MDPI

Article

Streamflow Forecasting via Two Types of Predictive Structure-Based Gated Recurrent Unit Models

Xuehua Zhao [1,*], Hanfang Lv [1], Yizhao Wei [1], Shujin Lv [2] and Xueping Zhu [1]

[1] College of Water Resources and Engineering, Taiyuan University of Technology, Taiyuan 030024, China; hbbdlhf@163.com (H.L.); weiyizhao0609@163.com (Y.W.); xpzhu01@163.com (X.Z.)

[2] Department of Data Science and Software Engineering, Baoding University, Baoding 071000, China; lshjin@sina.com

* Correspondence: zhaoxh688@126.com; Tel.: +86-13363435589

Abstract: Data-intelligent methods designed for forecasting the streamflow of the Fenhe River are crucial for enhancing water resource management. Herein, the gated recurrent unit (GRU) is coupled with the optimization algorithm improved grey wolf optimizer (IGWO) to design a hybrid model (IGWO-GRU) to carry out streamflow forecasting. Two types of predictive structure-based models (sequential IGWO-GRU and monthly IGWO-GRU) are compared with other models, such as the single least-squares support vector machine (LSSVM) and single extreme learning machine (ELM) models. These models incorporate the historical streamflow series as inputs of the model to forecast the future streamflow with data from January 1956 to December 2016 at the Shangjingyou station and from January 1958 to December 2016 at the Fenhe reservoir station. The IGWO-GRU model exhibited a strong ability for mapping in streamflow series when the parameters were carefully tuned. The monthly predictive structure can effectively extract the instinctive hydrological information that is more easily learned by the predictive model than the traditional sequential predictive structure. The monthly IGWO-GRU model was found to be a better forecasting tool, with an average qualification rate of 91.66% in two stations. It also showed good performance in absolute error and peak flow forecasting.

Keywords: gated recurrent unit; improved grey wolf optimizer; monthly streamflow forecasting; data-driven modeling

1. Introduction

Accurate and reliable streamflow forecasting is of great significance for water resource management [1] and optimal allocation [2], environmental protection [3], and flood control [4]. Thus, it is imperative to develop a robust and accurate streamflow forecasting method for hydrologists. However, the hydrological process is controlled by chaotic hydrological behavior presenting the complex features of nonlinearity and nonstationarity. Therefore, accurate and robust streamflow forecasting is expected to ensure the reliable operation of practical engineering systems. For this reason, constructing a new predictive model for streamflow series is essential.

To date, scholars have developed a number of forecasting models using data-driven approaches, such as ANN (artificial neural network) and GEP (gene expression programming), to emulate hydrological behavior because it requires the least information [5,6]. ANN is a widely used model with good performance in comparison to traditional regression models [7]. For example, the extreme learning machine (ELM), support vector machine (SVM), and least square support vector machine (LSSVM) have shown potential for hydrological prediction [8–10]. The ELM, a novel machine learning method, requires fewer iterations to adjust hidden neuron parameters in comparison with other single artificial neural networks [11]. The LSSVM reduces the computational time of the SVM using the least squares linear system as its loss function [10]. However, these traditional machine

learning algorithms belong to the theory of shallow learning. The method of time-series information mining refers to the time-embedding process to predict significant occurrences in the near future [4]. Thus, to overcome this problem, a recurrent neural networks (RNN) model has been proposed, and many studies have made use of this deep learning technique in which the intrinsic nature of the time-series can be learned [12]. The gated recurrent unit (GRU), as a special RNN that simplifies the structure of a long short-term memory model, gives the network a long memory for taking significant information from previous data, which can effectively model a complex and long time-series [13]. With the integration of deep learning theories and time-series simulations, the GRU has been successfully applied in many fields to solve prediction issues [14–16], such as miscellaneous electric loads, wind speed, and short-term photovoltaic power. However, there is scant research on streamflow forecasting.

To further enhance forecast accuracy, hybrid models are becoming popular alternatives because they can outperform single models [17]. Combining methods can build hybrid models with optimization algorithms, which can be used to identify the optimal parameter combination and calibrate models to enhance their robustness. Many algorithms are applied to improve single model forecasting abilities in hydrology, including the genetic algorithm (GA) [18], Fruit fly Optimization Algorithm (FOA) [19], grey wolf optimizer (GWO) [20], and particle swarm optimization (PSO) [21]. Based on the social hierarchy and hunting behaviors of grey wolves, Mirjalili et al. [22] proposed the GWO algorithm, one of the newest swarm intelligence algorithms. The GWO algorithm avoids several local solutions and balances between exploration and exploitation efficiency [20]. Moreover, it is simple, flexible, and has very few parameters [23]. Thus, the GWO algorithm is widely used to solve different problems, including global optimization problems [24,25], optimal reactive power dispatch problems [26], and route-planning problems [27]. Additionally, the improved GWO (IGWO) algorithm finds a new objective function to avoid under-fitting or over-fitting. Therefore, we exploited the IGWO method in this study to improve the robustness of the hybrid model.

Generally, there are two types of data widely used in the time-series analysis (dynamic and static variations), which inspired us to transform the streamflow time-series into two types of predictive structure: a sequential structure and a monthly structure. Many articles have investigated the prediction of time-series data with the traditional sequential series, such as rainfall-runoff [3], monthly streamflow [28], and sediment yield [29]. However, the streamflow series can be reconstructed into twelve groups of series for the same month, from January to December. Each series as the input can be used to build twelve models in streamflow forecasting, and each prediction results of each month can be used to construct the final one year and monthly streamflow forecasts. Thus, based on this method, we propose a new forecasting model to predict future streamflow from historical time-series data on a monthly basis, which is the monthly predictive structure-based model. Here, we used the integrated deep learning model IGWO-GRU and named the two types of predictive structure-based models as the "monthly IGWO-GRU" and "sequential IGWO-GRU", which are investigated for their abilities in solving benchmark and real problems. To the best of the authors' knowledge, the application of a monthly predictive-structure technique is yet to be explored in the forecasting of streamflow.

The purposes of this study are to (1) use structure-based models (sequential and monthly streamflow series) to form two types of time-series as inputs for the predictive method; (2) utilize the deep learning method GRU to learn the instinct relationship between the input and output; (3) to balance the generation and optimization capability of the model, using the IGWO method to tune the GRU model with suitable parameter combinations; (4) compare the prediction with the proposed model to that of the other state-of-the-art models that have been quantitatively analyzed. The results of the experiments of this study in the Fenhe River have demonstrated the monthly IGWO-GRU model can provide an effective modeling approach to capture the nonlinear characteristics of streamflow series, thus providing the most accurate prediction results among the models considered.

It must be declared that the application of GRU and its variants are very uncommon in streamflow forecasting studies. In addition, no such studies with the monthly predictive structure-based models have been performed so far, which is substantiated in this study.

The remainder of this paper is arranged as follows. Section 2 describes the methods of the proposed hydrological model. Section 3 introduces the two data structure-based models. Section 4 presents experiments and discusses the results of prediction in the case study. Section 5 is the last section, which outlines the conclusions of this paper.

2. Methodology

2.1. GRU

The traditional ANN has a simple corresponding relationship between inputs and outputs; however, RNN has an advanced structure that involves feedback connections that take previous content into account throughout the process [30]. In other words, RNN has the "memory" to retain information about the history of all steps of the series, as shown in Figure 1. The mathematical formulation of the RNN algorithm is as follows:

$$H_t = f(W \cdot H_{t-1}, W \cdot x_t) \tag{1}$$

where H_t and H_{t-1} denote the current state and the previous state, respectively; x_t is the inputs, and W represents the weights of calculations in the neuron. Every step of the process shares the same weights and reinforces learning from previous time steps.
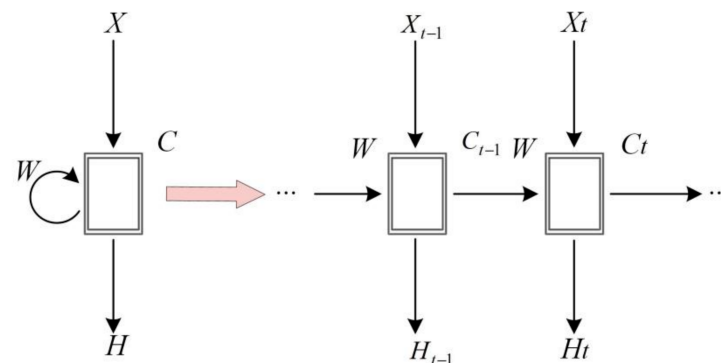

Figure 1. Structural schematic diagram of a recurrent neural network (RNN).

The RNN is considered an effective method to deal with problems of complex hydrological series but defective in processing long-term memory because of gradient vanishing and gradient explosion [13]. Long short-term memory (LSTM) and GRU are important enhanced versions of RNN. LSTM consists of three gate units, namely, the input gate, forget gate, and output gate. The input gate or output gate decides whether to store certain information in each step. The forget gate controls whether to keep or override the information from previous time steps [31].

Proposed by Cho et al. [32], the GRU method is a modern RNN and replaces the ordinary neuron in the RNN with a memory neuron that solves the vanishing or exploding gradient problem, thanks to the BPTT (back propagation through time) method. Figure 2 displays the structure of the neuron in the hidden layer of GRU. The neuron has two gates, simplifying the structure of LSTM: an update gate and a reset gate, which modulate the flow of information when the neural state is updated at each step. The update gate decides how much information from the previous step flows to the neuron and the reset gate controls whether to ignore the previous state and upset the current or not [17,32]. This allows the neuron to drop irrelevant information, which is not useful for mapping the input and output in the future and further helps the neuron remember long-term information by reducing the memory burden [13].

This is depicted in Figure 2, which shows the present step, *t*, the output of the previous state, H_{t-1}, combined with the input of the current state; X_t is the input of the update gate, Z_t, and reset gate, R_t.

$$Z_t = \sigma(W_Z \cdot [H_{t-1}, x_t]) \tag{2}$$

$$R_t = \sigma(W_R \cdot [H_{t-1}, x_t]) \tag{3}$$

where σ presents sigmoid function, W_Z and W_R are the weights of update gate and reset gate, respectively.

Then, the new hidden state $\widetilde{H}_t$ is calculated by the following:

$$\widetilde{H}_t = \varphi(W \cdot [R_t \cdot H_{t-1}, x_t]) \tag{4}$$

where φ presents the tanh function, W is the weight learned from the reset gate.

At last, the output of the present neuron is formulated as follows:

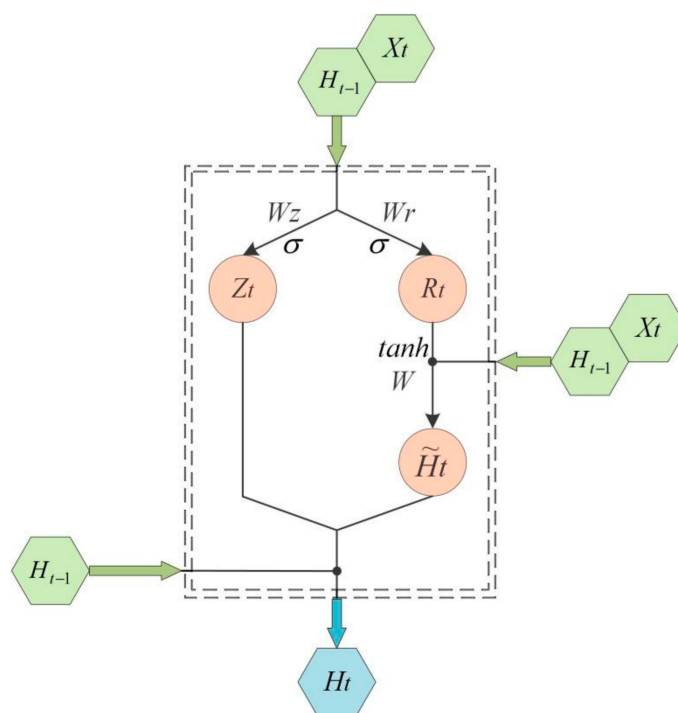$$H_t = Z_t \cdot \widetilde{H}_t + (1 - Z_t) \cdot H_{t-1} \tag{5}$$

Figure 2. The structure of the neuron in the gated recurrent unit (GRU).

2.2. IGWO

Generally, grey wolves mostly live in a pack and follow a strict social hierarchy. The social positions are divided into four groups: alpha, beta, delta, and omega. Alphas are the leaders, making decisions and commanding other wolves. Beta is the second level in the hierarchy, and they help the alpha with pack activities. The lowest ranking wolves are omegas, and they submit to all other wolves. The deltas are at the third level of the hierarchy, between omegas and betas [22]. Mathematical models of their pack hunting technique consider the alpha, beta, delta, and omega as the fittest solution (α), the second-best solution (β), third-best solution (σ), and the remaining candidate solutions (Ω), respectively.

The hunting behaviors of grey wolves consists of two main phases: encircling the prey and attacking prey. First, wolves encircle the prey reaching any positions. This behavior is mathematically modeled, as shown below:

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{6}$$

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \tag{7}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{8}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{9}$$

where t is the current iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_P$ indicates the position of the prey, and $\vec{X}$ is the position of a grey wolf.

Second, T_t supposes that the alpha, beta, and delta have better knowledge about the location of prey. The model saves the first three solutions and updates the positions of other agents according to the estimation of the position of the alpha, beta, and delta. This behavior is repeated until the best-fit solution is obtained. Figure 3 shows the process of position updating in the GWO, which can explain the following equations:

$$\vec{D}_a = \left| \vec{C}_1 \cdot \vec{X}_a - \vec{X} \right|, \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right|, \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \tag{10}$$

$$\vec{X}_1 = \vec{X}_a - \vec{A}_1 \cdot \left(\vec{D}_a \right), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left(\vec{D}_\beta \right), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \left(\vec{D}_\delta \right) \tag{11}$$

$$\vec{X}(t+1) = \frac{1}{3} \left(\vec{X}_1 + \vec{X}_2 + \vec{X}_3 \right) \tag{12}$$

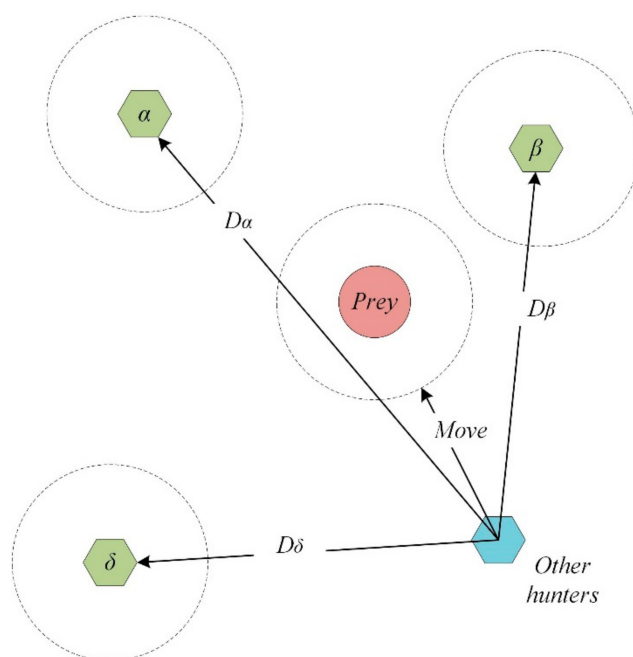where $\vec{X}_a, \vec{X}_\beta, \vec{X}_\delta$ are the updated positions.

Figure 3. Updating positions in the grey wolf optimizer (GWO).

2.3. Evaluation Indicators

This study evaluated model performance using the Nash–Sutcliffe efficiency coefficient (nse), root mean squared error (RMSE), correlation coefficient (r), mean absolute percentage error (MAPE), qualification rate (qr), absolute error (AE) between original streamflow series, and the predicted series:

$$\text{nse} = 1 - \frac{\sum\limits_{i=1}^{n}(x_i - x_{i0})^2}{\sum\limits_{i=1}^{n}(x_{i0} - x)^2} \tag{13}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{i0})^2} \tag{14}$$

$$\text{r} = \frac{\sum_{i=1}^{N}(x_i - \overline{x}_i)(x_{i0} - \overline{x}_{i0})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x}_i)^2}\sqrt{\sum_{i=1}^{N}(x_{i0} - \overline{x}_{i0})^2}} \tag{15}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - x_{i0}}{x_i}\right| \tag{16}$$

$$\text{qr} = \frac{k}{l} \times 100\% \tag{17}$$

$$\text{AE} = |x_i - x_{i0}| \tag{18}$$

where x_i and x_{i0} denote the original streamflow series and the predicted series, respectively; x represents the mean of the original streamflow series; n is the number of streamflow series; k denotes the number of the eligible prediction, which defined that the relative error of the forecast was less than 20% of the maximum amplitude of the streamflow series; l is the length of streamflow series at the validation period; AE is the difference between the predicted and original streamflow, which ideally must be 0 [33]. nse = 1, RMSE = 0, r = 1, MAPE = 0, qr = 100%, and AE = 0 are desirable for a perfect model in streamflow forecasting. In practical applications, nse $\geq$ 0.5 and qr $\geq$ 60% are considered to represent a satisfactory performance of the model [34].

3. Two Hybrid Forecasting Models

The procedures of the two structure-based predictive models for streamflow forecasting are described as follows:

The first model was the sequential time-series structure-based predictive model (namely, sequential IGWO-GRU). The streamflow series was divided into calibration, test, and validation datasets. Then, the IGWO method was used to tune the hyperparameters of the deep learning GRU model through the calibration and test period. Finally, the model's prediction for the two hydrological stations' outputs was predicted. The detailed steps of the sequential IGWO-GRU model were as follows:

(1) Let Y_i, $i = 1, 2, \ldots, n$, denote the streamflow time series.
(2) PACF (Partial Autocorrelation Function) was implemented to select four types of input and five types of output to create twenty combinations of input and output.
(3) The ELM, LSSVM, and GRU methods modeled the relationship between inputs and outputs.
(4) The dataset was divided into calibration, test, and validation datasets.
(5) IGWO was utilized to find the optimal combination of parameters for predictive models.
(6) Obtain forecasting results for the ELM, LSSVM, and GRU models, defined as $Pred_\text{E}$, $Pred_\text{L}$, and $Pred_\text{G}$, respectively.

The second model was based on the streamflow series for each month, called the monthly IGWO-GRU. The main difference between the first model and the second is that the streamflow series was divided into twelve groups (one for each month) to reconstruct the twelve predictions to the final forecasting results. First, the original streamflow series was divided into twelve sets of streamflow series for months ranging from January to December for the Shangjingyou station and the Fenhe reservoir station. Then, the GRU was utilized to forecast each series, and the model was optimized by IGWO, which selects the optimal parameter combination for modeling. After that, all of the predictions for each month's time-series composed the forecasting results. The steps of the monthly IGWO-GRU model were similar to those of the sequential IGWO-GRU model:

(1) Let Y_i, $i = 1, 2, \ldots, n$, denote the streamflow time-series.
(2) The streamflow series for the same month. Streamflow for January was considered as $Y_{JAN} = Y_1, Y_{13}, \ldots$. Similarly, streamflow for February was $Y_{FEB} = Y_2, Y_{14}, \ldots$, and for December was $Y_{DEC} = Y_{12}, Y_{24}, \ldots$
(3) Select the inputs and outputs for each of the monthly streamflow series through PACF.
(4) The dataset was divided into calibration, test, and validation datasets.
(5) IGWO was utilized to find the optimal combination of parameters for predictive models.
(6) Obtain forecasting results through the predictions of the GRU model and the predictions for each month, namely, $P_{JAN}, P_{FEB}, \ldots, P_{DEC}$.
(7) Reconstruct $P_{JAN}, P_{FEB}, \ldots, P_{DEC}$ into the final forecasting results.

Figure 4 shows the flowcharts of ELM, LSSVM, and traditional sequential IGWO-GRU models, and Figure 5 shows the monthly IGWO-GRU forecasting models.
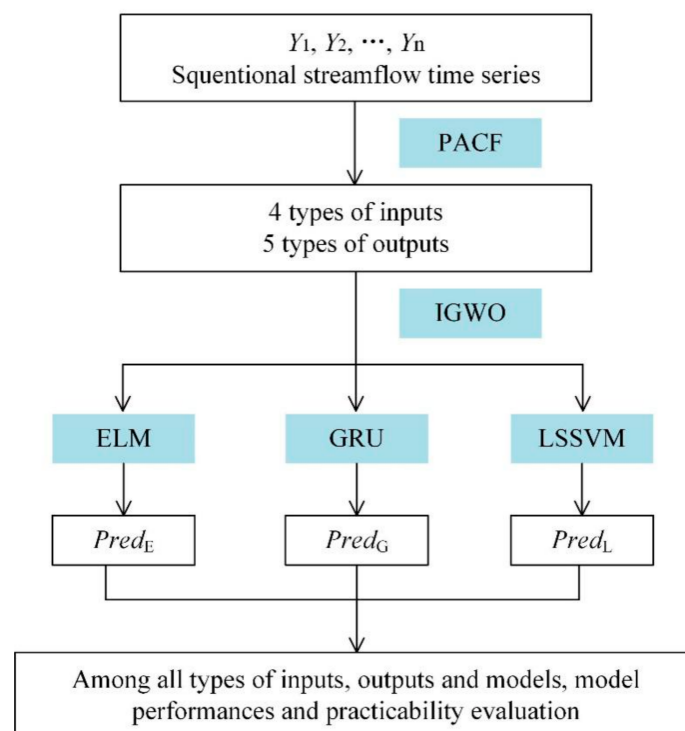

Figure 4. Flowchart of the least-squares support vector machine (LSSVM), extreme learning machine (ELM), and sequential IGWO-GRU model.
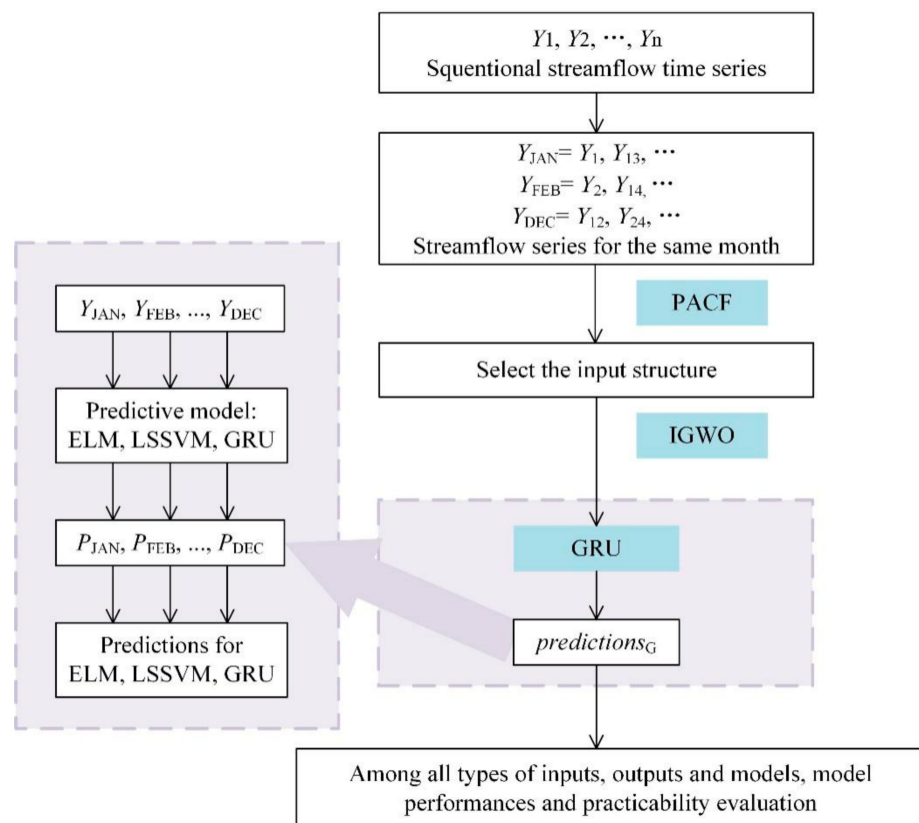
Figure 5. Flowchart of the monthly IGWO-GRU model.

4. Case Study

4.1. Study Area and Dataset

Fenhe River, the native river of Shanxi province, is the second-largest tributary of the Yellow River (Figure 6). It flows 716 km and has a total area of about 38,728 km^2. The basin contains the main industrial, urban residential, and agricultural areas of the Shanxi Province [35]. In this study, the hydrological stations at the Shangjingyou station and the Fenhe reservoir station are located in the upper reaches of the Fenhe River, which is an important water source and the largest functional area of drinking water in Taiyuan city. Therefore, developing a robust streamflow forecasting model with higher accuracy is essential for water resource allocation and sustainable economic growth in the Fenhe River upstream area.

There are 59 years of monthly streamflow data from the Fenhe reservoir station and 61 years for the Shangjingyou station available for the Fenhe River. Figure 7a,b shows the monthly series and boxplots of streamflow series for the same month at the Shangjingyou station, and Figure 7c,d shows the streamflow series from the Fenhe reservoir station. The streamflow time-series data for the same month exhibit greatly stochastic and nonlinear behavior, as shown in Figure 7b,d. In addition, the same month's streamflow data vary greatly from one month to another, demonstrating the difficulty of modeling it.
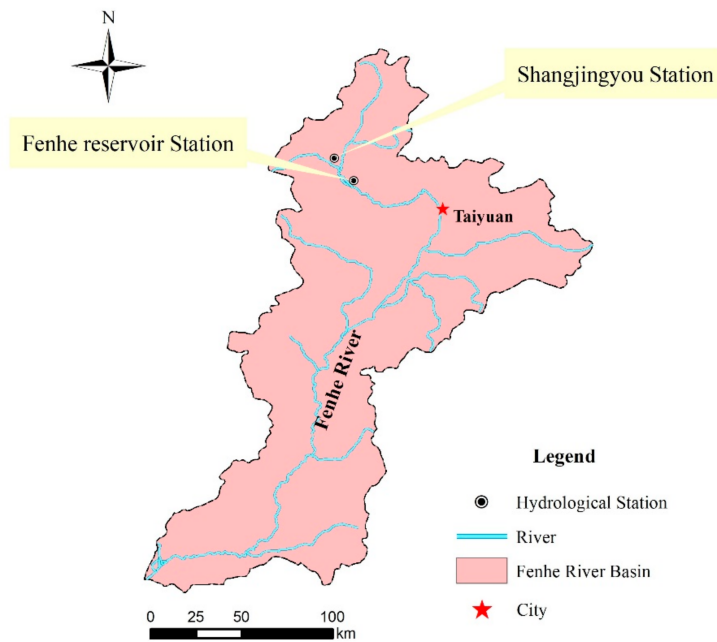
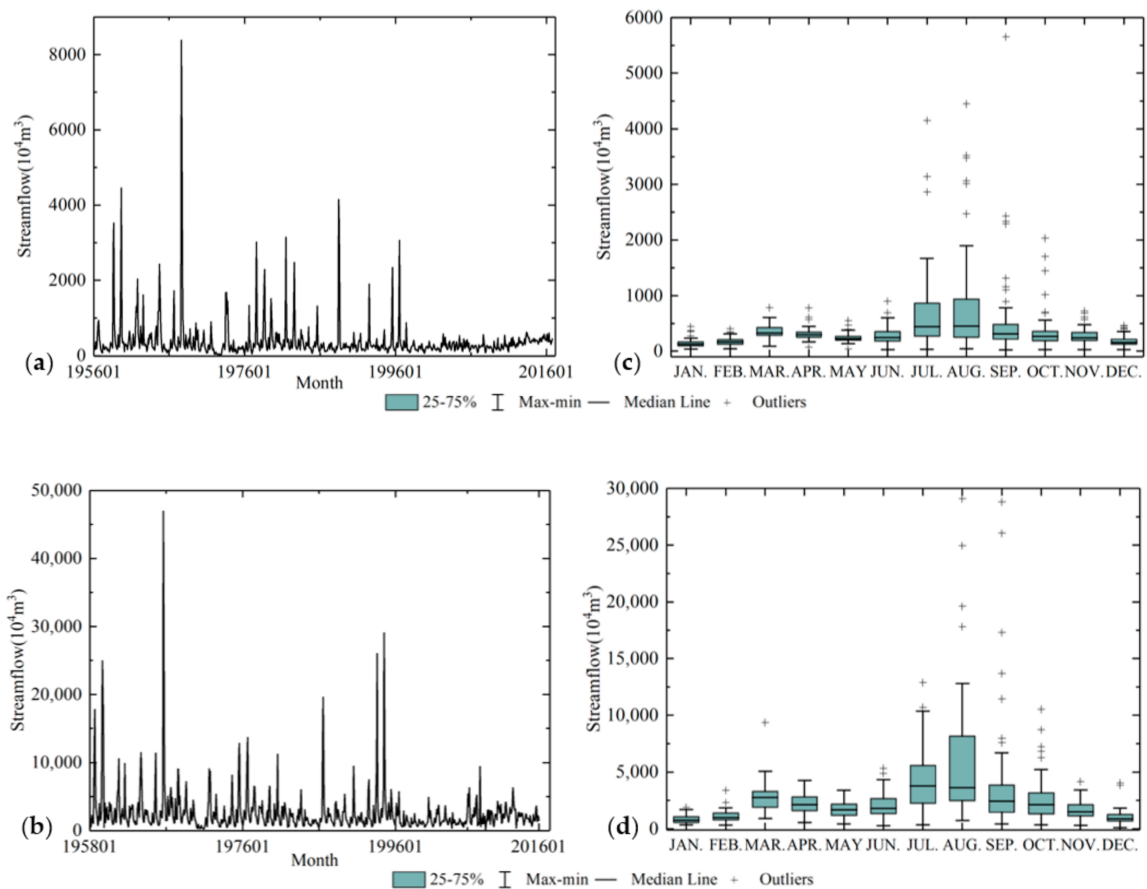**Figure 6.** The study area and location of the two hydrological stations.

Figure 7. Monthly streamflow data at the Shangjingyou station and the Fenhe station. (**a**) Monthly streamflow for Shangjingyou station; (**b**) Boxplots of streamflow for Shangjingyou station; (**c**) Monthly streamflow for Fenhe reservoir station; (**d**) Boxplots of streamflow for Fenhe reservoir station.

There is variation in the ranges of the streamflow for each month, which reflects the highly stochastic pattern of the streamflow. From Figure 7b,d, the range of outliers was larger between July and October, which means the streamflow series in the wet season exhibited greater stochasticity. The smaller distribution of streamflow between November and June showed a more stable change. Furthermore, Figure 7 presents two types of time-series data (sequential and monthly interval), which inspired us to build forecasting models based on these two types of predictive structure.

To enhance the learning process of the model and prevent information from being leaked from the calibration dataset, the streamflow records were split into three groups—the calibration dataset, test dataset, and validation dataset [36]. The traditional single ELM, LSSVM, and sequential IGWO-GRU models were used to predict future streamflow based on the streamflow data from January 1956 to December 2013 at the Shangjingyou station from January 1958 to December 2013 at the Fenhe reservoir station. The monthly IGWO-GRU model was employed with twelve bunches of streamflow data.

4.2. Model Development

In this section, detailed information about model development is provided. Every model used the same dataset structure and period division with the rolling learning-prediction procedure.

(1)   ELM and LSSVM model

In the single LSSVM model, the radial basis function was selected as the kernel function, and the grid search method was employed to identify the best parameter combination by comparing the mean absolute error (MAE) during the calibration period. The sigmoid function was chosen as the transfer function in the single ELM model, and the best number of neurons was determined through trial-and-error with the MAE. In addition, there were slight fluctuations in the ELM model. To ensure stability, each output was taken as the arithmetic average of 10 executions of the model. The ELM model and LSSVM model had 732 datasets from the Shangjingyou station and 708 datasets from the Fenhe reservoir station.

(2)   IGWO-GRU model

For the GRU method, if the learning rate of the deep learning model was large, the model took a long time to converge or had difficulty in reaching the convergence state in the end [37]. Moreover, the number of neurons in the hidden layer controls the activation functions for the "memory-forget" work at gates. Therefore, the learning rate and the number of hidden neurons in the GRU model were carefully tuned through several experiments. Hence, to increase the efficiency of the optimization, the IGWO was employed with GRU to optimize parameters. The initial parameters of the IGWO are listed in Table 1. Particularly, the objective function of the IGWO method is shown in Equation (13). To prevent the models from overfitting and underfitting, the objective function for improved GWO-GRU was to minimize the mean value of calibration loss and test loss when the difference of two losses was found within 0.005.

Table 1. Initial parameters of improved grey wolf optimizer (IGWO).

Models	Parameters	Value
IGWO	Population size	5
	Iterations number	10
	Range partitions (γ) [1]	(8, 200)
	Range partitions (lr) [2]	(0.0005, 0.1)
	Loss	MAE [3]

[1] γ means the number of neurons in the GRU hidden layer. [2] lr is the learning rate of the GRU model. [3] MAE is the mean absolute error.

In addition, the sequential IGWO-GRU model for five types of output (6, 12, 24, 36, and 72 streamflow values for output, respectively) forecasting with four types of inputs (Ma, Mb, Mc, and Md) was determined by ACF and PACF [38], as shown in Table 2. The main procedure of PACF and ACF are (1) Make PACF and ACF plot with 95% confidence bands. (2) The variable will be selected as an input when the PACF value or the ACF value is out of the confidence bands. (3) All the previous selected variables combine into different types of inputs–outputs models, as shown in Table 2. The goodness-of-fit (RMSE) between the predicted and original streamflow is displayed in the rose plots (Figure 8). It can be seen that the Md model for the Shangjingyou station and Ma model for the Fenhe reservoir station were the best types of input in terms of RMSE between predictions and original measured streamflow data. For the different stations, using different inputs–outputs models have different ranks of prediction accuracy—Ma model for the Shangjingyou station and Md for the Fenhe reservoir station. In addition, the best inputs models for the two stations achieved the highest accuracy whatever types of outputs, and considering the multi-step outputs, the Md model had 36 steps of outputs for the Shangjingyou station, and the Ma model had 12 steps of outputs for the Fenhe reservoir station; both had a good accuracy compared to the other input and output combinations. Therefore, the Md and Ma were selected as the input models for the sequential IGWO-GRU for the Shangjingyou station and the Fenhe reservoir station, respectively. What is more, the sequential IGWO-GRU model had 732 datasets in the Shangjingyou station and 708 datasets in the Fenhe reservoir station. The monthly IGWO-GRU model has 61 datasets in the Shangjingyou station and 59 datasets in the Fenhe reservoir station for each month, and a total of 732 datasets in the Shangjingyou station and 708 datasets in the Fenhe reservoir station in all.

Table 2. Different input combinations to forecast future streamflow.

Models	Inputs	Outputs
Ma	Q_{t-1}, Q_{t-2} [2]	Q_t [1]
Mb	Q_{t-1}, Q_{t-2}, Q_{t-11}, Q_{t-12}	Q_t
Mc	Q_{t-1}, Q_{t-2}, Q_{t-11}, Q_{t-12}, Q_{t-13}	Q_t
Md	Q_{t-1}, Q_{t-2}, Q_{t-12}, Q_{t-13}	Q_t

[1] Q_t is the target output variable [2] Q_{t-i} means the i antecedent variables of the target output variables.
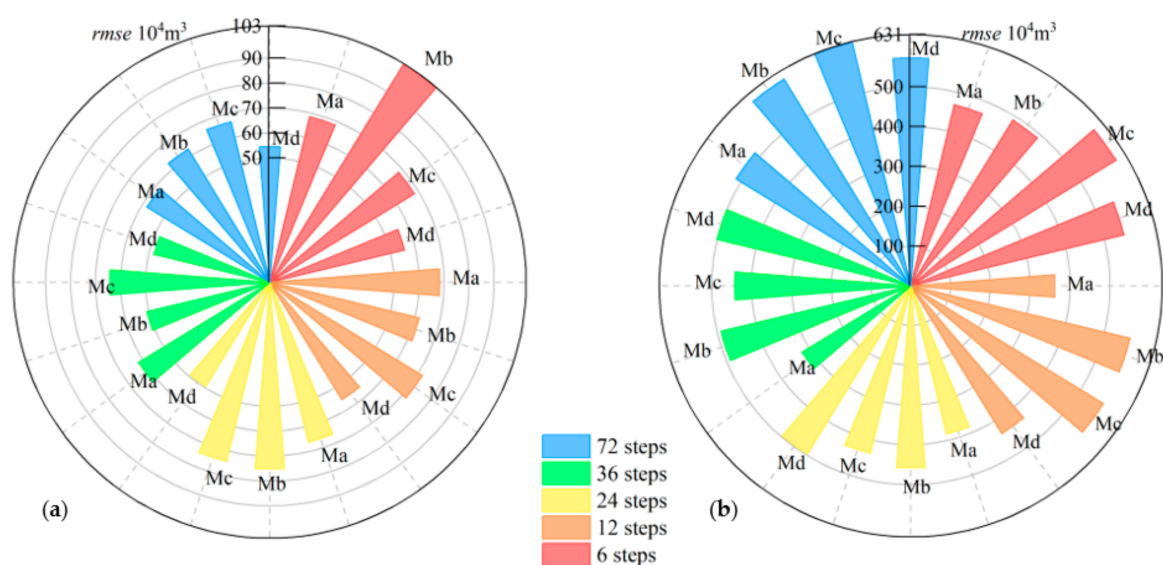

Figure 8. Rose plots for different combinations of inputs and outputs in the two hydrological stations for the sequential IGWO-GRU model. (**a**) Rose plots for Shangjingyou station; (**b**) Rose plots for Fenhe reservoir station.

4.3. Results and Discussion

The two types of predictive structure-based model (monthly IGWO-GRU and sequential IGWO-GRU) were appraised in comparison with single ELM and single LSSVM, using statistical metrics and error distributions between the predicted and original streamflow series.

The precision of the monthly IGWO-GRU model was compared to the sequential IGWO-GRU, single ELM and LSSVM models. The prediction results for each station are shown on the basis of r, nse, and RMSE in Table 3. The monthly IGWO-GRU model applied at the Shangjingyou station attained the highest r and nse and the smallest RMSE compared to the sequential IGWO-GRU, single ELM, and LSSVM model (Table 3). Similarly, considering the largest magnitudes of rand r and nse and the smallest magnitudes of RMSE among all of the forecasting models, the performance of the monthly IGWO-GRU was the best for the Fenhe reservoir station, followed by the sequential IGWO-GRU, single ELM, and LSSVM model. This emphasizes the forecasting ability of the proposed model for streamflow forecasting. The reason why the two station were so different is that the streamflow values between the Shangjingyou station and the Fenhe reservoir station are scaled. The Shangjingyou station is in the upper reaches of the Fenhe reservoir station, and after many rivers' confluence to the Fenhe River, the streamflow measured by the Fenhe reservoir station was much higher than the Shangjingyou station. Therefore, there were quite different evaluation indicators between the two stations.

Table 3 also presents MAPE and qr values of four types of models for two stations. It can be seen that the ELM model had poor performance compared with the other models. The LSSVM showed good performance, reducing the MAPE value by an average of 11.99% for the two stations, compared with the ELM model. The sequential IGWO-GRU model had a better performance and was superior to the ELM model, reducing the MAPE value by an average of 36.38%, compared with the ELM model. The monthly IGWO-GRU model showed the best performance among the four models and outperformed the ELM, LSSVM, and sequential IGWO-GRU model, reducing the MAPE values by an average of 55.8%, compared with the ELM model. Furthermore, the monthly IGWO-GRU model yielded a qr value of 97.22% in the Fenhe reservoir station and 86.11% in the Shangjingyou station, which falls in the "very good" and "good" grades, respectively [34]. This confirms the practicability of the monthly structure-based deep learning predictive model IGWO-GRU for forecasting streamflow series.

Table 3. Performance evaluation of the compared models applied in the forecasting of streamflow series in the validation period based on three evaluation indexes.

Sites	Performance Metrics	Model			
		ELM	LSSVM	Sequential IGWO-GRU	Monthly IGWO-GRU
Shangjingyou	r	0.458	0.108	0.708	0.896
	nse	0.135	0.261	0.421	0.675
	RMSE	63.315	55.519	35.862	19.615
	MAPE	0.168	0.153	0.109	0.060
	qr	50.00%	50.00%	75.00%	97.22%
Fenhe reservoir	r	0.727	0.850	0.792	0.971
	nse	0.414	0.356	0.589	0.922
	RMSE	361.564	346.713	361.999	147.666
	MAPE	0.093	0.079	0.058	0.049
	qr	66.67%	41.67%	66.67%	86.11%

Further analysis via boxplots (Figure 9) at each station was carried out to perform comparisons with different forecasting skills through absolute error. In addition, Figure 9 also exhibits the normal curve and dot matrix of value distribution. The smaller distribution of AE and a much smaller quartile range was acquired by the monthly IGWO-GRU model

for the Shangjingyou station, followed by the sequential IGWO-GRU model, single ELM, and LSSVM. Similarly, at the Fenhe reservoir station, AE distributions of the monthly model also achieved higher accuracy in relation to the other models. In addition, the results clearly indicated that the performance of models ranked from best to worst was the monthly IGWO-GRU, sequential IGWO-GRU, LSSVM, and ELM. This further confirms the suitability of the monthly IGWO-GRU model for forecasting streamflow values.
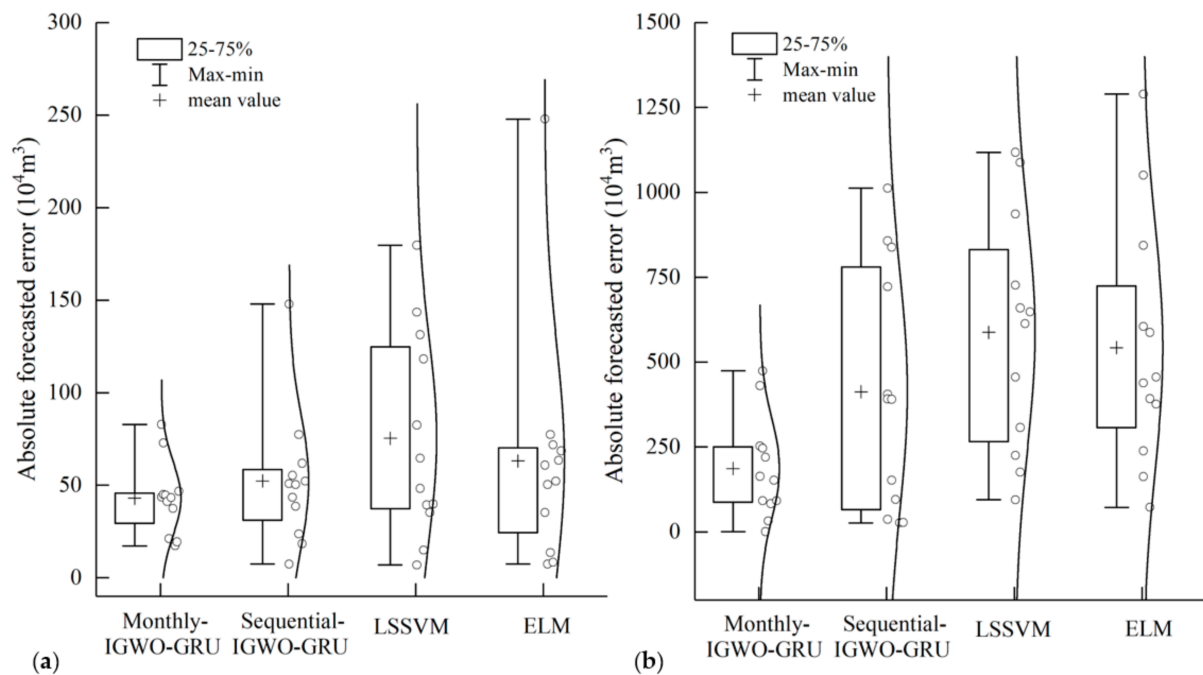

Figure 9. Boxplots of absolute error (AE) of models between forecasted and original streamflow for the Shangjingyou station (**a**) and the Fenhe reservoir station (**b**).

Peak flows, which are the extreme values of a streamflow series, make it difficult to model hydrological behaviors. The original and predicted peak flow during the validation period are presented in Figure 10. One can see that all of the forecasting models can trace the changes in peak flow, but the performance of the monthly IGWO-GRU model was superior to the compared models. For the Fenhe reservoir station, the average absolute relative error of the maximum peak flow of the monthly IGWO-GRU was about 7.34%, which was less than the values of 16.68% for sequential IGWO-GRU, 98.81% for LSSVM, and 69.67% for ELM. Again, for the Shangjingyou station, the estimation of monthly IGWO-GRU (11.74%) was less than the values of 13.41% for sequential IGWO-GRU, 90.91% for LSSVM, and 89.87% for ELM.

The relationship between historical streamflow and the future was learned by the GRU model, which has "memory" of the long time sequence in streamflow series [32]. The "memory" of GRU not only considers the current step but also the past connections between inputs and outputs, leading to the deep learning of the hydrological time-series [13]. The ELM model not only produces a fluctuating performance because of the empirical risk minimization principle but also belongs to the shallow network and cannot remember past cases [39]. In addition, the important hyperparameters in the GRU model are carefully tuned by the IGWO method, which is highly efficient in optimization and effectively reduces the influence of over-fitting or under-fitting [20,40]. Subsequently, the integration of the GRU model with the IGWO method makes the hybrid IGWO-GRU model well-suited for efficient streamflow forecasting and real-time applications. Therefore, the model integrated GRU and IGWO is a promising alternative tool in terms of capturing the peak flows of monthly streamflow at the two stations.
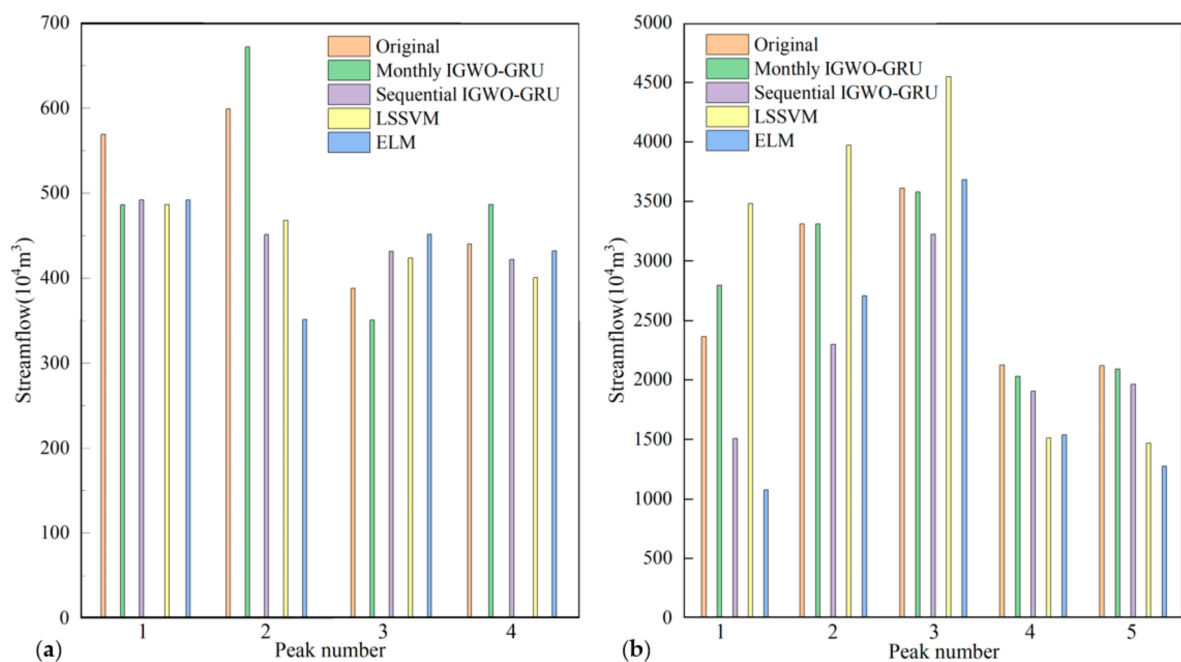
Figure 10. Peak flows of different models for the Shangjingyou station (**a**), and the Fenhe reservoir station (**b**) during the validation period.

From the comparison of streamflow predictions, it can be noted that the deep learning models (IGWO-GRU) based on two types of structure were superior to shallow neural network models (LSSVM and ELM). The deep learning models have shown better learning skills regarding the relationship between historical and future streamflow, which contains complex information about hydrological behavior. Through the detailed analysis of the two types of structure-based models at two stations, the monthly IGWO-GRU model was found to perform better than the sequential IGWO-GRU model. The streamflow in the same month had familiar environmental conditions, which showed the law of mid- to long-term interannual variation. Moreover, a good correlation allowed more information to be generated between the historical and the future streamflow values, thereby capturing more information that can be easily learned by deep learning models [41]. Thus, the monthly structure-based IGWO-GRU model exhibited higher predictive accuracy compared to the sequential structure-based predictive model.

Therefore, according to the results of Figures 6–10 and Tables 2 and 3, the integration of the IGWO with the deep learning GRU based on the monthly structure of streamflow time-series data enabled the new monthly IGWO-GRU model to emulate streamflow values effectively. In the process of streamflow forecasting, the main reasons for this model achieving the best performance are as follows: (1) The deep learning GRU accurately describes the complex behaviors of hydrological series and learns the connotative relationships between inputs and outputs; (2) the optimizer method IGWO finds better parameter combinations of the GRU model, reducing the computational burden and personal error effectively, and improves the generalization ability of the GRU model; (3) the monthly structure-based model reconstructs streamflow series and extracts the implied and useful hydrological information which can be learned easily by the deep learning method.

Despite the enhanced performance of the new monthly IGWO-GRU model, some limitations of the present undertaking could be explored in further studies. This study used only historical streamflow values to predict the future, which utilized a single input variable approach, but the hydrological process is highly complex and related to many climate driving factors, such as rainfall, land utilization, and temperature. Future models that include more factors in forecasting streamflow values will provide more information on the future hydrological process.

5. Conclusions

In this study, a monthly IGWO-GRU model was developed using historical streamflow data as the predictors to forecast the future streamflow at the Fenhe reservoir station and the Shangjingyou station. The hybrid monthly IGWO-GRU and sequential IGWO-GRU achieved a high level of accuracy compared to the single LSSVM and ELM models. The following results have been concluded:

(1) The optimizer IGWO can effectively find the best hyperparameters for the forecasting models. The superior ability of GRU is its deep learning of relationships between historical and future streamflow time-series compare to LSSVM and ELM models.

(2) The monthly structure-based model can effectively extract more information about the hydrological process than the traditional sequential predicting structure. As shown by its high accuracy, the performance of the monthly IGWO-GRU model was better than the sequential structure-based predictive model.

This study provides a newly developed model to learn historical streamflow data and predict future streamflow values. Forecasting streamflow at the mid- to long-term horizons could help the government and stakeholders with strategic future hydrological resource management. Moreover, the monthly structure-based predictive model can be further explored in predicting other variables if the data sets are suitable. More attention will be paid to the performances of multi-variable models in a more comprehensive study.

Author Contributions: Conceptualization, H.L.; Data duration, X.Z. (Xuehua Zhao); Formal analysis, X.Z. (Xuehua Zhao) and H.L.; Investigation, Y.W. and S.L.; Methodology, X.Z. (Xuehua Zhao) and S.L.; Resources, X.Z. (Xuehua Zhao) and H.L.; Supervision, X.Z. (Xuehua Zhao); Writing—original draft, X.Z. (Xuehua Zhao), H.L. and Y.W.; Writing—review and editing, X.Z. (Xuehua Zhao), S.L., and X.Z (Xueping Zhu). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to due the ownership is shared between all parties that contributed to the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahmed, J.A.; Sarma, A.K. Artificial neural network model for synthetic streamflow generation. *Water Resour. Manag.* **2007**, *21*, 1015–1029. [CrossRef]
2. Singh, S.K. Long-term streamflow forecasting based on ensemble streamflow prediction technique: A Case study in New Zealand. *Water Resour. Manag.* **2016**, *30*, 2295–2309. [CrossRef]
3. He, L.; Chen, Y.; Kang, Y.; Tian, P.; Zhao, H. Optimal water resource management for sustainable development of the chemical industrial park under multi-uncertainty and multi-pollutant control. *Environ. Sci. Pollut. Res.* **2018**, *25*, 27245–27259. [CrossRef] [PubMed]
4. Lv, N.; Liang, X.; Chen, C.; Zhou, Y.; Li, J.; Wei, H.; Wang, H. A long short-term memory cyclic model with mutual information for hydrology forecasting. *Adv. Water Resour.* **2020**, *141*, 103622. [CrossRef]
5. Mosavi, A.; Ozturk, P.; Chau, K. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [CrossRef]
6. Wang, Z.; Fatollahzadeh Attar, N.; Khalili, K.; Behmanesh, J.; Band, S.S.; Mosavi, A.; Chau, K. Monthly streamflow prediction using a hybrid stochastic-deterministic approach for parsimonious non-linear time series modeling. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 1351–1372.
7. Tan, Q.; Lei, X.; Wang, X.; Wang, H.; Wen, X.; Ji, Y.; Kang, A. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. *J. Hydrol.* **2018**, *567*, 767–780. [CrossRef]
8. Yaseen, Z.M.; Jaafar, O.; Deo, R.C.; Kisi, O.; Adamowski, J.; Quilty, J.; El-Shafie, A. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. *J. Hydrol.* **2016**, *542*, 603–614. [CrossRef]

9. Shamshirband, S.; Hashemi, S.; Salimi, H.; Samadianfard, S.; Asadi, E.; Shadkani, S.; Kargar, K.; Mosavi, A.; Nabipour, N.; Chau, K. Predicting standardized streamflow index for hydrological drought using machine learning models. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 339–350. [CrossRef]

10. Zhao, X.; Chen, X.; Xu, Y.; Xi, D.; Zhang, Y.; Zheng, X. An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* **2017**, *9*, 153. [CrossRef]

11. Yaseen, Z.M.; Sulaiman, S.O.; Deo, R.C.; Chau, K. 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.* **2019**, *569*, 387–408. [CrossRef]

12. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [CrossRef]

13. Zhang, D.; Lindholm, G.; Ratnaweera, H. Use long short-term memory to enhance internet of things for combined sewer overflow monitoring. *J. Hydrol.* **2018**, *556*, 409–418. [CrossRef]

14. Zhang, D.; Peng, Q.; Lin, J.; Wang, D.; Liu, X.; Zhuang, J. Simulating reservoir operation using a recurrent neural network algorithm. *Water* **2019**, *11*, 865. [CrossRef]

15. Wang, Y.; Liao, W.; Chang, Y. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies* **2018**, *11*, 2163. [CrossRef]

16. Yu, C.; Li, Y.; Bao, Y.; Tang, H.; Zhai, G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. *Energy Convers. Manag.* **2018**, *178*, 137–145. [CrossRef]

17. Ni, L.; Wang, D.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J.; Liu, J. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *J. Hydrol.* **2020**, *586*, 124901. [CrossRef]

18. Hong, M.; Wang, D.; Wang, Y.; Zeng, X.; Ge, S.; Yan, H.; Singh, V.P. Mid- and long-term runoff predictions by an improved phase-space reconstruction model. *Environ. Res.* **2016**, *148*, 560–573. [CrossRef]

19. Samadianfard, S.; Salar, J.; Ely, S.; Amir, M.; Akib, S.S.S. Support vector regression integrated with fruit fly optimization algorithm for river flow forecasting in lake urmia basin. *Water* **2019**, *11*, 1034. [CrossRef]

20. Tikhamarine, Y.; Souag-Gamane, D.; Najah Ahmed, A.; Kisi, O.; El-Shafie, A. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J. Hydrol.* **2020**, *582*, 124435. [CrossRef]

21. Ch, S.; Anand, N.; Panigrahi, B.K.; Mathur, S. Streamflow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputing* **2013**, *101*, 18–23. [CrossRef]

22. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]

23. Faris, H.; Aljarah, I.; Al-Betar, M.A.; Mirjalili, S. Grey wolf optimizer: A review of recent variants and applications. *Neural Comput. Appl.* **2018**, *30*, 413–435. [CrossRef]

24. Mirjalili, S. How effective is the Grey Wolf optimizer in training multi-layer perceptrons. *Appl. Intell.* **2015**, *43*, 150–161. [CrossRef]

25. Eswaramoorthy, S.; Sivakumaran, N.; Sekaran, S. Grey wolf optimization based parameter selection for support vector machines. *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.* **2016**, *35*, 1513–1523. [CrossRef]

26. Sulaiman, M.H.; Mustaffa, Z.; Mohamed, M.R.; Aliman, O. Using the gray wolf optimizer for solving optimal reactive power dispatch problem. *Appl. Soft Comput.* **2015**, *32*, 286–292. [CrossRef]

27. Zhang, S.; Zhou, Y.; Li, Z.; Pan, W. Grey wolf optimizer for unmanned combat aerial vehicle path planning. *Adv. Eng. Softw.* **2016**, *99*, 121–136. [CrossRef]

28. Wu, C.L.; Chau, K.W.; Li, Y.S. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour. Res.* **2009**, *45*, W08432. [CrossRef]

29. Misra, D.; Oommen, T.; Agarwal, A.; Mishra, S.K.; Thompson, A.M. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosyst. Eng.* **2009**, *103*, 527–535. [CrossRef]

30. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]

31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

32. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734.

33. Ali, M.; Prasad, R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renew. Sustain. Energy Rev.* **2019**, *104*, 281–295. [CrossRef]

34. Ministry of Water Resources. *GB/T 22482-2008, Standard for Hydrological Information and Hydrological Forecasting*; China Water and Power Press: Beijing, China, 2008.

35. Yang, Y.; Meng, Z.; Jiao, W. Hydrological and pollution processes in mining area of Fenhe River Basin in China. *Environ. Pollut.* **2018**, *234*, 743–750. [CrossRef] [PubMed]

36. Fang, W.; Huang, S.; Ren, K.; Huang, Q.; Huang, G.; Cheng, G.; Li, K. Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *J. Hydrol.* **2019**, *568*, 534–550. [CrossRef]

37. Yu, C.; Qi, X.; Ma, H.; He, X.; Wang, C.; Zhao, Y. LLR: Learning learning rates by LSTM for training neural networks. *Neurocomputing* **2020**, *394*, 41–50. [CrossRef]

38. Guo, Z.; Zhao, W.; Lu, H.; Wang, J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [CrossRef]

39. Sun, N.; Zhou, J.; Chen, L.; Jia, B.; Tayyab, M.; Peng, T. An adaptive dynamic short-term wind speed forecasting model using secondary decomposition and an improved regularized extreme learning machine. *Energy* **2018**, *165*, 939–957. [CrossRef]
40. Nourani, V.; Hosseini Baghanam, A.; Adamowski, J.; Kisi, O. Applications of hybrid wavelet—Artificial Intelligence models in hydrology: A review. *J. Hydrol.* **2014**, *514*, 358–377. [CrossRef]
41. Soltani, S.; Modarres, R.; Eslamian, S.S. The use of time series modeling for the determination of rainfall climates of Iran. *Int. J. Climatol.* **2007**, *27*, 819–829. [CrossRef]