Multi-step ahead daily inflow forecasting using ERA-Interim reanalysis dataset based on gradient boosting regression trees

Shengli Liao[1,2]**, Zhanwei Liu**[1,2]**, Benxi Liu**[1,2]**, Chuntian Cheng**[1,2]**, Xinfeng Jin**[1,2]**, and Zhipeng Zhao**[1,2]

[1]Institute of Hydropower System and Hydroinformatics, Dalian University of Technology, Dalian 116024, China
[2]Key Laboratory of Ocean Energy Utilization and Energy Conservation of Ministry of Education, Dalian University of Technology, Dalian 116024, China

Correspondence: Zhanwei Liu (337891617@qq.com)

Abstract. Inflow forecasting plays an essential role in reservoir management and operation. The impacts of climate change and human activities make accurate inflow prediction increasingly difficult, especially for longer lead times. In this study, a new hybrid inflow forecast framework with ERA-Interim reanalysis dataset as input, adopting gradient boosting regression trees (GBRT) and the maximal information coefficient (MIC) is developed for multi-step ahead daily inflow forecasting. Firstly, the ERA-Interim reanalysis dataset provides more information for the framework to discover inflow for longer lead times. Secondly, MIC can identify effective feature subset from massive features that significantly affects inflow so that the framework can reduce computational burden, distinguish key attributes with unimportant ones and provide a concise understanding of inflow. Lastly, the GBRT is a prediction model in the form of an ensemble of decision trees and has a strong ability to capture nonlinear relationships between input and output at longer lead times more fully. The Xiaowan hydropower station located in Yunnan Province, China is selected as the study area. Six evaluation criteria, the mean absolute error (MAE), the root mean squared error (RMSE), the Pearson correlation coefficient (CORR), Kling–Gupta efficiency scores (KGE), the percent bias in flow duration curve high-segment volume (BHV) and the Index of Agreement (IA) are used to evaluate the established models using historical daily inflow data (1/1/2017-31/12/2018). Performance of the presented framework is compared to that of artificial neural networks (ANN), support vector regression (SVR) and multiple linear regression (MLR) models. The results indicate that the reanalysis data enhances the accuracy of inflow forecasting for all lead times studied (1-10 days) and the developed method generally performs better than other models, especially for the extreme values and longer lead times (4-10 days).

1 Introduction

Reliable and accurate inflow forecasting 1-10 days in advance is significant for efficient utilization of water resources, reservoir operation and flood control, especially in areas with concentrated rainfall. Rainfall in southern China is usually concentrated for several days at a time due to strong convective weather, for example, typhoons. Low accuracy inflow predictions can easily cause the failure of power stations to make reasonable power generation plans 7-10 days ahead of disaster events and lead to unnecessary water abandonment and even substantial economic losses. Fig. 1 shows the losses of electric quantity due to discarded water (LEQDW) in Yunnan and Sichuan Provinces, China from 2011 to 2016 (Sohu, 2017; in en.com, 2018). The total amount of LEQDW in Yunnan and Sichuan Provinces increased from 1.5 billion kWh to 45.6 billion kWh from 2011 to 2016, with an average annual growth rate of 98.0%. In recent years, due to the increased number of hydropower stations and installed hydropower capacities, the problem of discarding water caused by inaccurate inflow forecasting is becoming increasingly serious, which has also produced a negative impact on the development of hydropower in China. The main challenge in inflow forecasting caused by climate change and human activities at present are low accuracy, especially for longer lead times (Badrzadeh et al., 2013; El-Shafie et al., 2007). Meanwhile, due to streamflow variation by reason of climate change and human activities, inflow forecasting model often needs to be rebuilt and the model parameters need to

be recalibrated according to the actual inflow and meteorological data within one or two years. To address these problems, a variety of models and approaches have been developed. These approaches can be divided into three categories: statistical methods (Valipour et al., 2013), physical methods (Duan et al., 1992; Wang et al., 2011; Robertson et al., 2013), and machine learning methods (Chau et al., 2005; Liu et al., 2015; Rajaee et al., 2019; Zhang et al., 2018; Yaseen et al., 2019; Fotovatikhah et al., 2018; Mosavi et al., 2018; Chau, 2017; Ghorbani et al., 2018). Each method has its own conditions and scope of application. Statistical methods are usually based on historical inflow records and mainly include the autoregressive model, the autoregressive moving average (ARMA) model and the autoregressive integrated moving average (ARIMA) model (Lin et al., 2006). Statistical methods assume that the inflow series is stationary and the relationship between input and output is simple. However, real inflow series is complex, nonlinear and chaotic (Dhanya and Kumar, 2011), making it difficult to obtain high-accuracy predictions using statistical models. Physical methods which have clear mechanisms are implemented using theories of inflow generation and confluence. These methods can reflect the characteristics of catchment but are very strict with initial conditions and input data (Bennett et al., 2016). Meanwhile, these methods used for flood forecasting have a shorter lead time and cannot be used to acquire long-term forecasting results due to input uncertainty. Machine learning methods, having a strong ability to handle the nonlinear relationship between input and output and recently shown excellent performance in inflow prediction, are widely used for medium and long-term inflow forecasts. In particular, several studies had shown that artificial neural networks (ANN) (Rasouli et al., 2012; Cheng et al., 2015; El-Shafie and Noureldin, 2011) and support vector regression (SVR) (Tongal and Booij, 2018; Luo et al., 2019; Moazenzadeh et al., 2018) are the two powerful models for inflow predicting. However, these models still have some inherent disadvantages. For example, ANN is prone to being trapped by local minima and both ANN and SVR suffer from over-fitting problems and reduced generalizing performance. In recent years, gradient boosting regression trees (GBRT) (Fienen et al., 2018; Friedman, 2001), a nonparametric machine learning method based on a boosting strategy and decision trees, was developed and had been used in traffic (Zhan et al., 2019) and environmental (Wei et al., 2019) field and proved to alleviate these problems mentioned above. Thus, GBRT is selected for daily inflow prediction with a lead times of 1-10 days in this paper. Compared with ANN and SVR, GBRT also has two other advantages. Firstly, GBRT can rank features according to their contribution to model scores, which is of great significance for reducing the complexity of the model. Secondly, GBRT is a white box model and can be easily interpreted. To the best of our knowledge, GBRT has not been used for daily inflow prediction with a lead times of 1-10 days before. For comparison purposes, ANN, SVR and multiple linear regression

(MLR) have been employed to forecast daily inflow and are considered as bench mark models in this study. In addition to forecasting models, a vital reason why many approaches cannot attain higher accuracy for inflow predictions is that inflow is influenced by various factors (Yang et al., 2019), such as rainfall, temperature, humidity, etc. Thus, it is very difficult to select appropriate features for inflow forecasting. Current feature selection methods for inflow forecasting mainly include two methodologies. The first method is the model-free method (Bowden, 2005; Snieder et al., 2020) which employs a measure of the correlation coefficient criterion (Badrzadeh et al., 2013;Siqueira et al., 2018; Pal et al., 2013) to characterise the correlation between a potential model input and the output variable. The second method is the model-based method (Snieder et al., 2020) which usually utilizes the model and search strategies to determine optimal input subset. Common search strategies include forward selection, backward elimination et al (May et al., 2011). The correlation coefficient has limited ability for capturing nonlinear relationships and exhaustive search tend to need the higher computation burden. In order to select effective inputs accurately and quickly, the maximal information coefficient (MIC) (Reshef et al., 2011) is used to select input factors for inflow forecasting. MIC is a robust measure of the degree of correlation between two variables and has attracted a lot attention from academia (Zhao et al., 2013; Ge et al., 2016; Lyu et al., 2017; Sun et al., 2018). In addition, sufficient potential input factors are the prerequisite for obtaining reliable and accurate prediction results and it is not enough to use only antecedent inflow series as the input of the model. To enhance the accuracy of inflow forecasting and acquire a longer lead time, increasing amounts of meteorological forecasting data have been used for inflow forecasting (Lima et al., 2017; Fan et al., 2015; Rasouli et al., 2012). However, with extended lead times, the errors of forecast data continuously increase because the variables obtained by numerical weather prediction (NWP) system are also affected by complex factors (Mehr et al., 2019). Moreover, with the continuous improvement of forecasting systems, it is difficult to obtain consistent and long series of forecasting data (Verkade et al., 2013). To mitigate these problems, the reanalysis data generated by ERA-Interim (European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis Interim) (Dee et al., 2011), which was proved to be one of the best methods for reanalysis of data describing atmospheric circulation and elements (Kishore et al., 2011), has been used as an input. The reanalysis data which has less error than observed data and forecast data is the result of assimilating observed data with forecast data. ERA-Interim shows the results of a global climate reanalysis from 1979 to date, which are produced by a fixed version of NWP system. The fixed version ensures that there are no spurious trends caused by an evolving NWP system. Therefore, meteorological reanalysis data satisfies the need for long sequences of consistent data and has been used for the prediction of wind speeds (Stopa
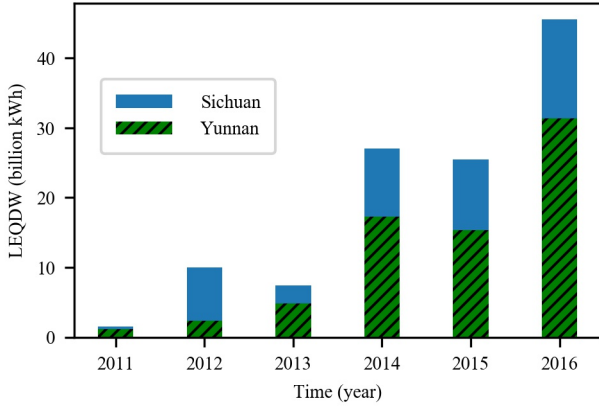
Figure 1. Losses of electric quantity due to discarded water (LEQDW) in the Sichuan and Yunnan province.

and Cheung, 2014) and solar radiation (Ghimire et al., 2019; Linares-Rodríguez et al., 2011). This study aims to provide a reliable inflow forecasting framework with longer lead times for daily inflow forecasting. The framework adopts the ERA-Interim reanalysis dataset as the input which ensured ample information is supplied to depict inflow. MIC is used to select appropriate features to avoid over-fitting and waste of computing resources caused by feature redundancy. GBRT, which is robust to outliers and has strong non-linear fitting ability, is used as the prediction model to improve inflow forecasting accuracy of longer lead times. This paper is organized as follows: Section 2 describes a case study and collected data. Section 3 introduces the theory and process of methods used, including MIC and GBRT. Section 4 shows the results and discussion of the data, followed by the conclusions in Section 5.

2 Data

2.1 Study area and collected data

The Xiaowan Hydropower Station in the lower reaches of the Lancang River is chosen as the study site (Fig. 3). The Xiaowan Hydropower Station is the main controlling hydropower station in the Lancang River and it is very meaningful to adopt the Xiaowan Hydropower Station as the case study. The Lancang River is approximately 2000 km long and has a drainage area of 113300 km^2 above the Xiaowan Hydropower Station. The Lancang River which is also known as the Mekong River originates in the Tibetan Plateau and runs through China, Myanmar, Laos, Thailand, Cambodia, and Vietnam. The major source of water flowing into the Lancang River in China comes from melting snow on the Tibetan plateau (Commission, 2005). We collected ERA-Interim reanalysis dataset, observed daily inflow and rainfall data for Xiaowan for 8 years (Jan-

uary 2011 to December 2018). Fig. 3 depicts the daily inflow series. The data from January 2011 to December 2014 (1461 days, approximately 50% of the whole dataset), from January 2015 to December 2016 (731 days, approximately 25% of the whole dataset) and from January 2017 to December 2018 (730 days, approximately 25% of the whole dataset) are used as training, validation and testing set, respectively. The reanalysis dataset can be downloaded from https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ and is provided every 12 hours on a spatial grid size of 0.25° × 0.25°. Based on the expert knowledge and on the basis of available literature, the near-surface 26 variables (Table A1) from the reanalysis data are considered as potential selected predictors for inflow forecasting, which include the total precipitation (tp), the 2 meter temperature (t2m), the total column water (tcw), etc. More details about ERA-Interim dataset are presented in the Appendix A.

2.2 Feature scaling and feature selection

Feature scaling is necessary for machine learning methods and all features are scaled to the range between 0 and 1 before taking part in the calculation, as follows:

$$x_{scale} = \frac{x_{original} - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_{scale} and $x_{original}$ indicate the scaled and original data, respectively. x_{max} and x_{min} represent the maximum and minimum of inflow series, respectively. Reasonable selection of input variables can reduce the computational burden and improve the prediction accuracy of the model by removing redundant feature information and reducing the dimensions of the features. If too many features are selected, model will become very complex, which will cause trouble when adjusting parameters, resulting in over-fitting and difficult convergence. Moreover, natural patterns in the data will be blurred by noise (Zhao et al., 2013). On the other hand, if irrelevant features are chosen, there will add noise into the model and also hinder the learning process. MIC is employed to select inputs from candidate predictors from reanalysis data. The lagged inflow and rainfall series are identified by partial autocorrelation function (PACF) and cross-correlation function (CCF). The corresponding 95% confidence interval is used to identify significant correlations. Furthermore, when correlation coefficient slowly declines and cannot fall into confidence interval, a trial and error procedure is used to determine the optimum lag, i.e., starting from one-lag and then modifying the external inputs by successively adding one more lagged time series into inputs (Amiri, 2015; Shoaib et al., 2015).

Table 1. The candidate inputs via PACF and CCF.

No.	Input
1	Q_{t-1}, Q_{t-4}
2	$Q_{t-1}, Q_{t-4}, R_{t-1}$
3	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}$
4	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}$
5	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}$
6	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}$
7	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}$
8	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}$
9	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}$
10	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, R_{t-9}$
11	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, R_{t-9}, R_{t-10}$
12	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, R_{t-9}, R_{t-10}, R_{t-11}$
13	$Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}, R_{t-7}, R_{t-8}, R_{t-9}, R_{t-10}, R_{t-11}, R_{t-12}$

**Table 2.** The candidate inputs from reanalysis data via MIC

No.	Input
1	$obs, stl3_{t-1}$
2	$obs, stl3_{t-1}, d2m_{t-1}$
3	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}$
4	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}$
5	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}$
6	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}$
7	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}$
8	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}$
9	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}$
10	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}, ro_{t-1}$
11	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}, ro_{t-1}, swvl1_{t-1}$
12	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}, ro_{t-1}, swvl1_{t-1}, swvl2_{t-1}$
13	$obs, stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}, ro_{t-1}, swvl1_{t-1}, swvl2_{t-1}, swvl3_{t-1}$
14	$obs, rea, t2m_{t-1}$
15	$obs, rea, t2m_{t-1}, swvl4_{t-1}$
16	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}$
17	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}$
18	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}$
19	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}$
20	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}$
21	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}$
22	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}, sd_{t-1}$
23	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}, sd_{t-1}, smlt_{t-1}$
24	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}, sd_{t-1}, smlt_{t-1}, istl1_{t-1}$
25	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}, sd_{t-1}, smlt_{t-1}, istl1_{t-1}, istl3_{t-1}$
26	$obs, rea, t2m_{t-1}, swvl4_{t-1}, mx2t_{t-1}, sf_{t-1}, cp_{t-1}, tp_{t-1}, rsn_{t-1}, lsp_{t-1}, sd_{t-1}, smlt_{t-1}, istl1_{t-1}, istl3_{t-1}, istl2_{t-1}$

None: obs represents the selected observed optimal input set, $obs = \{Q_{t-1}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}\}$. rea represents the selected a part of input set from reanalysis, $rea = \{stl3_{t-1}, d2m_{t-1}, tcwv_{t-1}, tcw_{t-1}, stl2_{t-1}, mn2t_{t-1}, tsn_{t-1}, stl4_{t-1}, stl1_{t-1}, ro_{t-1}, swvl1_{t-1}, swvl2_{t-1}, swvl3_{t-1}\}$.
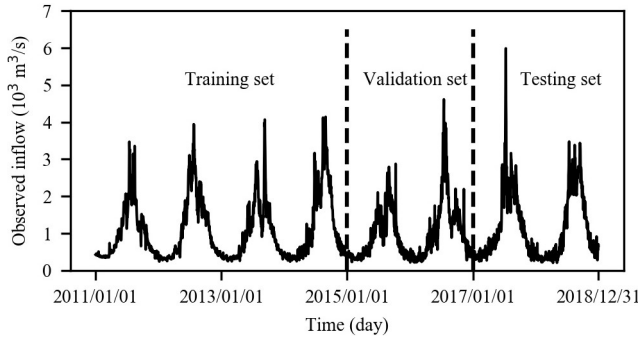
3   Methodology

3.1   Feature selection via maximal information coefficient

The calculation of MIC is based on concepts of the mutual information (MI) (Kinney and Atwal, 2014). For a random   5
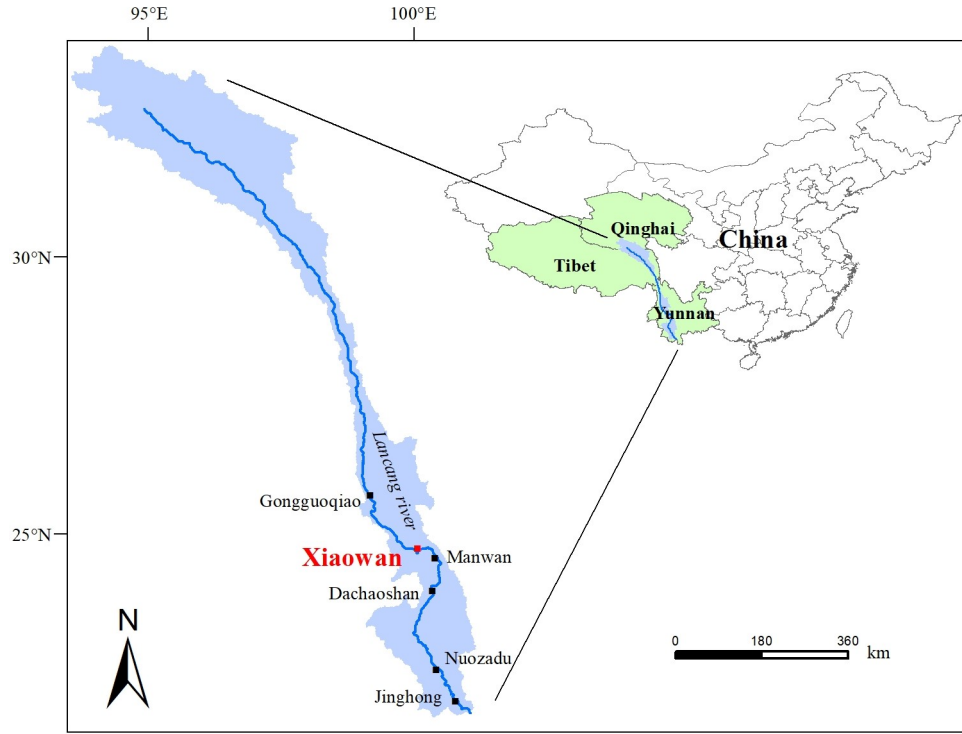
Figure 2. Location of the Xiaowan hydropower station.

Figure 3. Daily inflow series of the Xiaowan hydropower station.

variable X, such as observed inflow, the entropy of X is defined as

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{2}$$

where $p(x)$ is the probability density function of $X = x$. Furthermore, for another random variable Y, such as observed rainfall, the conditional entropy of X given Y may be evaluated from the following expression

$$H(X|Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y) \tag{3}$$

where $H(X|Y)$ is the uncertainty of X given knowledge, $p(x,y)$ and $p(x|y)$ are the joint probability density and the

conditional probability of $X = x$ and $Y = y$, respectively. The reduction of the original uncertainty of X, due to the knowledge of Y, is called the MI (Amorocho and Espildora, 1973; Chapman, 1986), defined by

$$MI(X,Y) = H(X) - H(X|Y)$$
$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{4}$$

The calculation of MIC is divided into three steps. Consider given a dataset D, including variable X and Y with a sample size n. Firstly, drawing scatter plots of X and Y and drawing grids for partitioning which is called an x-by-y grid. Let $D|G$ denote the distribution of D divided by one of x-by-y grids as G. $MI^*(D,x,y) = \max MI(D|G)$, where is the mutual information of $D|G$. Secondly, characteristic matrix is defined as

$$M(D)_{x,y} = \frac{MI^*(D,x,y)}{\log(\min(x,y))} \tag{5}$$

Lastly, MIC is introduced as the maximum value of characteristic matrix, that is, $MIC(D) = \max_{xy < B(n)} M(D)_{x,y}$, where $B(n)$ is the upper bound of the grid size which is a function of sample size, defined $B = n^{0.6}$. We perform feature selection from ERA-Interim reanalysis dataset in two steps via MIC. First, compute MIC of each reanalysis variables and observed inflow. Then, sort features based on MIC in a descending order and determine the optimum inputs by using a

trail-and-error procedure, i.e. starting from the top one feature and then modifying the external inputs by successively adding one more feature into model inputs. The selected k features from reanalysis data are used as part of input to the model.

3.2 Gradient boosting regression trees

GBRT is an ensemble model which mainly includes two algorithms: decision tree algorithm and the boosting algorithm. The decision tree robust to outliers is used as a primitive model and boosting algorithm as integration rule is used to improve inflow forecasting accuracy.

3.2.1 The decision tree

The decision tree in this paper refers to decision tree learning used in computer science, which is one of the predictive modelling approaches used in machine learning. A decision tree consists of branch nodes (the tree structure) and leaf nodes (the tree output). Supposing a training dataset is given in a feature space with N features and each feature with n samples, $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)(X_i = (x_1, x_2, \ldots, x_N), i = 1, 2, \ldots, n)$. In the input space where the training set is located, each region is recursively divided into two subregions and the output value of each subregion is used to construct a binary decision tree. The top-down cyclic branch learning of the decision tree adopts a greedy algorithm where each branch node only cares about its own objective function. By traversing all features and all segmentation points of each feature, the best feature j and segmentation points s can be found by minimizing squared loss:

$$\min_{j,s} \left[\min_{c_1} \sum_{X_i \in R_1(j,s)} (y_i - c_1)^2 \right.$$
$$\left. + \min_{c_2} \sum_{X_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{6}$$

where

$$R_1(j,s) = \left\{ X_i | x_i^{(j)} \leq s, i = 1, 2, \cdots, N \right\} \tag{7}$$

$$R_2(j,s) = \left\{ X_i | x_i^{(j)} > s, i = 1, 2, \cdots, N \right\} \tag{8}$$

$$c_m = \frac{1}{N_m} \sum_{X_i \in R_m(j,s)} y_i \quad (m = 1, 2) \tag{9}$$

y_i is the observed value and $R_1(j,s)$ and $R_2(j,s)$ are the results of partitioning. c_1 and c_2 are output values of $R_1(j,s)$ and $R_2(j,s)$, respectively. Fig. 4 shows an example of a decision tree model with a max depth and number of leaf nodes of 3 and 5, respectively. If the threshold of loss is set as the stopping condition of the decision tree, it will easily lead to over-fitting problems. Hence, we set the following parameters to alleviate the over-fitting problem of the decision tree
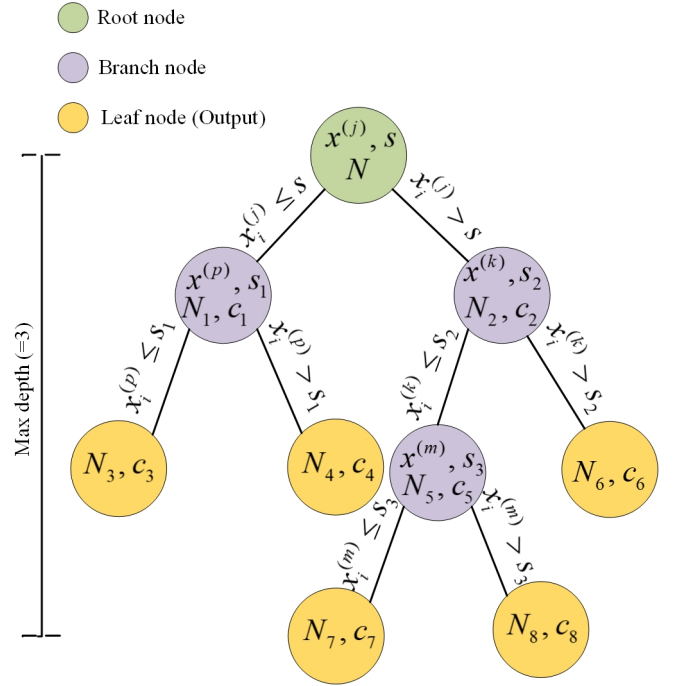

Figure 4. The structure of decision tree model.

model: the maximum depth of the tree, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node and the number of leaf nodes. These parameters are also the ones used for optimization when using the decision tree.

3.2.2 The boosting algorithm

The idea of gradient boosting originated in the observation by Breiman (Breiman, 1997) and can be interpreted as an optimization algorithm based on a suitable cost function. Explicit regression gradient boosting algorithms are subsequently developed (Friedman, 2001; Mason et al., 1999). The boosting algorithm used is described here. Supposing a training dataset with n sample $(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)$, a squared loss function is used to train the decision tree:

$$L(y, f(X)) = \sum_{i=1}^{n} (y - f(X_i))^2 \tag{10}$$

The core of the GBRT algorithm is the iterative process of training the decision with a residual method. The iterative training process of GBRT with M decision trees is as follows:

1) Initialization $f_0(x) = \arg\min_c \sum_{i=1}^{n} L(y_i, c)$.

2) For m-th $(m = 1, 2, \ldots, M)$ decision trees:

a) Operating i-th $(i = 1, 2, \ldots, n)$ sample points. Using the negative gradient of the loss function to replace the residual in the current model $r_{mi} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x) = f_{m-1}(x)}$.

b) Fitting a regression tree with $\{(x_i, r_{mi})\}$. The i-th regression tree with $(t = 1, 2, \ldots, T)$ as its corresponding leaf node region is obtained, where t is the number of leaf nodes of regression.

c) For each leaf region $t = 1, 2, \ldots, T$, and the best fitting value is calculated by $c_{mt} = \arg\min_{c} \sum_{x_i \in R_{mt}} L(y_i, f_{m-1}(x_i) + c)$.

d) The fitting results are updated by adding the obtained fitting values to the previous ones using

$$f_{mt}(x_i) = f_{m-1}(x_i) + \sum_{t=1}^{T} c_{mt} I \quad (x_i \in R_{mt}) .$$

3) Finally, a strong learning method is obtained

$$\hat{f}(x_i) = f_M(x_i) = \sum_{m=1}^{M} \sum_{t=1}^{T} c_{mt} I \quad (x_i \in R_{mt}) .$$

According to the above introduction to GBRT, the parameters of the GBRT can be divided into two categories: boosting parameters and tree parameters. The boosting parameters include the learning rate and the number of weak learners (learning_rate and n_estimators). The learning rate setting is used for reducing the gradient step. The learning rate influences the overall time of training, and the smaller the value is, the more iterations are required for training. There are four tree parameters: max_leaf_nodes, min_samples_leaf, min_samples_split and max_depth. Hence, GBRT has six parameters control model complexity (Fienen et al., 2018), which we adjusted for tuning by using a trial-and-error procedure.

3.3 Evaluation criteria of the models

It is critical to carefully define the meaning of performance and to evaluate the performance on the basis of the forecasting and fitted values of the model compared with historical data. The root mean squared error (RMSE) and mean absolute error (MAE) are the most commonly used criteria to assess model performance and are calculated using Eq. (11) and Eq. (12), respectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{Q}_i - Q_i)^2} \tag{11}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Q}_i - Q_i \right| \tag{12}$$

where $\hat{Q}_i$ and Q_i are the inflow estimation and observed value at time i, respectively and n is the number of samples. The RMSE is more sensitive to extremes in sample sets and thus it is used to evaluate the model's ability to simulate flood peaks. The Pearson correlation coefficient (CORR) is a measure of the strength of the association between observed inflow series and forecasted inflow series; it is calculated according to Eq. (13).

$$CORR = \frac{\sum_{i=1}^{n} (Q_i - \bar{Q})\left(\hat{Q}_i - \overline{\hat{Q}}\right)}{\sqrt{\sum_{i=1}^{n} (Q_i - \bar{Q})^2} \sqrt{\sum_{i=1}^{n} \left(\hat{Q}_i - \overline{\hat{Q}}\right)^2}} \tag{13}$$

where $\overline{\hat{Q}}$ is the mean of the estimation series. The range of the CORR is between 0 and 1 and values close to 1 demonstrate a perfect estimation result. Kling–Gupta efficiency scores (KGE) (Knoben et al., 2019) is also a widely used evaluation index. It can be provided as following Eq. (14) and (15).

$$KGE = 1$$
$$- \sqrt{(CORR - 1)^2 + \left(\frac{\hat{\sigma}}{\sigma} - 1\right)^2 + \left(\frac{\hat{Q}}{\bar{Q}} - 1\right)^2} \tag{14}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{Q}_i - \overline{\hat{Q}}\right)^2}, \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Q_i - \bar{Q})^2} \tag{15}$$

where σ is the standard deviation of the observed values, $\hat{\sigma}$ is the standard deviation of the inflow estimation, μ is the mean of the observed series and $\hat{\mu}$ is the mean of the inflow estimation series. The percent bias in flow duration curve high-segment volume (BHV) (Yilmaz et al., 2008; Vogel and Fennessey, 1994) is presented to estimate prediction performance of extreme value for model. It can be provided as following Eq. (16).

$$BHV = \frac{\sum_{k=1}^{H} \left(\hat{Q}_h - Q_h\right)}{\sum_{h=1}^{H} Q_h} \times 100 \tag{16}$$

where $h = 1, 2, \ldots, H$ are the inflow indices for inflows with exceedance probabilities lower than 0.02. In this paper, the inflow threshold of exceedance probabilities equalling 0.02 is $1722\ \mathrm{m}^3/\mathrm{s}$. The Index of Agreement (IA) (Willmott, 1981) plays a significant role in evaluating the degree of the agreement between observed series and inflow estimation series. Similar to CORR, its range is between 0 (no agreement at all) and 1 (perfect fit). It is given by:

$$IA = 1 - \frac{\sum_{i=1}^{n} (\hat{Q}_i - Q_i)^2}{\sum_{i=1}^{n} (\left|\hat{Q}_i - \bar{Q}\right| + \left|Q_i - \bar{Q}\right|)^2} \tag{17}$$

3.4 Overview of framework

Fig. 5 illustrates the overall structure of framework presented. This structure consists of two major models: GBRT

and GBRT-MIC. In GBRT, we measure the relevance of different lags observed inflow and rainfall with observed inflow at the time of forecast via partial autocorrelation function (PACF) and cross-correlation function (CCF) (Badrzadeh et al., 2013) and select appropriate lags as predictors of model by hypothesis test and trial-and-error procedures. Then, data pre-processing and feature scaling are carried out for selected predictors. Next, dividing the dataset into training set, validation set, and testing set according to the length of each data set specified in advance (in Section 2.2). A grid search algorithm, which is an exhaustive search all candidate parameter combination method, is guided to optimization model parameters by evaluation of validation set for each lead time (Chicco and Davide, 2017). Lastly, prediction results are evaluated based on testing set. Compared with GBRT, GBRT-MIC adds reanalysis data which are selected via MIC (in Section 3.1) as the input of the model. Moreover, GBRT-MIC also calculates the importance of features according to the prediction results and ranks the features (Louppe, 2014). It is difficult to perform multi-step forecasting by the reason of accumulation of errors, reduced accuracy, and increased uncertainty. The current state of multi-step ahead forecasting is reviewed, there are mainly two strategies that you can use for multi-step forecasting for single-output, namely, Static (Direct) multi-step forecast and Recursive multi-step forecast (Bontempi et al., 2012; Taieb et al., 2012). Recursive forecast strategy is biased when the underlying model is nonlinear and is sensitive to the estimation error, since estimated values, instead of actual ones, are more and more used when we get further in the future (Bontempi et al., 2012). Thus, the Static multi-step forecasting strategy is employed in this paper. Since the Static strategy does not use any approximated values to compute the forecasts, it is not prone to any accumulation of errors. The model structures of GBRT and GBRT-MIC are as follows:

$$
\begin{aligned}
\hat{Q}^I_{t+T} = & f(\boldsymbol{\theta}^I_t; Q_t, Q_{t-1}, \dots, Q_{t+1-p}, \\
& R_t, R_{t-1}, \dots, R_{t+1-q}) \quad (T=1,2,\dots,10)
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
\hat{Q}^{II}_{t+T} = & f(\boldsymbol{\theta}^{II}_t; Q_t, Q_{t-1}, \dots, Q_{t+1-p}, \\
& R_t, R_{t-1}, \dots, R_{t+1-q}, E^1_t, E^2_t, \dots, E^k_t) \\
& (T=1,2,\dots,10)
\end{aligned} \tag{19}
$$

where $\hat{Q}^I_{t+T}$ and $\hat{Q}^{II}_{t+T}$ are the forecasted value of GBRT and GBRT-MIC at the lead time T of current time t, respectively. $\boldsymbol{\theta}^I_t$ and $\boldsymbol{\theta}^{II}_t$ are parameters of GBRT and GBRT-MIC at the lead time T of current time t, respectively. p and q are lags of observed inflow and rainfall determined via PACF and CCF, respectively. E_t is the features from reanalysis data at the current time t and k is the number of features from reanalysis data determined via MIC.
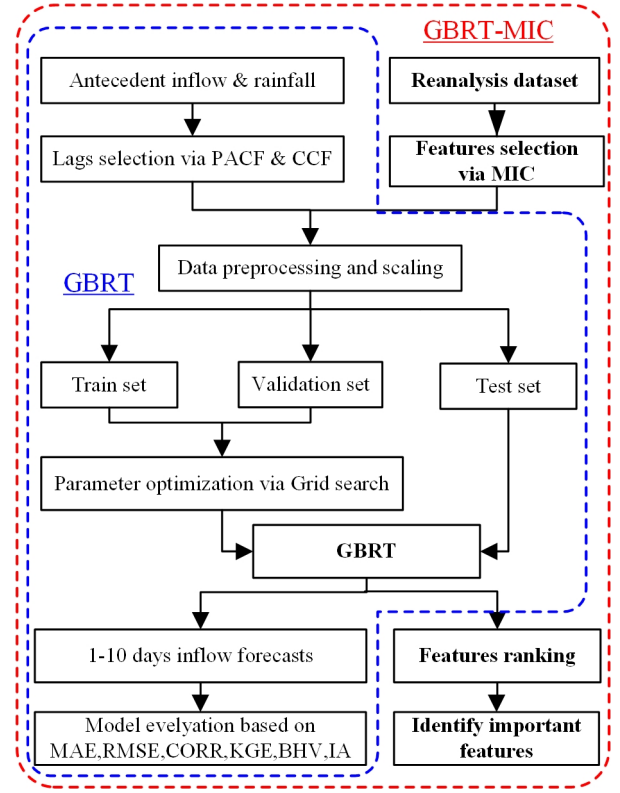

Figure 5. Overview of the framework.

4 Experimental results and discussion

In order to compare with GBRT-MIC, the ANN-MIC, SVR-MIC and MLR-MIC, obtained by replacing GBRT in the framework with ANN, SVR and MLR, respectively, are also employed for inflow forecasting with lead times of 1-10 days. As mentioned previously, six indices, i.e. the MAE, RMSE, CORR, KGE, BHV and IA, are calculated to evaluate the performance of models based on the testing set. We also explored the feature importance based on the GBRT-MIC model (Louppe, 2014). All computations of this paper are performed on a ThinkPad P1 workstation containing an Intel Core i7-9850H CPU with 2.60 GHz and 16.0 GB of RAM, using the version 3.7.10 of Python (Foundation, 2020), which is powerful, fast and open, and scikit-learn package (Pedregosa et al., 2011).

4.1 Feature selection

Fig. 6 shows the PACF, CCF and the corresponding 95% confidence interval from lag 1 to lag 12. The PACF shows significant autocorrelation at lag one and lag four, respectively (Fig. 6(a)), and thus, inflow series one and four-day lag are selected as the inputs of model. CCF between inflow and rainfall gradually decreases as increasing the time lag (Fig. 6(b)) and cannot fall into 95% confidence interval. Therefore, a trial-and-error procedure is used to determine

optimal selection of lagged rainfall series. 13 input structures are tried (Table 1) and the trial results are shown in Fig. A1. The results indicate that 7th input structure obtains best performance. Accordingly, rainfall series from one to six-day lag are selected as the inputs of model. As mentioned previously, based to MIC between inflow and the reanalysis variable (Table A1), a trial-and-error procedure is used to determine optimal input subset. 26 input structures are tried (Table 2) and the trial results are shown in Fig. A2. The results show that 8th input structure obtains best performance and thus the No.1 to 8 predictors in Table A1 are selected as the model input. Finally, a total of 16 variables including 8 observed variables and 8 reanalysis variables are selected as the model inputs (Table 3). As shown in Table 3, No. 9 to 18 are reanalysis variables and the range of MIC of the reanalysis variables selected is 0.643 to 0.847. Furthermore, No. 9 and No. 13 to 16 are variables related to temperature. Soil temperature level 3 (No. 9) is the temperature of the soil in layer 3 (28-100 cm, the surface is at 0 cm). The temperature of the snow layer (No. 13) gives the temperature of the snow layer from the ground to the snow-air interface. No. 10 to 12 are variables related to the water content of the atmosphere. 2 meter dewpoint temperature (No. 10) is a measure of the humidity of the air. Combined with temperature and pressure, it can be used to calculate the relative humidity. The total column water vapor (No. 11) is only the total amount of water vapor, which is a fraction of the total column water. Total column water (No. 12) is the sum of water vapor, liquid water, cloud ice, rain and snow in a column extending from the surface of the Earth to the top of the atmosphere. Volumetric soil water layer 1 (No. 19) is the volume of water in soil layer 1. In summary, all the selected predictors are interpretable and have a good physical connection with inflow.

4.2 Hyperparameter optimization

For machine learning methods, hyperparameters are a king of parameters that are set before training and cannot be directly learned from the regular training process. In order to improve the performance of models, it is imperative to tune the hyperparameters of models. Grid search is employed to tune the hyperparameters of GBRT, GBRT-MIC, ANN-MIC and SVR-MIC. Reviewing to the basis of available literature (Badrzadeh et al., 2013; Rasouli et al., 2012), an optimizer in the family of quasi-Newton methods, namely L-BFGS is used as the training algorithm of ANN and the number of hidden layers is fixed to 3. Another two parameters, namely activation function and the number of nodes of the hidden layer need to be adjusted. A range of 2-20 neurons and four commonly used activation functions (Table 4) are selected by grid search. To alleviate the influence of random initialization of weights, 50 ANN-MIC models are trained for each parameter combination. Optimal activation function and the number of nodes of the hidden layer are determined by selecting the minimal MAE of the validation set for each lead

time. The results of the trials show tanh and logistic function are two more robust activation function (Fig. 7) and ANN with fewer nodes is inclined to obtain lower error. The optimal parameter combination for each lead time is listed in Table 5. It can be seen that the optimal number of nodes is 2, 3 or 4 and the optimal activation function is either tanh or logistic function. For SVR, according to Lin et al. (2006) and Dibike et al. (2001), radial basis function (RBF) outperforms other kernel functions for runoff modelling and thus RBF is used as the kernel function in this study. There are three parameters need to be adjusted. Firstly, an appropriate tuning range of parameter is determined by a trial-and-error procedure. And then, to reach at an optimal choice of these parameters, the MAE is used to optimize the parameters by grid search. Optimal tuning parameters of SVR are shown in Table 5. As mentioned earlier, for GBRT, there are six parameters need to be adjusted. In order to obtain an optimal parameter combination as soon as possible, we optimize all parameters in two steps. Firstly, n_estimators and learning_rate are fixed to 100 and 0.1, respectively. The max_leaf_nodes, min_samples_leaf, max_depth and min_samples_split, four tuning parameters generate 40000 models at each lead time. Secondly, after the tree parameters are determined, learning_rate is modified to 0.01 and n_estimators is determined by grid search. To accommodate the computational burden, all models are distributed among about 12 central processing units (CPUs) and total wall time for the runs is about 7 hours for GBRT_MIC and GBRT. Table 6 lists optimal tuning parameters of GBRT and GBRT-MIC.

4.3 Inputs comparison

Fig. 8 illustrates performance indices of GBRT and GBRT-MIC on the testing set (2017/01/01-2018/12/31) at lead times of 1-10 days. It is obvious that the reanalysis data selected by MIC makes a great improvement on the GBRT forecasting at both short and long lead times. In particular, for the longer lead times prediction of GBRT-MIC is significantly outperform GBRT. For Fig. 8(a), the MAE of GBRT-MIC decreases from 175 to 172, a decrease of 1.74% for two-day ahead forecasting and decreases from 273 to 237, a decrease increasing to 13.18% for ten-day ahead forecasting compared with GBRT. For Fig. 8(b), the RMSE of GBRT-MIC achieves 1.4% and 10.6% reduction for two and ten-day ahead forecasting, respectively, compared with GBRT. For Fig. 8(c), 8(d) and 8(f), the CORR, KGE and IA of GBRT-MIC increase by 0.2%, 2.2%, 1.0% for two-day ahead forecasting and 3.4%, 7.8% and 2.2% for ten-day ahead forecasting, respectively. Fig. 8(e) compares the BHV of GBRT and GBRT-MIC which indicates reanalysis data can enhance forecasting of extreme values. Fig. 9(a) shows the five-day ahead forecasted inflow of GBRT-MIC and GBRT versus the observed inflow in the testing set. The slopes of fitting curve of GBRT-MIC and GBRT are 0.89 and 0.81, respectively, which also demonstrates that GBRT-MIC can obtain more accurate in-
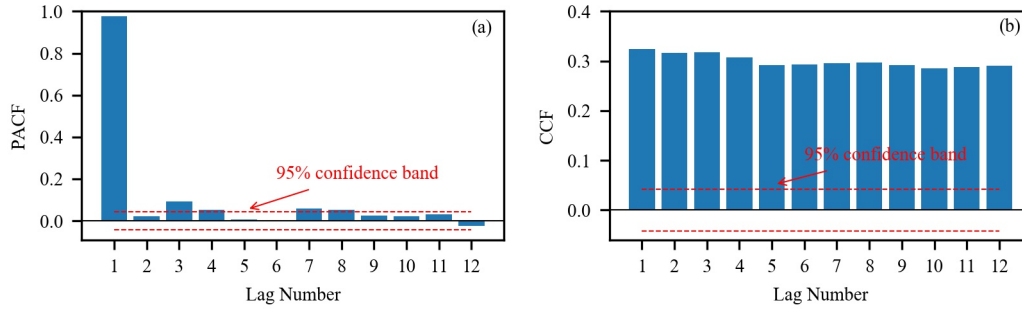
Figure 6. PACF plot of Xiaowan daily inflow and CCF of Xiaowan rainfall and inflow. (a) PACF (b) CCF.

Table 3. List of inputs of GBRT-MIC. There are of two types, observed and reanalysis variables. The reanalysis variables are available two time a day at 00:00 UTC and 12:00 UTC. The cumulative variable (e.g., Total column water) is the sum of two periods and the instantaneous variable (e.g. 2 meter dewpoint temperature) is the mean of two periods.

No.	Description	Index	Unit	MIC	Type
1	Inflow at day t – 1	Q_{t-1}	m^3/s	-	Obs.
2	Inflow at day t – 2	Q_{t-1}	m^3/s	-	Obs.
3	Rainfall at day t – 1	R_{t-1}	mm	-	Obs.
4	Rainfall at day t – 2	R_{t-2}	mm	-	Obs.
5	Rainfall at day t – 3	R_{t-3}	mm	-	Obs.
6	Rainfall at day t – 4	R_{t-4}	mm	-	Obs.
7	Rainfall at day t – 5	R_{t-5}	mm	-	Obs.
8	Rainfall at day t – 6	R_{t-6}	mm	-	Obs.
9	Soil temperature level 3	$stl3_{t-1}$	K	0.847	ERA-I.
10	2 meter dewpoint temperature	$2d_{t-1}$	K	0.781	ERA-I.
11	Total column water vapour	$tcwv_{t-1}$	kg/m^2	0.699	ERA-I.
12	Total column water	tcw_{t-1}	kg/m^2	0.699	ERA-I.
13	Soil temperature level 2	$stl2_{t-1}$	K	0.689	ERA-I.
14	Minimum temperature at 2 meters	$mn2t_{t-1}$	K	0.684	ERA-I.
15	Temperature of snow layer	tsn_{t-1}	K	0.664	ERA-I.
16	Soil temperature level 4	$stl4_{t-1}$	K	0.643	ERA-I.

**Table 4.** Four commonly used activation functions for ANN-MIC.

Name	Functional expression
Logistic	$f(x) = \frac{1}{1+e^{-x}}$
Tanh	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Identity	$f(x) = x$
Relu	$f(x) = max(0, x)$

flow forecasting than GBRT. Fig. 9(b) illustrates the distribution of the forecast errors of GBRT and GBRT-MIC. The results show the prediction error of two models approximate to normal distribution. It demonstrates that the prediction error contains less information that is not extracted by the model and more errors of forecasted inflow concentrate at 0 around by GBRT-MIC than GBRT. Fig. 9(c) provides forecasted in-

flow time series (from the testing set) of GBRT-MIC and GBRT at lead time of five-day. It can be seen that GBRT-MIC provides great performance compare to GBRT, especially for the extreme values. This reveals that the problem of inaccurate extreme value prediction arisen in areas with concentrated rainfall for the GBRT model could be mitigated by incorporating the reanalysis data identified by MIC.

4.4 Model comparison

GBRT-MIC, SVR-MIC, ANN-MIC with obtained optimal model parameters are employed for inflow forecasting of one to ten-day ahead. Summarized results for training and testing set are presented in Table 7 and Table 8, respectively. To avoid local minima problems, 50 ANN-MIC models are trained for each lead time and the median of the predictions of the 50 models gives the final prediction. It is clear from Table 7 that the GBRT-MIC are more efficient in the training set than other models at lead times of 1-10 days, which
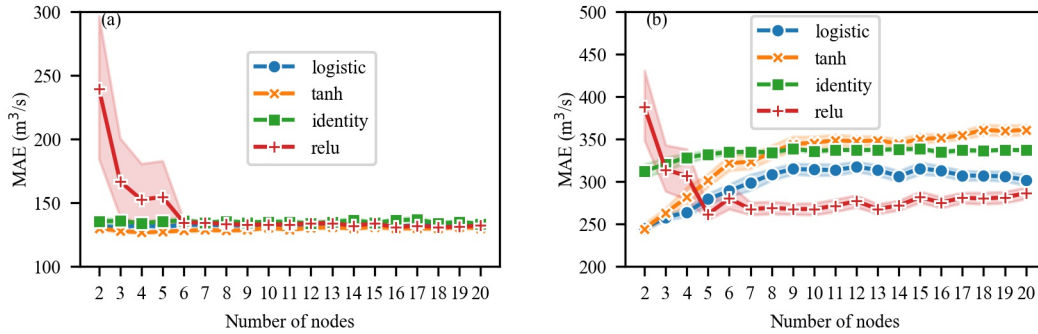
Figure 7. Sensitivity of the activation function and the number of nodes in the hidden layer on the MAE of ANN-MIC, the shadow part is 95% confidence interval obtained by bootstrap of 50 trials. (a) One-day ahead (b) Ten-day ahead.

Table 5. Tuning parameters of ANN-MIC and SVR-MIC.

Model	Tuning parameter	Tuning range	1	2	3	4	5	6	7	8	9	10
ANN-MIC	Structure	/	19-4-1	19-2-1	19-3-1	19-2-1	19-2-1	19-2-1	19-2-1	19-2-1	19-2-1	19-2-1
	Activate function	/	tanh	tanh	logistic	logistic	logistic	logistic	logistic	logistic	tanh	tanh
SVR-MIC	C	**(1, 100, 20)**	6.2105	1.0000	1.0000	1.0000	11.4211	1.0000	1.0000	6.2105	1.0000	6.2105
	ϵ	**(0.001, 0.1, 20)**	0.0069	0.0084	0.0017	0.0079	0.0017	0.0001	0.0022	0.0006	0.0048	0.0043
	γ	**(0.001, 0.1, 20)**	0.0323	0.0583	0.0844	0.0271	0.0062	0.0218	0.0375	0.0166	0.0687	0.0166

Note: The bold parts, (min, max, step) represent $[min + \frac{max-min}{step-1} \times 0, min + \frac{max-min}{step-1} \times 1, \dots, min + \frac{max-min}{step-1} \times (step-1)]$

demonstrates that GBRT-MIC has a powerful fitting ability. Meanwhile, all machine learning models obtain better forecasted results than MLR-MIC which cannot capture nonlinear relationship. It should be noted that ANN-MIC has best performance for extreme values in terms of BHV in the training set. As shown in Table 8, GBRT-MIC performs best for the testing set at lead times of 4-10 days in terms of six indices. At a lead time of ten days, the KGE of GBRT-MIC even reached 0.8317. At the lead times of 1-3 days, three machine learning models obtain approximate performance but outperform MLP-MIC. The machine learning models can acquire enough information to perform forecasting at the short lead time (1-3 days). The performance indices of these four models in the testing set (2017-2018) at the lead times of 1-10 days are presented in Fig. 10. The results indicate the performance of these four models decreases (higher MAE, RMSE and BHV, and lower CORR, KGE and IA) as the lead time increases. As mentioned earlier, the four models perform equally well for one- to three-day ahead forecasting, whereas significant differences among their performances are found as the lead time exceeds three days. It clearly indicates that the GBRT produce much higher CORR, KGE and IA, and lower MAE, RMSE and BHV than the other three models for four to ten-day ahead forecasting except that the ANN-MIC perform nearly to GBRT-MIC for ten-day ahead forecasting. It should be noted that SVR performs worst according to BHV and KGE, which demonstrates that SVR cannot capture extreme values. On the contrary, GBRT-MIC significantly outperform other models in terms of BHV at lead times of 1-10 days, which indicates that GBRT-MIC is able to obtain extreme values among all models developed in this paper.

4.5 Feature importance

A benefit of using gradient boosting is that after the boosted trees are constructed, relative importance scores for each feature can be acquired to estimate the contribution of each feature to inflow forecasting. Fig. 11 shows the feature importance based on GBRT-MIC for lead times of one and ten days. The one-day lag observed time series (Q_{t-1}) is more important for shorter lead times (Fig. 11(a)), which demonstrates that the historical observed values are essential to inflow forecasting at shorter lead times. The features (e.g., $stl3_{t-10}$ and $d2m_{t-10}$) from the reanalysis data have a high relative importance at longer lead times (Fig. 11(b)). Based on the analysis of the concepts of and (Section 4.1), we infer that the temperature near the ground effects the inflow by affecting the melting of snow which is consistent with the fact that the Lancang River is a snow-melt river. The ten-day lag observed time series (Q_{t-10}) is also very important which indicate the long memory of inflow series (Salas, 1993). Mean-

Table 6. Tuning parameters of GBRT and GBRT-MIC.

Tuning parameter	Tuning range	Optimal parameters (the lead times of 1-10 days)	
		GBRT	GBRT-MIC
max_leaf_nodes	[2, 4, 6,…, 40]	8, 4, 4,4, 4, 2, 4, 2, 2, 2	7, 9, 13, 7, 15, 4, 5, 4, 4, 17
min_samples_leaf	[1, 6, 11, …, 46]	6, 31, 1,1, 1, 31, 6, 1, 6, 1	2, 7, 2,4, 2, 1, 10, 10, 8, 1
max_depth	[1, 2, 3, …, 10]	3, 2, 2,2, 3, 1, 3, 1, 1, 1	4, 6, 8,5, 9, 9, 2, 2, 7, 2
min_samples_split	[2, 4, 6, …, 40]	18, 2, 16,16, 24, 2, 16, 2, 2, 2	18, 15, 12,13, 8, 3, 19, 3, 19, 8
n_estimators	[100, 200, 300, …, 4000]	1100, 900,1200, 700, 700, 1200, 600,1100, 900, 900	3800, 2700,1300, 900, 1000, 700, 1400,2000, 1300, 1200
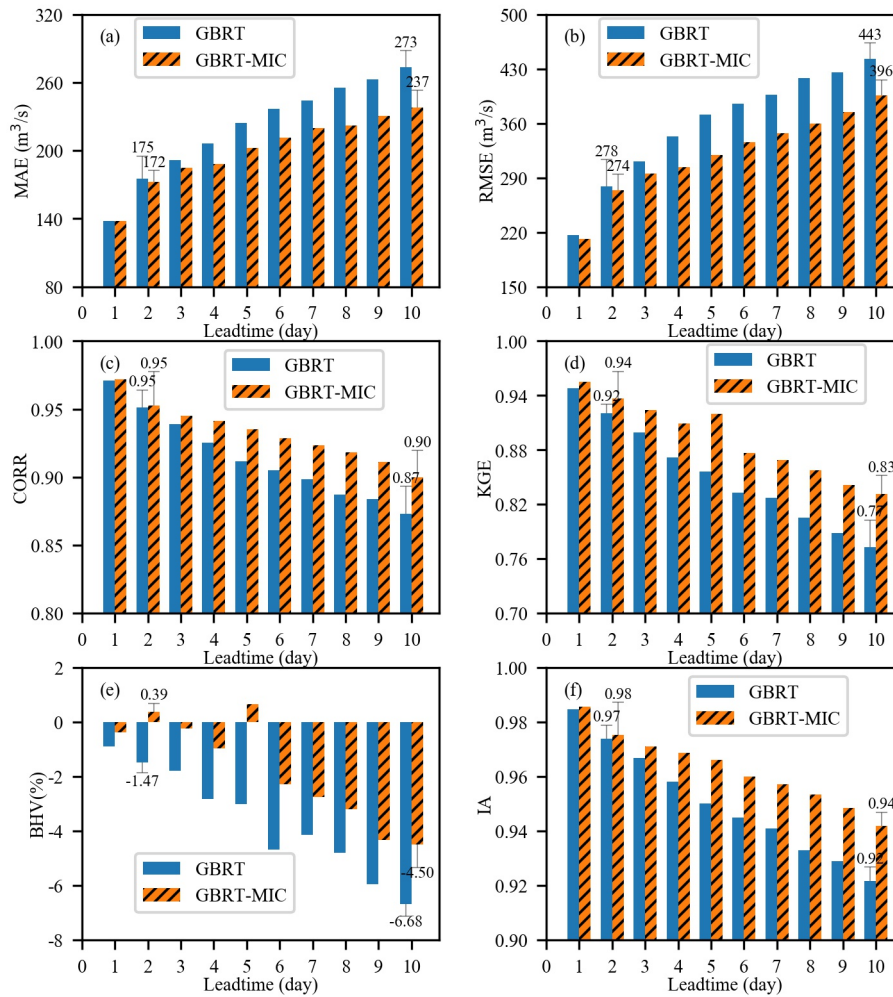
Figure 8. Performance of GBRT and GBRT-MIC for the testing set (2017-2018) in term of six indices. (a) MAE (b) RMSE (c) CORR (d) KGE (e) BHV (f) IA.

while, it is found that the reanalysis data provides important information for inflow forecasting at longer lead times.

5   Conclusions

In this study, GBRT-MIC is employed to make inflow forecasts for lead times of 1-10 days and ANN-MIC, SVR-MIC and MLR-MIC are developed to compare with GBRT-MIC. The reanalysis data selected by MIC, the antecedent inflow
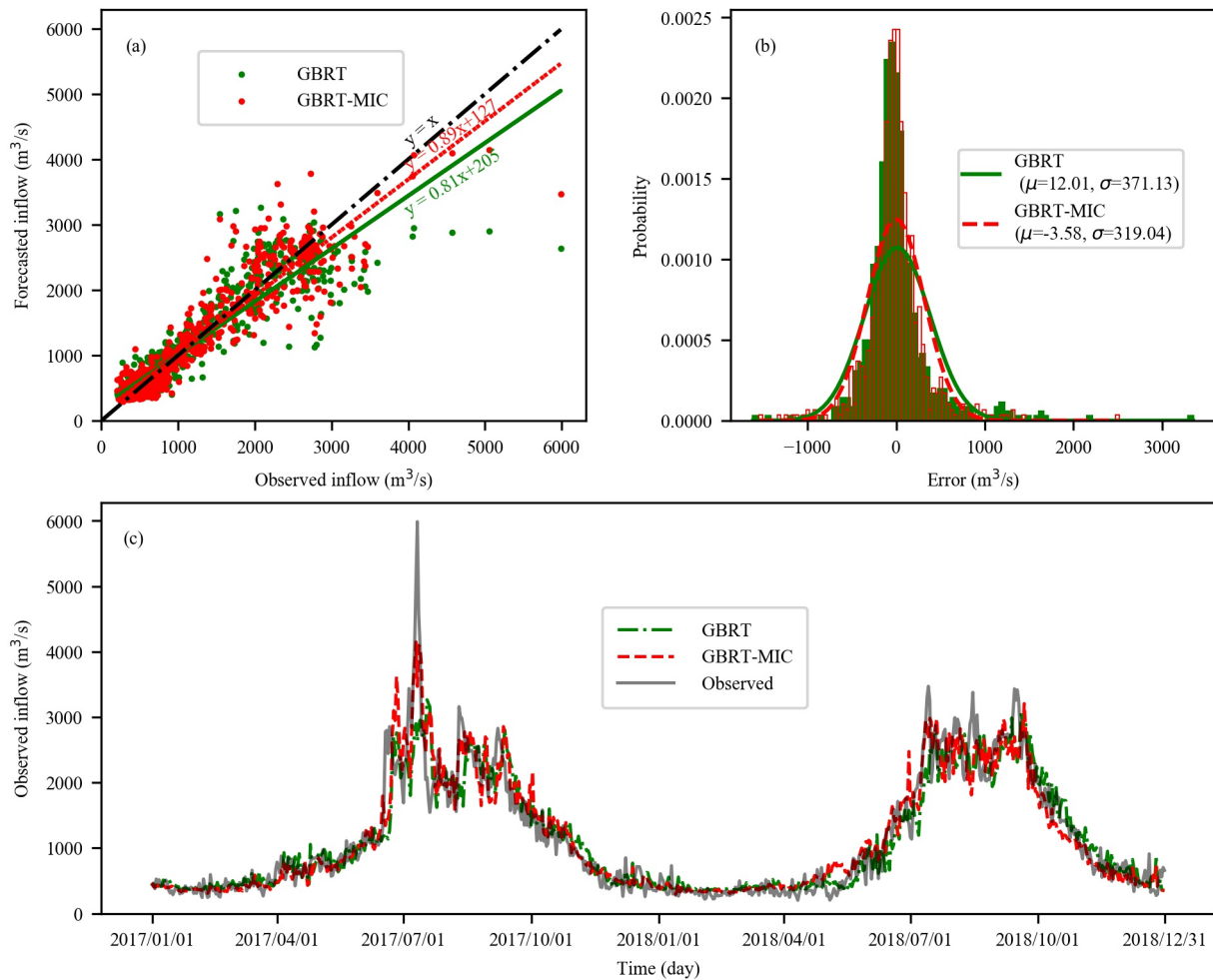
Figure 9. Five-day ahead inflow forecasts of the GBRT and GBRT-MIC for the testing set (2017-2018, 730 days). (a) Observed versus forecasted inflow. (b) The histogram of predicting error of testing set (c) Comparison of the observed and forecasted inflow.

and the rainfall records selected by PACF and CCF are used as predictors to drive the models. These models are compared using six evaluation criteria, the MAE, RMSE, CORR, KGE, BHV and IA. It is shown that GBRT-MIC, ANN-MIC and SVR-MIC outperform MLR-MIC at lead times of 1-10 days, and GBRT-MIC performs best at lead times of 4-10 days, especially for forecasting of extreme values. According to comparison the forecasted results of GBRT and GBRT-MIC, we conclude that GBRT-MIC can be used for more accurate and reliable inflow forecasting at lead times of 1-10 days and reanalysis data selected by MIC makes a great improvement on the GBRT forecasting, especially for lead times of 4-10 days. In addition, the feature importance achieved by GBRT-MIC demonstrates that soil temperature, the total amount of water vapour in a column and dewpoint temperature near the ground contribute to increase the prediction accuracy of inflow at longer lead times. In summary, the developed framework that integrates GBRT and reanalysis data selected MIC

and can well perform inflow forecasting at lead times of 1-10 days. The results of this study are of significance to assist power stations in making power generation plans 7-10 days in advance in order to reduce LEQDW and flood disasters. Another direction of improving the results could be considering heuristic methods (for example, Grey Wolf algorithm) to optimize model parameters, which could search for more wide range of hyper parameters and get optimization parameters more quickly.
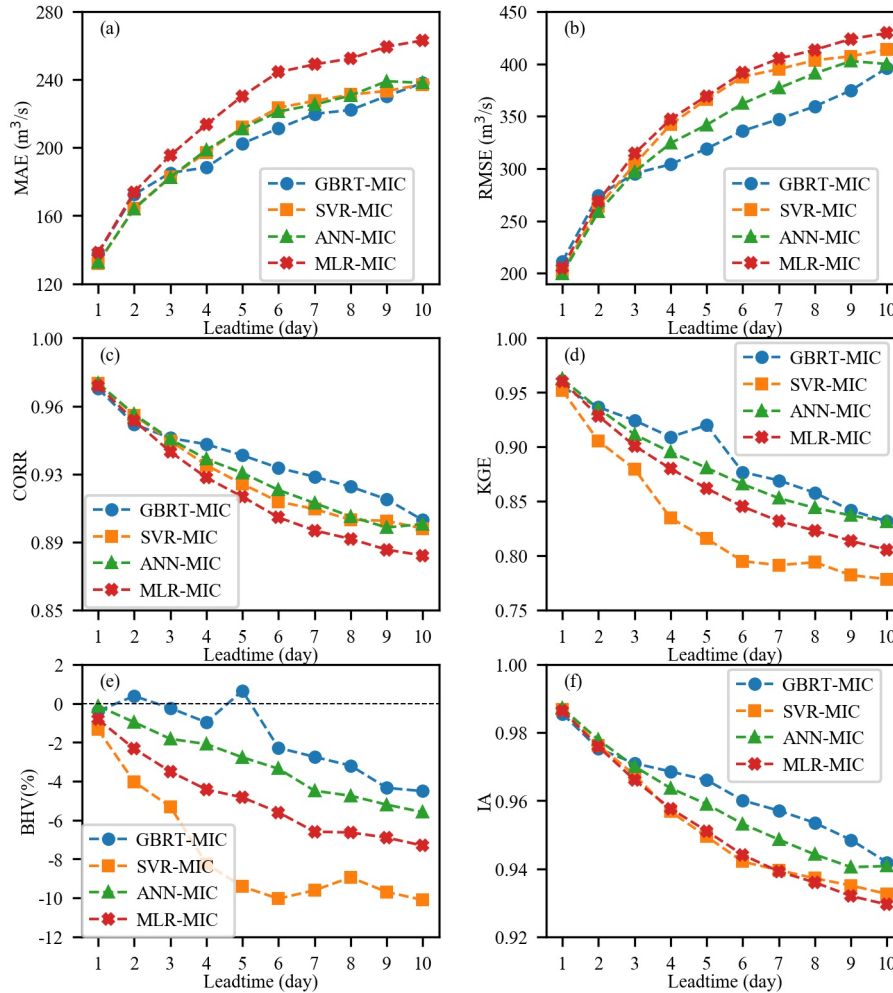
Figure 10. Performance of GBRT-MIC, SVR-MIC, ANN-MIC and MLR-MIC for the testing set (2017-2018) in term of six indices. (a) MAE (b) RMSE (c) CORR (d) KGE (e) BHV (f) IA.

Appendix A

The ERA-Interim is a reanalysis product of the global atmospheric forecasts at ECMWF which is produced through data assimilation system, called as the Integrated Forecast System (IFS). The system includes a 4-dimensional variational analysis (4D-Var) with a 12-hour analysis window. The spatial resolution of the data set is approximately 80 km (0.72°) on 60 levels in the vertical from the surface up to 0.1 hPa. (Berrisford et al., 2011). This reanalysis meteorological products of from 0.125° to 2.5° are generated by interpolation. This reanalysis meteorological products from the ERA-Interim such as rainfall, maximum and minimum temperatures, and wind speed at 0.25° (latitude) × 0.25° (longitude) spatial and 12-hour temporal resolutions for the study period 2011-2018 are downloaded from the ECMWF webpage.

13 input structures from observed data are tried and 50 trials are performed for each input structure. The results (Fig. A1) show 7th input structure is the optimal input subset for GBRT. 26 input structures from reanalysis data are tried and 50 trials are performed for each input structure. The results (Fig. A2) show 8th input structure is the optimal input subset for GBRT-MIC.

Table 7. Performance indices of the training set.

Indice	Model	1	2	3	4	5	6	7	8	9	10
MAE (m^3/s)	GBRT-MIC	**56**	**63**	**78**	**122**	**89**	**163**	**161**	**155**	**161**	**172**
	SVR-MIC	98	126	144	162	173	183	188	194	197	203
	ANN-MIC	99	129	148	162	172	184	192	196	203	205
	MLR-MIC	103	136	159	175	187	198	207	215	221	228
RMSE (m^3/s)	GBRT-MIC	**77**	**87**	**107**	**185**	**124**	**257**	**255**	**245**	**254**	**278**
	SVR-MIC	153	212	247	280	300	319	329	337	344	353
	ANN-MIC	151	206	240	264	284	304	318	328	334	339
	MLR-MIC	157	214	250	275	295	315	330	342	352	361
CORR	GBRT-MIC	**0.9952**	**0.9940**	**0.9908**	**0.9724**	**0.9877**	**0.9464**	**0.9468**	**0.9510**	**0.9476**	**0.9366**
	SVR-MIC	0.9811	0.9641	0.9511	0.9380	0.9286	0.9186	0.9126	0.9073	0.9039	0.8977
	ANN-MIC	0.9815	0.9653	0.9528	0.9424	0.9331	0.9232	0.9156	0.9101	0.9066	0.9036
	MLR-MIC	0.9801	0.9628	0.9485	0.9376	0.9278	0.9172	0.9090	0.9019	0.8959	0.8900
KGE	GBRT-MIC	**0.9884**	**0.9827**	**0.9738**	**0.9439**	**0.9642**	**0.9009**	**0.9069**	**0.9099**	**0.9002**	**0.8877**
	SVR-MIC	0.9618	0.9207	0.8982	0.8613	0.8445	0.8266	0.8223	0.8247	0.8149	0.8103
	ANN-MIC	0.9735	0.9508	0.9325	0.9177	0.9048	0.8907	0.8800	0.8724	0.8668	0.8611
	MLR-MIC	0.9718	0.9473	0.9272	0.9117	0.8979	0.8829	0.8713	0.8613	0.8528	0.8444
BHV (%)	GBRT-MIC	-0.3025	-0.6382	-0.8986	-1.3422	-1.4019	-1.5485	-1.7486	-1.7692	-2.6647	-3.0375
	SVR-MIC	-1.3488	-3.3959	-4.0686	-6.9058	-7.5421	-8.2216	-6.9950	-6.1996	-6.2406	-5.6687
	ANN-MIC	**-0.1814**	**-0.2586**	**-0.7710**	**-0.7723**	**-0.6249**	**-0.6815**	**-0.6878**	**-0.8821**	**-0.6487**	**-0.1239**
	MLR-MIC	-0.4668	-1.0527	-1.5863	-1.9709	-1.9477	-2.1634	-2.0182	-1.8074	-2.0454	-1.7473
IA	GBRT-MIC	**0.9976**	**0.9969**	**0.9952**	**0.9854**	**0.9935**	**0.9706**	**0.9712**	**0.9734**	**0.9712**	**0.9650**
	SVR-MIC	0.9902	0.9804	0.9727	0.9636	0.9574	0.9506	0.9472	0.9449	0.9421	0.9386
	ANN-MIC	0.9906	0.9820	0.9752	0.9695	0.9643	0.9586	0.9541	0.9509	0.9487	0.9468
	MLR-MIC	0.9898	0.9807	0.9729	0.9668	0.9613	0.9551	0.9502	0.9460	0.9423	0.9387

Note : The bold numbers represent the values of performance criterion for the best fitted models.
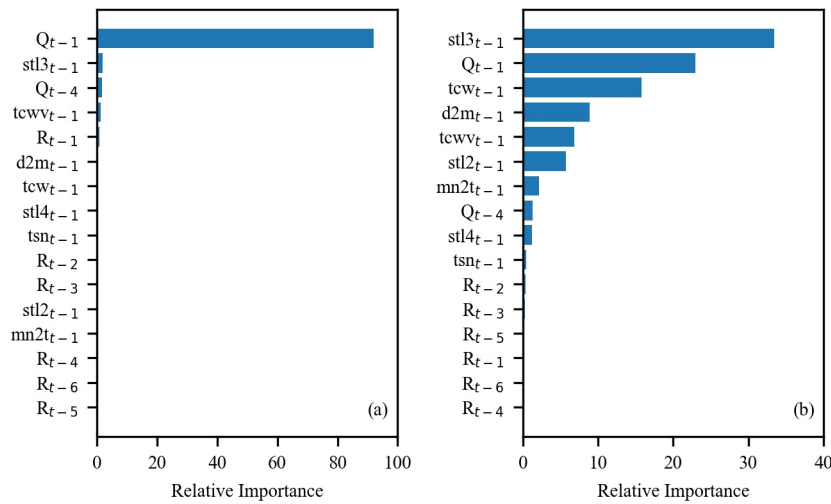

Figure 11. Feature importance obtained by GBRT-MIC. (a) One-day ahead (b) Ten-day ahead.

Table 8. Performance indices of the testing set.

Indice	Model	1	2	3	4	5	6	7	8	9	10
MAE (m³/s)	GBRT-MIC	137	172	185	**188**	**202**	**211**	**219**	**222**	**230**	**237**
	SVR-MIC	**131**	164	182	197	212	223	227	231	233	237
	ANN-MIC	132	**163**	**182**	198	211	221	225	230	239	238
	MLR-MIC	138	173	195	213	230	244	248	252	259	263
RMSE (m³/s)	GBRT-MIC	211	274	**295**	**304**	**319**	**336**	**347**	**359**	**374**	**396**
	SVR-MIC	200	263	303	342	366	387	395	403	407	413
	ANN-MIC	**199**	**258**	296	324	341	362	376	391	402	399
	MLR-MIC	205	268	314	347	369	391	404	413	423	429
CORR	GBRT-MIC	0.9722	0.9526	**0.9449**	**0.9414**	**0.9354**	**0.9285**	**0.9236**	**0.9181**	**0.9112**	**0.8997**
	SVR-MIC	0.9751	0.9575	0.9434	0.9300	0.9196	0.9099	0.9058	0.8999	0.8993	0.8950
	ANN-MIC	**0.9752**	**0.9580**	0.9444	0.9333	0.9257	0.9163	0.9091	0.9017	0.8956	0.8975
	MLR-MIC	0.9738	0.9545	0.9374	0.9231	0.9126	0.9012	0.8940	0.8893	0.8834	0.8802
KGE	GBRT-MIC	0.9550	**0.9367**	**0.9244**	**0.9092**	**0.9200**	**0.8769**	**0.8693**	**0.8580**	**0.8417**	**0.8317**
	SVR-MIC	0.9520	0.9055	0.8797	0.8347	0.8158	0.7950	0.7915	0.7941	0.7822	0.7786
	ANN-MIC	**0.9625**	0.9352	0.9115	0.8953	0.8808	0.8658	0.8530	0.8440	0.8371	0.8313
	MLR-MIC	0.9605	0.9284	0.9011	0.8800	0.8620	0.8452	0.8319	0.8232	0.8137	0.8054
BHV (%)	GBRT-MIC	-0.3826	**0.3880**	**-0.2319**	**-0.9629**	**0.6566**	**-2.2766**	**-2.7422**	**-3.1924**	**-4.3363**	**-4.5040**
	SVR-MIC	-1.3382	-4.0253	-5.3037	-8.2410	-9.4167	-10.0357	-9.6049	-8.9452	-9.6886	-10.1058
	ANN-MIC	**-0.1228**	-0.9608	-1.8150	-2.0839	-2.7642	-3.3509	-4.4831	-4.7424	-5.1999	-5.5886
	MLR-MIC	-0.8093	-2.3244	-3.4945	-4.4210	-4.8268	-5.5955	-6.5914	-6.6302	-6.8944	-7.3080
IA	GBRT-MIC	0.9856	0.9753	**0.9710**	**0.9686**	**0.9661**	**0.9601**	**0.9571**	**0.9535**	**0.9485**	**0.9419**
	SVR-MIC	0.9869	0.9763	0.9676	0.9568	0.9495	0.9421	0.9396	0.9372	0.9351	0.9326
	ANN-MIC	**0.9872**	**0.9779**	0.9701	0.9637	0.9590	0.9532	0.9486	0.9442	0.9405	0.9408
	MLR-MIC	0.9865	0.9759	0.9661	0.9577	0.9511	0.9441	0.9392	0.9360	0.9320	0.9295

Note : The bold numbers represent the values of performance criterion for the best fitted models.
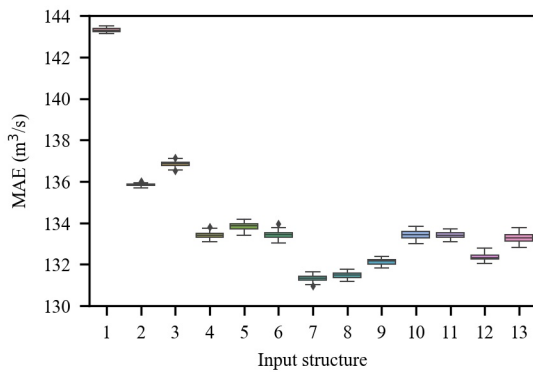
Figure A1. Trial results of 13 input structures from observed data.
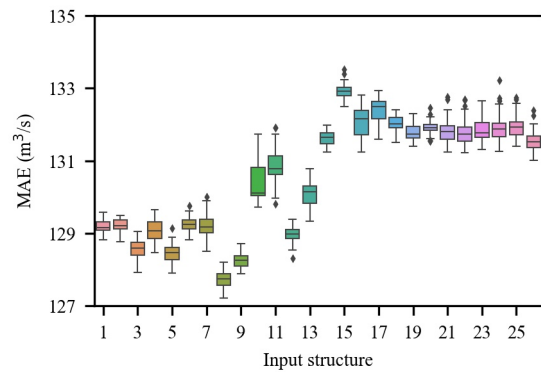
Figure A2. Trial results of 26 input structures from reanalysis data.

Table A1. Description and notations of the ECMWF Reanalysis Fields.

No.	Variable	MIC	Description	Units
1	$stl3$	0.847	Soil temperature level 3	K
2	$d2m$	0.781	2 metre dewpoint temperature	K
3	$tcwv$	0.699	Total column water vapour	$\mathrm{kg/m^2}$
4	tcw	0.699	Total column water	$\mathrm{kg/m^2}$
5	$stl2$	0.689	Soil temperature level 2	K
6	$mn2t$	0.684	Minimum temperature at 2 metres since previous post-processing	K
7	tsn	0.664	Temperature of snow layer	K
8	$stl4$	0.643	Soil temperature level 4	K
9	$stl1$	0.631	Soil temperature level 1	K
10	ro	0.619	Runoff	m
11	$swvl1$	0.614	Volumetric soil water layer 1	$\mathrm{m^3/m^3}$
12	$swvl2$	0.610	Volumetric soil water layer 2	$\mathrm{m^3/m^3}$
13	$swvl3$	0.610	Volumetric soil water layer 3	$\mathrm{m^3/m^3}$
14	$t2m$	0.571	2 metre temperature	K
15	$swvl4$	0.550	Volumetric soil water layer 4	$\mathrm{m^3/m^3}$
16	$mx2t$	0.539	Maximum temperature at 2 metres since previous post-processing	K
17	sf	0.470	Snowfall	m of water equivalent
18	cp	0.426	Convective precipitation	K
19	tp	0.416	Total precipitation	m
20	rsn	0.408	Snow density	$\mathrm{kg/m^3}$
21	lsp	0.358	Large-scale precipitation	m
22	sd	0.337	Snow depth	m of water equivalent
23	$smlt$	0.252	Snowmelt	m of water equivalent
24	$istl1$	0.112	Ice temperature layer 1	K
25	$istl3$	0.109	Ice temperature layer 3	K
26	$istl2$	0.100	Ice temperature layer 2	K

References

Amiri, E.: Forecasting daily river flows using nonlinear time series models, J. Hydrol., 527, 1054 – 1072, https://doi.org/10.1016/j.jhydrol.2015.05.048, 2015.

Amorocho, J. and Espildora, B.: Entropy in the assessment of uncertainty in hydrologic systems and models, water resour. res., 9, 1511–1522, https://doi.org/10.1029/WR009i006p01511, 1973.

Badrzadeh, H., Sarukkalige, R., and Jayawardena, A.: Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting, J. Hydrol., 507, 75–85, https://doi.org/10.1016/j.jhydrol.2013.10.017, 2013.

Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E., and Schepen, A.: Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, water resour. res., 52, 8238–8259, https://doi.org/10.1002/2016WR019193, 2016.

Berrisford, P., Kållberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A. J., Poli, P., and Sato, H.: Atmospheric conservation properties in ERA-Interim, Q. J. Roy. Meteor. Soc., 137, 1381–1399, https://doi.org/10.1002/qj.864, 2011.

Bontempi, G., Taieb, S. B., and Le Borgne, Y.-A.: Machine learning strategies for time series forecasting, in: European business intelligence summer school, pp. 62–77, Springer, https://doi.org/https://doi.org/10.1007/978-3-319-61164-8, 2012.

Bowden, G.J., D.-G. M. H.: Input determination for neural network models in water resources applications. Part 1 - background and methodology, J. Hydrol., 301, 75–92, https://doi.org/10.1016/j.jhydrol.2004.06.021, 2005.

Breiman, L.: Arcing the edge, Report, Statistics Department, University of California,Technical Report 486, https://statistics.berkeley.edu/sites/default/files/tech-reports/486.pdf, 1997.

Chapman, T. G.: Entropy as a Measure of Hydrologic Data Uncertainty and Model Performance, J. Hydrol., 85, 111–126, https://doi.org/10.1016/0022-1694(86)90079-X, 1986.

Chau, K., Wu, C., and Li, Y.: Comparison of several flood forecasting models in Yangtze River, J. Hydrol. eng., 10, 485–491, https://doi.org/10.1061/(ASCE)1084-0699(2005)10:6(485), 2005.

Chau, K.-w.: Use of meta-heuristic techniques in rainfall-runoff modelling, Water, 9, 3, https://doi.org/10.3390/w9030186, 2017.

Cheng, C. T., Feng, Z.-K., Niu, W.-J., and Liao, S.-L.: Heuristic Methods for Reservoir Monthly Inflow Forecasting: A Case Study of Xinfengjiang Reservoir in Pearl River, China, Water, 7, 4477–4495, https://doi.org/10.3390/w7084477, 2015.

Chicco and Davide: Ten quick tips for machine learning in computational biology, Biodata Min., 10, 35, https://doi.org/10.1186/s13040-017-0155-3, 2017.

Commission, M. R.: Overview of the Hydrology of the Mekong Basin, Report, Mekong River Commission, Vientiane, http://www.mekonginfo.org/assets/midocs/0001968-inland-waters-overview-of-hydrology-of-the-mekong-basin.pdf, 2005.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, d. P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dhanya, C. T. and Kumar, D. N.: Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach, water resour. res., 47, W06 507, https://doi.org/10.1029/2010WR010173, 2011.

Dibike, Y. B., Velickov, S., Solomatine, D., and Abbott, M. B.: Model Induction with Support Vector Machines: Introduction and Applications, J. Comput. Civil Eng., 15, 208–216, https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208), 2001.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, water resour. res., 28, 1015–1031, https://doi.org/10.1029/91WR02985, 1992.

El-Shafie, A. and Noureldin, A.: Generalized versus non-generalized neural network model for multi-lead inflow forecasting at Aswan High Dam, Q. J. Roy. Meteor. Soc., 15, 841–858, https://doi.org/10.5194/hess-15-841-2011, 2011.

El-Shafie, A., Taha, M. R., and Noureldin, A.: A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam, Water Resour. Manag., 21, 533–556, https://doi.org/10.1007/s11269-006-9027-1, 2007.

Fan, F. M., Schwanenberg, D., Collischonn, W., and Weerts, A.: Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil, J. Hydrol., 4, 196–227, https://doi.org/10.1016/j.ejrh.2015.05.012, 2015.

Fienen, M. N., Nolan, B. T., Kauffman, L. J., and Feinstein, D. T.: Metamodeling for groundwater age forecasting in the Lake Michigan Basin, water resour. res., 54, 4750–4766, https://doi.org/10.1029/2017WR022387, 2018.

Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K.-w., Faizollahzadeh Ardabili, S., and Piran, M. J.: Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work, Eng. Appl. Comp. Fluid., 12, 411–437, https://doi.org/10.1080/19942060.2018.1448896, 2018.

Foundation, P. S.: Python Language Reference, version 3.7, available at http://www.python.org, 2020.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Ann. Stat., 29, 1189–1232, https://doi.org/10.2307/2699986, 2001.

Ge, R., Zhou, M., Luo, Y., Meng, Q., Mai, G., Ma, D., Wang, G., and Zhou, F.: McTwo: a two-step feature selection algorithm based on maximal information coefficient, Bmc Bioinformatics, 17, 142, https://doi.org/10.1186/s12859-016-0990-0, 2016.

Ghimire, S., Deo, R. C., Downs, N. J., and Raj, N.: Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cites of queensland Australia, J. Clean. Prod., 216, 288–310, https://doi.org/10.1016/j.jclepro.2019.01.158, 2019.

Ghorbani, M. A., Kazempour, R., Chau, K.-W., Shamshirband, S., and Ghazvinei, P. T.: Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran, Eng. Appl. Comp. Fluid., 12, 724–737, https://doi.org/10.1080/19942060.2018.1517052, 2018.

in en.com: https://www.in-en.com/article/html/energy-2266458.shtml, last access: 9 February 2020, 2018.

Kinney, J. B. and Atwal, G. S.: Equitability, mutual information, and the maximal information coefficient, P. Natl. Acad. Sci. USA, 111, 3354–3359, https://doi.org/10.1073/pnas.1309933111, 2014.

Kishore, P., Ratnam, M. V., Namboothiri, S., Velicogna, I., Basha, G., Jiang, J., Igarashi, K., Rao, S., and Sivakumar, V.: Global (50°S–50°N) distribution of water vapor observed by COSMIC GPS RO: Comparison with GPS radiosonde, NCEP, ERA-Interim, and JRA-25 reanalysis data sets, J. Atmos. Sol-Terr. Phy., 73, 1849–1860, https://doi.org/10.1016/j.jastp.2011.04.017, 2011.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sc., 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Lima, A. R., Hsieh, W. W., and Cannon, A. J.: Variable complexity online sequential extreme learning machine, with applications to streamflow prediction, J. Hydrol., 555, 983–994, https://doi.org/10.1016/j.jhydrol.2017.10.037, 2017.

Lin, J. Y., Cheng, C.-T., and Chau, K.-W.: Using support vector machines for long-term discharge prediction, Hydrolog. Sci. J., 51, 599–612, https://doi.org/10.1623/hysj.51.4.599, 2006.

Linares-Rodríguez, A., Ruiz-Arias, J. A., Pozo-Vázquez, D., and Tovar-Pescador, J.: Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks, Energy, 36, 5356–5365, https://doi.org/10.1016/j.energy.2011.06.044, 2011.

Liu, Z., Zhou, P., and Zhang, Y.: A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input*, J. Hydrometeorol., 16, 2209–2229, https://doi.org/10.1175/JHM-D-14-0210.1, 2015.

Louppe, G.: Understanding Random Forests: From Theory to Practice, https://doi.org/10.13140/2.1.1570.5928, 2014.

Luo, X., Yuan, X., Zhu, S., Xu, Z., Meng, L., and Peng, J.: A hybrid support vector regression framework for streamflow forecast, J. Hydrol., 568, 184–193, https://doi.org/10.1016/j.jhydrol.2018.10.064, 2019.

Lyu, H., Wan, M., Han, J., Liu, R., and Cheng, W.: A filter feature selection method based on the Maximal Information Coefficient and Gram-Schmidt Orthogonalization for biomedical data mining, Comput. Biol. Med., 89, 264–274, https://doi.org/10.1016/j.compbiomed.2017.08.021, 2017.

Mason, L., Baxter, J., Bartlett, P. L., and Frean, M. R.: Boosting algorithms as gradient descent, in: Advances in neural information processing systems, vol. 12, pp. 512–518, https://doi.org/10.1016/0012-365X(80)90239-3, 1999.

May, R., Dandy, G., and Maier, H.: Review of Input Variable Selection Methods for Artificial Neural Networks, in: Artificial Neural Networks, edited by Suzuki, K., chap. 2, IntechOpen, Rijeka, https://doi.org/10.5772/16004, 2011.

Mehr, A. D., Jabarnejad, M., and Nourani, V.: Pareto-optimal MPSA-MGGP: A new gene-annealing model for monthly rainfall forecasting, J. Hydrol., 571, 406–415, https://doi.org/10.1016/j.jhydrol.2019.02.003, 2019.

Moazenzadeh, R., Mohammadi, B., Shamshirband, S., and Chau, K.-w.: Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran, Eng. Appl. Comp. Fluid., 12, 584–597, https://doi.org/10.1080/19942060.2018.1482476, 2018.

Mosavi, A., Ozturk, P., and Chau, K.-w.: Flood prediction using machine learning models: Literature review, Water, 10, 1536, https://doi.org/10.3390/w10111536, 2018.

Pal, I., Lall, U., Robertson, A. W., Cane, M. A., and Bansal, R.: Predictability of Western Himalayan river flow: melt seasonal inflow into Bhakra Reservoir in northern India, Q. J. Roy. Meteor. Soc., 17, 2131–2146, https://doi.org/10.5194/hess-17-2131-2013, 2013.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V.: Scikit-learn: Machine learning in Python, J. Mach. Learn Res., 12, 2825–2830, https://arxiv.org/pdf/1201.0490.pdf, 2011.

Rajaee, T., Ebrahimi, H., and Nourani, V.: A review of the artificial intelligence methods in groundwater level modeling, J. Hydrol., 572, 336–351, https://doi.org/10.1016/j.jhydrol.2018.12.037, 2019.

Rasouli, K., Hsieh, W. W., and Cannon, A. J.: Daily streamflow forecasting by machine learning methods with weather and climate inputs, J. Hydrol., 414, 284–293, https://doi.org/10.1016/j.jhydrol.2011.10.039, 2012.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C.: Detecting novel associations in large data sets, science, 334, 1518–1524, https://doi.org/10.1126/science.1205438, 2011.

Robertson, D., Pokhrel, P., and Wang, Q.: Improving statistical forecasts of seasonal streamflows using hydrological model output, Q. J. Roy. Meteor. Soc., 17, 579–593, https://doi.org/10.5194/hess-17-579-2013, 2013.

Salas, J. D.: Analysis and modelling of hydrological time series, in: Stochastic Water Resources Technology, pp. 20–66, Springer, https://doi.org/10.1007/978-1-349-03467-3_2, 1993.

Shoaib, M., Shamseldin, A. Y., Melville, B. W., and Khan, M. M.: Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach, J. Hydrol., 527, 326 – 344, https://doi.org/10.1016/j.jhydrol.2015.04.072, 2015.

Siqueira, H., Boccato, L., Luna, I., Attux, R., and Lyra, C.: Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants, Appl. Soft. Comput., 68, 494–506, https://doi.org/10.1016/j.asoc.2018.04.007, 2018.

Snieder, E., Shakir, R., and Khan, U.: A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models, J. Hydrol., 583, 124 299, https://doi.org/10.1016/j.jhydrol.2019.124299, 2020.

Sohu: https://www.sohu.com/a/209379703_357198, last access: 9 February 2020, 2017.

Stopa, J. E. and Cheung, K. F.: Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis, Ocean Model., 75, 65–83, https://doi.org/10.1016/j.ocemod.2013.12.006, 2014.

Sun, G., Li, J., Dai, J., Song, Z., and Lang, F.: Feature selection for IoT based on maximal information coefficient, Future Gener. Comp. Sy., 89, 606–616, https://doi.org/10.1016/j.future.2018.05.060, 2018.

Taieb, S. B., Bontempi, G., Atiya, A. F., and Sorjamaa, A.: A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition, Expert Syst. Appl., 39, 7067 – 7083, https://doi.org/10.1016/j.eswa.2012.01.039, 2012.

Tongal, H. and Booij, M. J.: Simulation and forecasting of streamflows using machine learning models coupled with base flow separation, J. Hydrol., 564, 266–282, https://doi.org/10.1016/j.jhydrol.2018.07.004, 2018.

Valipour, M., Banihabib, M. E., and Behbahani, S. M. R.: Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, J. Hydrol., 476, 433–441, https://doi.org/10.1016/j.jhydrol.2012.11.017, 2013.

Verkade, J., Brown, J., Reggiani, P., and Weerts, A.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, J. Hydrol., 501, 73–91, https://doi.org/10.1016/j.jhydrol.2013.07.039, 2013.

Vogel, R. M. and Fennessey, N. M.: Flow-Duration Curves. I: New Interpretation and Confidence Intervals, J. Water Res. Pl., 120, 485–504, https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485), 1994.

Wang, E., Zhang, Y., Luo, J., Chiew, F. H., and Wang, Q.: Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data, water resour. res., 47, W05 516, https://doi.org/10.1029/2010WR009922, 2011.

Wei, Z., Meng, Y., Zhang, W., Peng, J., and Meng, L.: Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau, Remote Sens. Environ., 225, 30–44, https://doi.org/10.1016/j.rse.2019.02.022, 2019.

Willmott, C. J.: On the validation of models, Phys. Geogr., 2, 184–194, https://doi.org/10.1080/02723646.1981.10642213, 1981.

Yang, Q., Zhang, H., Wang, G., Luo, S., Chen, D., Peng, W., and Shao, J.: Dynamic runoff simulation in a changing environment: A data stream approach, Environ. Modell. softw., 112, 157–165, https://doi.org/10.1016/j.envsoft.2018.11.007, 2019.

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., and Chau, K.-W.: An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction, J. Hydrol., 569, 387–408, https://doi.org/10.1016/j.jhydrol.2018.11.069, 2019.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, water resour. res., 44, https://doi.org/10.1029/2007WR006716, 2008.

Zhan, X., Zhang, S., Szeto, W. Y., and Chen, X.: Multi-step-ahead traffic speed forecasting using multi-output gradient boosting regression tree, J. Intell Transport. S., pp. 1547–2442, https://doi.org/10.1080/15472450.2019.1582950, 2019.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., and Zhuang, J.: Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm, J. Hydrol., 565, 720–736, https://doi.org/10.1016/j.jhydrol.2018.08.050, 2018.

Zhao, X., Deng, W., and Shi, Y.: Feature selection with attributes clustering by maximal information coefficient, Procedia Comput. Sci., 17, 70–79, https://doi.org/10.1016/j.procs.2013.05.011, 2013.