

Research papers

Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm

Di Zhang^a, Junqiang Lin^{a,*}, Qidong Peng^a, Dongsheng Wang^b, Tiantian Yang^c, Soroosh Sorooshian^c, Xuefei Liu^a, Jiangbo Zhuang^a

^a State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, China

^b China Renewable Energy Engineering Institute, Beijing, China

^c Department of Civil and Environmental Engineering, Center for Hydrometeorology and Remote Sensing [CHRS], University of California-Irvine, Irvine, CA, USA

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:

Reservoir operation
Artificial intelligence
BP neural network
SVR
LSTM

ABSTRACT

Reservoirs and dams are vital human-built infrastructures that play essential roles in flood control, hydroelectric power generation, water supply, navigation, and other functions. The realization of those functions requires efficient reservoir operation, and the effective controls on the outflow from a reservoir or dam. Over the last decade, artificial intelligence (AI) techniques have become increasingly popular in the field of streamflow forecasts, reservoir operation planning and scheduling approaches. In this study, three AI models, namely, the backpropagation (BP) neural network, support vector regression (SVR) technique, and long short-term memory (LSTM) model, are employed to simulate reservoir operation at monthly, daily, and hourly time scales, using approximately 30 years of historical reservoir operation records. This study aims to summarize the influence of the parameter settings on model performance and to explore the applicability of the LSTM model to reservoir operation simulation. The results show the following: (1) for the BP neural network and LSTM model, the effects of the number of maximum iterations on model performance should be prioritized; for the SVR model, the simulation performance is directly related to the selection of the kernel function, and sigmoid and RBF kernel functions should be prioritized; (2) the BP neural network and SVR are suitable for the model to learn the operation rules of a reservoir from a small amount of data; and (3) the LSTM model is able to effectively reduce the time consumption and memory storage required by other AI models, and demonstrate good capability in simulating low-flow conditions and the outflow curve for the peak operation period.

1. Introduction

Half of the major global river systems are affected by reservoirs and dams, and human beings manage and utilize water resources through reservoirs for power generation, water supply, navigation, disaster prevention, flood control and mitigation, drought relief (Dynesius and Nilsson, 1994; WCD, 2000; ICOLD, 2011; Lehner et al., 2011; Shang et al., 2018). In recent years, many countries (including China) have also actively adopted reservoir operations to mitigate the adverse effects of reservoirs and maintain the health of river ecosystems. The scientific calculation, simulation and prediction of reservoir storage or release, as well as the development of proper reservoir operation plans are important to achieve all types of reservoir functions and to avoid danger to humans and river ecology (Loucks and Sigvaldason, 1981)..

Starting in the 1980s, with the development of hydrology, hydraulics and river dynamics, conceptual or physical-based models (such

as HEC-ResSim, WEAP21, etc.) have been proposed and are widely used in reservoir hydrological process simulation and reservoir operation decisions (Klipsch and Hurst, 2003; Yates et al., 2005). Such models transform the empirical, mechanical, and blind operation patterns of early reservoir operations that were based on historical hydrological statistics, operated by so-called rule curves. Physical-based models provide a more practical physical and mathematical basis for the calculation of controlled releases or storage (See Table 1).

However, the practical application scenarios of reservoir operation are extremely complex and involve multiple time scales and multiflow regimes, often accompanied by occasional emergencies. A reservoir should undertake the medium- and long-term (seasonal and monthly scale) operation task of managing downstream water supply and optimization of economic benefit. Reservoirs should also undertake short-term (daily and hourly scale) operation tasks of managing power grid load, water demand, navigation and stimulation of fish breeding,

* Corresponding author.

E-mail address: linjq@iwhr.com (J. Lin).

Table 1
Detailed information of the reservoir operation data.

Operation data name	Unit	Resolution	Normal value of accumulated year	Normal value of flood season
Reservoir inflow	m^3/s	Two or four hours	13,368	22,368
Reservoir outflow	m^3/s	Two or four hours	13,269	22,495
water level upstream of the dam	m	Two or four hours	65.06	65.40
water level downstream of the dam	m	Two or four hours	43.49	46.92

disaster prevention, emergency operations during floods, droughts. These various scheduling scenarios illustrate that the actual operation process of a reservoir is rapidly changing and often deviates from the operation plan. These deviations often make it difficult for the physical model based on the operation rule to accurately simulate reservoir operation and predict the reservoir controlled releases (Johnson et al., 1991; Oliveira and Loucks, 1997). In addition, when the physical model needs to be rebuilt with a new scheduling rule, the demand for the professional expertise of the reservoir operator is high, and the calculation time of the model cannot meet the requirements of emergency operation. Reservoir operation is the result of multiple factors with strongly nonlinear interactions, which are influenced by natural conditions, such as precipitation, runoff, agricultural irrigation and human needs, such as industrial production water consumption, power grid peak shaving, flood peak shaving. These complex factors have uncertainty and increase the difficulty of using physical-based models.

In recent years, with the development of artificial intelligence (AI) and big data mining technology, data-driven AI models have become important in various fields. This kind of model does not heavily rely on physical meaning, but is good at solving nonlinear simulation and prediction problems that are influenced by multiple complex factors. At present, AI models have been successfully extended to the reservoir operation field. In contrast to physical-based models, AI models have the ability to autonomously learn the various reservoir operation rules from a large amount of hydrological data and the real-time reservoir operation data. Moreover, AI models need low professional requirements from operators and have fast response speeds (Hejazi and Cai, 2009).

Among the many AI models, artificial neural networks (ANN) and support vector machine or regression (SVM or SVR) are the two most typical models in the field of reservoir operation. ANN models benefit from the proposed backpropagation algorithm (BP). The BP solves the training problem of the neural network, which gives the ANN models good nonlinear prediction ability. Many scholars have successfully promoted ANN in the reservoir operation field (Thirumalaiah and Deo, 1998; Jain et al., 1999; Chaves and Chang, 2008). Then, to further improve the accuracy of the ANN model, some scholars coupled the ANN algorithm with other AI algorithms and explored the application of the improved ANN algorithm in reservoir management. For example, Chaves and Chang (2008) improved ANN by combining them with a genetic algorithm and verified the applicability of the improved ANN in reservoir operation simulation. Chen and Chang (2009) combined evolutionary algorithm and ANN and proposed a new evolutionary-ANN algorithm for reservoir inflow prediction.

With increased ANN model research, the limitations of ANN have been highlighted, such as local optimal solutions and gradient disappearance, which limit the application of the model (Yang et al 2017a). At this time, the SVM algorithm invented by Cortes and Vapnik (1995) is better than ANN in many aspects, with fast training speed and global optimal solutions. The SVR algorithm is derived from SVM, which is similar to the SVM algorithm, and it is one of the most widely

used AI models in the reservoir operation field (Lin et al., 2006; Hipni et al., 2013; Yang et al., 2017b). Meanwhile, some scholars coupled the SVR algorithm with other AI algorithms and explored the application of the improved SVR algorithm in reservoir management (Khalil et al., 2005; Su et al., 2013; Ji et al., 2014; Aboutalebi et al., 2015). For example, Aboutalebi et al. (2015) coupled the nondominated sorting genetic algorithm and SVR algorithm and applied the coupled model to optimize reservoir operation rules.

In addition to the above two classic AI algorithms, many other AI algorithms have been successfully applied to the reservoir operation field, such as genetic algorithm (GA), adaptive network-based fuzzy inference system (ANFIS), decision tree (DT). Chang and Chang (2001) and Chang et al. (2005) coupled the GA and ANFIS and applied the coupled model to estimate reservoir storage or release. Yang et al. (2016) used the improved DT algorithm, classification and regression tree, to reasonably estimate the storage or release of 9 reservoirs in California.

Although the above AI algorithms have been proved to be applicable to the estimation of reservoir storage or release, those algorithms still have some shortcomings, such as insufficient feature extraction capability and longer time consumption. In recent years, a new type of machine learning method, i.e., deep learning, has gradually become the frontier of computer science and technology and has achieved great success in the fields of computer vision, speech recognition and natural language processing. Deep learning, derived from ANN, is a new field in machine learning research. This algorithm has been proven as an abstract, high-level representation of attribute categories or characteristics through the combination of low-level features and can significantly improve recognition accuracy (Girshick et al., 2014; Lecun et al., 2015). LSTM model is a widely used deep learning model, which is applied to hydrological forecasting because of its ability to solve complex scheduling problems (Zhang et al., 2018). Zaytar and Amrani (2016) and Zhang et al. (2018) applied the LSTM model to forecast weather and urban sewage pipeline overflow, respectively. They obtained satisfactory results and verified the validity of LSTM in the prediction of timing problems. Shi et al. (2015) improved the traditional LSTM model, proposed a convolutional LSTM (ConvLSTM) and used it to build an end-to-end trainable model for the precipitation nowcasting problem on the spatial and temporal scale, and the application of the LSTM model has been extended from a one-dimension temporal sequence to a two-dimension spatial and temporal sequence. Because LSTM is a new type of deep learning model, it has few reports in the field of reservoir operation.

In recent years, research on AI models in the field of reservoir operation has developed rapidly, but there are still many shortcomings. First, at present, AI model research focuses on a specific case problem (often a single time scale or flow regime) and lacks a systematic comparison of the simulation effect of the model with complex operation scenarios (multiscale and multiflow regime). Second, the deep learning model as a popular AI model, has a strong ability to address the time series problem, but whether the model can address the reservoir operation problem effectively and accurately is unknown. Third, the parameter setting is the key technology of AI model building. However, investigations of different parameters among those models and comprehensive comparison studies are rarely reported.

Therefore, in this study, we selected three AI models, (1) a benchmark three-layer backpropagation (BP) neural network, (2) an SVR technique, and (3) the long short-term memory (LSTM) model, and constructed a reservoir operation model with three time scales including hourly, daily, and monthly scale to analyze the sensitivity of applying AI models to reservoir operation. For case study, we choose Gezhouba (GZB) reservoir in China (which had relatively complete, detail and long sequence operation records) to test the simulation performance of three models at various flow regimes, including (1) low flow, (2) intermediate flow, and (3) high flow. In summary, the goals of this study are (1) to summarize the influence of the parameter settings

on model performance and propose parameter settings for different AI models in assisting reservoir operation; (2) to compare the simulation results under different time scales and flow regimes and propose suggestions for the applicability of AI models under different inflow scenarios; and (3) to explore the superiority of the LSTM model over traditional AI models in assisting reservoir operations, and investigate the improvement of prediction accuracy and calculation speed.

2. Methodology

2.1. BP neural network

ANN are mathematical models of biologically motivated computation (Haykin, 1994), and they are known as flexible modeling tools with the ability to provide a neural computing approach for processing complex problems that might otherwise not be solved with a mathematical formula (Yang et al., 2017a). ANN application in hydrology started in the early 1990s. In the last decade, ANN have been applied to hydrology, including rainfall-runoff modeling (Wu and Chau, 2011; Chiang et al., 2004), streamflow forecasting (Moradkhani et al., 2004; Anctil et al., 2004), ground water (Johnson and Rogers, 2000) reservoir operation (Chaves et al., 2004; Jain et al., 1999; Cheng et al., 2015).

In this study, a feed-forward neural network is used in combination with an error backpropagation training algorithm, namely, the BP neural network. The BP neural network consists of one input layer, one or more hidden layers and one output layer (Fig. 1). The important issues in the establishment of topological structure include the determination of the number of hidden layers, the number of hidden neurons (nodes) and the transfer function. The Kolmogorov theorem has certified that a single hidden layer is competent for ANN to approximate any complicated nonlinear function and establish a nonlinear mapping between the input and output layers. Therefore, this study establishes a three-layer BP neural network; the topological configuration and the selection of the activation function are illustrated

in Fig. 1. In contrast to the signal, the error is backward propagated, and in the process of backward propagation, the weights and deviations of the network are gradually adjusted to complete the training. The interested readers are referred to Thirumalaiah and Deo, 1998, Jain and Srinivasulu, 2004, Fernando and Shamseldin, 2009, Senthil kumar et al., 2012 for further details.

2.2. SVR

The basic concept of SVR is to nonlinearly map the initial data in a higher dimensional feature space and to solve the linear regression problem in the feature space (Fig. 2). Therefore, SVR usually needs to build a suitable function $f(x)$ to describe the nonlinear relationship between feature x_i and target value y_i , as shown in the following Eq. (4):

$$f(x_i) = w \cdot \varphi(x_i) + b \quad 1$$

where w is the coefficient vector, $\varphi(x_i)$ is the transformation function, and w and b represent the weight and bias, respectively. w and b are estimated by minimizing the so-called regularized risk function, as shown in the following Eq. (5):

$$R(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n L_\epsilon(y_i, f(x_i)) \quad 2$$

where $\frac{1}{2} \|w\|^2$ is the regularization term; C is the penalty coefficient; $L_\epsilon(y_i, f(x_i))$ is the ϵ -insensitive loss function, which is calculated according to Eq. (6):

$$L_\epsilon(y_i, f(x_i)) = \max\{0, |y_i - f(x_i)| - \epsilon\} \quad 3$$

where ϵ denotes the permitted error threshold, so that ϵ will be ignored if the predicted value is within the threshold; otherwise, the loss equals a value greater than ϵ .

To solve the optimization boundary, two slack factors ξ^+ and ξ^- are introduced:

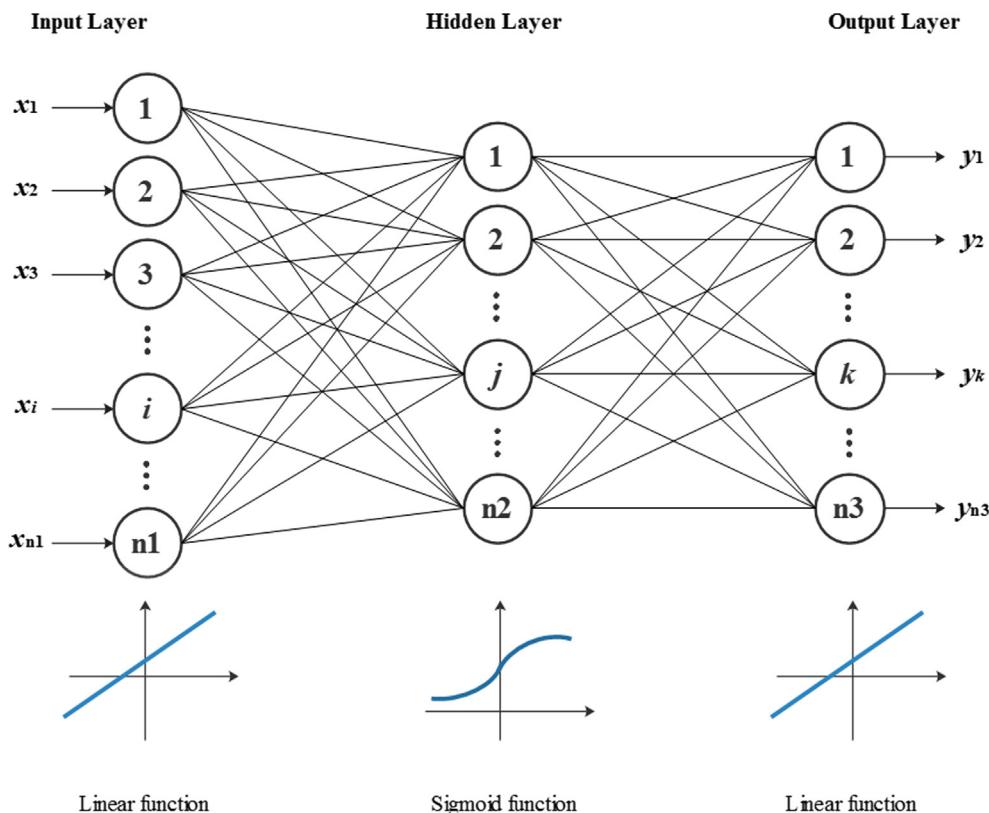


Fig. 1. Schematic diagram of the three-layer BP neural network.

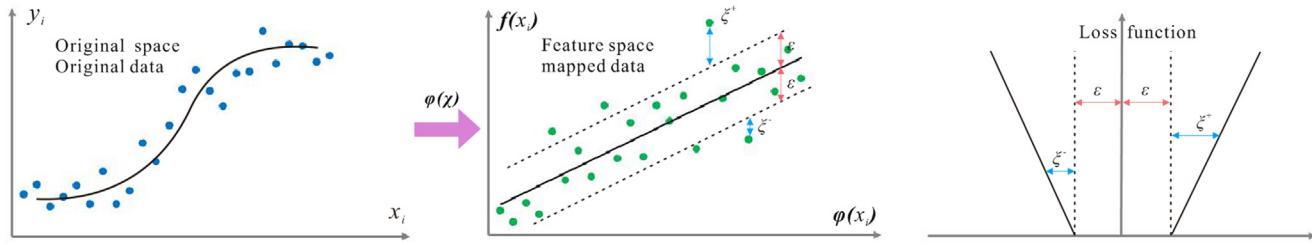


Fig. 2. SVR schematic diagram.

$$\min(w, \xi^-, \xi^+) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi^-, \xi^+)$$

Subject to

$$\begin{cases} y_i - [w \cdot \varphi(x_i)] - b \leq \varepsilon + \xi^-, \xi^- \geq 0 \\ [w \cdot \varphi(x_i)] + b - y_i \leq \varepsilon + \xi^+, \xi^+ \geq 0 \end{cases}$$

The key is to develop a Lagrangian function according to the objective function and the corresponding constraint conditions.

$$\begin{aligned} \max H(\partial_i^-, \partial_i^+) &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\partial_i^- - \partial_i^+) (\partial_j^- - \partial_j^+) K(x_i, x_j) + \sum_{i=1}^n y_i (\partial_i^- - \partial_i^+) \\ &\quad - \varepsilon \sum_{i=1}^n y_i (\partial_i^- + \partial_i^+) \end{aligned}$$

Subject to

$$\sum_{i=1}^n (\partial_i^- - \partial_i^+) = 0, \partial_i^-, \partial_i^+ \in [0, C]$$

Hence, the regression function is as follows:

$$f(x) = \sum_{i=1}^n (\partial_i^- - \partial_i^+) K(x_i, x_j) + b$$

where $K(x_i, x_j)$ is the kernel function. In this study, the simulation effect of kernel function is tested, include the Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid kernels, as shown in the following equation:

$$\begin{cases} \text{Linear kernel: } K(x, x_i) = x \cdot x_i \\ \text{Polynomial kernel: } K(x, x_i) = (\gamma(x \cdot x_i) + b)^d \\ \text{Radial Basis Function kernel: } K(x, x_i) = \exp(-\frac{\|x - x_i\|}{2\sigma^2}) \\ \text{Sigmoid kernel: } K(x, x_i) = \tanh(\gamma(x \cdot x_i) + v) \end{cases}$$

where d is the degree of the polynomial term, v represents the residuals, and γ is the structural parameter in the polynomial, RBF and sigmoid kernels. Different d , γ , and penalty coefficient C values will be tested in the paper.

2.3. Lstm

Long short-term memory is a complex recurrent model developed by Hochreiter and Schmidhuber (1997) to address the deficiencies of RNNs. In contrast to RNNs, LSTMs apply memory blocks to replace the hidden layer of RNNs. In this paper, the LSTM network consists of parallel memory blocks recurrently connected to each other, lying between an input layer and an output layer, and the conceptual illustration is shown in Fig. 3. The LSTM block was obtained from Gers et al. (2000), and it consists of an input gate, a memory cell, a forget gate, and an output gate, as shown in Fig. 4. The corresponding forward propagation equations are presented below. For our LSTM, the first step is to decide what information to remove from the cell state by the forget gate. The forget gate is essentially a sigmoid function. The forget gate will generate an f_t value between 0 and 1, based on the previous moment output h_{t-1} and current input x_t to decide whether to let the

information C_{t-1} that is produced in the previous moment pass or partially pass, which is depicted by Eq. (13):

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \quad 10$$

The next step is to decide what new information will be stored in the cell state. The step is divided into two parts: first, the input gate determines which values will be updated. Second, the memory cell generates a vector of new candidate values C_b , which can be added to the cell state. In the next step, we combine Eqs. (14) and (15) to update the state:

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \quad 11$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \quad 12$$

Finally, we decide what information we are going to output. First, we decide what parts of the cell state to output by running a sigmoid layer. Then, we resize the C_t value to -1 to 1 through \tanh and multiply it by the output of the sigmoid gate so that we output only the target information.

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \quad 13$$

$$h_t = o_t \cdot \tanh(C_t) \quad 14$$

In Eqs. (13)–(17), f , i , C , o and h are the forget gate, input gate, cell, output gate and hidden output, respectively. W_x and W_h in Eqs. (15)–(17) are the input and hidden weights for the gates or cells with the corresponding subscripts; b in Eqs. (15)–(17) represents learnable biases for the gates; and σ and \tanh represent the gate activation function (the logistic sigmoid in this paper) and the hyperbolic tangent activation function, respectively.

In this study, we establish a three-layer LSTM network that consists of one input layer, one LSTM hidden layer and one output layer, and we train the network with an algorithm termed backpropagation through time (BPTT), which has a similar principle to the BP algorithm, by using a backpropagation network to continuously adjust the weights and thresholds (Werbos, 1990).

3. Data and processing

3.1. Case reservoir and operation data

The Gezhouba (GZB) dam is the first dam across Yangtze River which built after the founding of New China (Fig. 5). It located in Yichang city of Hubei province, approximately 38 km downstream of the Three Gorges dam, plays a role in power generation, flood control, navigation, and fisheries, and provides other comprehensive benefits. GZB is a large-scale water project that controls a drainage area of $1 \times 10^6 \text{ m}^2$, accounting for more than half of the Yangtze River basin. The mean annual discharge of this reservoir is $14,300 \text{ m}^3/\text{s}$, the design-flood discharge is $86,000 \text{ m}^3/\text{s}$, the normal water level is 66.0 m , the maximum dam height is 53.8 m , and the total storage is $15.8 \times 10^8 \text{ m}^3$. GZB reservoir has relatively mature operating rules, and complete, detail and long sequence operation records. These records cover all kinds of operating scenarios, such as the flood control, drought relief, peaking generation, navigation and fish spawning stimulation etc., and

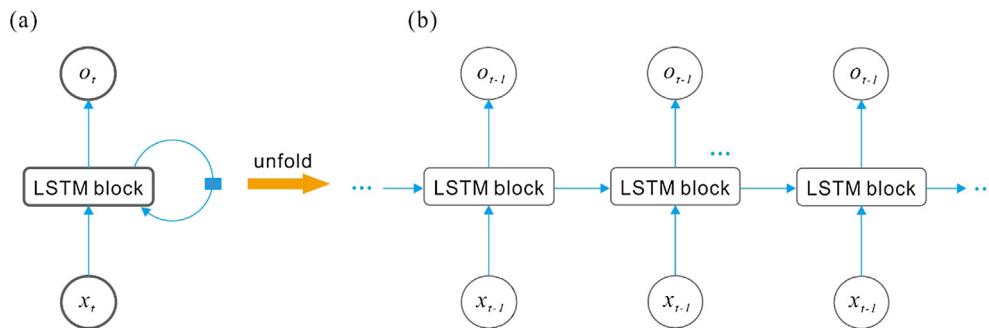


Fig. 3. The illustration of an LSTM network with one hidden layer. (a) Folded computational diagram, where the blue square indicates a delay of 1 time step. (b) Unfolded computational diagram, where each node is associated with one particular time instance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

provide a large number of operating data for machine learning in this study. The reservoir hydrological data and operation data were obtained from the official website of China Three Gorges Corporation (<http://www.ctg.com.cn/>). The GZB data span a period of more than 30 years, from 1982 to 2015. The data include the inflow and outflow of the reservoir and the water level upstream and downstream of the dam measured every six hours (or four hours, when the flood peak passes through the reservoir, all reservoir operating indicators are more intensive monitored). These data were aggregated into daily and monthly time scales.

3.2. Model building and inputs

In this study, we use selected artificial intelligence and data mining tools, namely, the BP neural network, SVR and LSTM, to simulate the GZB reservoir operation at monthly, daily, and hourly scales. The hourly data refer to the reservoir operation data collected every two or four hours. GZB reservoir operation data span the period from 27 June

1981 to 7 September 2015. Based on the commonly accepted “80/20” split rule, the data from 1 January 2008 to 7 September 2015 are used as the test period and the remainder are used for training, as shown in Table 2. The reservoir operation data are imported as model input (decision variables) and output (target variable). Specifically, the inputs include current and previous inflow (1-moment lag), previous outflow (1-moment lag), the current and previous water level in front of the dam (1-moment lag), the current and previous water level of the downstream region (1-moment lag), and the current time (month of year). Therefore, 9 inputs are included in the model. The model output is the current outflow of the reservoir. To match the consistency of the model, all source data were normalized in the range 0.0–1.0 and were then transformed to original values after the simulation using, the normalized equation is as follows:

$$X = \frac{X_{\text{ori}} - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad 15$$

where X is the normalized value; X_{ori} is the original value; and X_{min} and

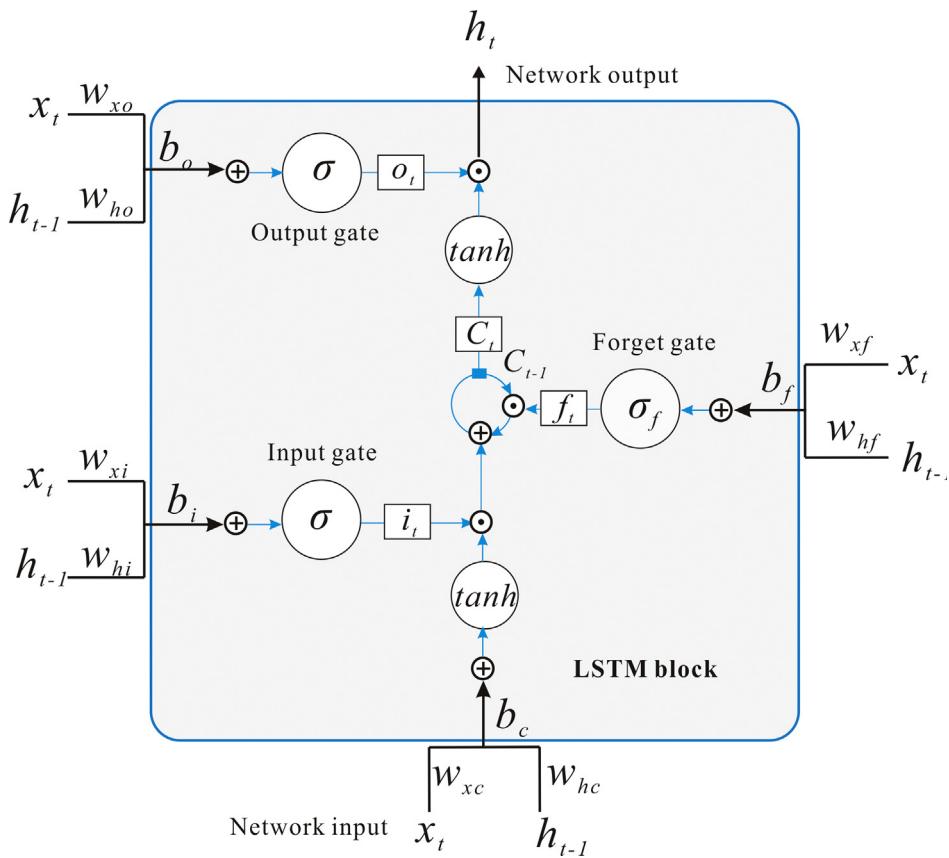


Fig. 4. The conceptual illustration of the LSTM memory block. The f , and σ_h represent the activation functions used for different gates. σ_f represent the “sigmoid” activation function σ_g and σ_h both represent the “tanh” activation function in this study.

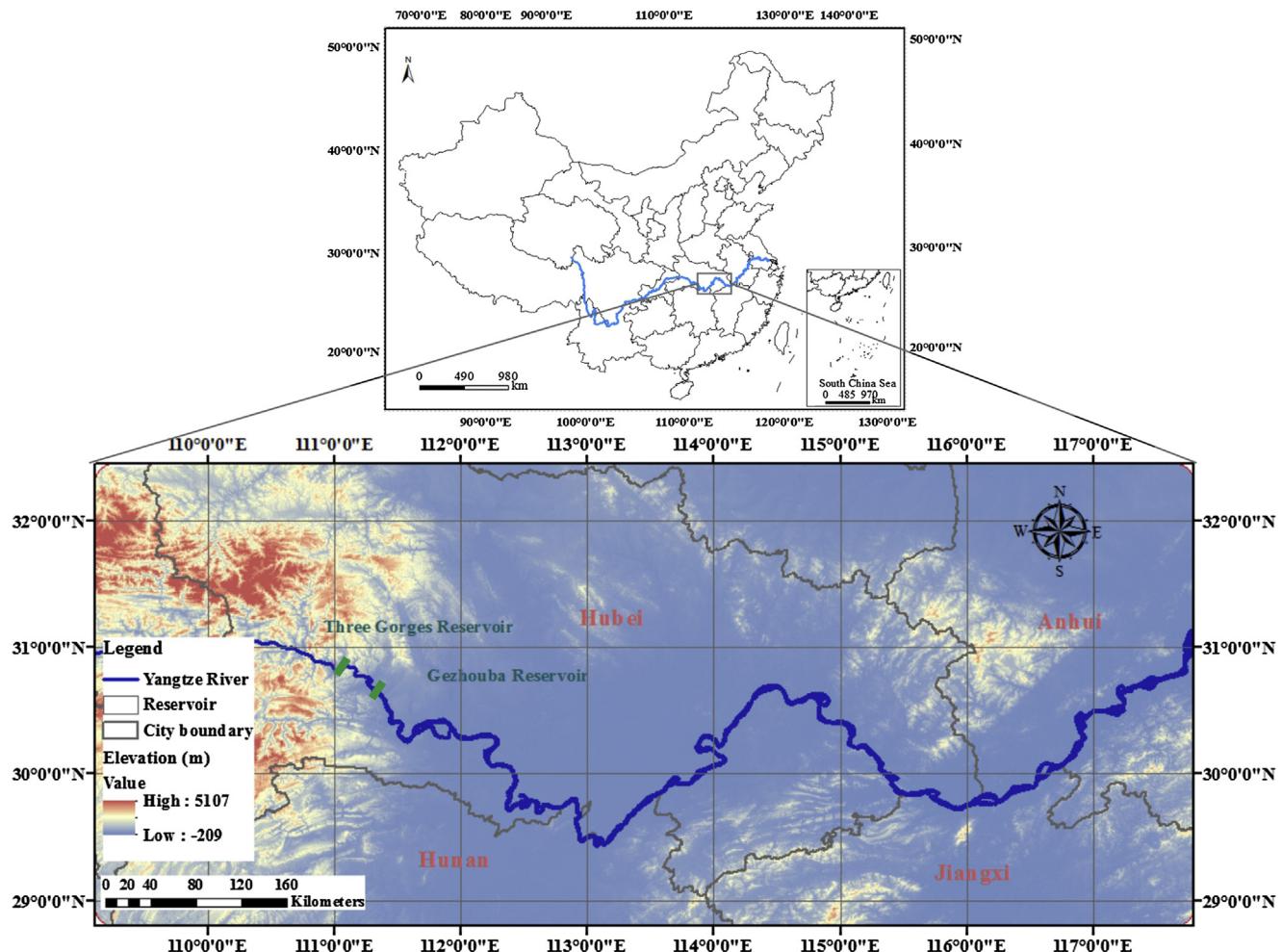


Fig. 5. The location of Gezhouba (GZB) Reservoir.

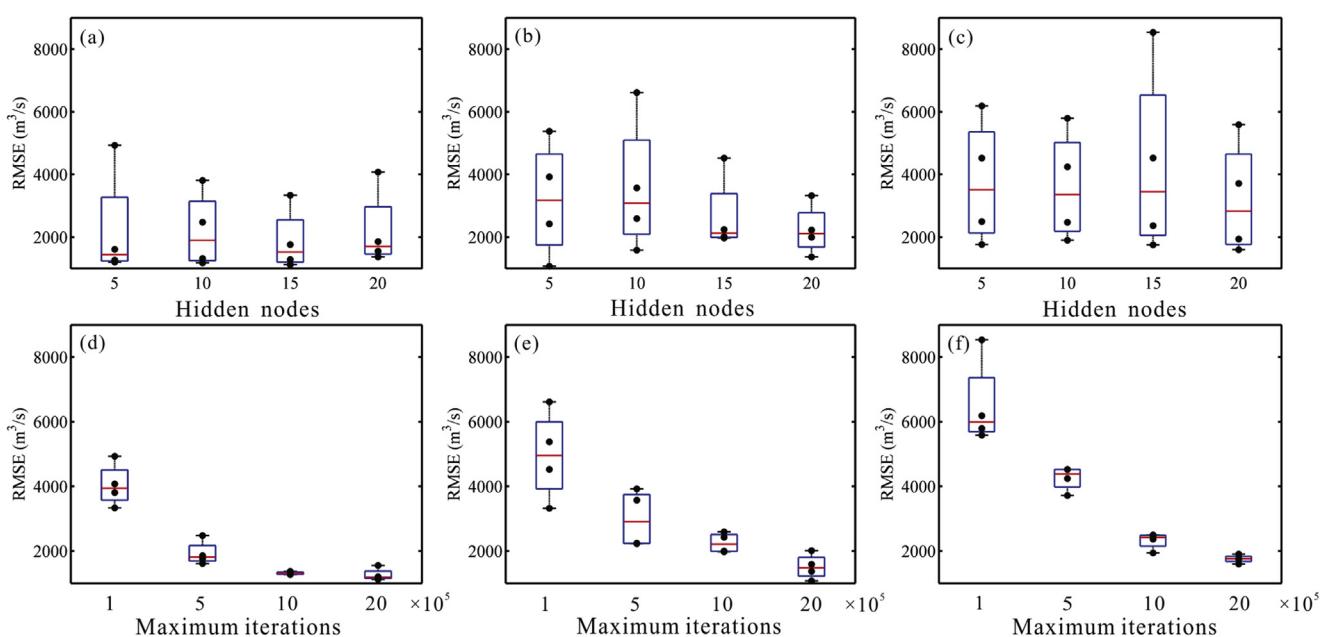


Fig. 6. BP neural network performance varies with the number of hidden nodes and maximum iterations. (a) and (d) represent the monthly statistical results; (b) and (e) represent the daily statistical results; (c) and (f) represent the hourly statistical results.

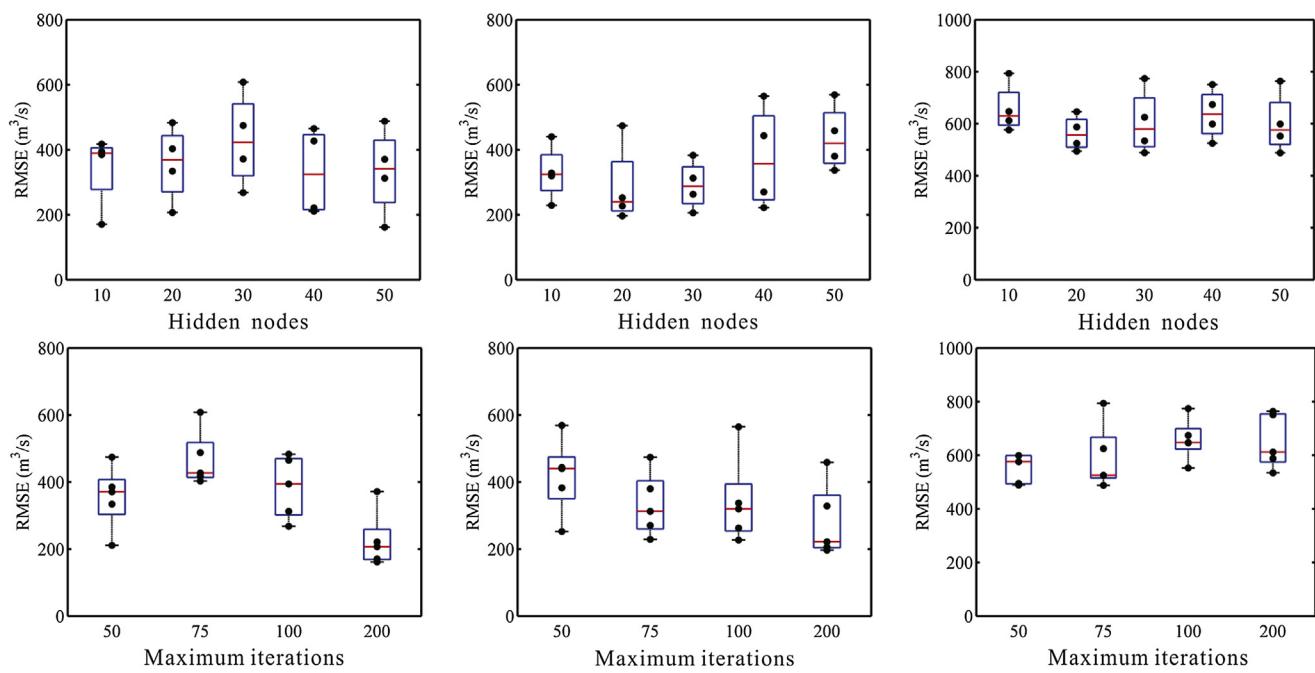


Fig. 7. LSTM model performance varies with the number of hidden nodes and maximum iterations. (a) and (d) represent the monthly statistical results; (b) and (e) represent the daily statistical results; (c) and (f) represent the hourly statistical results.

X_{max} are the respective minimum and maximum original values. The input construction was standardized for all methods, i.e., BP, SVR, and the LSTM model, to ensure an impartial comparison.

3.3. Parameterization and setting

To illustrate the strengths and weaknesses of the different AI methods, different parameterizations are used for BP, SVR and LSTM. In BP, four maximum iteration (MI) numbers are tested, i.e., 1×10^5 , 5×10^5 , 1×10^6 , and 2×10^6 , in combination with different numbers of hidden nodes, i.e., 5, 10, 15, and 20. In SVR, four types of kernel functions are set, i.e., linear, RBF, polynomial, and sigmoid kernel functions with different penalty coefficients (C , 20 values are selected from 0.0001 to 10,000, according to the geometric progression), gamma (γ , 20 values are selected from 0.0001 to 10,000, according to the geometric progression), and degree (d , i.e., 2, 3, 4). Therefore, for linear kernel function, we test 20 set of parameterizations, for RBF and sigmoid kernel function, we test 400 set of parameterizations, and for polynomial kernel function, we test 1200 set of parameterizations. Then, a performance comparison is conducted among various types of kernel functions in combination with the optimal parameters. With respect to the LSTM model, different MI numbers are tested, including 50, 75, 100, and 200 iterations, in combination with various numbers of hidden nodes, i.e., 10, 20, 30, 40 and 50. Specific parameter

configurations are described in the results and discussion sections.

3.4. Model evaluation statistics

To mathematically quantify the skill of the model simulations, four statistical measures are calculated: the root mean square error ($RMSE$), $RMSE$ -observation standard deviation ratio (RSR), and the Nash-Sutcliffe model efficiency coefficient (NSE) (Nash and Sutcliffe, 1970). The indices $RMSE$ and RSR are valuable because they indicate error in the units (or squared units) of the constituent of interest, which contributes to result analysis, and $RMSE$ and RSR values of 0 indicate a perfect fit. The NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970). NSE values range from negative infinity to 1, with 1 indicating an exact match between simulated and observed values. As suggested by previous studies, an $RMSE$ value less than half the standard deviation of the observed data may be considered low, and if $RSR < 0.7$ and $NSE > 0.5$, model performance can be considered satisfactory (Singh et al., 2010; Moriasi et al., 2007; Yang et al., 2017b). The equations of the selected statistics are shown below (Eqs. (19)–(20)):

$$RMSE = \sqrt{\frac{\sum_{i=0}^n (s_i - o_i)^2}{n}}$$

16

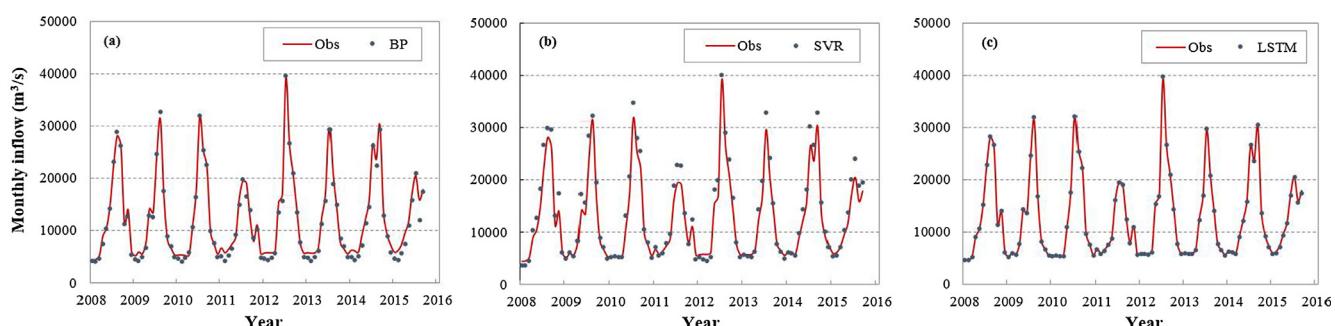


Fig. 8. Comparison of predicted and observed monthly outflow using BP neural network (a), SVR (b), and LSTM (c) model.

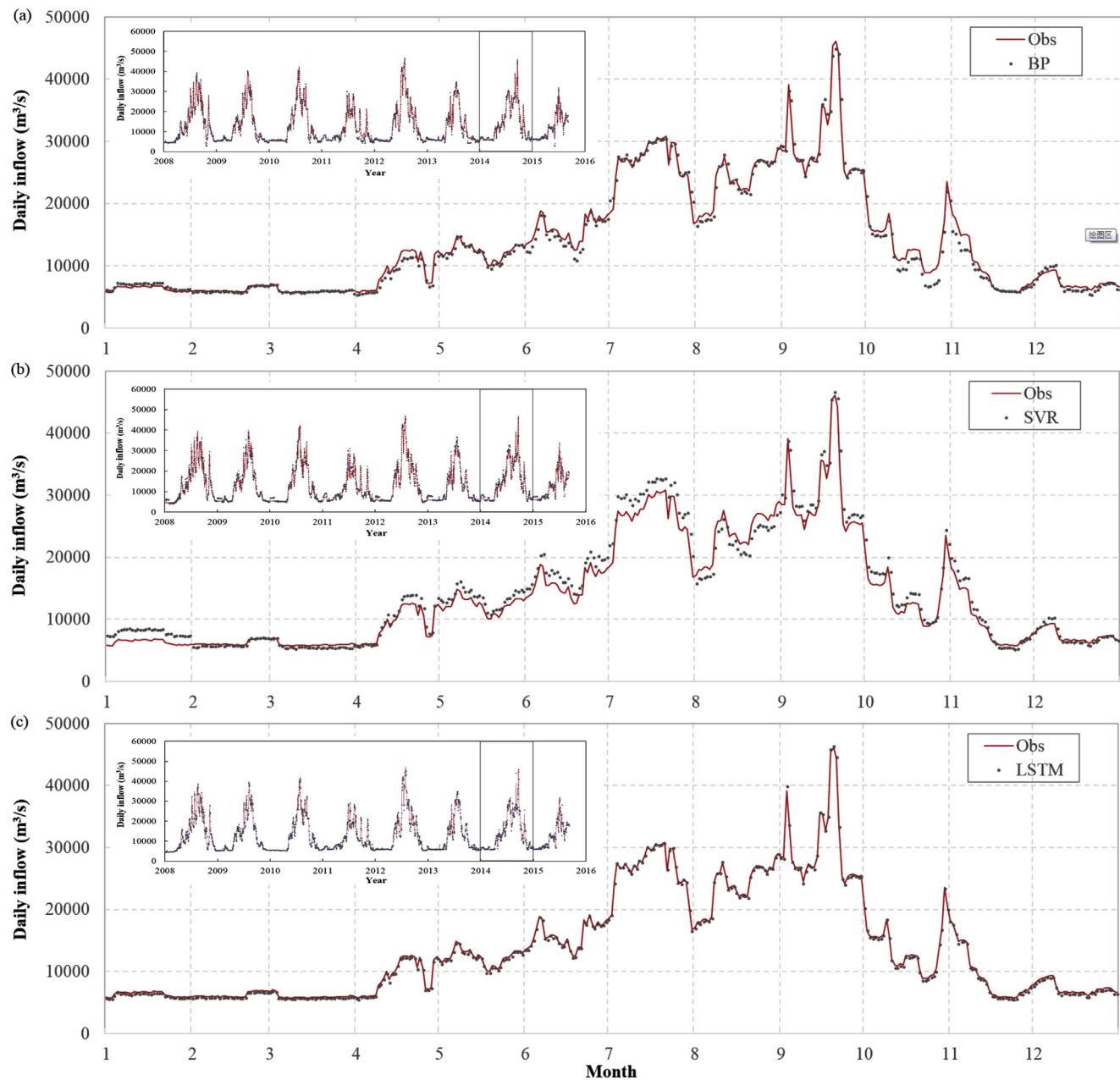


Fig. 9. Comparison of predicted and observed daily outflow using BP neural network (a), SVR (b), and LSTM (c) model.

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{[\sqrt{\sum_{i=0}^n (s_i - o_i)^2}]}{[\sqrt{\sum_{i=0}^n (o_i - \bar{o}_i)^2}]} \quad 17$$

$$NSE = 1 - \frac{\sum_{i=0}^n (o_i - s_i)^2}{\sum_{i=0}^n (o_i - \bar{o}_i)^2} \quad 18$$

where o_i and s_i are the respective observed and simulated values for outflows; \bar{o}_i and \bar{s}_i are the respective average observed and simulated inflows; and n is the total number of observations. In Eq. (21), r , α and β represent the correlation coefficient, variability and bias ratio between the simulations and observations, respectively.

To estimate the uncertainty associated with model simulations, the residuals of testing sets are computed and analyzed. The analysis consists of three components: independence analysis, heteroscedastic analysis, and normality analysis; the implementation methods consist of plotting graphs of residual autocorrelation, residual variation relative to observed values, and residual probability distributions (Figs. 11–13).

The residual independent analysis is based on residual autocorrelation: if the residual sequence is autocorrelated, then the model fails to fully explain the variation rule of the variable, and there is some regularity that is not explained, which eventually leads to a larger prediction deviation of the model. On the other hand, low residual heteroscedasticity and a close approximation to the normal distribution indicate the model is closer to unbiased estimation and has low uncertainty.

The so-called residual is the difference between the predicted value and the actual observed value, which is calculated according to Eq. (22):

$$e_i = o_i - s_i \quad 19$$

where e_i represents residuals, o_i represents observed values, and s_i represents predicted values. In this paper, the model residuals were standardized before the residual analysis, and the process is as follows:

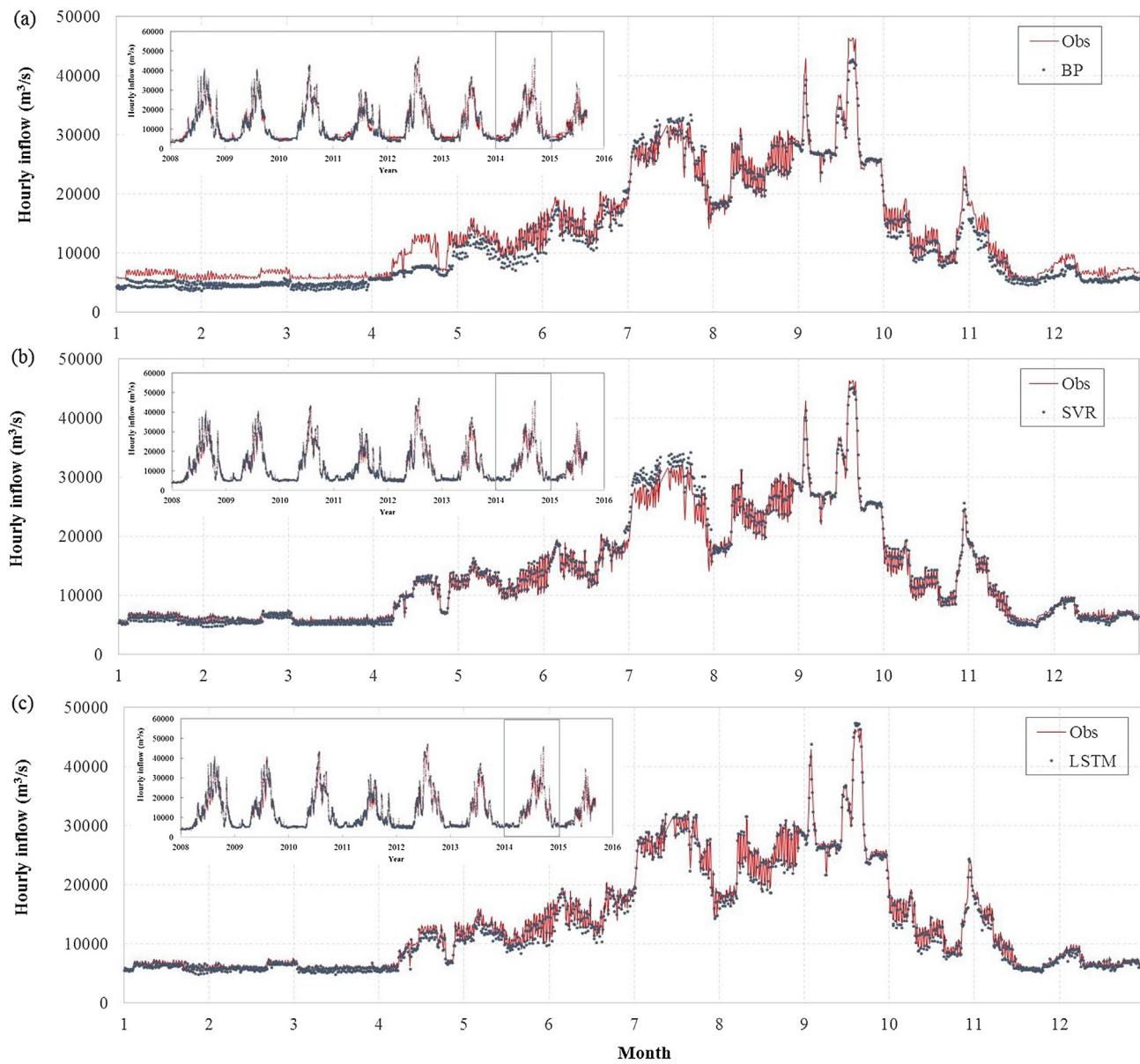


Fig. 10. Comparison of predicted and observed hourly outflow using BP neural network (a), SVR (b), and LSTM (c) model.

Table 2

The data sets sizes of training and testing the model.

	Training set	Testing set	Total set
Monthly scale	319	93	412
Daily scale	9989	2806	12,795
Hourly scale	90,327	24,431	114,758

$$r_{s,i} = \frac{e_i}{\sigma} \quad 20$$

where r_s represents standardized residuals, and σ represents the standard deviation.

All the experiments presented in this study were performed on an Inter(R) Xeon(R) CPU E5-2660 v3 @ 2.60 GHz with 64 GB DDR3 1600 MHz memory.

4. Results

4.1. Comparison of simulation results for different time scales

In this section, the observed GZB reservoir outflows are compared with simulated results based on the three different AI models, i.e., the BP neural network, SVR, and LSTM model, combined with various model parameters. First, our results show that, BP neural network and LSTM, the number of hidden nodes and maximum iterations are two key parameters that affect the simulation accuracy. Moreover, there is no obvious regularity about the effect of the number of hidden nodes on simulation accuracy, but the increase of the number of maximum iterations can significantly improve the simulation accuracy of the two models (Figs. 6 and 7). With respect to SVR, the choice of kernel function can directly affect the model accuracy. Our results show that, at monthly scale, the simulation accuracy is ranked as Sigmoid > Polynomial > RBF > Linear among the different kernel functions, at daily scale, the best accuracy ranking is Sigmoid > RBF > Linear > Polynomial; and at hourly scale, the best accuracy ranking is Sigmoid > RBF > Polynomial > Linear (Table 4).

Table 3

Statistical performance on different time scales of BP neural network with different maximum iteration (MI) and different numbers of hidden nodes (H). The bold and underlined values indicate the best statistics at same time scales.

	Monthly scale				Daily scale				Hourly scale			
	RMSE (m^3/s)	RSR (-)	NSE (-)	Time	RMSE (m^3/s)	RSR (-)	NSE (-)	Time	RMSE (m^3/s)	RSR (-)	NSE (-)	Time
MI = 1×10^5 H = 5	4931.537	0.6199	0.6118	00:36:55	5377.449	0.6278	0.6057	60:48:04	6189.017	0.7310	0.4656	78:39:56
MI = 1×10^5 H = 10	3806.961	0.4785	0.7686	00:37:31	6614.750	0.7723	0.4034	57:09:55	5792.710	0.6842	0.5318	71:55:17
MI = 1×10^5 H = 15	3335.441	0.4193	0.8224	00:38:16	4524.253	0.5282	0.7209	59:19:19	8535.343	1.0081	0.0163	80:03:27
MI = 1×10^5 H = 20	4075.991	0.5123	0.7348	00:37:07	3323.053	0.3880	0.8494	62:49:06	5587.853	0.6600	0.5644	75:49:24
MI = 5×10^5 H = 5	1610.882	0.2025	0.9587	6:20:47	3923.759	0.4581	0.7901	130:16:19	4522.122	0.5341	0.7147	158:58:52
MI = 5×10^5 H = 10	2478.664	0.3116	0.9019	6:35:29	3571.767	0.4170	0.8260	150:21:40	4244.474	0.5013	0.7487	160:57:31
MI = 5×10^5 H = 15	1762.991	0.2216	0.9504	6:36:39	2243.612	0.2619	0.9314	121:26:55	4527.507	0.5348	0.7140	162:58:32
MI = 5×10^5 H = 20	1855.010	0.2332	0.9451	6:35:25	2227.102	0.2600	0.9324	129:48:06	3714.903	0.4388	0.8074	159:08:51
MI = 1×10^6 H = 5	1269.123	0.1595	0.9743	15:34:42	2424.844	0.2831	0.9198	259:26:56	2497.333	0.2950	0.9130	334:13:15
MI = 1×10^6 H = 10	1319.689	0.1659	0.9722	15:17:25	2590.966	0.3025	0.9085	257:34:26	2472.812	0.2921	0.9147	386:25:26
MI = 1×10^6 H = 15	1283.529	0.1613	0.9737	15:12:46	1973.271	0.2304	0.9469	231:23:29	2364.631	0.2793	0.9220	397:42:26
MI = 1×10^6 H = 20	1366.860	0.1718	0.9702	15:11:55	1997.889	0.2333	0.9456	244:26:09	1938.357	0.2289	0.9476	357:05:48
MI = 2×10^6 H = 5	1202.594	0.1512	0.9769	74:44:14	1073.402	0.1253	0.9843	519:31:23	1765.321	0.2085	0.9565	799:25:59
MI = 2×10^6 H = 10	1179.157	0.1482	0.9778	74:30:39	1587.474	0.1853	0.9656	526:59:02	1901.642	0.2246	0.9495	805:22:17
MI = 2×10^6 H = 15	1122.827	0.1411	0.9799	74:05:56	2012.372	0.2349	0.9448	533:44:21	1752.630	0.2070	0.9571	792:02:28
MI = 2×10^6 H = 20	1551.823	0.1951	0.9616	74:22:54	1369.356	0.1599	0.9744	503:44:17	1596.052	0.1885	0.9645	786:13:09

On the other hand, the prediction performances of the BP neural network, SVR, and LSTM model are shown in Tables 3–5 and Figs. 8–10. The calculation results show that at the monthly scale, the best RMSE values obtained by the BP neural network, SVR, and LSTM model are 1122.827, 1932.814, and 161.712; the best RSR values are 0.1411, 0.2429, and 0.0203; and the best NSE values are 0.9799, 0.9404, and 0.9996, respectively (Tables 3–5). The best simulation results are shown in Fig. 8. When the best statistical results are achieved, the MI and H of the BP neural network are 2×10^6 and 15, respectively; the kernel function of SVR is sigmoid, the respective values of γ and C are 0.1265 and 3.728; and the MI and H of the LSTM model are 200 and 50, respectively (Tables 3–5). According to the comparison among different models, the best accuracy ranking is the

LSTM model > BP neural network > SVR.

With respect to the daily time scale, the best RMSE values generated using the BP neural network, SVR, and LSTM model are 1073.402, 1332.720, and 663.723; the best RSR values are 0.1253, 0.1556, and 0.0230; and the best NSE values are 0.9843, 0.9758, and 0.9935, respectively (Tables 3–5). The best simulation results are shown in Fig. 9. When the best statistical results are obtained, the MI and H of the BP neural network are 2×10^6 and 5, respectively (Table 3), the kernel function of SVR is sigmoid, and the values of γ and C are 0.0869 and 2.560, respectively (Table 4). In addition, the MI and H of the LSTM model are 200 and 50, respectively (Table 5). According to our experimental results, the best accuracy ranking is the LSTM model > BP neural network > SVR (Tables 3–5, Fig. 9).

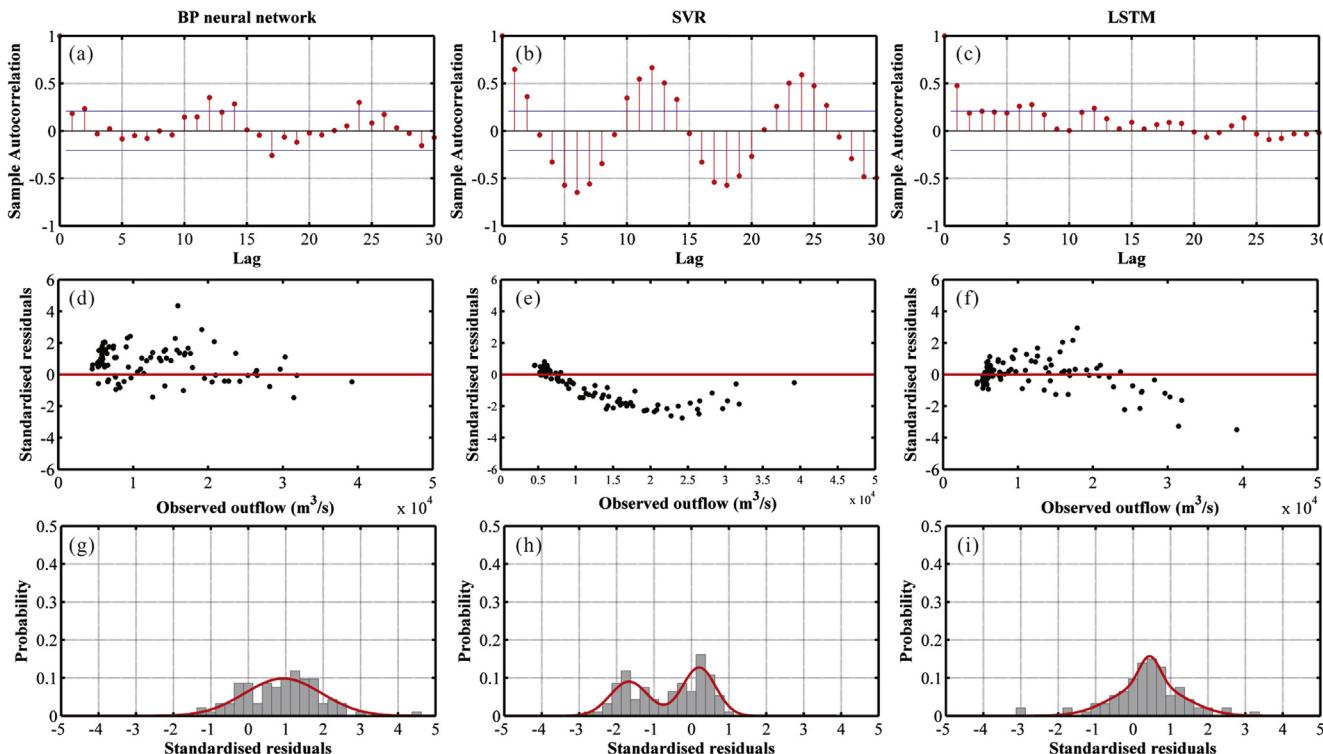


Fig. 11. Investigation of residuals of BP neural network (Column1), SVR (Column2), and LSTM (Column3) model at monthly scale. (a–c) Autocorrelation function (ACF) plots of r_s at with 95% significance levels. (d–f) r_s as a function of observed reservoir outflows. (g–i) Fitted (solid line) and actual (bars) probability density

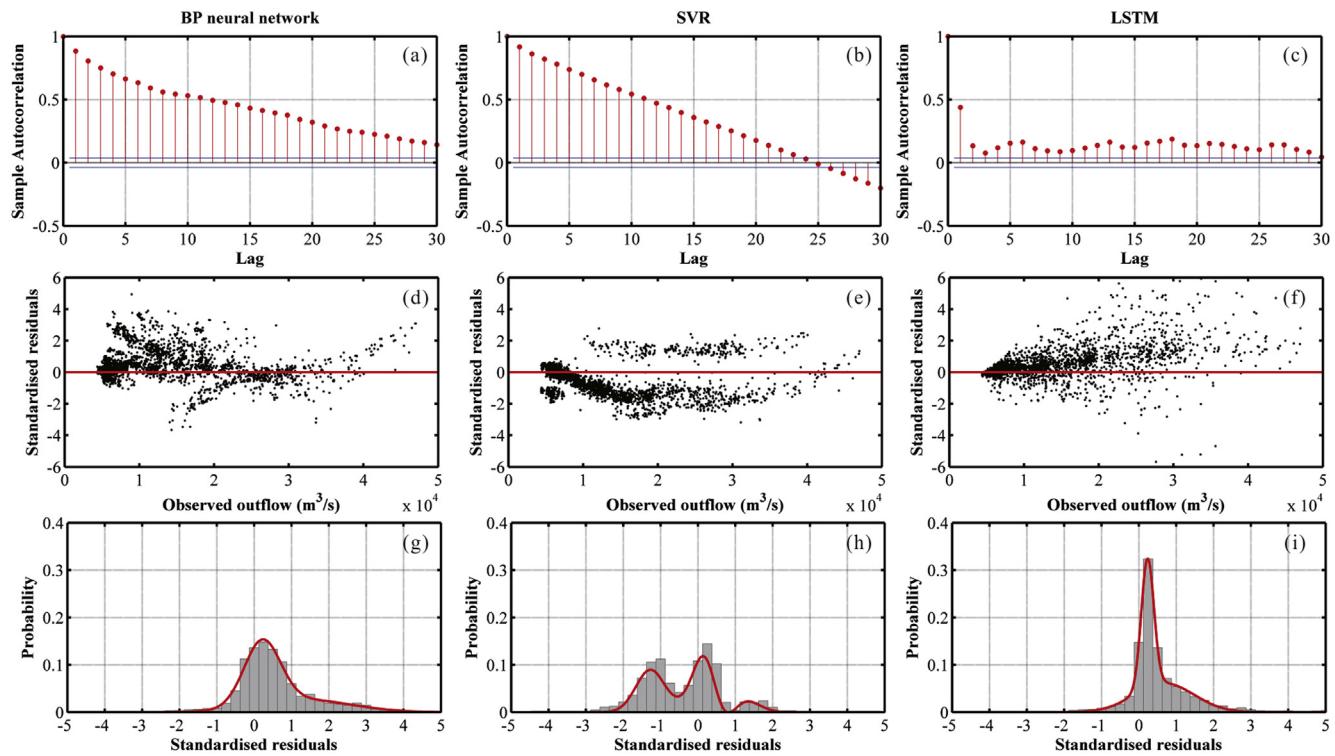


Fig. 12. Investigation of residuals of BP neural network (Column1), SVR (Column2), and LSTM (Column3) model at daily scale. (a–c) Autocorrelation function (ACF) plots of r_s at with 95% significance levels. (d–f) r_s as a function of observed reservoir outflows. (g–i) Fitted (solid line) and actual (bars) probability density function (PDF) of r_s .

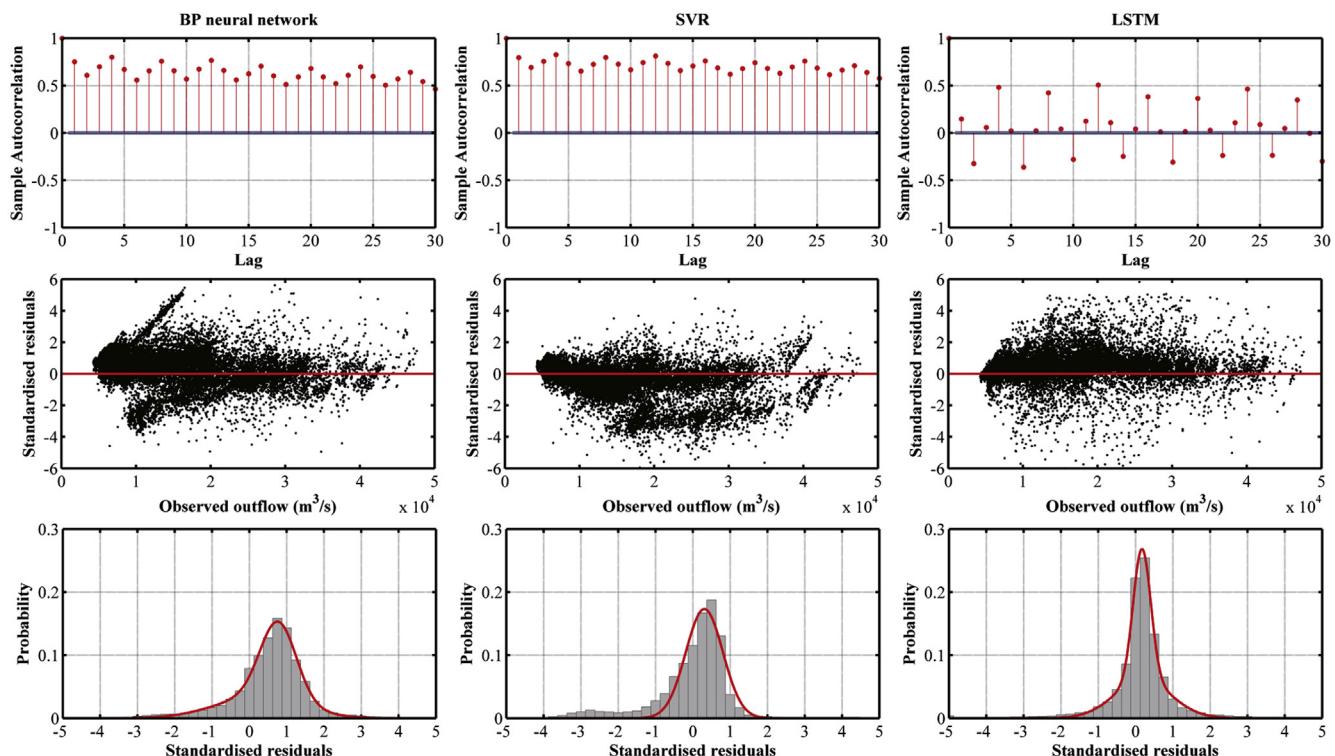


Fig. 13. Investigation of residuals of BP neural network (Column1), SVR (Column2), and LSTM (Column3) at hourly scale. (a–c) Autocorrelation function (ACF) plots of r_s at with 95% significance levels. (d–f) r_s as a function of observed reservoir outflows. (g–i) Fitted (solid line) and actual (bars) probability density function (PDF) of the r_s .

Table 4

Statistical performance on different time scales of SVR using different kernel functions with different structural parameter (γ), penalty coefficients (C) and degree (d, for polynomial). The bold and underlined values indicate the best statistics at same time scales.

			RMSE (m ³ /s)	RSR (-)	NSE (-)	Time (-)
Monthly scale	RBF	$\gamma = 0.0336$ C = 1.6238	2961.197	0.3722	0.8600	00:05:38
	Sigmoid	$\gamma = 0.0886$ C = 4.2813	<u>1932.814</u>	<u>0.2429</u>	<u>0.9404</u>	<u>00:05:47</u>
	Polynomial	$d = 3$ $\gamma = 0.6158$ C = 0.2336	2770.405	0.3482	0.8775	00:17:59
	Linear	C = 0.0336	3180.128	0.3997	0.8386	00:01:36
Daily scale	RBF	$\gamma = 0.0001$ C = 1438.4499	3322.261	0.3879	0.8495	45:29:31
	Sigmoid	$\gamma = 0.0886$ C = 1.6237	<u>1332.720</u>	<u>0.1556</u>	<u>0.9758</u>	<u>49:26:03</u>
	Polynomial	$d = 3$ $\gamma = 0.0886$ C = 78.4760	3480.387	0.4063	0.8348	152:06:59
	Linear	C = 0.2336	3371.827	0.3937	0.8499	6:01:07
Hourly scale	RBF	$\gamma = 0.0001$ C = 1438.4499	1104.779	0.1305	0.9830	163:09:45
	Sigmoid	$\gamma = 0.03360$ C = 4.2813	<u>1051.484</u>	<u>0.1242</u>	<u>0.9846</u>	<u>204:16:42</u>
	Polynomial	$d = 3$ $\gamma = 0.6158$ C = 78.4760	1652.226	0.1951	0.9614	635:15:19
	Linear	C = 0.6158	1774.222	0.2096	0.9560	12:26:14

Table 5

Statistical performance on different time scales of LSTM with different maximum iteration (MI) and different numbers of hidden nodes (H). The bold and underlined values indicate the best statistics at same time scales.

	Monthly scale			Daily scale			Hourly scale			Time		
	RMSE (m ³ /s)	RSR (-)	NSE (-)	Time	RMSE-D (m ³ /s)	RSR (-)	NSE (-)	Time	RMSE (m ³ /s)	RSR (-)	NSE (-)	
MI = 50H = 10	385.096	0.0484	0.9976	00:00:24	440.376	0.0514	0.9974	00:10:04	576.407	0.0681	0.9954	0:33:25
MI = 50H = 20	334.178	0.0420	0.9982	00:00:24	252.441	0.0295	0.9991	00:10:18	494.814	0.0584	0.9966	0:39:05
MI = 50H = 30	474.592	0.0597	0.9964	00:00:24	383.048	0.0447	0.9980	00:10:24	488.872	0.0577	0.9967	0:35:27
MI = 50H = 40	210.880	0.0265	0.9993	00:00:25	443.702	0.0518	0.9973	00:10:17	599.306	0.0708	0.9950	0:36:22
MI = 50H = 50	370.906	0.0466	0.9978	00:00:24	569.247	0.0665	0.9956	00:10:47	599.306	0.0708	0.9950	0:36:22
MI = 75H = 10	417.693	0.0525	0.9972	00:00:35	228.924	0.0267	0.9993	00:13:05	793.922	0.0938	0.9912	1:04:33
MI = 75H = 20	403.289	0.0507	0.9974	00:00:35	474.111	0.0554	0.9969	00:13:09	524.723	0.0620	0.9962	1:06:53
MI = 75H = 30	608.502	0.0765	0.9941	00:00:36	312.815	0.0365	0.9987	00:13:14	625.140	0.0738	0.9945	1:03:48
MI = 75H = 40	427.246	0.0537	0.9971	00:00:37	270.263	0.0316	0.9990	00:13:20	525.245	0.0620	0.9955	1:03:48
MI = 75H = 50	487.979	0.0613	0.9962	00:00:36	380.130	0.0444	0.9980	00:13:25	<u>488.121</u>	<u>0.0577</u>	<u>0.9967</u>	<u>1:05:21</u>
MI = 100H = 10	394.358	0.0496	0.9975	00:00:48	319.959	0.0374	0.9986	00:18:33	647.549	0.0765	0.9941	1:39:12
MI = 100H = 20	483.471	0.0608	0.9963	00:00:49	227.077	0.0265	0.9993	00:19:04	646.599	0.0764	0.9942	1:33:48
MI = 100H = 30	268.045	0.0337	0.9989	00:00:50	263.070	0.0307	0.9991	00:19:05	774.390	0.0915	0.9916	1:40:29
MI = 100H = 40	465.524	0.0585	0.9965	00:00:52	565.103	0.0660	0.9956	00:19:15	674.315	0.0796	0.9936	1:40:39
MI = 100H = 50	312.662	0.0393	0.9984	00:00:51	337.258	0.0394	0.9984	00:19:17	552.746	0.0653	0.9957	1:41:00
MI = 200H = 10	170.775	0.0215	0.9995	00:01:51	328.510	0.0384	0.9985	00:36:33	612.316	0.0723	0.9948	3:18:27
MI = 200H = 20	207.015	0.0260	0.9993	00:01:47	<u>196.697</u>	<u>0.0230</u>	<u>0.9995</u>	<u>00:37:04</u>	587.953	0.0694	0.9952	3:17:21
MI = 200H = 30	371.593	0.0467	0.9978	00:01:49	206.057	0.0241	0.9994	00:35:05	534.235	0.0631	0.9960	3:13:45
MI = 200H = 40	221.356	0.0278	0.9992	00:01:55	221.957	0.0259	0.9993	00:36:05	751.348	0.0887	0.9920	3:14:29
MI = 200H = 50	<u>161.712</u>	<u>0.0203</u>	<u>0.9996</u>	<u>00:01:56</u>	458.931	0.0536	0.9971	00:36:17	764.212	0.0903	0.9919	3:15:12

The results of the hourly time scale suggest that the best RMSE values produced by the BP neural network, SVR, and LSTM model are 1596.052, 1051.484, and 488.121; the best RSR values are 0.1885, 0.1242, and 0.0577; and the best NSE values are 0.9645, 0.9846, and 0.9967, respectively (Tables 3–5). The best simulation results are shown in Fig. 10. When the best statistical results are achieved, the MI and H of the BP neural network are 2×10^6 and 20, respectively (Table 3); the kernel function of SVR is sigmoid, the values of γ and C are 0.0464 and 2.783, respectively (Table 4), and the MI and H of the LSTM model are 75 and 50, respectively (Table 5). Notably, within a day, Gezhouba peak-load-dispatching operation according to the inflow conditions and peak frequency modulation to meet the demand of the power grid and to stabilize the flow regime of the reservoir. Therefore, compared with the monthly and daily scales, the hourly model needs to consider the predicted accuracy of outflow for the peak operation period. Fig. 10 shows that the simulation accuracy of the LSTM model is higher than that of the BP neural network and SVR during the peak-load-dispatching operation period. Overall, the best accuracy ranking is the LSTM model > SVR > BP neural network.

In addition to the simulation accuracy, the calculation speed is also an important index to measure the performance of a model. In this paper, the time consumption is used as an evaluation index to compare the calculation speed of the three models. The experimental results

show that there are significant differences in the calculation speed among the three models and the three time scales. In general, the time consumption is ranked as the BP neural network > SVR > LSTM model among the different models and as the hourly scale > daily scale > monthly scale among the different time scales. The BP neural network is the longest-running model, and the time consumption increases as MI increases (Table 3). The time consumption associated with optimal results at the monthly, daily, and hourly scales is 74:05'56", 519:31'23" and 786:13'09", respectively (Table 3). In addition to the time scale, the time consumption of SVR is also related to the selected kernel function, and the time consumption is ranked as polynomial > sigmoid > RBF > linear. The time consumption associated with optimal results at the monthly, daily, and hourly scales is 54'47", 49:26'03" and 204:16'42", respectively (Table 4). Similar to the BP neural network, the time consumption of the LSTM model is related to the time scale and MI, which manifests as an increase in the time consumption with an increase in MI, and the time consumption associated with optimal statistical results for the LSTM algorithm is 6", 37'04", and 1:5'21" for the monthly, daily, and hourly scales, respectively (Table 5). It should be noted that the time consumption of the LSTM model is significantly lower than that of the BP neural network and SVR (Tables 3–5).

4.2. Residuals analysis

As shown in Figs. 11–13, to evaluate the uncertainty of the models, residual analysis is performed on the best statistical results of the three models at different time scales. At the monthly scale, the experimental results show that the r_s of the SVR model has a significant autocorrelation, and the ACF presents a trend of alternating positive and negative variations as the lag changes (Fig. 11b). Compared with SVR, the autocorrelation of r_s for the BP neural network and LSTM model is weak, and the ACF lies mainly in the 95% confidence interval (Fig. 11a–c). Fig. 9d–f show the scatter points of r_s as a function of observed reservoir outflows. It is clear that the r_s values do not appear to be randomly distributed over the flow interval for each model, and the r_s of the LSTM model shows an obvious increasing trend with an increase in outflow. Fig. 11g–i display the probability density distribution of r_s for the three models. The results show that the probability density distribution curve of r_s for the BP neural network is flat with obvious skewness, and the values of r_s are mainly distributed between –1 and 3 (Fig. 11g). With respect to the SVR model, the probability density of r_s shows a bimodal distribution, ranging mainly between –2 and 0 (Fig. 11h). The r_s of the LSTM model presents a unimodal distribution with a sharp peak, and the r_s values are mainly distributed between –2 and 2 (Fig. 11i).

At the daily scale, the autocorrelation of r_s for the LSTM model is the weakest, whereas the r_s values of the BP neural network and SVR model show remarkable autocorrelation (Fig. 12a–c). The r_s values of the three models exhibit heteroscedasticity as the observed outflow changes. Compared with the BP neural network and SVR, the spatial distribution of r_s with observed outflow for the LSTM model is relatively uniform (Fig. 12d–f). The probability density of r_s for the BP neural network displays a unimodal distribution, ranging mainly between –1 and 1 (Fig. 12g). The r_s of SVR has a multimodal distribution, with three peaks (two high and one low) distributed at –1.2, 0 and 1.4, respectively (Fig. 12h). The r_s of the LSTM model presents a unimodal distribution with a sharp peak, and r_s is mainly distributed between –1 and 2 (Fig. 12i).

The experimental results show that at the hourly scale, autocorrelation of r_s is found in all three models (Fig. 13a–c). The ACF values of the BP neural network and SVR are positive, with periodic increases or decreases as the lag changes (Fig. 13a and b). The ACF values of the LSTM model present a trend of alternating positive and negative variations as the lag changes and also exhibit the phenomenon of periodic increases or decreases. The r_s values of all three models exhibit heteroscedasticity (Fig. 13d–f). The probability density curve of r_s for the BP neural network model has a single peak, ranging mainly between –1 and 2, and obvious (mainly positive) skewness is exhibited (Fig. 13g). The distribution of r_s for the SVR model is also unimodal, and a wide distribution is evident between –4 and 2 (Fig. 13h). With respect to the LSTM model, the probability density curve of r_s is unimodal, and r_s is densely distributed, mainly between –1 and 1 (Fig. 13i).

4.3. Comparison of simulation results for different flow regimes

The experimental results of Section 4.2 show that model uncertainty varies with observed flow. To better determine the ability of the models to reproduce different flow regimes, the models with the best simulation results are evaluated for three specific regimes, i.e., low, intermediate and high flow conditions. The division of these flow regimes is based on the calculation of the inflow frequency distribution curve. The flow regimes are categorized by using the 25th and 75th percentiles, and the flow limit is 5985 m³/s and 18,050 m³/s, respectively. The model accuracy for the different flow regimes is graphically represented in three scatter plots (Figs. 14–16) and is measured in terms of RMSE, RSR, and NSE (Table 6). The experimental results show that, in general, the LSTM model has a significant advantage with respect to simulation

accuracy, and this model can produce more accurate simulation results for each time scale and flow regime. Specifically, at the monthly scale, the comparison results show that the evaluation index values of the LSTM model are the best and the simulation accuracy is the highest; the comparison results of the two other models show that, under low flow conditions, the simulation accuracy of these two models is not ideal, i.e., the RSR value is greater than 1, and the NSE value is negative. Under intermediate and high flow conditions, the simulation accuracy of the two latter models meets the evaluation criteria, and the accuracy of the BP neural network simulation is higher than that of SVR (Table 6, Fig. 14). At the daily time scale, the simulation accuracy of the LSTM model is significantly higher than that of the two other models. Under low flow conditions, poor simulation is observed for the BP neural network and SVR, and it is difficult to obtain satisfactory results. Under intermediate and high flow conditions, both of these models exhibit satisfactory simulation results; the simulation accuracy of the SVR model is slightly higher than that of the BP neural network under intermediate flow conditions, whereas the simulation accuracy of the BP neural network is higher than that of SVR under high flow conditions (Table 6, Fig. 15). At the hourly scale, the LSTM model also shows higher prediction accuracy than the BP neural network and SVR (Table 6). The simulation accuracy of the BP neural network and SVR is poor under low flow conditions. Under intermediate and high flow conditions, the simulation accuracy of the two models meets the evaluation criteria; the simulation accuracy of SVR is higher than that of the BP neural network under intermediate flow conditions, whereas the simulation accuracy of the BP neural network is slightly higher than that of SVR under high flow conditions (Table 6, Fig. 16).

5. Discussion

5.1. Suggestions for model parameterization

Parameter setting has always been the focus of the performance research of AI models. In this study, we test the influence of different parameter combinations on the performance of three selected AI models. First, we explore the effect of the maximum number of iterations and the number of hidden nodes on the accuracy of a BP neural network. According to the statistics shown in Table 3 for the BP neural network, (1) in general, an increase in the number of maximum iterations can improve the simulation accuracy, and (2) the effects of the number of hidden nodes on the simulation accuracy are uncertain and irregular. In addition, selection of the appropriate number of hidden nodes can help improve the statistics. As shown in Table 3, the best BP neural network statistics for the three time scales are consistently achieved by a maximum number of iterations equal to 2×10^6 . In the use of the BP neural network in our study, an error backpropagation combined with a gradient descent algorithm is employed. As the number of maximum iterations increases, the gradient optimization scheme continuously explores the response surface of the objective function (Eq. (3)), and the evolution is not completed until the algorithm reaches the preset error value or the allowable number of maximum iterations. However, as the search evolves with more iterations, the gradient of the objective function will decrease; consequently, any further increase in the maximum iteration number results in less improvement in the objective function value, as shown for the 1×10^6 and 2×10^6 maximum iteration scenarios in Table 3. In addition, to ensure that the network training reaches the maximum number of iterations, the error value is set as 0.0001, which is significantly lower than that of the model. Our results show that choosing the appropriate number of hidden nodes can significantly improve model accuracy, but the effect of the number of hidden nodes on accuracy is unpredictable. The best BP neural network statistics for the monthly, daily, and hourly time scales are obtained when the number of hidden nodes is 15, 5, and 20, respectively (Table 3). As mentioned in Yao (1999), the number of hidden nodes in ANN is crucial for model performance and should be

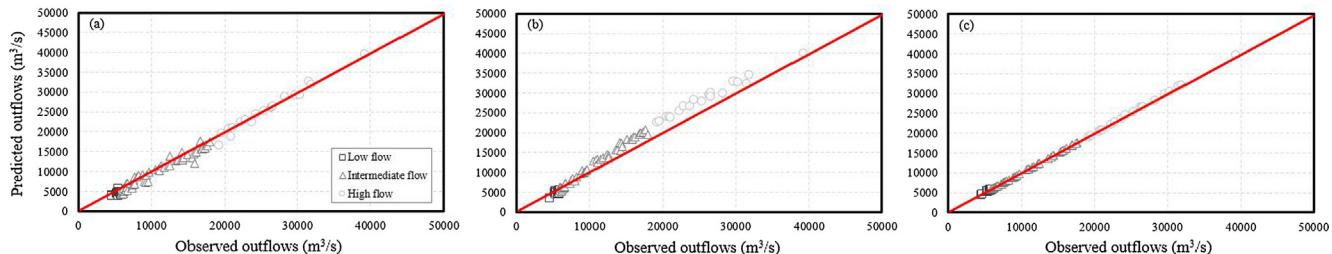


Fig. 14. Scatter plots of predicted and observed hourly outflow for BP neural network (a), SVR (b), and LSTM (c) model. Different colors are used to represent the flow regimes.

jointly designed and optimized with a proper training algorithm. As such, the number of hidden nodes is one of the research hotspots in the ANN field. It is generally believed that determination of the number of hidden nodes is related to the number of input and output nodes (Lippmann, 1987; Chen, 1996; Moody and Antsaklis, 1996). Nevertheless, studies have not adopted a universally accepted method for choosing the number of hidden nodes. At present, the number of hidden nodes is usually determined by trial and error with the objective of minimizing the cost function.

With respect to SVR, we examine the influence of the kernel function on the simulation performance based on an optimal structural parameter (γ), penalty coefficients (C) and degree (d , for the polynomial). The kernel function is introduced to map the linear non-separable training sample from the input space to the feature space, thereby realizing the linear separability of the training samples in the feature space. In this way, a linear classifier can be used to divide the training samples after mapping in the feature space. As shown in Table 4, SVR models are able to produce satisfactory results, but the simulation accuracy and time consumption differ significantly among different kernel functions. Our results show that, regardless of the time scale, the best simulation accuracy is consistently generated by the sigmoid kernel function; however, the time consumption of this function is relatively long (Table 4). This is not consistent with previous research, and most reviewed studies have suggested that RBF is the most appropriate kernel function that tends to exhibit satisfactory performance (Yaseen et al., 2015). However, technological research conducted on the kernel function shows that any kernel function has its own advantages and limitations because of different training samples, and kernel functions have different classification abilities (Amari and Wu, 1999; Asefa et al., 2006; Yang et al., 2017b). The sigmoid kernel function is derived from ANN and has been proved to have good global classification performance (Lin and Lin, 2005). Our results also show that the classification performance of the sigmoid kernel function is better than that of other kernel functions for our training sample.

With respect to the LSTM model, we mainly consider the influence of the number of maximum iterations and hidden nodes on model performance, and the main results in this paper are as follows: (1) an increase in the number of maximum iterations improves the precision of the LSTM model, and (2) a change in the number of hidden nodes affects the simulation accuracy, but the function is weak and irregular.

As shown in Table 5, when the number of maximum iterations increases from 50 to 75, the model accuracy is effectively increased, but increasing the number of maximum iterations to greater than 100 does not improve accuracy further. Based on an advanced investigation, we deduce that model accuracy is not improved further because the training process of the LSTM model adopts the BPTT algorithm, which has a similar principle as the classical BP algorithm. The BPTT algorithm uses a forward algorithm to calculate the output value, backward calculates the error of each individual LSTM cell, and continuously updates the network weight, extending the direction to reduce the error. With the narrowing of the gap between the model's output value and the desired output value, the decline rate of the model error tends to slow down, so an increase in the number of maximum iterations does not significantly improve model precision when the number of maximum iterations reaches a certain limit. However, unlike the BP neural network, the number of iterations required for LSTM model convergence is much smaller than the BP neural network. Our analysis indicates that the reason for this result is that the LSTM model has strong feature extraction capabilities, which ensures that the model can extract the characteristic information of the data more efficiently and complete the convergence process quickly. According to our experimental results, we cannot specify how to achieve the best statistics by choosing the number of hidden nodes. However, it is worth noting that while the number of iterations is the same, the changes of number of hidden node have only a limited effect on the simulation accuracy. At present, studies that examine the influence of the number of hidden nodes on the precision of LSTM models are rarely reported. Wielgosz et al. (2017) tested the influence of the number of hidden nodes on simulation accuracy and concluded that, for their test data, the simulation accuracy was highest when the number of hidden nodes was 32, but the effect of different numbers of hidden nodes on model accuracy was not discussed in detail.

In conclusion, we believe that priority should be given to the number of maximum iterations in the BP neural network and LSTM model building, and a reasonable increase the number of maximum iterations can significantly improve the model accuracy. In contrast, the number of hidden nodes has a weak effect on model accuracy. For the SVR model, the selection of kernel function is the key to model construction. Combined with previous research results, we believe that sigmoid and RBF kernel function can be considered the priority object.

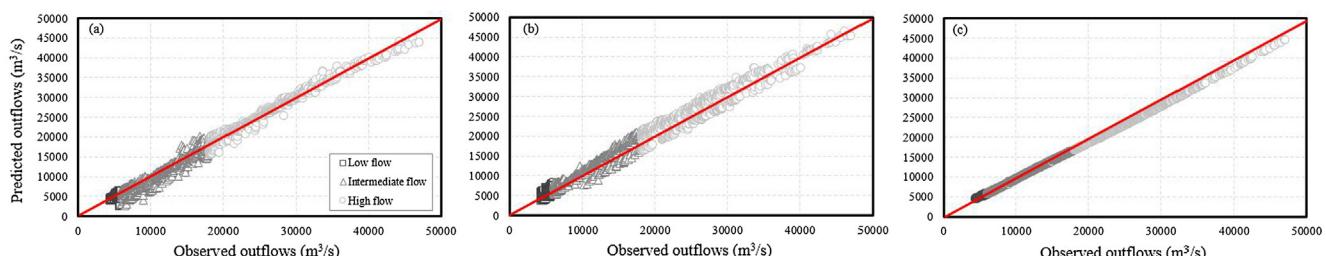


Fig. 15. Scatter plots of predicted and observed daily outflow for BP neural network (a), SVR (b), and LSTM (c) model. Different colors are used to represent the flow regimes.

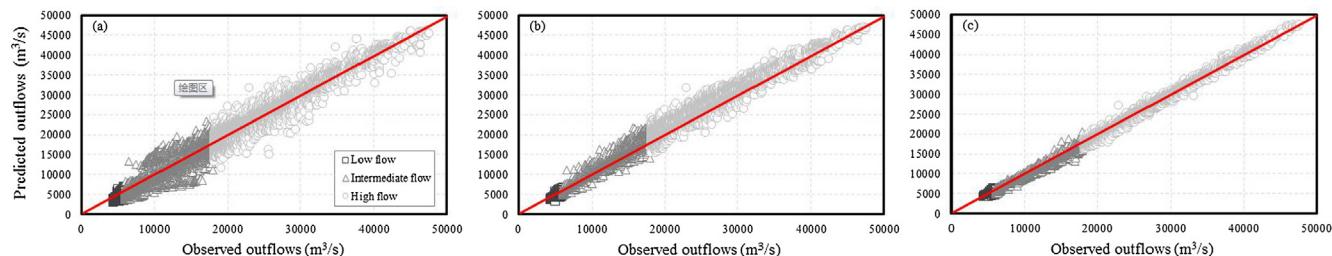


Fig. 16. Scatter plots of predicted and observed hourly outflow for BP neural network (a), SVR (b), and LSTM (c) model. Different colors are used to represent the flow regimes.

Meanwhile, in the process of model construction, due to differences in data volume and structure, model parameters have different influences on model performance. Therefore, we suggest that the model should be repeatedly trained before practical application to determine the optimal parameters and ensure the prediction ability of the model.

5.2. Suggestions for the applicability of AI models under different scenarios

The AI model has the ability to address complex nonlinear prediction problems, which is widely applied in the field of reservoir operation simulation. At present, reports on the performance comparison and analysis of AI models are common, but such reports mainly focus on the global precision comparison analysis of a single time scale. However, analyzing the global and extreme event simulation performance of the model from multiple time scales is the key to the model comprehensive learning operation rule and generating long and short-term operation plans for different scenarios. Therefore, in this study, we analyzed the simulation results of the BP neural network, SVR and LSTM under different time scales and flow regimes from three aspects of model accuracy, uncertainty and calculation speed. Based on the above analysis, we explored the guiding significance of three AI models for reservoir long and short-term operation, and the coping capacity of three AI models for extreme inflow conditions such as drought and flood, and we summarized the application scenarios of each model in order to provide a reference for the practical application of these model.

First, the application effect of the BP neural network in reservoir operation was studied. In the 1990s, the emergence of BP algorithm greatly facilitated the development of ANN and effectively promoted the research of ANN algorithms in the field of reservoir operation. In our study, the BP neural network was applied to the reservoir operation simulation, and the results showed that, for reservoir long-term operation (monthly scale), the simulation accuracy of the ANN model is good, the uncertainty is weak, and the time consumption is long; at the daily scale, the accuracy of the BP model still met the evaluation criteria, but the residuals were autocorrelated and exhibited heteroscedasticity, and the time consumption increased. At the hourly scale,

due to the further increase in data volume, the disadvantages of the model regarding the uncertainty and time consumption are more obvious (Table 3). Thus, we consider that if the model is able to learn the operation rule of a reservoir from a small amount of data (such as the research of reservoir long-term operations or the influence factors of reservoir operations), the BP neural network can obtain satisfactory simulation accuracy and the time consumption is not too long, so it can be used for the simulation of reservoir operations. However, for large reservoirs and the short-term fine operation of reservoirs, the data that are required by the model learning reservoir operation rules are large, due to the large number of influencing factors. In this case, the time consumption of the BP neural network is too long, and the applicability is weak. On the other hand, for different flow regimes, the results show that under intermediate and high flow conditions, the BP neural network can obtain satisfactory results, but under low flow conditions, the BP neural network has poor simulation results and is not suitable for simulation of low inflow scenarios.

Because the feature extraction ability is poor, the training time of the BP neural network is too long, and the practical application is limited, the emergence of the SVR algorithm partly compensates for the deficiency in BP neural networks. The SVR algorithm introduces the Lagrangian method to simplify the quadratic optimization problem in SVR calculations and introduces the kernel function to reduce the complexity of high-dimensional computations, thus allowing the SVR model to calculate the high efficiency characteristics. This paper first compares the processing power of the model to different time scale problems. The results show that the SVR algorithm can attain satisfactory statistics, but compared with the BP neural network, the two models have advantages and disadvantages. At monthly and daily time scales, the accuracy of the BP neural network model is higher than SVR, and at the hourly time scale, the calculation accuracy of the SVR model is higher than that of the BP neural network (Tables 3 and 4). In this study, the simulation ability of the model to different time scale problems reflects the processing power of the model to different data volumes to some extent. Therefore, the BP neural network has a stronger processing capability than SVR when the data volume is small, whereas

Table 6

Statistical performance on different flow regimes of BP neural network, SVR, and LSTM. The bold and underlined values indicate the best statistics for each flow regimes at same time scales.

		Low flow			Intermediate flow			High flow		
		BP	SVR	LSTM	BP	SVR	LSTM	BP	SVR	LSTM
Monthly scale	RMSE (m^3/s)	920.057	578.508	75.719	1262.017	1739.038	150.391	912.0822	3101.806	243.847
	RSR (-)	2.3805	1.4968	0.1959	0.3213	0.4428	0.0383	0.1768	0.6011	0.0473
	NSE (-)	-4.9370	-1.3472	0.9598	0.8947	0.8001	0.9985	0.9670	0.6186	0.9976
Daily scale	RMSE (m^3/s)	475.169	784.946	118.550	1350.407	1255.688	354.916	963.313	1897.362	179.351
	RSR (-)	1.2718	2.1009	0.2850	0.3805	0.3538	0.1000	0.1509	0.2973	0.0281
	NSE (-)	-0.6195	-3.4194	0.9174	0.8551	0.8747	0.9900	0.9772	0.9115	0.9992
Hourly scale	RMSE (m^3/s)	1058.031	496.4571	102.934	1861.919	899.2275	527.392	1606.689	1711.116	478.904
	RSR (-)	2.7868	1.3077	0.2710	0.5227	0.2524	0.1480	0.2584	0.2752	0.0770
	NSE (-)	-6.76746	-0.71019	0.9265	0.7267	0.9363	0.9781	0.9332	0.9243	0.9941

the advantages of the SVR model are gradually revealed when the data volume is large. Next, we compared the differences in calculation speed and the uncertainty of the two models, and the results show that, compared with the BP neural network, the computing speed of the SVR model significantly increased, but the time was still very long, and the uncertainty of the SVR model is high (Tables 3 and 4). Finally, under different flow regimes, our research results show that SVR still does not overcome the problem that the BP neural network cannot simulate low flow conditions (Table 6, Figs. 14–16).

In view of the above problems, namely, that the simulated speed of the traditional AI models is slow, the accuracy needs to be increased, and it is difficult to address extreme events, we turned our attention to the LSTM model, which performed well in solving the time series problems. Taking the GZB reservoir as an example, this paper constructs an LSTM model to predict the outflow of the GZB reservoir and explores the application of LSTM in reservoir operations. The results show that the LSTM model effectively makes up for the deficiency of the traditional AI models. Whether from the accuracy, uncertainty, calculation speed, extreme conditions processing, the LSTM has significant advantages over the BP neural network and SVR (Tables 3–6, Figs. 8–16). Especially at the hourly scale, facing vast amounts of data, LSTM performance is outstanding, and the training and prediction process takes only approximately an hour. Meanwhile, LSTM is the only one of the three models that can accurately simulate the outflow curve of GZB for the peaking operation period (Fig. 10). The experimental results for show that the LSTM model overcomes the disadvantage of previous models, i.e., poor simulation accuracy for extreme events, and the LSTM model can accurately predict reservoir outflow under low and high inflow conditions. In conclusion, compared with the BP neural network and SVR, the LSTM model has significant advantages in terms of simulation accuracy, stability and computing speed, and we believe that LSTM, as a deep learning model, has a strong sequential predictive ability and can be used for reservoir operation simulations. At present, the application of the LSTM model in reservoir operation has been rarely reported, but Zhang et al. (2018) arrived at similar conclusions in the study of sewage overflow prediction, which can provide some reference for us. Zhang et al. (2018) compared the predictive effects of the LSTM and SVR models on sewage discharge, and the results showed that the simulation accuracy of the LSTM model was significantly higher than that of the SVR model, which was similar to our conclusion.

In summary, with respect to the AI model applicability, the BP neural network is suitable if the model is able to learn the operation rule of a reservoir from a small amount of data (such as the research of reservoir long-term operations or if the influence factors of reservoir operation are small). For large reservoirs, short-term fine operation of reservoirs and low inflow conditions, the applicability of the BP neural network is weak. The main limitation of the BP neural network is that the independent feature extraction ability of the network is poor, so the time consumption required for satisfactory results is too long. The applicability of the SVR model and the BP neural network is similar. Although the SVR model improves computing speed, its ability to address massive data and low inflow conditions is still insufficient. By contrast, the LSTM model has obvious advantages and can quickly and accurately simulate the reservoir operation under various time scales and flow conditions. Therefore, LSTM can be used for medium- and long-term reservoir operation and short-term refinement operation simulation, as well as to address various emergencies. Meanwhile, the LSTM effectively solves the time consumption problems of the BP neural network and SVR model corrects for the fact that traditional AI cannot simulate low flow conditions or the outflow curve for the peak operation period.

6. Conclusions

Reservoir operation is an important part of reservoir management, and the theory and method of reservoir operation have been gradually

developed with the construction of reservoirs and hydropower stations in the early 20th century. At present, according to the theoretical basis of the model, reservoir operation models are divided into two main categories: models based on physical concepts and AI models based on data mining technology. However, a physical-based model is useful only if the operating rules incorporated in the simulation can realistically reflect the actual operation. In practice, the operation of a reservoir is affected by many uncertain factors, such as natural conditions and artificial demand, and the operation often deviates from the operating rules, which limits the application of such models. AI models, or data-driven models, are able to autonomously learn the operating rules from the historical operation data of a reservoir and thus have greater robustness and are good at dealing with complex factors.

This paper investigated the usefulness of two traditional AI models (BP neural network and SVR) and a new deep learning model (LSTM model) in assisting reservoir operation. Detailed discussion and recommendation are made with respect to the process of model parameter settings, the simulation performances, and the applications of employed AI models under different flow regimes and temporal resolution. The main conclusions are as follows:

- (1) With respect to parameter setting, our results show for the BP neural network and LSTM model, the effects of the number of maximum iterations on model performance should be prioritized. The effects of the number of hidden nodes on model performance are limited. For the SVR model, the simulation performance is directly related to the selection of the kernel function. Combined with previous research results, we consider that sigmoid and RBF kernel functions should be prioritized in uses of SVR model. Meanwhile, in the process of model construction, due to differences in data volume and structure, model parameters have a different influence on model performance. Therefore, we suggest that the model should be repeatedly trained before practical application to determine the optimal parameters and ensure the prediction ability of the model.
- (2) With respect to the AI model applicability, the BP neural network is suitable if the model is able to learn the operation rule of a reservoir from a small amount of data (such as the research of reservoir long-term operation or if the influence factors of reservoir operation are small). For large reservoirs, short-term fine operation of reservoirs and low inflow conditions, the applicability of the BP neural network is weak. The main reason to limit the application of the BP neural network is that the independent feature extraction ability of the network is poor, so the time consumption required for satisfactory results is long. The applicable conditions of the SVR model and BP neural network are similar. Although the SVR model improves computing speed, its ability to address massive data and low inflow conditions is still insufficient. By contrast, the LSTM model has obvious advantages and can quickly and accurately simulate the reservoir operation under various time scales and flow conditions. Therefore, LSTM can be used for medium- and long-term reservoir operation and short-term refinement operation simulation, as well as to address various emergencies.
- (3) The LSTM model can effectively solve the time consumption problem associated with the BP neural network and SVR model, and it has superior performances over other AI models in simulating reservoir operation during low-flow conditions or the outflow curve for the peak operation period, whereas traditional BP neural network and SVR model tend to fail.

Acknowledgement

The data presented in this paper are available at the official website of the China Three Gorges Corporation (CDEC): <http://www.ctg.com.cn/>. This paper is a part of a joint cooperation between the China Institute of Water Resources and Hydropower Research and the University of California, Irvine. This work was supported by the

National Key Research and Development Program of China [No. 2016YFE0102400], the Research Program of Songhua River Hydroelectric Power Generation Co. Ltd., the Fengman Dam Reconstruction Program Construction Bureau [JSFW[2015]198], the Scientific Research Program for the China Three Gorges Corporation [0799564], the Fundamental Research Funds for the China Institute of Water Resources and Hydropower Research [SD0145B162017], the U.S. Department of Energy [DOE Prime Award # DE-IA00000018], and California Energy Commission [CEC Award # 300-15-005].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jhydrol.2018.08.050>.

References

- Aboutalebi, M., Haddad, O.B., Loaiciga, H., 2015. Optimal monthly reservoir operation rules for hydropower generation derived with svr-nsgaii. *J. Water Resour. Plan. Manage.* 141 (11), 04015029.
- Amari, S., Wu, S., 1999. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks the Official J. Int. Neural Network Soc.* 12 (6), 783–789.
- Anctil, F., Michel, C., Perrin, C., Andréassian, V., 2004. A soil moisture index as an auxiliary ANN input for stream flow forecasting. *J. Hydrol.* 286 (1–4), 155–167.
- Asefa, T., Kembrowski, M., McKee, M., Khalil, A., 2006. Multi-time scale stream flow predictions: the support vector machines approach. *J. Hydrol.* 318 (1–4), 7–16.
- Chang, L.C., Chang, F.J., 2001. Intelligent control for modelling of real-time reservoir operation. *Hydrol. Process.* 15 (9), 1621–1634.
- Chang, Y.T., Chang, L.C., Chang, F.J., 2005. Intelligent control for modeling of real-time reservoir operation, part II: artificial neural network with operating rule curves. *Hydrolog. Process.: An Int. J.* 19 (7), 1431–1444.
- Chaves, P., Chang, F.J., 2008. Intelligent reservoir operation system based on evolving artificial neural networks. *Adv. Water Resour.* 31 (6), 926–936.
- Chaves, P., Tsukatani, T., Kojiri, T., 2004. Operation of storage reservoir for water quality by using optimization and artificial intelligence techniques. *Math. Comput. Simulation* 67 (4–5), 419–432.
- Chen, C.P., 1996. A rapid supervised learning neural network for function interpolation and approximation. *IEEE Trans. Neural Networks* 7 (5), 1220–1230.
- Chen, Y.H., Chang, F.J., 2009. Evolutionary artificial neural networks for hydrological systems forecasting. *J. Hydrol.* 367 (1–2), 125–137.
- Cheng, C.T., Feng, Z.K., Niu, W.J., Liao, S.L., 2015. Heuristic methods for reservoir monthly inflow forecasting: a case study of Xinfengjiang reservoir in Pearl river, china. *Water* 7 (8), 4477–4495.
- Chiang, Y.M., Chang, L.C., Chang, F.J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *J. Hydrol.* 290 (3–4), 297–311.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learn.* 20 (3), 273–297.
- Dynesius, M., Nilsson, C., 1994. Fragmentation and flow regulation of river systems in the northern third of the world. *Science* 266, 753–762. <https://doi.org/10.1126/science.266.5186.753>.
- Fernando, D.A.K., Shamseldin, A.Y., 2009. Investigation of internal functioning of the radial-basis-function neural network river flow forecasting models. *J. Hydrol. Eng.* 14 (3), 286–292.
- Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with LSTM. *Neural Comput.* 12 (10), 2451–2471.
- R.B. Girshick, J. Donahue, T. Darrell, J. Malik, In IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580–587.
- Haykin, S., 1994. Neural networks: a comprehensive foundation. *Neural Networks A Comprehensive Foundation* 71–80.
- Hejazi, M.J., Cai, X.M., 2009. Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRM) algorithm. *Adv. Water Resour.* 32 (4), 582–593.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hipni, A., El-Shafie, A., Najah, A., Karim, O.A., Hussain, A., Mukhlisin, M., 2013. Daily forecasting of dam water levels: comparing a support vector machine SVM model with adaptive neuro fuzzy inference system ANFIS. *Water Resour. Manage.* 27 (10), 3803–3823.
- ICOLD The world registers of dams 2011 International Commission on Large Dams Paris.
- Jain, A., Srinivasulu, S., 2004. Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. *Water Resour. Res.* W04302, 1–12.
- Jain, S.K., Das, A., Srivastava, D.K., 1999. Application of ANN for reservoir inflow prediction and operation. *J. Water Resour. Plann. Manage.* 125 (5), 263–271.
- Ji, C.M., Zhou, T., Huang, H.T., 2014. Operating rules derivation of jinsha reservoirs system with parameter calibrated support vector regression. *Water Resour. Manage.* 28 (9), 2435–2451.
- Johnson, V.M., Rogers, L., 2000. Accuracy of neural network approximators in simulation-optimization. *J. Water Resour. Plann. Manage.* 126 (2), 48–56.
- Johnson, S.A., Stedinger, J.R., Staschus, K., 1991. Heuristic operating policies for reservoir system simulation. *Water Resour. Res.* 27 (5), 673–685.
- Khalil, A., McKee, M., Kembrowski, M., Asefa, T., 2005. Sparse bayesian learning machine for real-time management of reservoir releases. *Water Resour. Res.* 41 (11), 4844–4847.
- Klipsch, J.D., Hurst, M.B., 2003. HEC- ResSim: Reservoir System Simulation 2003 User's manual US Army Corps of Engineers. Hydrologic Engineering Centre, Davis, CA.
- Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436.
- Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* 9 (9), 494–502.
- Lin, H.T., Lin, C.J., 2005. A study on sigmoid kernels for SVM and the training of non-PSD Kernels by SMO-type methods. Submitted to. *Neural Comput.* 27 (1), 15–23.
- Lin, J.Y., Cheng, C.T., Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. *Int. Assoc. Sci. Hydrol. Bullet.* 51 (4), 599–612.
- Lippmann, R.P., 1987. An introduction to computing with neural nets. *IEEE Acoustics Speech Signal Process. Magazine* 16 (2), 4–22.
- Loucks, D.P., Sigvaldason, O.T., 1981. Multiple-reservoir operation in North America. *Surface Water Impoundments*. ASCE 6774, 711–728.
- Moody, J.O., Antsaklis, P.J., 1996. The dependence identification neural network construction algorithm. *IEEE Trans. Neural Networks* 7 (1), 3–15.
- Moradkhani, H., Hsu, K.L., Gupta, H.V., Sorooshian, S., 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *J. Hydrol.* 295 (1–4), 246–262.
- Moriatis, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50 (3), 885–900.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I: a discussion of principles. *J. Hydrol.* 10 (3), 282–290.
- Oliveira, R., Loucks, D.P., 1997. Operating rules for multi-reservoir systems. *Water Resour. Res.* 33 (4), 839–852.
- Senthil Kumar, A.R., Ojha, C.S.P., Goyal, M.K., Singh, R.D., Swamee, P.K., 2012. Modeling of suspended sediment concentration at Kasol in India using ANN, Fuzzy Logic and Decision Tree algorithms. *J. Hydrol. Eng.* 17 (3), 394–404.
- Shang, Y., Lu, S., Ye, Y., Liu, R., Shang, L., Liu, C., et al., 2018. China's energy-water nexus: hydropower generation potential of joint operation of the three gorges and qingjiang cascade reservoirs. *Energy* 142, 14–32.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *arXiv* 1506.04214v2.
- Singh, J., Knapp, H.V., Arnold, J.G., Demissie, M., 2010. Hydrological modeling of the iroquois river watershed using HSPF and SWAT. *Jawra J. Am. Water Resour. Assoc.* 41 (2), 343–360.
- Su, J., Wang, X., Liang, Y., Chen, B., 2013. Ga-based support vector machine model for the prediction of monthly reservoir storage. *J. Hydrol. Eng.* 19 (7), 1430–1437.
- Thirumalaiah, K., Deo, M.C., 1998. River stage forecasting using artificial neural networks. *J. Hydrol. Eng.* 3 (1), 26–32.
- WCDM Dams and development: a new framework for decisionmaking: the report of the world commission on dams 2000 Earthscan Publications Ltd London and Sterling.
- Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. *Proc. IEEE* 78 (10), 1550–1560.
- Wielgosz, M., Skoczeń, A., Mertlik, M., 2017. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. *Nucl. Instr. Methods Phys. Res.* 867, 40–50.
- Wu, C.L., Chau, K.W., 2011. Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. *J. Hydrol.* 399 (3), 394–409.
- Yang, T., Gao, X., Sorooshian, S., Li, X., 2016. Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resour. Res.* 52, 1626–1651.
- Yang, T., Asanjan, A.A., Faridzad, M., Hayatbini, N., Gao, X., Sorooshian, S., 2017a. An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis. *Inf. Sci.* 418, 302–316.
- Yang, T., Asanjan, A.A., Welles, E., Gao, X., Sorooshian, S., Liu, X., 2017b. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resour. Res.* 53 (4), 2786–2812.
- Yao, X., 1999. Evolving artificial neural networks. *Proc. IEEE* 87 (9), 1423–1447.
- Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J. Hydrol.* 530, 829–844.
- Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21-A demand-, priority-, and preference-driven water planning model: Part1: model characteristics. *Water Int.* 30 (4), 487–500.
- Zhang, D., Lindholm, G., Ratnaweera, H., 2018. Use long short-term memory to enhance internet of things for combined sewer overflow monitoring. *J. Hydrol.* 556, 409–418.
- Zaytar, A.M., Amrani, E.C., 2016. Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. *Int. J. Computer Appl.* 143 (11), 7–11.