner of the figure in this study. It can be seen from this figure that only a few high-flow observation values at both stations exceed the confidence interval, while most of the observed values are distributed within the confidence intervals. This result demonstrates that the LSTM model has the ability to reliably realize the streamflow prediction at these two stations.
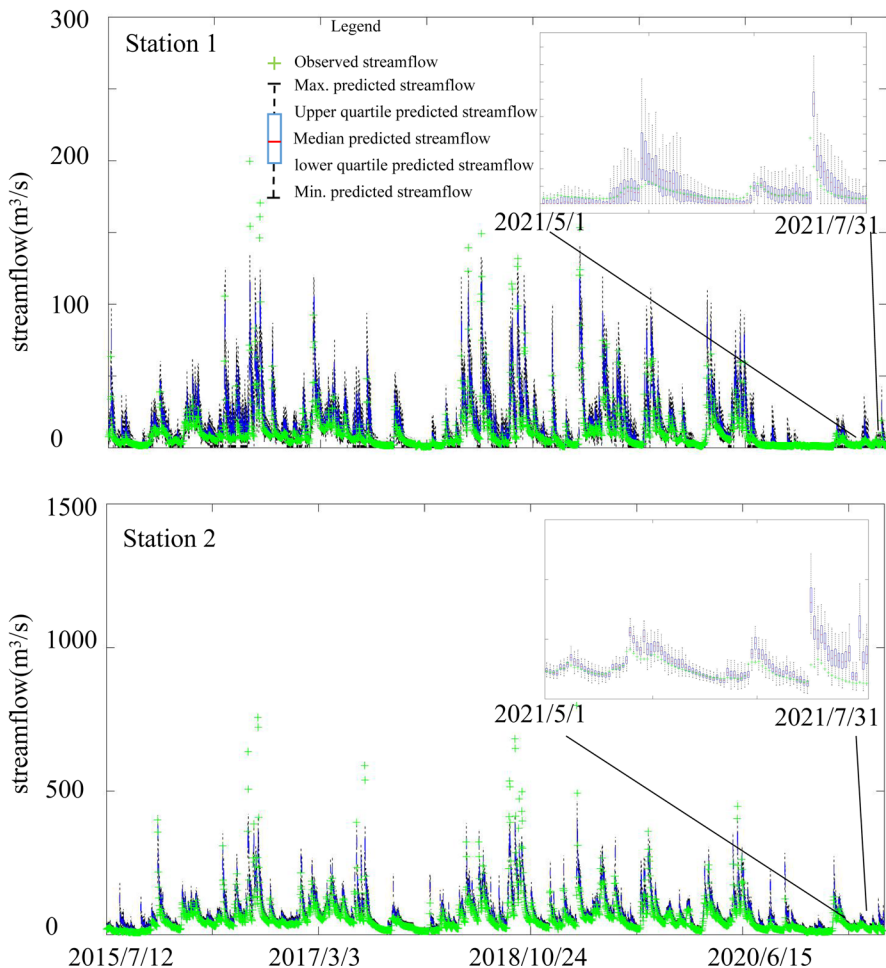
Fig. 6 Streamflow forecasting and confidence interval compared to the observations

5 Conclusion

In this study, LSTM model was proposed for daily streamflow prediction, and the approach was tested at two stations in the Mississippi River basin in Iowa, USA. The potential of LSTM models for daily streamflow prediction was explored and compared with the MLR, GRNN and SVR models. The impact of the parameters and input structure on model performance was also explored in the process of LSTM modelling. The results showed that the LSTM model outperformed the MLR, GRNN and SVR models with an improved performance of approximately 10%. LSTM models have a high POD and low FAR for high-flow events, which demonstrates that the LSTM model achieves a relatively good performance, especially for high flows.

The number of neurons and the period have a great influence on the model performance of the LSTM model, and it is essential to optimize these parameters during the modelling process. Four input combination scenarios were compared on the forecast performance of the LSTM model, and LSTM with selected local weather information and global climate indices had the best performance. The results from this comparison indicated that local weather information and global climate indices should be selected and considered in daily streamflow prediction; it not only extracts the main characteristic relationship between the streamflow and predictors but also reduces the noise impact of other factors.

Then, bootstrap was used to generate training data scenarios for evaluating the forecast uncertainty based on LSTM. The LSTM-bootstrap approach assesses the reliability and confidence interval of the streamflow prediction, which are of particular importance for reducing the risk and improving the management efficiency.

The stations in the paper are located in humid regions, sufficient information on the response relationship between precipitation and streamflow can be extracted from the historical observation data, and the LSTM models perform well for streamflow prediction at these two stations. In the future, LSTM should be applied in more regions with different climatic characteristics, especially arid areas with limited data.

Author Contribution Zhuoqi Wang: methodology, data curation, programming fine-tuning, writing of the original draft; Yuan Si: validation, data curation; Haibo Chu: writing, reviewing and editing the manuscript, methodology.

Availability of Data and Materials Available from the corresponding author upon reasonable request.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests None.

References

Alemu TE, Palmer NP, Polebitski A, Meaker B (2011) Decision support system for optimizing reservoir operations using ensemble streamflow predictions. J Water Resour Plan Manag 137(1):72–82

Andrew RB, William HF, Lauren EH (2018) Quantifying uncertainty in simulated streamflow and runoff from a continental-scale monthly water balance model. Adv Water Resour 122:166–175

Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos Res 172:37–47

Chu HB, Wei JH, Qiu J (2018) Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection. Water 10(10):1–15

Cho K, Kim Y (2022) Improving streamflow prediction in the WRF-Hydro model with LSTM networks. J Hydrol 605:1–12

Chen L, Singh VP, Lu WW, Zhang JH, Zhou JZ, Guo SL (2016) Streamflow forecast uncertainty evolution and its effect on real-time reservoir operation. J Hydrol 540:712–726

Cheng M, Fang F, Kinouchi T, Navon IM, Pain CC (2021) Long lead-time daily and monthly streamflow forecasting using machine learning methods. J Hydrol 590:1–13

Chu HB, Wei JH, Jiang Y (2021) Middle- and long-term streamflow forecasting and uncertainty analysis using Lasso-DBN-Bootstrap model. Water Resour Manag 35(8):2617–2632

Dehghani M, Saghafian B, Saleh FN, Farokhnia A, Noori R (2014) Uncertainly analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. Int J Climatol 34(4):1169–1180

Eum H-I, Kim YO (2010) The value of updating ensemble streamflow prediction in reservoir operations. Hydrol Process 24:2488–2499

Engeland K, Steinsland I, Johansen SS, Petersen-Overleir A, Kolberg S (2016) Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway. J Hydrol 536:147–160

Gopala SP, Kawamura A, Amaguchi H, Takasaki T, Azhikodan G (2019) A bootstrap approach for the parameter uncertainty of an urban-specific rainfall-runoff model. J Hydrol 579:1–18

Hu CH, Wu Q, Li H, Jian SQ, Li N, Lou ZZ (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water 10(11):1–16

Kan GY, Yao C, Li QL, Li ZJ, Yu ZB, Liu ZY, Ding LQ, He XY, Liang K (2015) Improving event-based rainfall-runoff simulation using an ensemble artificial neural network based hybrid data-driven model. Stoch Environ Res Risk Assess 29(5):1345–1370

Kim T, Yang TT, Gao S, Zhang LJ, Ding ZY, Wen X, Gourley JJ, Hong Y (2021) Can artificial intelligence and data-driven machine learning models match or even replace process-driven hydrologic models for streamflow simulation: a case study of four watersheds with different hydroclimatic regions across the CONUS. J Hydrol 589:1–20

Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci 22(11):6005–6022

Kasiviswanathan KS, He JX, Sudheer KP, Tay JH (2016) Potential application of wavelet neural network ensemble to forecast streamflow for flood management. J Hydrol 536:161–173

Ni LL, Wang D, Singh VP (2020) Streamflow and rainfall forecasting by two long short-term memory-based models. J Hydrol 583:1–10

Nourani V, Komasi M (2013) A geomorphology-based ANFIS model for multi-station modeling of rainfall-runoff process. J Hydrol 409:41–55

Pandey A, Srinivas VV (2015) Use of data driven techniques for short lead time streamflow forecasting in Mahanadi basin. Aquatic Procedia 4:972–978

Parisouj P, Mohezadeh H, Lee T (2020) Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zone in the United States. Water Resour Manag 34(13):4113–4131

Rahimzad M, Nia AM, Zolfonoon H, Soltani J, Mehr AD, Kwon HH (2021) Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. Water Resour Manag 35(12):4167–4187

Sammen SS, Ehteram M, Abba SI, Abdulkadir RA, Ahmed AN, El-Shafie A (2021) A new soft computing model for daily streamflow forecasting. Stoch Environ Res Risk Assess 35(12):2479–2491

Sahoo B, Jha R, Singh A, Kumar D (2019) Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys 67(5):1471–1481

Saraiva SV, Carvalho FD, Santos CAG, Barreto LC, Freire PKDM (2021) Daily streamflow forecasting in Sobradinho Reservoir using machine learning models coupled with wavelet transform and bootstrapping. Appl Soft Comput 102:1–11

Xu SC, Chen YB, Xing LX, Li C (2021) Baipenzhu reservoir inflow flood forecasting based on a distributed hydrological model. Water 13(3):1–16

Xiang ZR, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resour Res 56(1):1–17

Zhang D, Holland E, Lindholm G (2019) Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer. J Hydrol 567:792–802

Zhang ZH, Zhang Q, Singh VP, Shi PJ (2018) River flow modelling: Comparison of performance and evaluation of uncertainty using data-driven models and conceptual hydrological model. Stoch Env Res Risk A 32(9):2667–2682

Zhang Z, Lu WX, Chu HB, Cheng WG, Zhao Y (2014) Uncertainty analysis of hydrological model parameters based on the bootstrap method: a case study of the SWAT model applied to the Dongliao River Watershed, Jilin Province, Northeastern China. Sci China Technol Sci 57(1):219–229