Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-020-01766-4

ORIGINAL PAPER

An improved long short-term memory network for streamflow forecasting in the upper Yangtze River

Shuang Zhu[1] · Xiangang Luo[1] · Xiaohui Yuan[2] · Zhanya Xu[1]

Abstract
Characterized by essential complexity, dynamism, and dynamics, streamflow forecasting presents a great challenge to hydrologists. Long short-term memory (LSTM) streamflow forecast model has received a lot of attention in recent years due to its powerful non-linear modeling ability. But probabilistic streamflow forecasting has rarely been addressed by the LSTM approach. In this study, a probabilistic Long Short-Term Memory network coupled with the Gaussian process (GP) is proposed to deal with the probabilistic daily streamflow forecasting. Moreover, considering that changing mean and variance over time exist in the daily streamflow time series, the heteroscedastic Gaussian process regression is adopted to produce a varying prediction interval. The proposed method encapsulates the inductive biases of the LSTM recurrent network and retains the non-parametric, probabilistic property of Gaussian processes. The performance of the proposed model is investigated by predicting the daily streamflow time series collected from the upper Yangtze River and its tributaries. Artificial neuron network, generalized linear model, heteroscedastic GP, and regular LSTM models are also developed for comparison. Results indicated that the performance of the proposed model is satisfying. It improves prediction accuracy as well as provides an adaptive prediction interval, which is of great significance for water resources management and planning.

Keywords Streamflow forecasting · LSTM · Heteroscedastic Gaussian process regression · Prediction interval

1 Introduction

Streamflow forecasts play an extremely important role in the optimal management of water conservancy during flood events as well as in arid regions (Kisi and Cimen 2011). Hydrological processes are influenced by complex weather and non-linear infiltration mechanisms, which are difficult to model (Wang et al. 2009) and thus, reliable forecasting streamflow remains a challenge.

Streamflow forecasting methods mainly include physical/conceptual models, time series models and machine learning models (ML) (Chen et al. 2015). The physical/conceptual models provide an in-depth investigation of the behavior of the catchment on the basis of physical laws that govern water flow or simplified descriptions of hydrological processes. The disadvantage of the physical/conceptual models is the time-space complexity and model biases that cannot be ignored. It is well acknowledged that statistical streamflow prediction is convenient, effective, and highly accurate (Papacharalampous et al. 2018b). Time series models for streamflow forecasting can be classified into exponential smoothing, AutoRegressive Integrated Moving Average (ARIMA), seasonal models, long-range dependence models (Papacharalampous et al. 2018a, b), etc. Alternatively, machine learning (ML) algorithms are widely used to learn and model the high dimensional and non-linear relationships of hydrological variables (Yuan and Sarma 2010; El-Shafie et al. 2013; Karimi et al. 2018; Lu et al. 2018). The most popular ML models are Neural Networks (NN) Yuan et al. (2018b), Random Forests (RF), and Support Vector Machines (SVM) (Fu et al. 2019b, a). Researchers conducted large-scale temperature, precipitation and river discharge

✉ Xiaohui Yuan
xiaohui.yuan@unt.edu

[1] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China

[2] Department of Computer Science and Engineering, University of North Texas, Denton, TX 76210, USA

Published online: 04 February 2020

forecasting by using various stochastic and ML methods, indicated that both stochastic and ML methods may produce useful forecasts (Papacharalampous et al. 2018c, 2019). Recently, it was demonstrated that deep learning techniques have superior performance in numerous applications such as streamflow forecasting (Akram and El 2016; Shen 2018; Shen et al. 2018; Siqueira et al. 2018; Yaseen et al. 2019; Yuan and Abouelenien 2015). Recurrent Neural Network (RNN) includes feedback connections in the framework that takes into account the current information as well as other adjoining information in the data. Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber 1997; Gers et al. 2002) overcomes the problem of the traditional RNN of learning long-term dependencies representing e.g. storage effects within hydrological catchments, which may play an important role for hydrological processes (Kratzert et al. 2018). LSTM has no explicit internal representation of the water balance. However, it might be possible to analyze the behavior of the cell-states and link them to basic hydrological patterns (Kratzert et al. 2018). Zhang et al. (2018) developed LSTM to predict water table depth and achieved higher prediction accuracy when compared with the results of traditional feed-forward neural network (FFNN), proved that LSTM model can preserve and learn previous information well. Yuan et al. (2018a) investigated the accuracy of hybrid LSTM and parameter optimizer model in prediction of monthly runoff, indicated that the LSTM model has higher accuracy than that of other models. The competitive performance and high-stability of LSTM streamflow forecasting are demonstrated in these studies.

However, due to the intricacy of water dynamics and the spatial heterogeneity of the pertinent variables, the hydrological process is plagued by uncertainty, which makes daily streamflow forecasts in a probabilistic framework an appealing option (Moradkhani 2015). Modeling the prediction uncertainty across time has rarely been addressed by the LSTM approach. Most existing studies rely on ensemble techniques such as bagging and random forest, which use sampling with replacement to improve consistency. Gaussian process (GP) regression model is a non-parametric kernel-based probabilistic model (Rasmussen 2004), it was firstly introduced in the hydrology simulation and forecasting by Sun et al. (2014). GP constructs joint distribution by assuming that the variables used to model the sample follow the prior Gaussian distribution. The increasing popularity is due to the fact that non-linear problems can be solved in a principled Bayesian framework for learning and uncertainty estimation. With the convenience properties of GP, it is reasonable to couple GP with LSTM for probabilistic time series forecast. By combing LSTM and GP, Wu et al. (2016) achieved wind power forecasting and acquired prediction interval by implementing conditional error analysis. Gruet et al. (2018) used the deterministic results of the LSTM as the mean function of the GP model to derive the confidence interval for Dst geomagnetic index. Zhang et al. (2019) used GP to analyze the LSTM forecast errors of wind turbine power. Essentially, in these studies, GP is served as a post-processing analysis technique that is implemented on the deterministic forecast. Thus, a more direct, reliable and probabilistic LSTM streamflow forecasting research needs to be extended. On the other hand, a typical setting in the standard GP is the assumption of constant variance. As changing mean and variance over time exist in the hydrological time series, it is highly desirable to consider models with input-dependent variance (Lázaro-Gredilla and Titsias 2011). This leads to the need to consider the heteroscedastic Gaussian process regression.

In this paper, we propose a novel hybrid model that integrates the heteroscedastic GP model into the inner structure of LSTM. The proposed method encapsulates the inductive biases of LSTM recurrent networks, retains the non-parametric, probabilistic property of Gaussian processes, and naturally produces prediction interval without any post-processing. The performance of the proposed model is investigated by predicting the daily streamflow time series collected from the upper Yangtze River and its tributaries. To make a comparison, a widely used ANN, Generalized Linear Model (GLM), heteroscedastic GP model, and regular LSTM models are also developed to implement daily streamflow forecasting in this study. The rest of the article is organized as follows. The study area and data are detailed in Sect. 2. The modeling framework and forecasting evaluation indicators are explained in Sect. 3. Section 4 presents our experimental results and discussion. Section 5 concludes the paper with a summary.

2 Study area and runoff data

The length of the Yangtze River is 6380 km, and the watershed area is 1,808,500 km^2. The drainage basin is located between 91°E to 122°E and 25°N to 35°N. Due to its abundant water resources, the river has been used for irrigation, industry, power generation, sanitation and shipping, and plays a vital role in the development of China's socio-economic and ecological environment.

We consider daily streamflow forecasting for the upper Yangtze River. A schematic of the upper Yangtze River, regional main tributaries, and the gauging stations are shown in Fig. 1. Yichang is the outlet of the upper reaches of the Yangtze river basin, with a length of 4504 km and a drainage area of 1,000,000 km^2. The main tributaries from Pingshan to Yichang include the Min River and the Jialing

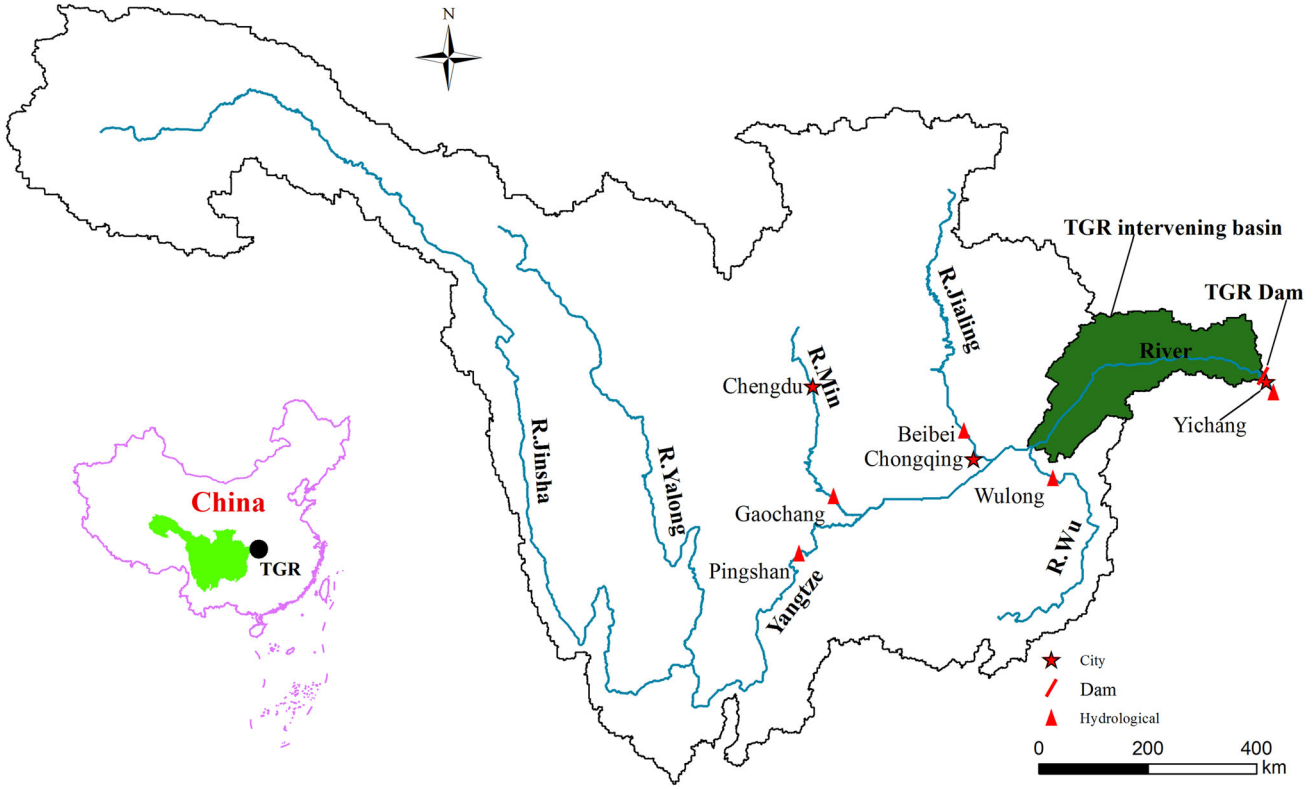

Fig. 1 A schematic of upper Yangtze River, main tributaries, and the gauging stations

River on the north bank and the Wu River on the south bank. From the source of the river to Pingshan, it is called the Jinsha River, which is 3481 km long and has a drop of about 5100 m, accounting for about 95% of the total Yangtze river drop. On the left bank of the Jinsha River, there is a large tributary Yalong River, which has an elevation of over 4000 m above sea level. Runoff of the Jinsha River is mainly supplemented by snow. Gaochang is the outlet of the Min River. The total length of the main stream is 711 km, the drainage area is 135,400 km^2, the average annual flow is 2830 m^3/s, and the total drop is 3560 m. Beibei is the outlet of Jialing River, with a drainage area of 157,900 km^2 and a total length of 1571 km. The average annual flow is 1860 m^3/s, with a total drop of 2300 m. The source areas of the Min River and the Jialing River have a high terrain, with an altitude of 3000 to 4000 m, and the terrain near the Sichuan Basin suddenly drops to 200–600 m. In addition to the Jinsha River in the upper reaches of the Yangtze River, heavy rains frequently occurred in the rest of the areas. Heavy rains are the main supply of runoff in the Jialing River and the Min River.

The statistical features, involving catchment area, river length, mean streamflow and available data of the mainstream and tributaries were given in Table 1. As the most important hydrological control site of the upper Yangtze River, Yichang is included for streamflow forecasts in this study. Three secondary control sites, Pingshan, Gaochang, and Beibei are used in our experiments. Daily streamflow series (2004/1/1-2010/12/31) recorded at Yichang, Pingshan, Gaochang, and Beibei is provided and checked by the Yangtze River Waterway Bureau, China (http://www.cjw.gov.cn/). Figure 2 is a daily time series plot of Yichang, Beibei, Gaochang and Pingshan. It is clear that the streamflow in Beibei, Pingshan, Gaochang and Yichang stations has consistent annual and interannual trends. Due to the influence of the subtropical monsoon climate, streamflow in the upper Yangtze River has an uneven distribution and severe fluctuations, which makes it difficult to obtain an accurate runoff forecast.

3 Improved long short-term memory network

An improved long short-term memory network coupled with heteroscedastic GP, labeled LSTM-HetGP, is developed for probabilistic streamflow forecasting in this paper. The overview streamflow forecasting model and the architecture of LSTM-HetGP are shown in Fig. 3. It has a network structure of four types of layers: an input layer, numbers of recurrent hidden layers, a Gaussian process

Table 1 Sub-basin characteristics and streamflow statistics in the upper reaches of the Yangtze River

Rivers	Area (km^2)	Length (km)	Annual streamflow (m^3/s)	Record length
Yalong River	144,200	1571	1860	1950–2010
Jinsha River	485,099	3481	4750	1950–2010
Min River	135,400	711	2830	1950–2010
Jialing River	157,900	1062	2100	1950–2010
Wu River	87,920	1037	1650	1950–2010
Upper Yangtze River	1,005,501	4504	13,600	1950–2010

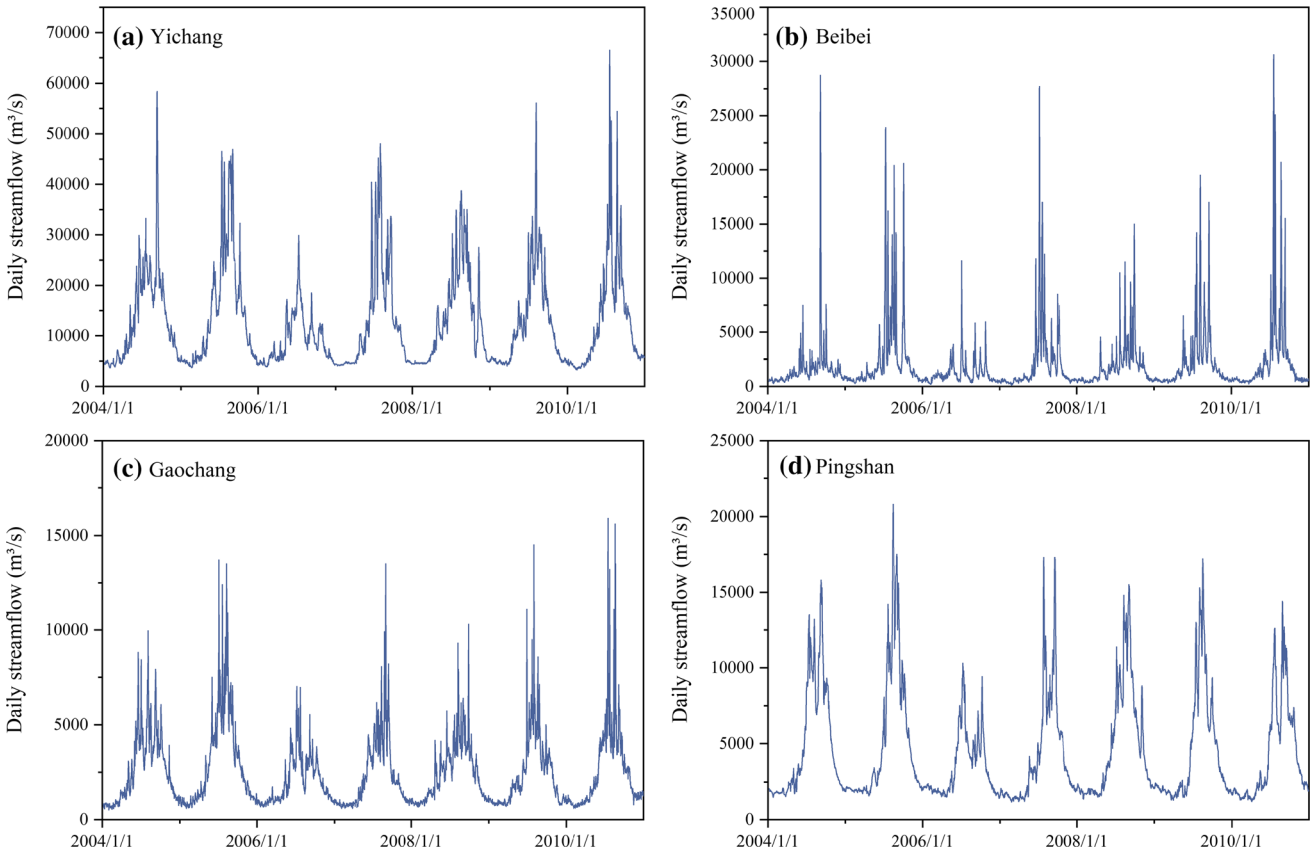

Fig. 2 Average daily streamflow series for upper Yangtze River and the tributaries

layer, and an output layer. In the procedure, input variables are mapped into a single hidden vector through the recurrent hidden layers. Instead of connecting directly to the output layer, it is projected into a high dimensional feature space with the kernel function of the heteroscedastic GP, the prediction interval is derived by the GP posterior distribution over the target. A semi-stochastic alternating gradient descent optimization procedure is applied to carry out weight updates and fully joint training of the hybrid model. Python programming environment (Guttag 2016), the packages of Keras (Chollet et al. 2015) and heteroscedastic Gaussian process regression (Binois et al. 2018) were used to implement the algorithms.

3.1 Long short-term memory network

Long Short-Term Memory network (LSTM) is evolved from Recurrent Neural Networks (RNN). RNN is a circular network in which an additional input is added to represent the state of the neuron in the hidden layer at the previous time steps (Elman 1990; Ishak et al. 2003). Figure 4 shows the repeated module in a recurrent network with a single layer. Given an input sequence $X = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_T})$, the state of the neuron at the current time step h_t is computed as follows:

$$h_t = tanh(W_{\mathbf{x}h}\mathbf{x_t} + W_{hh}h_{t-1} + b_h), \quad (1)$$

where h_t, h_{t-1} represent the hidden neuron states at the time step t and $t-1$ respectively, $W_{\mathbf{x}h}$ and W_{hh} mean weight

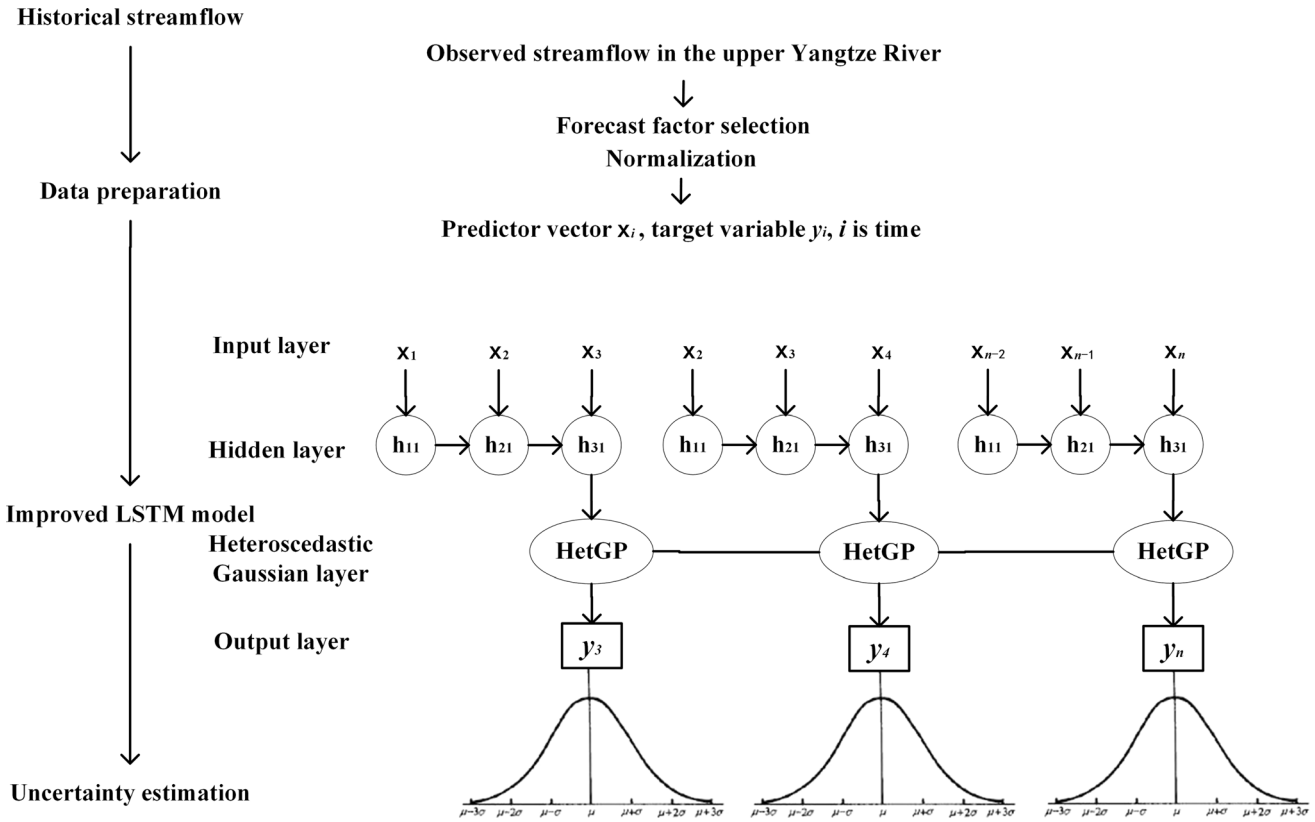

Fig. 3 An overview of the streamflow forecasting model and the architecture of LSTM-HetGP

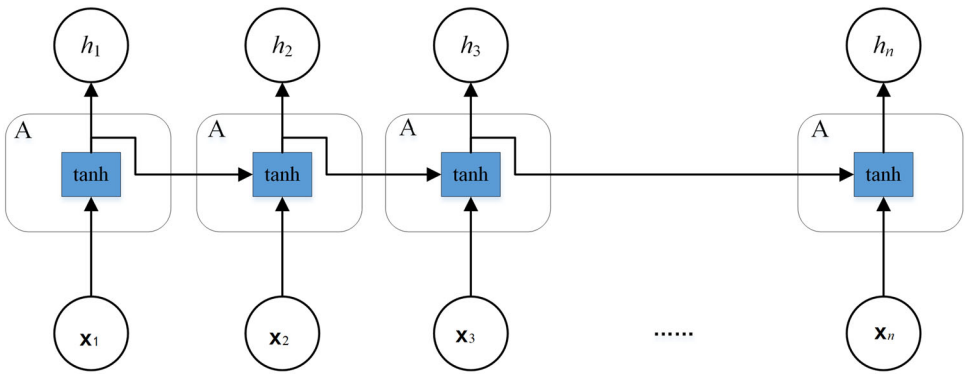

Fig. 4 The repeated module in a RNN with a single layer

matrices of input-hidden neurons and hidden-hidden neurons, respectively; b_h is bias term, $tanh(\cdot)$ is element-wise hyperbolic tangent function, $\mathbf{x_t}$ vector is model input at the time step t.

In Eq. (1), learning the parameters of the earlier neurons is difficult because the error of partial derivative accumulates and the gradients can easily become extremely large or vanish. LSTM is designed to solve this problem by making the regular RNN out of little modules designed to remember values for a long time. Therefore, a basic LSTM consists of an input layer, a hidden layer, and an output layer, and the hidden layer of LSTM has a different structure to optionally let information through. Three extra modules are added to control the preservation, reading, and modification of the memory cell of the LSTM forecasting model.

Suppose the input at the time step t is $\mathbf{x_t}$, the updated hidden layer of LSTM is calculated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, \mathbf{x_t}] + b_f) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, \mathbf{x_t}] + b_i) \tag{3}$$

$$C_t' = tanh(W_C \cdot [h_{t-1}, \mathbf{x_t}] + b_C) \tag{4}$$

$$C_t = f_t C_{t-1} + i_t C_t' \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, \mathbf{x_t}] + b_o) \tag{6}$$

$$h_t = o_t * tanh(C_t) \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid function; W_f, W_i, W_C, W_o are weight matrices; b_f, b_i, b_C, b_o are bias vector; f_t, i_t and o_t are the values of forget gate, update gate and output gate. The output of LSTM is

$$y = \psi(W_{hy}^T h_t) \tag{8}$$

Where ψ is fixed element-wise function, W_{hy} is weights of output layer.

3.2 Heteroscedastic Gaussian process regression

Gaussian process regression (GPR) is firstly proposed by Gibbs (1998) and lately extended by Kersting et al. (2007) and Tolvanen et al. (2014). Due to the advantage of handling uncertainty, it has begun to be applied for probabilistic streamflow forecasting in the last few years (Sun et al. 2014; Zhu et al. 2018). A detailed description of GPR prediction theory can be found in Rasmussen (2003, pp. 8–13). Given an input vector $\mathbf{x}_*$, the Gaussian posterior distribution of the predicted value y_* is

$$p(y_*|\mathbf{x}_*, \mathbf{x}, y) = N\left(\frac{1}{\sigma^2}\mathbf{x}_*^T A^{-1}\mathbf{x}y, \mathbf{x}_*^T A^{-1}\mathbf{x}_*\right) \tag{9}$$

where $A = \sigma^{-2}\mathbf{x}\mathbf{x}^T + \sum_p^{-1}$, $\sum_p$ is covariance matrix, σ is constant variance.

The predictive distribution is Gaussian, with a mean given by the posterior mean of the weights $\frac{1}{\sigma^2}A^{-1}\mathbf{x}y$ multiplied by the test input $\mathbf{x}_*^T$. The variance is a quadratic form of the inputs with the posterior covariance matrix. The above equation is based on the Bayesian assumption and is suitable for linear models. A simple idea to overcome this problem is to first project the inputs into some high dimensional space using a kernel function. Then apply the linear model in the high dimensional space instead of directly on the inputs themselves.

Function ϕ is introduced to project the input $\mathbf{x}$ into a high dimensional feature space to implement polynomial regression. The Gaussian posterior can be written as

$$p(y_*|\mathbf{x}_*, \mathbf{x}, y) = N\left(\frac{1}{\sigma^2}\phi(\mathbf{x}_*^T)A^{-1}\phi(\mathbf{x})y, \phi(\mathbf{x}_*^T)A^{-1}\phi(\mathbf{x}_*)\right) \tag{10}$$

To simplify the calculation, kernel function $K(\mathbf{x}, \mathbf{x})$ is used to directly obtain the inner product of the complex nonlinear transformation. The probability distribution is

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathbf{x}, y) = & N(K(\mathbf{x}_*, \mathbf{x})(K(\mathbf{x}, \mathbf{x}) + \sigma^2 I_n)^{-1} y, K(\mathbf{x}_*, \mathbf{x}_*) \\ & - K(\mathbf{x}_*, \mathbf{x})(K(\mathbf{x}, \mathbf{x}) + \sigma^2 I_n)^{-1} K(\mathbf{x}, \mathbf{x}_*)) \end{aligned} \tag{11}$$

where $A = \phi^T \sum_p \phi$.

Assume that the training samples follow a normal distribution with a zero mean, a unique GPR model is obtained with the kernel function K, that is $f(\mathbf{x}) \sim GP(0, K_f(\mathbf{x}, \mathbf{x}_*))$. For a heteroscedastic GP (HetGP) model, observation noise has a possibly different variance $r(\mathbf{x})$ at each input point $\mathbf{x}$.

$$f(\mathbf{x}) \sim GP(0, K_f(\mathbf{x}, \mathbf{x}_*)), \epsilon \sim N(0, r(\mathbf{x})) \tag{12}$$

The model is fully specified and depends only on the kernel functions $K_f(\mathbf{x}, \mathbf{x}_*)$ and $r(\mathbf{x})$ Lázaro-Gredilla and Titsias (2011). The predictive distribution of a heteroscedastic GP is expressed as:

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathbf{x}, y) = & N(K(\mathbf{x}_*, \mathbf{x})(K(\mathbf{x}, \mathbf{x}) + R(\mathbf{x}))^{-1} y, K(\mathbf{x}_*, \mathbf{x}_*) \\ & + R(\mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{x})(K(\mathbf{x}, \mathbf{x}) + R(\mathbf{x}))^{-1} K(\mathbf{x}, \mathbf{x}_*)) \end{aligned} \tag{13}$$

where $R(\mathbf{x}) = diag(r(\mathbf{x}))$, and $R(\mathbf{x}_*) = diag(r(\mathbf{x}_*))$.

3.3 Probabilistic LSTM coupled with heteroscedastic GP

To produce a probabilistic LSTM forecasting, a heteroscedastic GP layer is introduced after the hidden layer of the LSTM. In this new architecture, input variables are embedded into a single hidden vector in the hidden space, H. Then it is transformed into a high dimensional space with a kernel function. Target distribution is learned in the GP layer, as shown in Fig. 3. For this hybrid feed-forward network, Wilson et al. (2016) proposed to use full-batch algorithms to jointly optimize the hyperparameters of the kernel function and the network weights. In our case, the network is recurrent. Stochastic updates should be considered since it allows for the efficient topology-independent implementation of backpropagation. Therefore, a semi-stochastic alternating gradient descent optimization procedure proposed by Al-Shedivat et al. (2016) is applied to carry out weight updates as well as fully joint training of the LSTM-HetGP hybrid model, by alternately updating the hyperparameters of GPR on the full training data first and then updating the weights of LSTM on the mini-batch data using stochastic steps. Instead of using all training data items (full-batch) or using a single training item (as in stochastic training), mini-batch training uses a specified number of training items to compute gradients to get faster convergence. The algorithm steps are shown in Fig. 5.

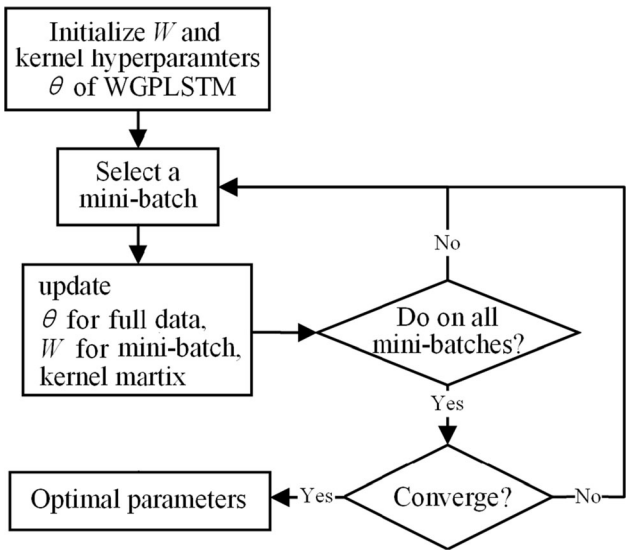

Fig. 5 The semi-stochastic alternating gradient optimization procedure of LSTM-HetGP

The parameters of the LSTM-HetGP model include weight parameters of W and GP hyperparameters θ. The model training procedure is to map the input time series into a vector in a hidden layer with the w, transfer the vector into a high dimensional space with the kernel function and corresponding θ, compute the kernel matrix and then get the Gaussian posterior distribution on target. The kernel matrix of LSTM-HetGP is computed with an initial θ and full training data set. For a fixed kernel matrix, update W on a mini-batch using the derivatives of the negative log marginal likelihood of LSTM-HetGP with respect to W. Then update the θ on the full vectors in the hidden layer due to the updated W. In this step, θ of the heteroscedastic GP is estimated by the maximum likelihood method and replicates the strategy proposed by Binois et al. (2018). Repeat the process for the mini-batches until convergence.

Once the parameters of LSTM-HetGP are known, the Gaussian posterior distribution of LSTM-HetGP over target y_* is

$$p(y_*|\mathbf{x}_*, \mathbf{x}, y) = N(\mu_*, \sigma_*) \tag{14}$$

The expectation is

$$\mu_* = K(\mathbf{x}'_*, \mathbf{x}')(K(\mathbf{x}', \mathbf{x}') + R(\mathbf{x}'))^{-1} y \tag{15}$$

The variance is

$$\sigma_* = K(\mathbf{x}'_*, \mathbf{x}'_*) + R(\mathbf{x}'_*) - K(\mathbf{x}'_*, \mathbf{x}')(K(\mathbf{x}', \mathbf{x}') + R(\mathbf{x}'))^{-1} K(\mathbf{x}', \mathbf{x}'_*) \tag{16}$$

where $\mathbf{x}' = W_{xh}\mathbf{x}$, $\mathbf{x}'_* = W_{xh}\mathbf{x}_*$, W_{xh} is weight matrix of input-hidden neurons. Based on Bayes theorem, the expectation is used for prediction, and the uncertainty interval is derived with mean and standard deviation (Tyralis and Koutsoyiannis 2014).

4 Results and discussion

Four sets of historical daily streamflow series (from 1/1/2004 to 31/12/2010) of the upper Yangtze River and its tributaries (daily streamflow in the Beibei, Gaochang, Pingshan, and Yichang gauging stations) were investigated by the proposed LSTM-HetGP model. In addition, streamflow forecast model ANN, GLM, heteroscedastic GP, and regular LSTM models are also included to perform comprehensive analysis and comparison. The first five-year data (2004 to 2008) of daily streamflow are used as training sets and the next two-year data (2009 to 2010) are used for testing.

4.1 Evaluation metrics

The root mean squared error (RMSE) criterion and the mean relative error (MRE) are the two criteria most widely used for calibration and evaluation of streamflow forecasting models. While the MRE gives the same weight to all errors, the RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values (Chai and Draxler 2014). Nash-Sutcliffe efficiency (NSE) (Gupta et al. 2009) calculates the proportion of total variation, considering the error in the variability and dynamics. It varies on the interval [-inf to 1.0], having an ability to measure the error of process simulation. The mean squared logarithmic error (MSLE) (De Vos and Rientjes 2008; Hogue et al. 2000) is a less used evaluation indicator, it more emphasizes on the error of low flows due to the logarithmic transformation. The percentage of coverage (POC) and the average interval width (AIW) (Lei et al. 2014) are two important attributes of the probabilistic streamflow forecasting. The POC calculates the percentage of the target points within the range of prediction intervals at all target points. AIW calculates the average width of all prediction intervals. Higher POC means greater probabilities of predictive streamflow range correctly covering the target. If the POC values are close, then a better streamflow forecasting is detected with a lower AIW. The formulas of NSE, RMSE, MRE, M4E, MSLE, POC and AIW are shown as follows.

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{N}(Q_i - \bar{Q})^2} \tag{17}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2} \tag{18}$$

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{Q}_i - Q_i|}{Q_i} \tag{19}$$

$$MSLE = \frac{1}{N}\sum_{i=1}^{N}(lnQ_i - ln\hat{Q}_i)^2 \tag{20}$$

$$g_i = \begin{cases} 1, & L_i \le Q_i \le U_i \\ 0, & otherwise \end{cases} \tag{21}$$

$$POC = \left(\frac{1}{N}\sum_{i=1}^{N} g_i\right) * 100\% \tag{22}$$

$$AIW = \left(\frac{1}{N}\sum_{i=1}^{N}(U_i - L_i)\right) * 100\% \tag{23}$$

where N is the sample length, Q_i is the i-th observed runoff, $\hat{Q}_i$ is the i-th simulated runoff, $\bar{Q}$ is the mean of observed runoff series, U_i and L_i are the estimated upper and lower bounds of the i-th sample,

4.2 Models and candidate predictors

Using the training data, AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) were used to analyze the correlation of streamflow time series. PACF is a conditional correlation by taking into account the impacts of some other variables and eliminating their redundancy. The previous discharge at stations in Beibei, Gaochang, Pingshan, and Yichang with a lag of 1 to 16 days was taken as candidate inputs of models. The ACF and PACF between target streamflow and the antecedent streamflow are presented in Fig. 6. Taking Beibei as an example, it shows that the ACF is very high and decreases slowly as time lag increases, and the PACF absolute value is higher with lags of one, two, three days, and decreases quickly. Therefore, streamflow with lags of one, two, three days were used as predictors. Previous streamflow data at time steps t-1, t-2 and t-3 were used as predictors for Beibei and Pingshan. Streamflow data at time steps t-1, t-2, t-3, t-4 and t-5 were included as predictors for Gaochang and Yichang.

Table 2 lists the stations, models used, input data, and output in our experiments. Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5} are previous discharge with a lag of 1 to 5 days.

4.3 Daily streamflow forecasting

In the LSTM-HetGP model, the squared exponential kernel function and *tanh* activation function are adopted for the model calibration. Determining the hidden structure (number of hidden layers and hidden units) is the top priority of training the LSTM-HetGP model using the semi-stochastic optimization algorithm. Using the training dataset, we conducted a number of experiments, by training the model with one and two hidden layers and 3 to 100 hidden units. To achieve better generalization, cross-validation was employed in this paper. A compound RMSE, rather than the simple early stopping method was used as the objective function in cross-validation. Meanwhile, dropout is a regularization method that can be used during training. In the process of feedforward and weight updates, it probabilistically inactivates the inputs and recursive connections of LSTM neurons. We used this method to effectively avoid LSTM overfitting problems and improve model performance. Figure 7 plots the experiments result for the case of Yichang streamflow forecasting. The left chart shows the convergence process on one and two hidden layers. An epoch is one forward pass and one backward pass of all the training examples. After 30 epochs, MSE values approach stability, one hidden layer is better than two hidden layers due to smaller MSE. Then the structure of one hidden layer and 3 to 100 hidden units are tested, shown in the right chart. MSE gets its minimum when the number of hidden units is 50, so we determined the structure of the LSTM-HetGP model with 1 recurrent hidden layer and 50 hidden units. With the same procedure, structures and parameters of forecast models for Beibei, Pingshan, Gaochang and Yichang are optimal.

Figure 8 shows the LSTM-HetGP streamflow forecasting. The yellow line represents the observed value, the colored dot means predictive mean, the gray band is the prediction interval with 95% confidence. Figure 8(a) is the forecasts for Yichang, (b) is for Beibei, (c) is for Gaochang and (d) is for Pingshan. It can be observed that the predictive means follow the observed values very well, and the prediction interval can effectively cover the observations. Further, in the mainstream and tributaries of the upper reaches of the Yangtze River, the daily streamflow forecasts of Pingshan and Yichang are better than Beibei and Gaochang. For Beibei and Gaochang, the catchment areas are relatively small, and heavy rain is the main source of floods. These two characteristics make the seasonal and annual variations of streamflow greater than Pingshan and Yichang. As shown in Fig. 8b, c, in the dry season, the streamflow is very stable and small, while in the rainy season, the floods in Beibei and Gaochang steeply rise and fall. It makes forecasting difficult. Therefore, for Beibei and Gaochang, using more meteorological predictors improves the accuracy of the forecast.

Moreover, it can be observed that the width of the prediction interval varies with the predictive mean. For small streamflow value, the corresponding 95% confidence interval is narrower, and vice versa. The varying prediction interval based on the heteroskedastic Gaussian process regression has a good practical value.

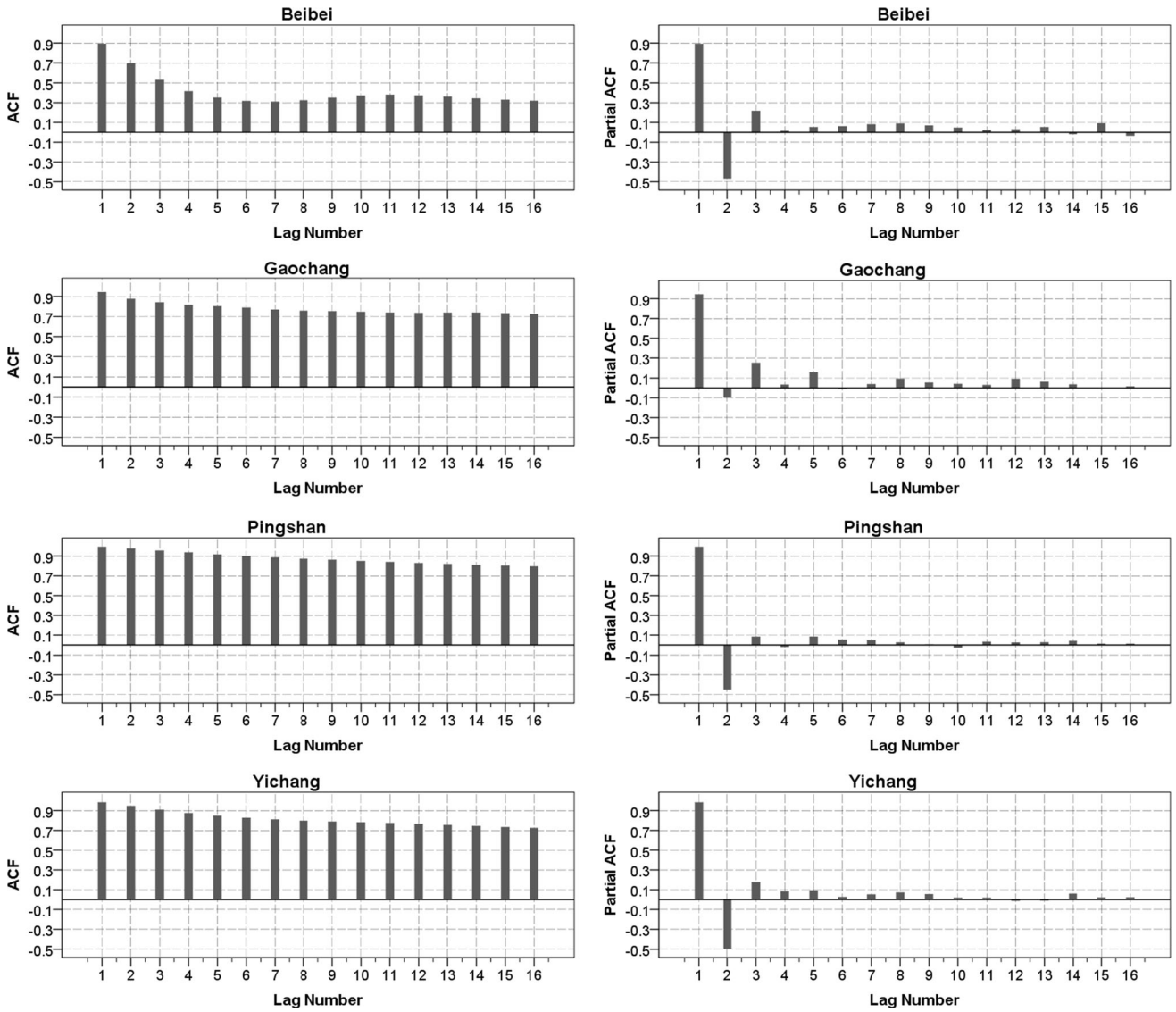

Fig. 6 ACF and PACF statistics for Beibei, Gaochang, Pingshan and Yichang streamflow

Table 2 Models, input data, and output used in our experiments

Station	Model	Inputs	Output
Beibei, Pingshan	GLM	Q_{t-1}, Q_{t-2}, Q_{t-3}	Q_t
	ANN	Q_{t-1}, Q_{t-2}, Q_{t-3}	Q_t
	LSTM	Q_{t-1}, Q_{t-2}, Q_{t-3}	Q_t
	HetGP	Q_{t-1}, Q_{t-2}, Q_{t-3}	Q_t
	LSTM-HetGP	Q_{t-1}, Q_{t-2}, Q_{t-3}	Q_t
Gaochang, Yichang	GLM	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Q_t
	ANN	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Q_t
	LSTM	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Q_t
	HetGP	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Q_t
	LSTM-HetGP	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Q_t

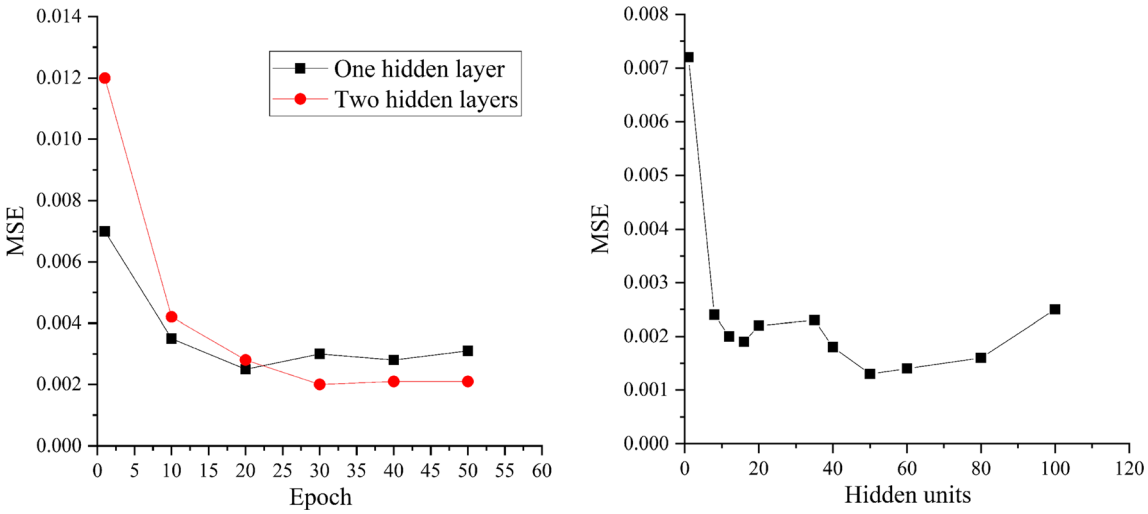

Fig. 7 MSE versus epoch number for one to two hidden layer and the number of hidden units per layer

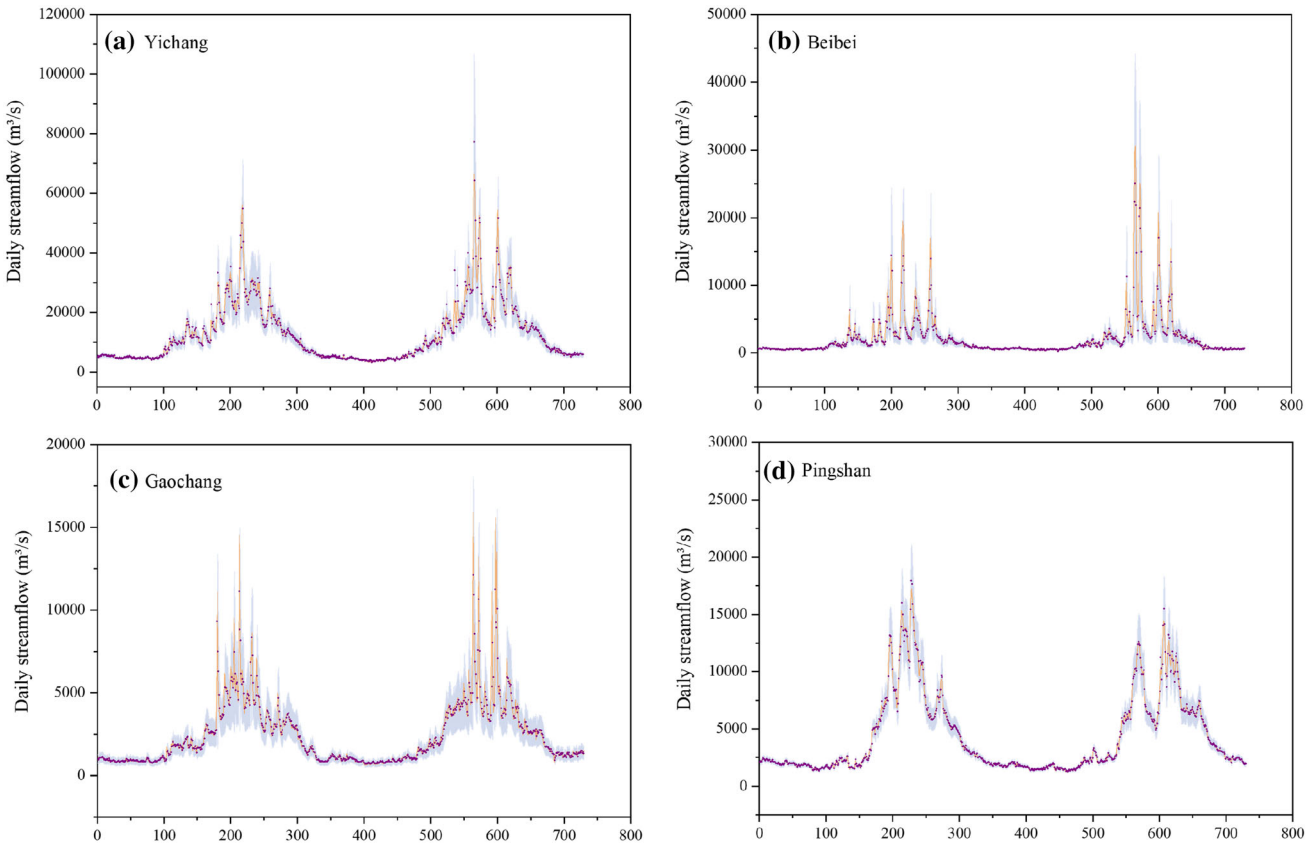

Fig. 8 LSTM-HetGP streamflow forecasting value and observed value; the yellow line represents the observed value, the colored dot means predictive mean, the gray band is the prediction interval with 95% confidence

4.4 Prediction interval analysis

The proposed method gives a variable prediction interval by combining the heteroscedastic Gaussian process regression algorithm, which greatly expands the reliability of the forecast. To evaluate LSTM-HetGP in case of changing mean and variance over time of hydrologic time series, we adopted the extra segment strategy (described in the follows) and one-parameter Box-Cox data transformation (Zhu et al. 2018) to normalize the daily streamflow time series and combined LSTM with a traditional homoscedastic Gaussian process regression (HomGP).

Here we take Yichang as an example to investigate their different predictive uncertainty.

For segment strategy, we divided a year into four periods (March to May, June to September, October to November, December to next year February), and in each segment, daily streamflow series has a normal distribution. The forecasting range within 95% confidence interval is presented in Fig. 9. The yellow solid line represents the true value, the purple dot represents the predicted value, and the gray band represents the confidence interval. Figure 9a–d are the daily streamflow prediction interval from March to May, June to August, September to November, and December to February of the next year in the testing period. In Fig. 9a, the time-axis sequentially includes March-2009, April-2009, May-2009, March-2010, April-2010, May-2010, total 184 days. The streamflow values on May 31, 2009, and 1 March 2010 vary greatly and therefore produce an abrupt falling of the line chart (time axis 92). The 95% confidence interval in the figure can perfectly cover the true value of the streamflow, the width of the forecast interval is constant and symmetrical towards predicted value. But during the flood season, specific from June to August in the Yangtze River region, the forecast interval cannot cover individual extreme floods, shown in Fig. 9b, it is caused by the constant mean variance of the GP model in each segment. The high flow events are relatively rare and flow values are very unstable, what lead to accurate intervals difficult, even though, prediction intervals can still provide much more information than single-point forecast.

We also used a one-parameter Box-Cox data transformation to normalize the daily streamflow time series. The maximum likelihood method is used to estimate conversion parameters using the training dataset. Then test dataset is transformed with the estimated parameters (Papacharalampous et al. 2018b). Figure 10a plots the entire forecasting interval in the testing period. It can be observed that the forecast interval is asymmetrical and the coverage width varies, the reason is due to the non-linear transform of Box-Cox and its inverse. The confidence interval can perfectly cover the real value, even the extreme flood event, but there is still a disadvantage that the confidence interval at a low flow period has a risk of nearly missing the true runoff for the same reason, as shown in Fig. 10b. It plots the forecasting interval at the low flow period (2009/11/15 to 2010/4/15), the interval fails to cover the true value, especially near 2010/2/15.

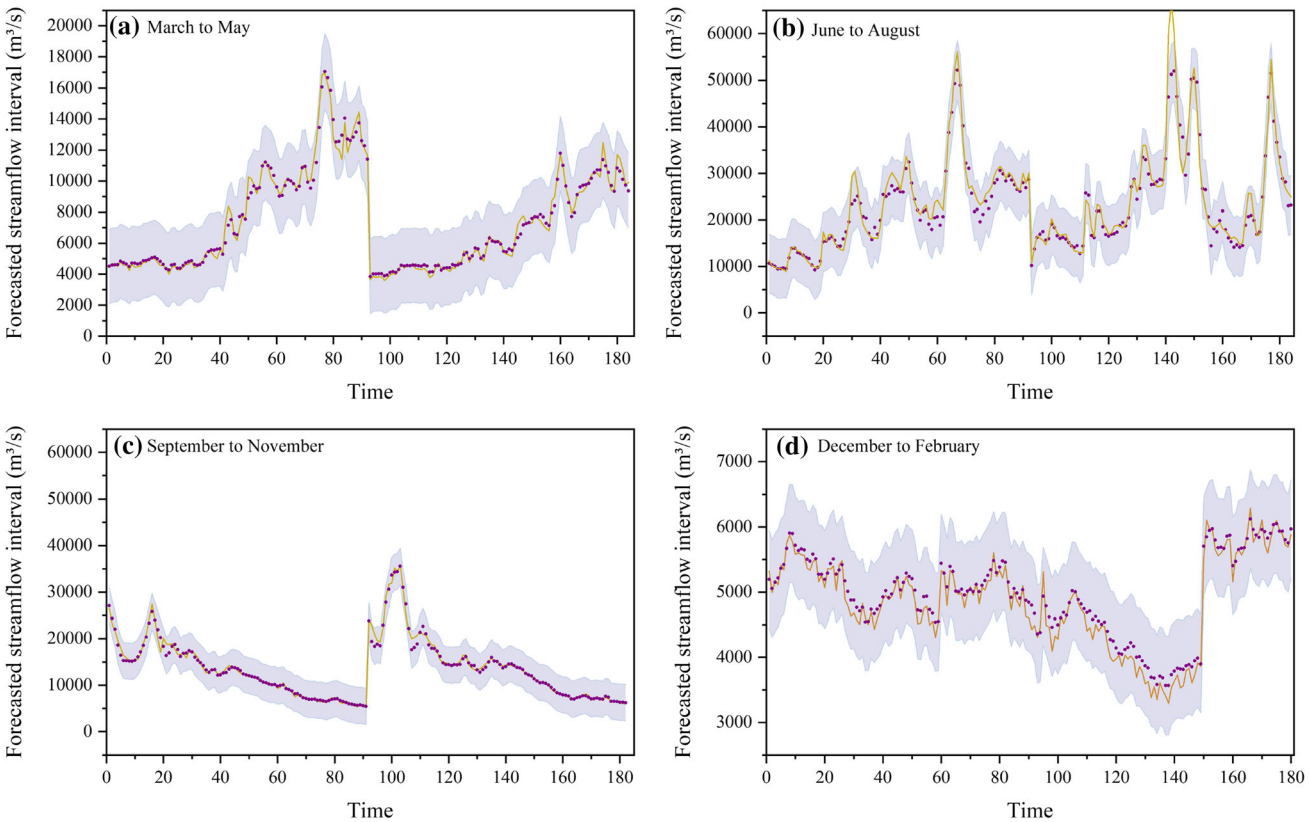

Fig. 9 Daily streamflow forecast interval using segment strategy (the yellow solid line represents the true value, the purple dot represents the predicted value, and the gray block represents the forecast interval)

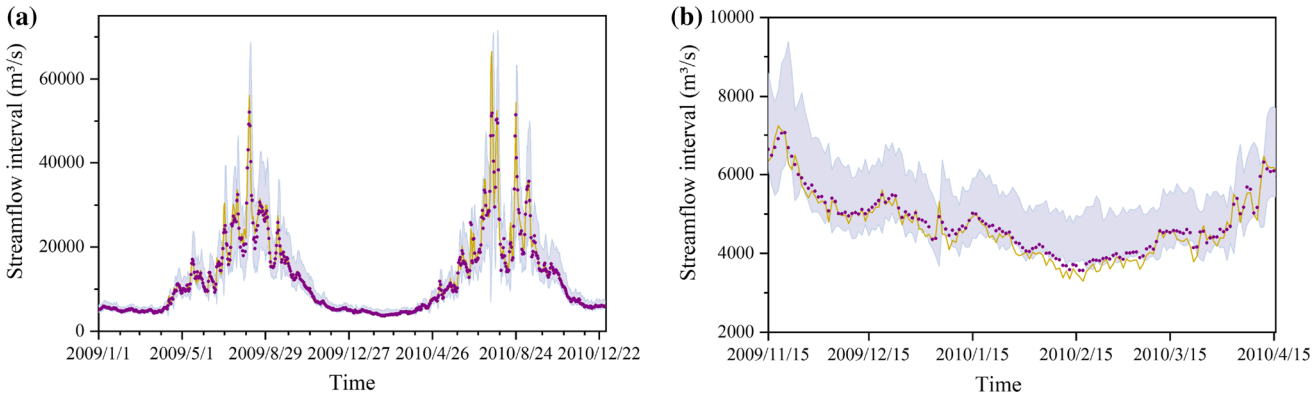

Fig. 10 Daily streamflow forecast interval using Box-Cox transformation (the yellow solid line represents the true value, the red dot represents the predicted value, and the gray block represents the forecast interval)

The performance of methods of constructing variable prediction interval is evaluated with the POC and the AIW. Better probabilistic forecasts are detected with a higher POC and a lower AIW. The POC and AIW values of HetGP, HomGP with segment strategy and HomGP with Box-Cox transformation are shown in Table 3. POC and AIW of LSTM-HetGP are 0.94 and 5615 m^3/s, respectively. The POC of LSTM-HetGP is comparable to that of Box-Cox transformation, and AIW is much lower than segment strategy and Box-Cox transformation. Therefore, aggregating LSTM with HetGP for streamflow interval forecasts is convenient and useful.

4.5 Comparison study

Figure 11 shows the streamflow forecasting error statistics of LSTM-HetGP, ensemble ANN, HetGP, GLM and regular LSTM with a radar chart. Each row of graphs sequentially represents the error statistics for Beibei, Gaochang, Pingshan and Yichang. Each column sequentially represents the values of RMSE, MRE, MSLE, NSE and POC. ANN and GLM are benchmark models for seeing the improvement of the proposed model over the traditional model. HetGP and LSTM are developed to investigate the real benefit of the hybrid model. The ensemble technique is adopted for probabilistic ANN forecasting Yuan et al. (2018b). HetGP is within the Bayesian framework and has probabilistic property. GLM is specified by univariate independent response variables and the canonical link function. The confidence interval is derived based on the asymptotic variance of the maximum likelihood estimate. Therefore, LSTM-HetGP, ensemble ANN, HetGP, GLM can provide prediction interval and regular LSTM has only point forecast. The statistical indicators RMSE, MRE, NSE, MSLE are used to evaluate point forecasting and POC is used to evaluate interval forecasting. A model shows satisfying results with low RMSE, MRE, MSLE values, and high NSE and POC values. For Beibei, RMSE, MRE, MSLE values of LSTM-HetGP are lowest and NSE value is highest in all developed models. POC of LSTM-HetGP is as high as that of HetGP. Therefore, the prediction accuracy of LSTM-HetGP is better than other models, and the prediction interval is comparable to HetGP. For Gaochang, Pingshan and Yichang streamflow forecasting, MRE, NSE, MSLE values are much better than LSTM, ANN, GLM and HetGP. The accuracy of LSTM-HetGP is demonstrated well. Moreover, it is more valuable than LSTM because it has a reliable prediction interval. Tyralis and Koutsoyiannis (2014) and Tyralis and Koutsoyiannis (2011) proved that predictive uncertainties are higher when the degree of modeled autocorrelation increases. Special LSTM structure brings in the advantages of solving the gradient explosion problem as well as decreasing uncertainties by reducing long-term sequence dependence. The aggregation of the GP model provides LSTM-HetGP additional interval information. Results indicate that the performance of LSTM-HetGP is satisfying, it improves the accuracy of simple benchmark models due to the advanced LSTM structure and also maintains the superior prediction interval of GPR.

Table 3 The performance of constructing prediction interval by using HetGP, HomGP with segment strategy and HomGP with Box-Cox transformation

Method	POC	AIW (m^3/s)
HetGP	0.94	5615
HomGP with segment strategy	0.86	6686
HomGP with Box-Cox transformation	0.95	7159

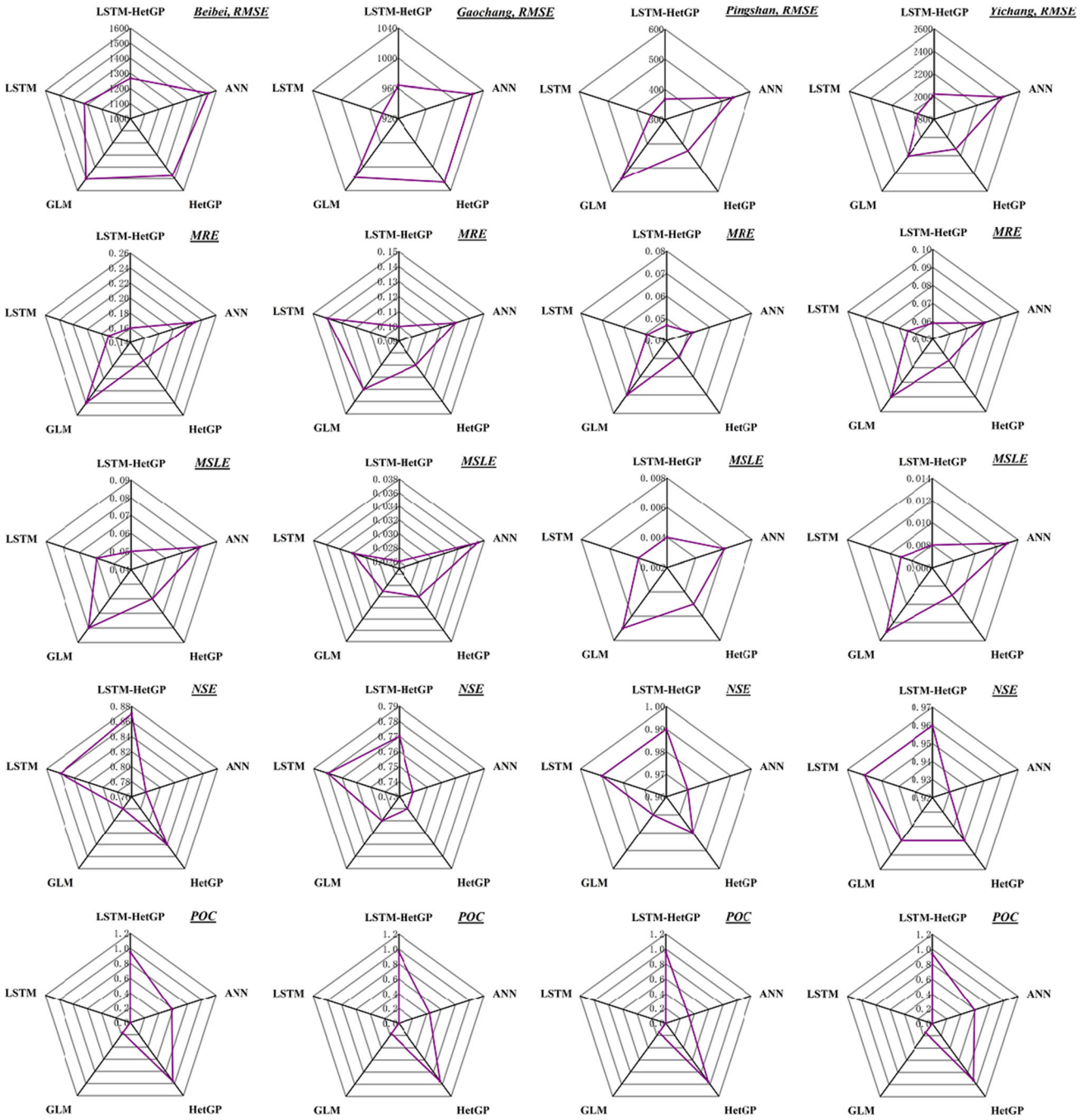

Fig. 11 Streamflow forecast error using LSTM-HetGP, ANN, HetGP, GLM and regular LSTM, error indicators include RMSE (m^3/s), MRE, MSLE, NSE, and POC

4.6 High flow forecasting

High flow forecasting is a difficult issue in streamflow forecasting due to complex uncertainty and randomness. Previous analysis on streamflow series shows that maximum daily streamflow occurs mainly in July, August, and September. To further evaluate the performance of the proposed model during high flow seasons, here we extracted forecasting error series (July 15 to September 15, 2009; July 15 to September 15, 2010), the boxplots of forecast errors produced by LSTM-HetGP, ensemble ANN, HetGP, GLM and regular LSTM are illustrated in Fig. 12. Figure 12(a) is for Yichang, (b) is for Beibei, (c) is for Gaochang and (d) is for Pingshan. The green block represents 0.25 and 0.75 quantiles of the error distribution. Black dots represent the error extremes. Figure 12 shows

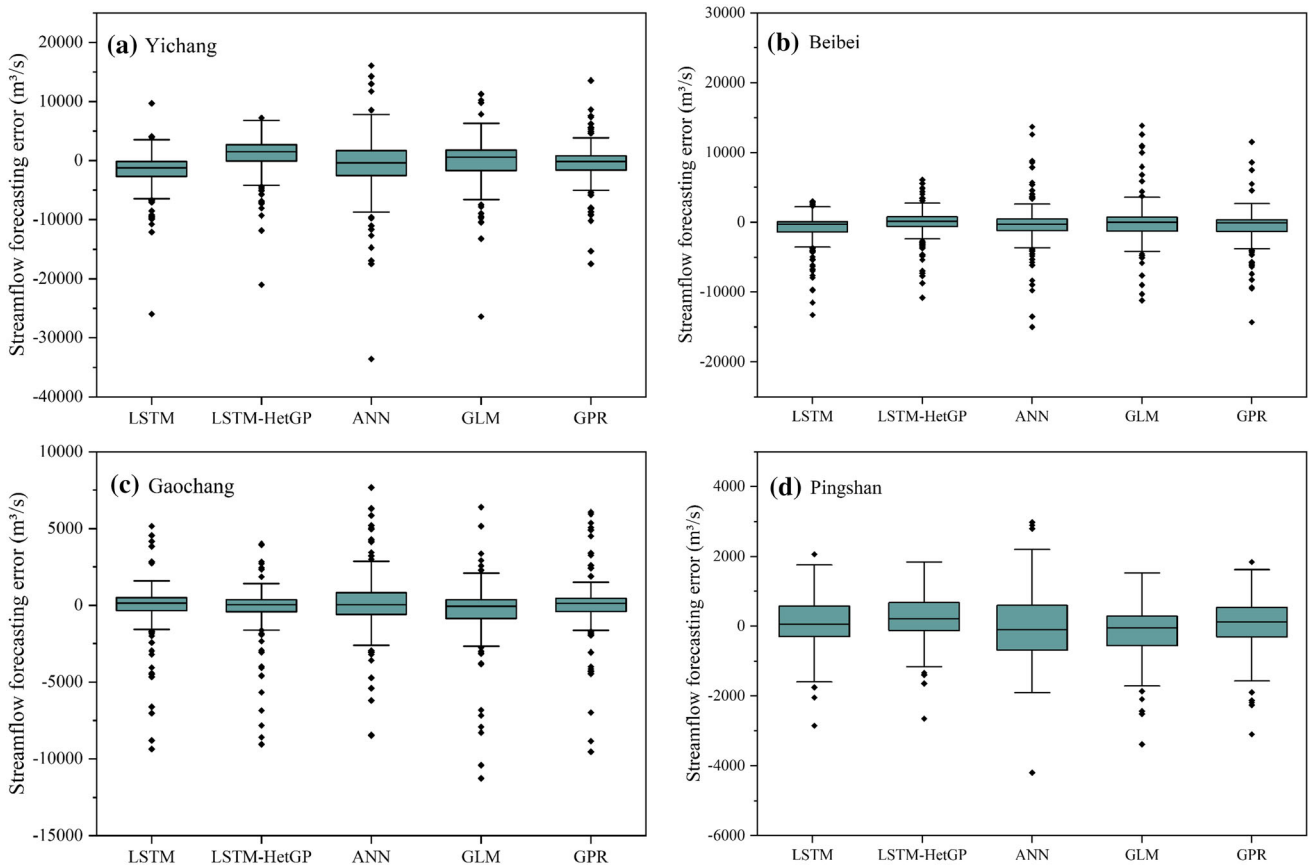

Fig. 12 The boxplots of forecast errors produced by LSTM-HetGP, ANN, HetGP, GLM and regular LSTM

that, generally, the ranges of forecast errors of LSTM-HetGP are narrower than the others, the forecast errors of ANN is the most decentralized. In Fig. 12a, d, the error ranges of 0.25 and 0.75 quantiles produced by LSTM-HetGP are similar to other models, but extreme errors produced by LSTM-HetGP is fewer. From the above analysis, the performance of LSTM-HetGP for high flow is accepted for the cases in the upper Yangtze River.

4.7 Discussion

A comparison of LSTM-HetGP and ANN, HetGP, GLM, regular LSTM streamflow forecasts were provided in the above analysis. Compared with the traditional ANN models, the stochastic models and the linear models, the results show that the deep learning models demonstrate excellent streamflow prediction ability. This attribute makes deep learning have great potential in hydrological prediction and analysis. The hydrological process is featured with great randomness and uncertainty. The LSTM-HetGP model proposed in this paper combines the advantages of deep learning and traditional stochastic models, has a fairly high prediction accuracy, and also provides a reliable prediction interval. Additionally, results indicate that using multiple metrics to assess forecast performance is important. Relying on a single metric for evaluating forecast performance can lead to sub-optimal conclusions. For example, if one considers the RMSE metric alone, the LSTM model yields the almost best performance. Note that the RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values (Chai and Draxler 2014), it reflects the prediction accuracy of high flow. While the performance metrics based on the RMSE, MRE, NSE, and MSLE comprehensively reflect a combination of reliability, sharpness, and bias characteristics of the forecasts. The computation time metric needed to optimize the model is not considered here, but it is especially significant for the case that multiple time series should be forecasted (Papacharalampous et al. 2018c). These findings highlight the value of multiple independent performance metrics (Clark et al. 2012). When demonstrating an operational forecast model for locations with streamflow regimes diverse and variable, it is essential to thoroughly evaluate multiple modeling methods over multiple locations to ensure the findings are sufficiently robust and general (Charles et al. 2018). We considered the mainstream and tributaries of the Yangtze River in this study. This set-up yields valuable insights into spatial

patterns in the forecast performance of the proposed model. It is found that streamflow forecasting models perform relatively well in catchments in the Jinsha River, and relatively worse in Min River and Jialing River. The evaluation over a larger number of catchments in different hydro-climatic regions is clearly beneficial to establish the robustness of streamflow forecasting methods. Papacharalampous et al. (2018a) and Papacharalampous et al. (2019) conducted larger scale studies, using thousands of time series, multistep ahead forecasting and numerous statistical and machine learning models. Therefore, with sufficient data support, systematic experiments in different hydro-climatic regions in the Yangtze River Basin could be investigated in the future. The results clearly show that LSTM-HetGP improves forecast precision while maintaining forecast accuracy and reliability. This improvement in forecast quality offers an opportunity to improve operational planning and management of water resources, which requires an understanding of the current and future availability of water. There are several opportunities to further improve the daily streamflow forecasting model. In this study, the parameters are estimated using a training data set of five years, and hence could be subject to the overfitting problem. To achieve better generalization, cross-validation was employed in this paper. A compound RMSE, rather than the simple early stopping method was used as the objective function in cross-validation. Meanwhile, the dropout regularization method was used during training. The impact of default values of the hyperparameters on the case studies needs more attention. As many sites of potential application may lack the data length available in this work, the sensitivity of forecast performance to the length of the calibration period warrants further investigation. The use of "pre-trained LSTM" is a promising way to reduce the large data-demand for an individual basin.

5 Conclusion

Daily streamflow forecasting plays an extremely important role in the optimal management of water resources, particularly for operational hydrology purposes. Long short-term memory (LSTM) streamflow forecast model has received a lot of attention in recent years due to its powerful non-linear simulation ability. But the probabilistic streamflow forecasting based on LSTM is still insufficient as large scale ensemble technique based on resampling is difficult to implement. This paper proposed a more direct and reliable probabilistic streamflow forecast model LSTM by coupling Gaussian process regression (GPR) to the internal structure of LSTM. Importantly, considering that changing mean and variance over time exist in the daily streamflow time series, and the distribution of streamflow tends to be skewed, the heteroscedastic GPR (HetGP) was adopted to produce a varying predictive streamflow interval. As HetGP acted as an internal unit of the LSTM network, the proposed models would maintain the advanced prediction ability of deep learning and provide a direct confidence interval of prediction. It enriches the information content of streamflow prediction and makes the daily streamflow prediction result more meaningful.

The performance of the LSTM-HetGP model was demonstrated by studying the daily streamflow forecasting in the mainstream and tributaries of the upper Yangtze River; Beibei, Gaochang, Pingshan, and Yichang. Pingshan is located in the outlet of the Jinsha River, which has a high altitude and dry climate. Snow is the main supplement of streamflow. In addition to the Jinsha River in the upper reaches of the Yangtze River, heavy rains frequently occurred in the rest of the areas. Heavy rains are the main supply of runoff in the Jialing River and the Min River. The proposed model is compared with ensemble ANN, HetGP, GLM, and regular LSTM.

The main findings are as follows:

1. The streamflow forecasting results of the Pingshan are the best among the four selected stations. The forecast results of Gaochang and Beibei are relatively poor. The analysis is that the floods caused by heavy rains in Beibei and Gaochang steeply rise and fall, it makes forecasting difficult. Using more meteorological predictors can improve the accuracy of the forecast.
2. The statistical indicators for deterministic streamflow forecasting and associated prediction interval estimation were used to evaluate the performance of LSTM-HetGP, ensemble ANN, HetGP, GLM for comparison. For the streamflow forecasting in the upper Yangtze River, it was found that the performance of LSTM-HetGP is satisfying, it improves the accuracy of simple benchmark models due to the advanced LSTM structure and also maintains the superior prediction interval of GPR. High flow forecasting is a difficult issue in streamflow forecasting due to complex uncertainty and randomness. The performance of LSTM-HetGP is also accepted for the high flow forecasting in the upper Yangtze River.
3. The important advantage of the proposed LSTM-HetGP lies in the adaptive prediction interval, which is greatly significant for water resources decision making. To demonstrate the benefit of the HetGP for constructing interval range, traditional segment strategy, and Box-Cox data transformation were also studied. POC and AIW analysis show that aggregating LSTM with HetGP for streamflow interval forecasts is convenient and useful.

LSTM will always strongly rely on the available data for calibration. For the usage within a single basin with limited data, the selection of a training data set is crucial for LSTM or any other data-driven model. Reliance on just one split sample of the training set may not yield a parameter set with good generalization capability. To achieve better generalization, cross-validation and dropout regularization were used during the training period to effectively avoid overfitting problems and improve model performance. However, there are still limitations to this study. The computation time needed to optimize the model could be considered. The impact of default values of the hyperparameters on the case studies also needs more attention. The forecasting approaches developed in this work can support improved water management in the drainage system considered. It becomes possible to produce streamflow forecasts with high reliability and reduced bias, and provide useful probabilistic streamflow in the next day.

Acknowledgements This work is supported by the National Natural Science Foundation of China (51809242), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (G1323541875, G1323519436)

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Akram M, El C (2016) Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int J Comput Appl 143(11):7–11

Al-Shedivat M, Wilson AG, Saatchi Y, Hu Z, Xing EP (2016) Learning scalable deep kernels with recurrent structure. arXiv preprint arXiv:161008936

Binois M, Gramacy RB, Ludkovski M (2018) Practical heteroscedastic gaussian process modeling for large simulation experiments. J Comput Graph Stat 27(4):808–821

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding rmse in the literature. Geosci Model Dev 7(3):1247–1250

Charles SP, Wang QJ, Ahmad MUD, Hashmi D, Schepen A, Podger G, Robertson DE (2018) Seasonal streamflow forecasting in the upper indus basin of pakistan: an assessment of methods. Hydrol Earth Syst Sci 22(6):3533–3549

Chen L, Singh VP, Guo S, Zhou J, Zhang J (2015) Copula-based method for multisite monthly and daily streamflow simulation. J Hydrol 528:369–384

Chollet F et al (2015) Keras

Clark MP, Kavetski D, Fenicia F (2012) Reply to comment by K. J. Beven et al. on pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour Res 48(11):802–808

De Vos NJ, Rientjes THM (2008) Multiobjective training of artificial neural networks for rainfall runoff modeling. Water Resour Res 44(8):134–143

El-Shafie A, Alsulami HM, Jahanbani H, Najah A (2013) Multi-lead ahead prediction model of reference evapotranspiration utilizing ann with ensemble procedure. Stoch Environ Res Risk Assess 27(6):1423–1440

Elman JL (1990) Finding structure in time. Cogn Sci 14(2):179–211

Fu W, Wang K, Li C, Tan J (2019a) Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. Energy Convers Manag 187:356–377

Fu W, Wang K, Zhang C, Tan J (2019b) A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine. Trans Inst Meas Control 41(15):4436–4449

Gers FA, Schmidhuber JA, Cummins FA (2002) Learning to forget: continual prediction with LSTM. In: Artificial neural networks, 1999. ICANN 99. Ninth international conference on, p 2451

Gibbs MN (1998) Bayesian gaussian processes for regression and classification. Ph.D. thesis, Citeseer

Gruet MA, Chandorkar M, Sicard A, Camporeale E (2018) Multiple-hour-ahead forecast of the dst index using a combination of long short-term memory neural network and gaussian process. Space Weather 16(11):1882–1896

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and nse performance criteria: implications for improving hydrological modelling. J Hydrol 377(1):80–91

Guttag J (2016) Introduction to computation and programming using python: with application to understanding data. MIT Press, Cambridge

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hogue TS, Sorooshian S, Gupta H, Holz A, Braatz D (2000) A multistep automatic calibration scheme for river forecasting models. J Hydrometeorol 1(6):524–542

Ishak S, Kotha P, Alecsandru C, Student G (2003) Optimization of dynamic neural network performance for short-term traffic prediction. Transp Res Rec 1836(1):27–31

Karimi S, Shiri J, Kisi O, Xu T (2018) Forecasting daily streamflow values: assessing heuristic models. Hydrol Res 49(3):658–669

Kersting K, Plagemann C, Pfaff P, Burgard W (2007) Most likely heteroscedastic gaussian process regression. In: Proceedings of the 24th international conference on Machine learning, ACM, pp 393–400

Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399(1):132–140

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

Lázaro-Gredilla M, Titsias MK (2011) Variational heteroscedastic gaussian process regression. In: ICML, pp 841–848

Lei Y, Zhou J, Zeng X, Guo J, Zhang X (2014) Multi-objective optimization for construction of prediction interval of hydrological models based on ensemble simulations. J Hydrol 519:925–933

Lu X, Wang X, Zhang L, Zhang T, Yang C, Song X, Yang Q (2018) Improving forecasting accuracy of river flow using gene expression programming based on wavelet decomposition and de-noising. Hydrol Res 49(3):711–723

Moradkhani H (2015) Statistical-dynamical drought forecast within bayesian networks and data assimilation: how to quantify drought recovery. In: EGU general assembly conference abstracts, vol 17

Papacharalampous G, Tyralis H, Koutsoyiannis D (2018a) One-step ahead forecasting of geophysical processes within a purely statistical framework. Geosci Lett 5(1):12

Papacharalampous G, Tyralis H, Koutsoyiannis D (2018b) Predictability of monthly temperature and precipitation using automatic time series forecasting methods. Acta Geophys 66(4):807–831

Papacharalampous G, Tyralis H, Koutsoyiannis D (2018c) Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from greece. Water Resour Manag 32(15):5207–5239

Papacharalampous G, Tyralis H, Koutsoyiannis D (2019) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Stoch Environ Res Risk Assess 33(2):1–34

Rasmussen CE (2003) Gaussian processes in machine learning. In: Summer school on machine learning, Springer, Berlin, pp 63–71

Rasmussen CE (2004) Gaussian processes in machine learning. In: Advanced lectures on machine learning, Springer, Berlin, pp 63–71

Shen C (2018) A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour Res 54(11):8558–8593

Shen C, Laloy E, Elshorbagy A, Albert A, Bales J, Chang FJ, Ganguly S, Hsu KL, Kifer D, Fang Z, Fang K, Li D, Li X, Tsai WP (2018) Hess opinions: incubating deep-learning-powered hydrologic science advancesas a community. Hydrol Earth Syst Sci 22(11):5639–5656

Siqueira H, Boccato L, Luna I, Attux R, Lyra C (2018) Performance analysis of unorganized machines in streamflow forecasting of brazilian plants. Appl Soft Comput 68:494–506

Sun AY, Wang D, Xu X (2014) Monthly streamflow forecasting using gaussian process regression. J Hydrol 511:72–81

Tolvanen V, Jylänki P, Vehtari A (2014) Expectation propagation for nonstationary heteroscedastic gaussian process regression. In: Machine learning for signal processing (MLSP), 2014 IEEE international workshop on, IEEE, pp 1–6

Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. Stoch Environ Res Risk Assess 25(1):21–33

Tyralis H, Koutsoyiannis D (2014) A bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. Clim Dyn 42(11–12):2867–2883

Wang W, Jin J, Li Y (2009) Prediction of inflow at three gorges dam in Yangtze River with wavelet network model. Water Resour Manag 23(13):2791–2803

Wilson AG, Hu Z, Salakhutdinov R, Xing EP (2016) Deep kernel learning. In: Artificial intelligence and statistics, pp 370–378

Wu W, Chen K, Qiao Y, Lu Z (2016) Probabilistic short-term wind power forecasting based on deep neural networks. In: 2016 International conference on probabilistic methods applied to power systems (PMAPS), IEEE, pp 1–8

Yaseen ZM, Sulaiman SO, Deo RC, Chau KW (2019) 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 569:387–408

Yuan X, Abouelenien M (2015) A multi-class boosting method for learning from imbalanced data. Int J Granul Comput Rough Sets Intell Syst 4(1):13

Yuan X, Sarma V (2010) Automatic urban water-body detection and segmentation from sparse alsm data via spatially constrained model-driven clustering. IEEE Geosci Remote Sens Lett 8(1):73–77

Yuan X, Chen C, Lei X, Yuan Y, Adnan RM (2018a) Monthly runoff forecasting based on lstm-alo model. Stoch Environ Res Risk Assess 32(8):2199–2212

Yuan X, Xie L, Abouelenien M (2018b) A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recognit 77:160–172

Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol 561:918–929

Zhang J, Yan J, Infield D, Liu Y, Lien F (2019) Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl Energy 241:229–244

Zhu S, Luo X, Xu Z, Ye L (2018) Seasonal streamflow forecasts using mixture-kernel GPR and advanced methods of input variable selection. Hydrol Res 50(1):200–214