Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques

Yanlai Zhou[1,2], Zhen Cui[1], Kangling Lin[1], Sheng Sheng[1], Hua Chen[1,3], Shenglian Guo[1,3], Chong-Yu Xu[2]

[1.] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China.

[2.] Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, N-0316 Oslo, Norway.

[3.] Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan 430072, China.

Corresponding authors: yanlai.zhou@whu.edu.cn (Yanlai Zhou); c.y.xu@geo.uio.no (Chong-Yu Xu).

Abstract

Making accurate and reliable probability density forecasts of flood processes is fundamentally challenging for machine learning techniques, especially when prediction targets are outside the range of training data. Conceptual hydrological models can reduce rainfall-runoff modelling errors with efficient quasi-physical mechanisms. The Monotone Composite Quantile Regression Neural Network (MCQRNN) is used for the first time to make probability density forecasts of flood processes and serves as a benchmark model, whereas it confronts the drawbacks of overfitting and biased-prediction. Here we propose an integrated model (i.e. XAJ-MCQRNN) that incorporates Xinanjiang conceptual model (XAJ) and MCQRNN to overcome the phenomena of error propagation and accumulation encountered in multi-step-ahead flood probability density forecasts. We consider flood forecasts as a function of rainfall factors and runoff data. The models are evaluated by long-term (2009-2015) 3-hour streamflow series of the Jianxi River catchment in China and rainfall products of the European Centre for Medium-Range Weather Forecasts. Results demonstrated that the proposed XAJ-MCQRNN model can not only outperform the MCQRNN model but also prominently enhance the accuracy and reliability of multi-step-ahead probability density forecasts of flood process. Regarding short-term forecasts in testing stages at four horizons, the XAJ-MCQRNN model achieved higher Nash-Sutcliffe Efficiency but lower Root Mean Square Error values, while improving Coverage Ratio and Relative Bandwidth values in comparison to the MCQRNN model. Consequently, the improvement can benefit the mitigation of the impacts associated with uncertainties of extreme flood and rainfall events as well as promote the accuracy and reliability of flood forecasting and early warning.

Keywords: Short-term flood forecast; Probability density forecast; Monotone composite quantile regression; Conceptual model; Machine learning

1 Introduction

Floods are among the world's costliest disasters, and climate and anthropogenic changes have made catastrophic floods much more likely happen. Over the past century, floods accounted for about 30% of natural disasters, claiming more than 19% of the total fatalities and more than 48% of the total number of people affected (Adikari and Yoshitani, 2009; Takeuchi et al., 2018). Flood forecasting and early warning play a pivotal role in geohazard mitigation, floodplain management, design and management of infrastructure projects, agricultural cultivation, and human daily life (Liu et al., 2018; Giuliani et al., 2019). Therefore, accurate and reliable forecasting of flood processes during rainstorm events is extremely crucial and beneficial for a country or catchment, whose prosperity is largely dependent on the optimum use of water resources and flood level control.

To date, a wide variety of modelling frameworks have been introduced to model the complicated nonlinear rainfall-runoff process (Sood and Smakhtin, 2015; Shen, 2018; Kan et al., 2020). In general, these modelling frameworks can be categorized into three groups: conceptual, physically-based and machine learning models. A conceptual hydrological model typically includes prior-knowledge as a necessary part of conceptual components, possessing quasi-physical mechanisms (Li et al., 2020; Tarek et al., 2020). There are all sorts of spatially lumped and distributed conceptual models available with deterministic and stochastic parameter values for a given point or region. A physically-based model usually establishes a simplified catchment system and expresses the internal hydrological processes through several mathematical equations considering the spatial variability of rainfall and parameters (Li et al., 2019; Li and Willems, 2020). However, physically-based models encounter a much higher challenge when trying to achieve the required prediction accuracy due to limited watershed information. An alternative, the machine learning model, constructs a direct mapping between rainfall and runoff variables and extracts their relationship according to the historical observations by machine learning algorithms, without a piece of prior knowledge in relation to internal hydrological processes (Feng et al., 2020; Kao et al., 2020; Li et al., 2020; Lin et al., 2020; Nearing et al., 2020; Xiang et al., 2020).

To the best of our knowledge, the growth in machine learning for hydrological sciences is tied to advances in computing power as well as the vast increase of hydrological observations. However, machine learning models are purely data-driven and do not have any physical mechanisms, which easily lead to overfitting and biased-predictions, especially when prediction targets are unobserved and outside the range of rainfall-runoff data adopted for model training (Hunter et al., 2018; Xie et al., 2021). Typically, at longer forecast horizons, the non-availability of antecedent streamflow data makes it fundamentally challenging to maintain the forecast accuracy of data-driven models. Such weakness imposes limitations on the applicability of data-driven models in rainfall-runoff studies (Schmidt et al., 2020; Zahura et al., 2020). In comparison to physically-based models, conceptual models have been widely applied in rainfall-runoff modelling owing to the practical and simple manner. Formulation of conceptual models signifies only a partial set of the real hydrological cycle processes.

This triggers several constraints like model calibration and parameter transferability difficulties (Humphrey et al., 2016; Kumanlioglu and Fistikoglu, 2019), in cases where forecasters and decision-makers usually require prior knowledge for calibrating and running conceptual models to achieve reasonable accuracy of flood forecasts (Bandai and Ghezzehei, 2020; Gu et al., 2020). Although conceptual and data-driven models are derived from different philosophies, they can help remedy and enhance each other with regard to their inherent limitations and strengths (Tian et al., 2018; Sun et al., 2019; Chadalawada et al., 2020). Hence, a hybridization of both models is an attractive and effective approach for rainfall-runoff modelling (Yong et al., 2017; Kurian et al., 2019; Ghaseminejad and Uddameri, 2020). For instance, Yang et al. (2019) evaluated the integration of machine learning and conceptual models for flood simulation at 1032 streamflow gauging stations worldwide. Farfan et al. (2020) adopted streamflow series forecasts made by a conceptual model as input data of back-propagation neural networks to forecast streamflow data. Hitokoto and Sakuraba (2020) integrated a rainfall-runoff model and a feed-forward artificial neural network to predict real-time water level processes. Hosseiny et al. (2020) integrated a two-dimensional hydraulic model, a random forest classification and a multilayer perceptron for modeling flood depth of the river segment of the Flaming Gorge Dam in the northeast corner of Utah in the US. Konapala et al. (2020) verified hybrid models constructed by combining a conceptual model with a long short-term memory neural network on their capability to simulate streamflow series in hundreds of catchments across the US. The above-mentioned studies explored hybrid models for flood prediction in watersheds with diverse climate conditions by means of point forecasts. This study is the first to apply conceptual models with artificial neural networks incorporating flood probability density (FPD) forecasting for a humid subtropical climate. Exploration of conceptual and machine learning models for FPD forecasting in this study is owing to the following reasons: first, conceptual models possess superior operability while machine learning models possess efficient computation power; and second, physically-based models usually have bottlenecks of low prediction accuracy and time-consuming due to the limitation of available data or quality on the physical properties of watersheds (Liu et al., 2018; Li et al., 2020; Cui et al., 2021; Kao et al., 2021).

Previous research regarding hybrid models mainly concentrated on point forecasts or interval predictions of flood processes by integrating conceptual models with single-output artificial neural networks, whereas to date no study has involved FPD forecasting. Point forecasting, the most frequently used approach, can provide a point estimate of the future flood as precisely as possible for each forecast horizon. Rather than offering single-valued forecasts, interval prediction approaches attempt to create well-calibrated lower and upper bounds of each future prediction subject, to a prescribed confidence level. Furthermore, FPD forecasting can provide flood forecasts in the form of quantile point forecasts, confidence interval forecasts and probability density functions. Although it needs extra efforts to calculate the probability for each possible prediction, additional information gained is highly helpful to promote the full understanding of flood predictability. FPD forecasting not only effectively quantifies the uncertainties in input datasets but also has the ability to construct the complete conditional probability density curves of future flood processes. Consequently, there is a noticeable migration from point forecasts and interval predictions towards FPD forecasts where modelers seek to characterize the full probability distribution of flood events. Among machine learning techniques, the Monotone Composite Quantile Regression Neural Network (MCQRNN) is a special type of multi-output artificial neural networks (ANN) that can effectively estimate multiple non-crossing and nonlinear quantile functions simultaneously. It is a novel and appealing approach for probability density forecasting, as compared with other multi-output ANNs like support vector machine and long short-term memory (Cannon, 2018). Since MCQRNN applications do not exist in FPD forecasts, it would be important and interesting to explore in-depth conceptual hydrological models to decrease the uncertainty of multi-output ANN (e.g. MCQRNN) for FPD forecasting.

The novelty of this study relies on the integration of the Xinanjiang (XAJ) conceptual model and the innovative MCQRNN with a multi-output task. Meanwhile, this is the first application of the integrated XAJ-MCQRNN model for FPD forecasting. There are three main steps in this approach. First, the conceptual hydrological model for rainfall-runoff simulation was constructed to create point forecasts of flood processes and the outputs were used as inputs of the machine learning model. Second, the multi-output machine learning model for FPD forecasting was built to further create interval forecasts of flood processes after being supplied with point forecasts by the conceptual model. Third, the training process of the hybrid model was optimized by an evolutionary optimization algorithm, and the predictability of the hybrid model was tested by using medium-range numeric rainfall forecast products. The Jianxi River catchment of China constituted the case study.

2 Methods

This study aims to integrate the XAJ and MCQRNN models for raising the accuracy and reliability of short-term FPD forecasts. Fig. 1 presents the main architectures utilized in this study: the MCQRNN (Fig. 1(a)); and the proposed XAJ-MCQRNN model that combines XAJ with MCQRNN models (Fig. 1(b)).

2.1 Monotone Composite Quantile Regression Neural Network (MCQRNN)

The MCQRNN has been developed to effectively estimate multiple quantile functions for making probabilistic density forecasts of time series (Cannon, 2018). In comparison to the Quantile Regression Neural Network (QRNN) (Taylor, 2000), the monotone QRNN (Cannon, 2011) and the composite QRNN (Xu et al., 2017), the merit of the MCQRNN model is that it can estimate multiple non-crossing, non-linear conditional quantile functions, can allow for the constraints of optional monotonicity, positivity/non-negativity and generalized additive model, and can be applied to estimating standard least-squares regression (Cannon, 2018).

The structure of the MCQRNN model is composed of input, hidden, and output layers (Fig. 1(a)). Regarding the *k*th quantile τ_k, the outputs of hidden and output layers in association with multi-input variables at horizon T are described below.

$$h_j(t+\mathrm{T}) = f\left(\sum_{s=0}^{\mathrm{T}} \widehat{\mathrm{P}}(t+s)\cdot \exp\left(w_{sj}^{(\mathrm{h})}\right) + \sum_{m=0}^{\mathrm{p}} \mathrm{P}(t-m)\cdot \exp\left(w_{mj}^{(\mathrm{h})}\right) + \sum_{n=0}^{\mathrm{q}} \mathrm{Q}(t-n)\cdot w_{nj}^{(\mathrm{h})} + b_j^{(\mathrm{h})}\right) \tag{1a}$$

$$\widehat{\mathrm{Q}}^{\tau_k}_{\mathrm{MCQRNN}}(t+\mathrm{T}) = \mathrm{g}\left[\sum_{j=1}^{\mathrm{J}} h_j(t+\mathrm{T}) \cdot \exp\left(w_{jk}^{(\mathrm{o})}\right) + b_k^{(\mathrm{o})}\right] \tag{1b}$$

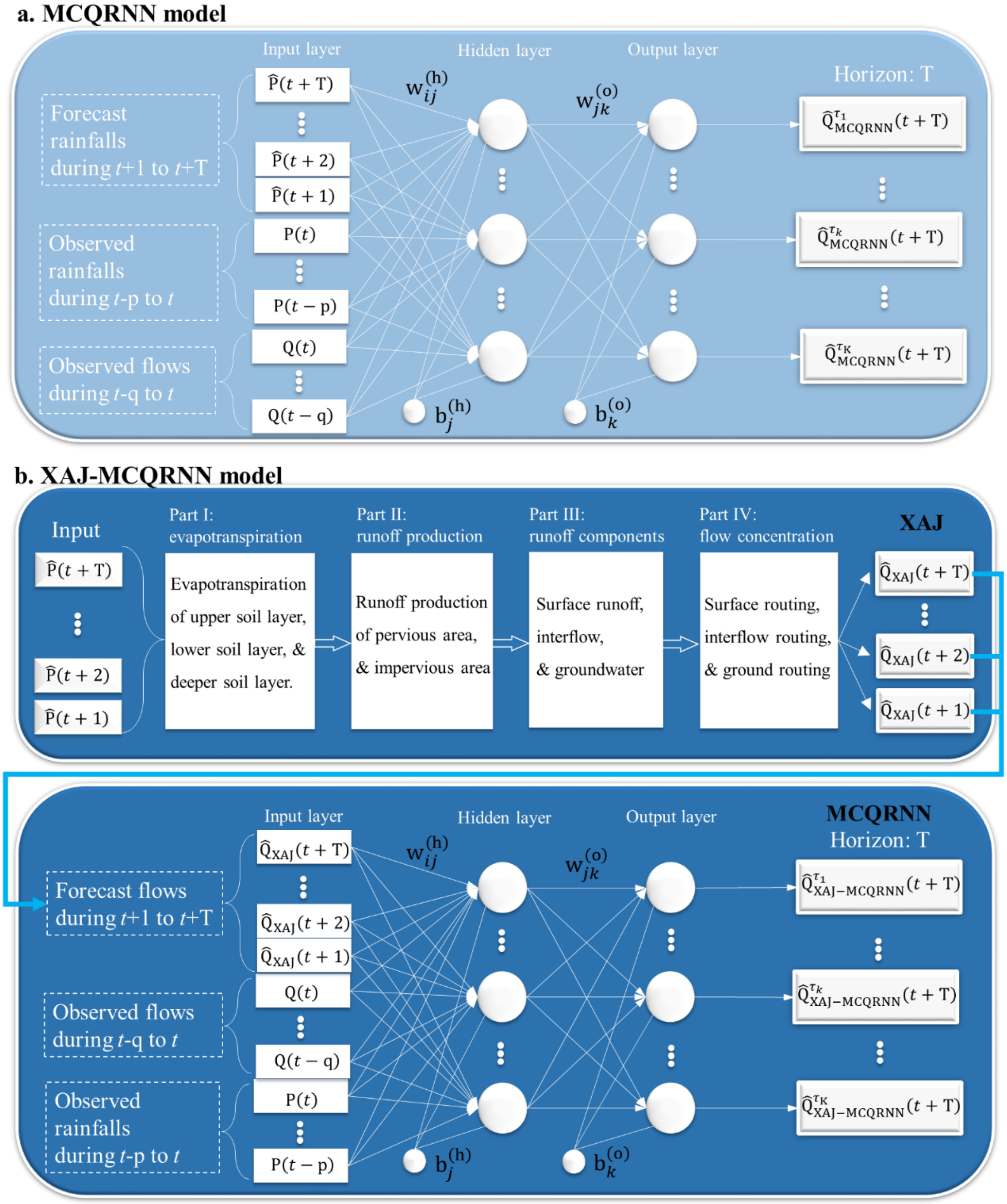

Fig. 1 Main architectures adopted in this study. a. Monotone composite QRNN model (MCQRNN). b. Proposed XAJ-MCQRNN model. $\mathrm{w}_{ij}^{(\mathrm{h})}$, $\mathrm{b}_j^{(\mathrm{h})}$, $\mathrm{w}_{jk}^{(\mathrm{o})}$ and $\mathrm{b}_k^{(\mathrm{o})}$ are the parameters in hidden and output layers of MCQRNN model. $\widehat{\mathrm{Q}}_{\mathrm{MCQRNN}}(t+\mathrm{T})$ is the forecast streamflow of MCQRNN model at horizon *t*+T. $\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+1),\ldots,\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+\mathrm{T})$ are the forecast streamflows of XAJ model during horizons *t*+1 up to *t*+T. $\widehat{\mathrm{Q}}^{\tau_k}_{\mathrm{XAJ-MCQRNN}}(t+\mathrm{T})$ is the forecast streamflow of XAJ-MCQRNN model with τ_k quantile at horizon *t*+T. *t* and T are the current time and forecast horizon, respectively.

where $h_j(t+\mathrm{T})$ and $\widehat{\mathrm{Q}}_{\mathrm{MCQRNN}}^{\tau_k}(t+\mathrm{T})$ are the *j*th node's output transformed from input variables in the hidden layer and the model output with the τ_k quantile (1≤*k*≤K) in the output layer at horizon T, respectively. $\widehat{\mathrm{P}}(t+s)$ is the forecast rainfall at horizon *s* (1≤*s*≤T). $\mathrm{P}(t-m)$ and $\mathrm{Q}(t-n)$ are the *m*th antecedent rainfall and the *n*th antecedent streamflow, respectively. *f*(·) is the hyperbolic tangent function in the hidden layer. g(·) is the sigmoid transfer function in the output layer. *k* and *j* are the indexes of quantile and node, respectively. p and q are the numbers of antecedent rainfall and antecedent streamflow variables at horizon T, respectively. $w_{sj}^{(\mathrm{h})}$ is the weight parameter which connects the *s*th forecast rainfall and the *j*th node in the hidden layer, where the superscript (h) represents the hidden layer. $w_{mj}^{(\mathrm{h})}$ and $w_{nj}^{(\mathrm{h})}$ are the weight parameters which connects the *m*th (*n*th) antecedent rainfall (streamflow) input variable and the *j*th node in the hidden layer, respectively. $b_j^{(\mathrm{h})}$ is the intercept parameter of the *j*th node in the hidden layer. $w_{jk}^{(\mathrm{o})}$ and $b_k^{(\mathrm{o})}$ are the *j*th weight and the intercept parameter of the *k*th quantile in the output layer, respectively, where the superscript (o) represents the output layer. J is the number of nodes.

To enhance the stability of model training, regularization terms are integrated into the error function to penalize the magnitude of the weight parameters within hidden and output layers, as described below.

$$\rho_{\tau_k}(\varepsilon)=\begin{cases}\tau_k\cdot\varphi(\varepsilon) & \varepsilon\geq 0\\ (\tau_k-1)\cdot\varphi(\varepsilon) & \varepsilon\geq 0\end{cases} \tag{2a}$$

$$\varphi(\varepsilon)=\begin{cases}\frac{\varepsilon^2}{2\alpha} & 0\leq|\varepsilon|\leq\alpha\\ |\varepsilon|-\frac{\alpha}{2} & |\varepsilon|>\alpha\end{cases} \tag{2b}$$

$$\mathrm{E}=\sum_{k=1}^{\mathrm{K}}\omega_{\tau_k}\cdot\sum_{t=1}^{\mathrm{D}}\rho_{\tau_k}\left(\mathrm{Q}(t+\mathrm{T})-\widehat{\mathrm{Q}}_{\mathrm{MCQRNN}}^{\tau_k}(t+\mathrm{T})\right) \tag{2c}$$

$$\widetilde{\mathrm{E}} = \mathrm{E} + \lambda^{(\mathrm{h})} \cdot \frac{1}{(\mathrm{T+p+q}) \cdot \mathrm{J}} \sum_{i=1}^{(\mathrm{T+p+q})} \sum_{j=1}^{\mathrm{J}} \left(w_{ij}^{(\mathrm{h})} \right)^2 + \lambda^{(\mathrm{o})} \cdot \frac{1}{\mathrm{J}} \sum_{j=1}^{\mathrm{J}} \left(w_{jk}^{(\mathrm{o})} \right)^2 \qquad (2\mathrm{d})$$

where $\rho_{\tau_k}(\varepsilon)$ and $\varphi(\varepsilon)$ are the Huber-norm function (Chen, 2007) with the τ_k quantile and the Huber function (Huber, 1964) with the variable ε, respectively. $|\varepsilon|$ and α are the absolute error and the given positive value, respectively. The Huber function transforms smoothly from the squared error, which is applied around the origin (α) to accomplish differentiability, and the absolute error. ω_{τ_k} is the weight that represents the contribution of the forecast streamflow with the τ_k quantile to the total error. $\mathrm{Q}(t + \mathrm{T})$ is the observed streamflow. K is the number of quantiles. $\lambda^{(\mathrm{h})}$ and $\lambda^{(\mathrm{o})}$ are the coefficients (i.e., regularization items) that denote the penalties applied to the weight parameters in hidden and output layers, respectively. *i* is the index of input variable. $w_{ij}^{(\mathrm{h})}$ is the weight parameter for connecting the *i*th input variable and the *j*th node in the hidden layer. D is the number of input data. E and $\widetilde{\mathrm{E}}$ are the error functions without (the former) and with (the latter) regularization items considering the Huber-norm approximation, respectively.

In the case of the specified model architecture (i.e. three layers, shown in Fig. 1(a)) with the numbers of input variables (V), hidden nodes (J), and output variables (O), the total number of weight and intercept parameters in the MCQRNN model is the sum of (V×J+ J×O) weight parameters and (J+ O) intercept parameters.

2.2 Xinanjiang (XAJ) conceptual hydrological model

The XAJ model is a conceptual rainfall-runoff model suitable for flood forecasts of humid and semi-humid catchments (Zhao, 1992). From the viewpoint of model architecture, the XAJ model has four implementation phases (Fig. 1(b)). Part Ⅰ is to compute evapotranspiration by three representative soil layers. Part Ⅱ is to calculate runoff production under runoff formation on repletion of storage. Part Ⅲ is to separate total runoff into the components of surface runoff, interflow and groundwater. Part Ⅳ is to compute flow concentrations by means of the Nash Unit Hydrograph and Muskingum routing. The advantage of the XAJ model is that the model has not only the explicit physical meaning of rainfall-runoff processes but also a simple manner for flood forecasting. The mathematical equation of the rainfall-runoff process in the XAJ model during horizons *t*+1 up to *t*+T is depicted as follows.

$$\widehat{\mathbf{Q}}_{\mathbf{XAJ}}^{\mathbf{T}} = \phi\left[\widehat{\mathrm{P}}(t+1), \widehat{\mathrm{P}}(t+2), \cdots, \widehat{\mathrm{P}}(t+\mathrm{T})\right] \quad (3a)$$

$$\widehat{\mathbf{Q}}_{\mathbf{XAJ}}^{\mathbf{T}} = \left[\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+1), \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+2), \cdots, \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+\mathrm{T})\right] \quad (3b)$$

where $\widehat{\mathbf{Q}}_{\mathbf{XAJ}}^{\mathbf{T}}$ is the vector of the XAJ model output with lead time T. $\widehat{\mathrm{P}}(t+\mathrm{T})$ and $\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+\mathrm{T})$ are the forecast rainfall (i.e., mode input) and flow (i.e., model output) at horizon *t*+T, respectively. $\phi[\cdot]$ is the transformation function of rainfall-runoff processes including four parts (Fig. 1(b)).

The XAJ model has 15 parameters, including 4 evapotranspiration parameters, 3 runoff production parameters, 4 runoff separation parameters, and 4 flow concentration parameters. Owing to the effectiveness and practicality, the Shuffled Complex Evolution method developed at The University of Arizona (SCE-UA) (Duan et al., 1994) was adopted to optimize the XAJ model parameters in this study. More details on the XAJ model can be found in Zhao (1992).

2.3 Hybridization of the XAJ and MCQRNN (XAJ-MCQRNN)

The comparison between MCQRNN (Fig. 1(a)) and XAJ-MCQRNN (Fig. 1(b)) models established in this study is summarized as follows: when making FPD forecasts, the former adopts the forecast rainfall data ($\widehat{\mathrm{P}}(t+1), \widehat{\mathrm{P}}(t+2), \cdots, \widehat{\mathrm{P}}(t+\mathrm{T})$) provided by available numerical forecast products as input variables to test the model during horizons t+1 up to t+T, while the latter adopts the forecast streamflow data $\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+1), \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+2), \cdots, \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+\mathrm{T})$ made by the XAJ model during horizons t+1 up to t+T as input variables to test the model.

From the perspective of model complementarity, the XAJ model can produce stable point forecasts of flood hydrographs so that the MCQRNN model can create more accurate and reliable FPD forecasts after being supplied with point forecasts by the XAJ model. The general equation of the proposed hybrid model for the quantile τ_k is described below.

$$\widehat{\mathrm{Q}}_{\mathrm{XAJ-MCQRNN}}^{\tau_k}(t+\mathrm{T}) = \psi\big[\widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+\mathrm{T}), \cdots, \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+2), \widehat{\mathrm{Q}}_{\mathrm{XAJ}}(t+1), \mathrm{Q}(t),$$

$$\mathrm{Q}(t-1), \cdots, \mathrm{Q}(t-\mathrm{q}), \mathrm{P}(t), \mathrm{P}(t-1), \cdots, \mathrm{P}(t-\mathrm{p})\,\big] \tag{4}$$

where $\widehat{\mathrm{Q}}_{\mathrm{XAJ-MCQRNN}}^{\tau_k}(t+\mathrm{T})$ is the forecast streamflow of the XAJ-MCQRNN model with the quantile τ_k at horizon t+T. $\psi[\cdot]$ is the quantile regression function that is the combination of the hyperbolic tangent function $f(\cdot)$ in the hidden layer and the sigmoid transfer function g(·) in the output layer.

2.4 Training process of the MCQRNN model

The gradient-based nonlinear optimization algorithm (Kingma and Ba, 2015; Cannon, 2018) was used to optimize the weight ($w_{sj}^{(\mathrm{h})}$, $w_{mj}^{(\mathrm{h})}$, $w_{nj}^{(\mathrm{h})}$ and $w_{jk}^{(\mathrm{o})}$ in Eqs. 1(a) and 1(b)) and intercept ($b_j^{(\mathrm{h})}$ and $b_k^{(\mathrm{o})}$ in Eqs. 1(a) and 1(b)) parameters of the MCQRNN while the training process of the XAJ-MCQRNN was performed after being supplied with point forecasts by the XAJ at the same time. Furthermore, the Genetic Algorithm (GA) (Goldberg and Deb, 1991; Zhou et al., 2019) was used to optimize the hyperparameters (the maximal iteration I_{max}, the number of nodes J, the regulation coefficients $\lambda^{(h)}$ and $\lambda^{(o)}$, and the learning rate η). Fig. 2 illustrates the flow chart of the XAJ-MCQRNN training process. The implementation procedure is described as follows.

Step 1: Randomly generate an initial population X_0 of size I_{pop} with respect to the hyperparameters. Integer-coded and real-coded solutions are adopted for hyperparameters I_{max} and J as well as for hyperparameters $\lambda^{(h)}$, $\lambda^{(o)}$ and η, respectively. The search spaces of I_{max}, J, $\lambda^{(h)}$, $\lambda^{(o)}$ and α are set to be [200, 500], [1, 30], [0.001, 0.01], [0.001, 0.01] and [0.001, 0.05], respectively.

Step 2: Perform the gradient-based nonlinear optimization algorithm for optimizing weight and intercept parameters of the MCQRNN model, and execute the tournament selection, the crossover operator with probability (P_c) and the mutation operator with probability (P_m) to create an offspring population Y_0 of size I_{pop}.

Step 3: Evaluate the fitness values of Y_u for the *u*th generation, combine Y_{u-1} and Y_u into an intermediate population X_u of size $2 * I_{pop}$, partition this combined population into different ranks according to fitness values, and store the best solution using the elitism strategy.

Step 4: Implement the tournament selection to select a new parent population X_{u+1} of size I_{pop} from X_u, create an offspring population Y_{u+1} using crossover and mutation operators, and evaluate their fitness values.

Step 5: Repeat Steps 3 and 4 when the generation number u is smaller than the maximal generation (G_{max}). Otherwise, output the optimal results after stopping iteration.

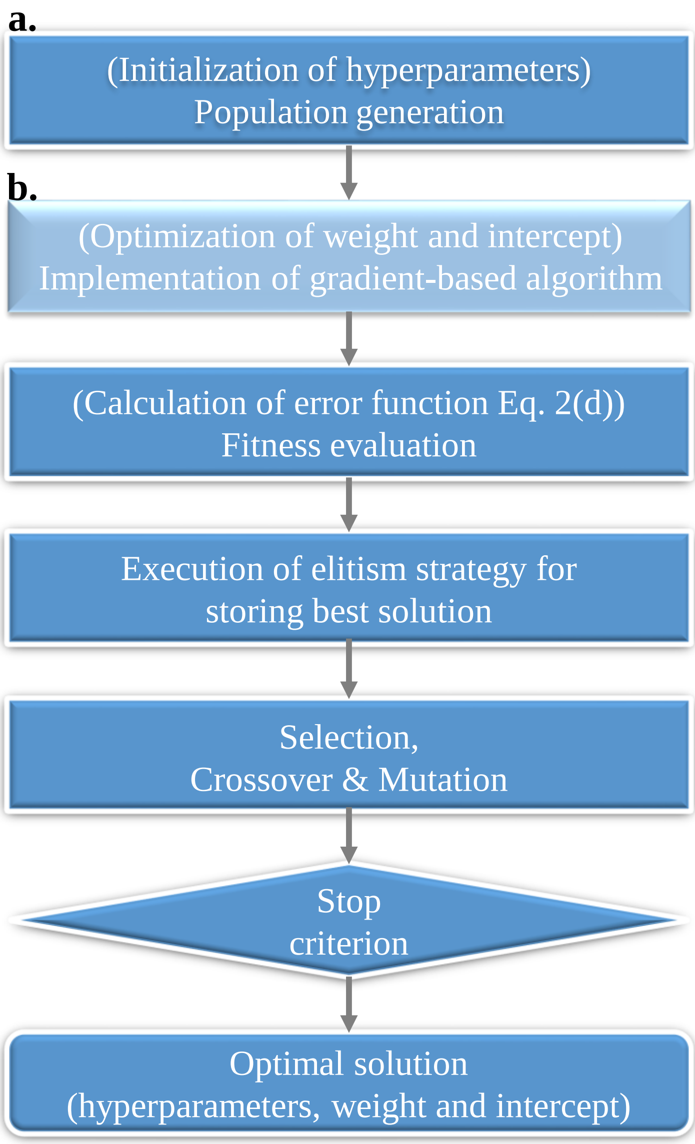

Fig. 2 Flow diagram of the XAJ-MCQRNN model trained simultaneously by: a. the Genetic Algorithm applied to the optimization of hyperparameters; b. the gradient-based nonlinear optimization algorithm applied to the optimization of weight and intercept parameters.

In this study, the population size (I_{pop}), the maximal generation (G_{max}), the crossover probability (P_c) and the mutation probability (P_m) of GA were set to be 200, 100, 0.85 and 0.1, respectively. The aforementioned computations regarding the GA, XAJ model and MCQRNN model are conducted in R software (https://www.r-project.org/), where the computation of the MCQRNN is based on the freely available QRNN package (https://CRAN.R-project.org/package=qrnn).

As river streamflow data are time series, this study applied the Root Mean Square Error (RMSE), the Nash-Sutcliffe Efficiency (NSE) and the Mutual Information (MI) to evaluate the point forecast results of the two MCQRNN models. The RMSE value indicates the error between forecasts and observations, and its value ranges from 0 to infinity (Legates and McCabe, 1999). The NSE is commonly employed to evaluate the forecast accuracy of hydrological models and its value ranges from negative infinity to 1 (Nash and Sutcliffe, 1970). Considering that the upper bound (=1) of the NSE doesn't leave much room for improvement, a logarithm conversion for the NSE (LNSE) is used instead of an unbounded quality indicator to evaluate model accuracy. The MI value suggests the goodness-of-fit between forecasts and observations, and its value ranges from 0 to infinity (Shannon, 1948). A model with low RMSE and high LNSE and MI values suggests it can produce high forecast accuracy. The computations of RMSE, LNSE and MI values are described below.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{D}}\sum_{t=1}^{\mathrm{D}}\left(\mathrm{Q}(t+\mathrm{T}) - \widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})\right)^2} \tag{5}$$

$$\mathrm{LNSE} = -\ln(1-\mathrm{NSE}) = -\ln\left(\frac{\sum_{t=1}^{\mathrm{D}}\left(\mathrm{Q}(t+\mathrm{T})-\widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})\right)^2}{\sum_{t=1}^{\mathrm{D}}(\mathrm{Q}(t+\mathrm{T})-\overline{\mathrm{Q}}_{\mathrm{T}})^2}\right) \tag{6}$$

$$\mathrm{MI} = \sum_{t=1}^{\mathrm{D}}\sum_{t=1}^{\mathrm{D}} p\left(\mathrm{Q}(t+\mathrm{T}), \widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})\right) log_2\left(\frac{p\left(\mathrm{Q}(t+\mathrm{T}), \widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})\right)}{p(\mathrm{Q}(t+\mathrm{T}))f\left(\widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})\right)}\right) \tag{7}$$

where $\widehat{\mathrm{Q}}_{\mathrm{Model}}^{\tau_k}(t+\mathrm{T})$ is the forecast value of a model (MCQRNN or XAJ-MCQRNN) with the τ_k quantile at horizon T. $\overline{\mathrm{Q}}_{\mathrm{T}}$ is the average of observed values at horizon T.

$p(\cdot,\cdot)$ is the joint probability density function of observations and forecasts. $p(\cdot)$ is the probability density function of observations or forecasts. The median (i.e., τ_k=0.5) forecasts of a model are commonly employed to evaluate the prediction accuracy of point forecast results (Zhou et al., 2020).

The Coverage Ratio (CR) of the forecast interval relative to observations and the Relative Bandwidth (RB) of the forecast interval (Xiong and O'Connor, 2008) were used to evaluate the accuracy of the forecast interval results in this study. The CR represents the coverage interval of the forecasts and its value ranges from 0 to 100%. The RB represents the width of the forecast interval and its value ranges from 0 to infinity. A model with its CR value closer to 100% and its RB closer to 0 implies it can forecast more accurately. The computations of CR and RB values are depicted as follows.

$$\mathrm{CR} = \frac{1}{\mathrm{D}} \times \sum_{t=1}^{\mathrm{D}} \theta_t \times 100\% , \quad \theta_t = \begin{cases} 1, & \mathrm{Q}(t+\mathrm{T}) \in \left[\widehat{\mathrm{Q}}_{\mathrm{L}}(t+\mathrm{T}), \widehat{\mathrm{Q}}_{\mathrm{U}}(t+\mathrm{T})\right] \\ 0, & \mathrm{Q}(t+\mathrm{T}) \notin \left[\widehat{\mathrm{Q}}_{\mathrm{L}}(t+\mathrm{T}), \widehat{\mathrm{Q}}_{\mathrm{U}}(t+\mathrm{T})\right] \end{cases} \tag{8}$$

$$\mathrm{RB} = \frac{1}{\mathrm{D}} \times \sum_{t=1}^{\mathrm{D}} \left(\frac{\widehat{\mathrm{Q}}_{\mathrm{U}}(t+\mathrm{T}) - \widehat{\mathrm{Q}}_{\mathrm{L}}(t+\mathrm{T})}{\mathrm{Q}(t+\mathrm{T})} \right) \times 100\% \tag{9}$$

where $\widehat{\mathrm{Q}}_{\mathrm{L}}(t+\mathrm{T})$ and $\widehat{\mathrm{Q}}_{\mathrm{U}}(t+\mathrm{T})$ are the lower and upper boundaries of forecast interval results at horizon T with respect to a given confidence level (e.g. 90%), respectively (e.g. L=5%, U=95%). The θ_t values are either one or zero, in which one implies the observed data fall within its forecast intervals while zero implies the observed data fall outside of its forecast intervals at *t* time.

3 Study area and materials

3.1 Study area

The Jianxi River catchment with an area of 14787 km^2 is located in southern China (Fig. 3). The catchment climate is affected by the southwest Indian Ocean and southeast Pacific Ocean subtropical monsoons and regional landforms (Lin et al., 2020). The annual rainfall ranges from 1800 mm to 2200 mm and the annual runoff ranges from 950 to 1700 mm, the highest amount (74%) of which occurs within six months (April–September). Due to the moist and rainy features as well as red, yellow, and paddy soils, runoff production of this catchment is dominated by the runoff generation on repletion of storage, which is in line with the typical rainfall-runoff characteristics in southern China.

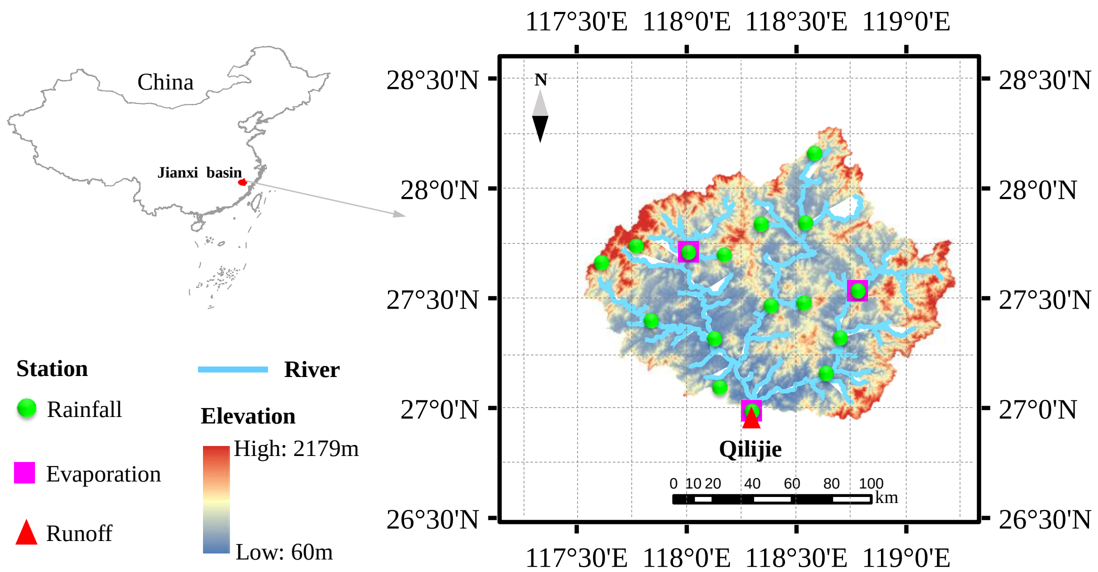

Fig. 3 Locations of the Jianxi River catchment and hydro-meteorological (rainfall, evaporation and runoff) gauge stations.

3.2 Materials

This study collected the data associated with flood events during the flood season (April 1st–September 30th), including the 3-hour (time step) rainfall data of 16 gauge stations, the 3-hour evaporation of 3 gauge stations and the 3-hour streamflow data of the Qilijie hydrological station from 2009 to 2015, and the 3-hour forecast rainfall of ERA-Interim datasets of the European Centre for Medium-Range Weather Forecasts (ECMWF) from 2014 to 2015. The spatial resolution of each grid in ERA-Interim is 0.125°E×0.125°N, while 32 grid points are required to fully cover the Jianxi River catchment (Fig. 3). More details on the ERA-Interim datasets of the ECMWF can be found in Berrisford et al. (2011). The areal mean value of observed rainfall data was calculated by using the datasets of 16 gauge stations colored in green (Fig. 3), the areal mean value of observed evaporation data was calculated by using the datasets of 3 gauge stations colored in purple (Fig. 3), and the areal mean value of forecast rainfall data was calculated by using the datasets of 32 grid points through the Thiessen polygon method (Thiessen, 1911).

A total of 20448 (=8 time step × 365 days × 6 non-leap years + 8 time step × 366 days × 1 leap year) time series values were divided into three datasets for model training (8760 from 2009-2011), validation (5848 from 2012-2013), and testing (5840 from 2014-2015). The training dataset was applied to adjusting the parameters (weights and intercepts) of the XAJ model (the MCQRNN model). The validation dataset was applied to checking whether a prediction model is overfitted or undertrained. The test dataset was applied to evaluating the model prediction accuracy.

3.3 Model construction

A 12-hour forecast horizon has been considered in this study since the longest transform time in the rainfall-runoff processes is 12 hours in this catchment. Besides, the current forecast horizon at the Qilijie station is only 6 hours, inducing the demand for improving the lead time and forecast accuracy. In the light of the observed 3-hour rainfall-runoff datasets of the Jianxi River catchment, the model output was set to be *t*+1 (Horizon=3 hours) up to *t*+4 (Horizon=12 hours) time-step-ahead streamflow, where *t* is the current time.

As aforementioned, the GA was performed to determine the optimal hyperparameters in association with MCQRNN models at four horizons. The optimal values of the maximal iteration (I_{max}), the number of nodes (J), the regulation coefficients ($\lambda^{(h)}$ and $\lambda^{(o)}$) and the learning rate (η) were 300, 12, 0.006, 0.004 and 0.009, respectively, while the weight and intercept parameters optimized by the gradient-based nonlinear optimization algorithm within the QRNN package were adopted in this study.

This study adopted the Partial Mutual Information and Partial Weights (PMI-PW) method (Sharma et al., 2016) to quantify the contribution of input combination to model performance (Table 1). To estimate the input variables weights for constructing the MCQRNN and XAJ-MCQRNN models, the following framework was adopted. Apart from the observed streamflow and rainfall data (traced back to the previous 12 hours), the rainfall of the ERA-Interim during horizon *t*+1 up to *t*+4 was used to train, validate and test the former, while the forecast streamflow of the XAJ model during horizon *t*+1 up to *t*+4 was used to train, validate and test the latter.

Table 1 Weights of input variables for constructing models by the Partial Mutual Information and Partial Weights (PMI-PW) method

Variable	Input combination	Contribution			
		Horizon *t*+1	Horizon *t*+2	Horizon *t*+3	Horizon *t*+4

		MCQRNN	XAJ-MCQRNN	MCQRNN	XAJ-MCQRNN	MCQRNN	XAJ-MCQRNN	MCQRNN	XAJ-MCQRNN
Rainfall	P(*t*+4)							0.03	
	P(*t*+3)					0.04		0.06	
	P(*t*+2)			0.05		0.06		0.08	
	P(*t*+1)	0.05		0.06		0.09		0.10	
	P(*t*)	0.07	0.07	0.09	0.08	0.11	0.09	0.17	0.15
	P(*t*-1)	0.09	0.09	0.11	0.11	0.15	0.14	0.09	0.10
	P(*t*-2)	0.12	0.10	0.17	0.13	0.08	0.09	0.06	0.06
	P(*t*-3)	0.18	0.14	0.08	0.09	0.04	0.04	0.02	0.05
Forecast streamflow of XAJ model	$\hat{Q}_{XAJ}(t+4)$								0.25
	$\hat{Q}_{XAJ}(t+3)$						0.23		0.19
	$\hat{Q}_{XAJ}(t+2)$				0.22		0.17		0.10
	$\hat{Q}_{XAJ}(t+1)$		0.20		0.14		0.11		0.06
Observed streamflow	Q(*t*)	0.20	0.18	0.21	0.11	0.24	0.09	0.25	0.04
	Q(*t*-1)	0.15	0.09	0.13	0.08	0.12	0.04	0.09	
	Q(*t*-2)	0.09	0.07	0.07	0.04	0.05		0.04	
	Q(*t*-3)	0.05	0.06	0.03		0.02		0.01	
Sum		1	1	1	1	1	1	1	1

4 Results

To improve the predictability and reliability of FPD forecasts, this study intends to explore and evaluate the accuracy of the XAJ model coupled with the MCQRNN on short-term (12 hours) flood forecasts at different horizons. The results, findings and discussion were as follows.

4.1 Reliability of XAJ and MCQRNN models on point forecasts of flood processes

The point forecasts of floods under the circumstance of the quantile τ_k=0.5 (i.e. median forecasts) were specified to test the reliability of the constructed models. Table 2 summarizes the RMSE, LNSE and MI values of the XAJ, MCQRNN and XAJ-MCQRNN models in training, validation and testing stages at four horizons. It is apparent from this table that the XAJ-MCQRNN model performs better than the XAJ and MCQRNN models in the three stages at four horizons. Take the testing stage for example, the LNSE and MI values of the XAJ-MCQRNN model exceed 0.90 and 22 respectively while the RMSE values are less than 140 m^3/s.

Two main results are acquired from the Table 2. Firstly, the comparison of the XAJ-MCQRNN and XAJ models in the testing stage reveals that there are significant differences in the improvement rates of RMSE and NSE values at four horizons. The improvement rates of RMSE (LNSE & MI) for median flood forecasting reach 4.7% (9.0% & 4.3%) and 5.2% (17.0% & 6.1%) at horizons t+1 and t+2, respectively, while they increased to 6.6% (22.2% & 8.8%) and 8.1% (28.4% & 11.2%) at horizons t+3 and t+4, respectively. Secondly, the comparison of the XAJ-MCQRNN and MCQRNN models in the testing stage points out that the former can provide stable forecast results after being supplied with point forecasts by the XAJ model either at shorter horizons (t+1 and t+2) or at longer horizons (t+3 and t+4). That is to say, the XAJ-MCQRNN model can effectively extract the dependence between rainfall and runoff processes even in the case of the12-hour forecast horizon. The fluctuations of forecast rainfalls are stronger at long horizons (t+3 and t+4) than at short horizons (t+1 and t+2). Higher fluctuations of forecast rainfall data are easy to induce overfitting problems, which would demand more complex models to mimic the relationship between rainfall and runoff. This is considered as a driver to improve the forecast of the XAJ-MCQRNN model.

Table 2 Results of Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE) and Mutual Information (MI) values acquired from the XAJ, MCQRNN and XAJ-MCQRNN models in training, validation and testing stages

Stage	Horizon	RMSE (m^3/s)	LNSE	MI

		XAJ	MCQRNN	XAJ-MCQRNN	XAJ	MCQRNN	XAJ-MCQRNN	XAJ	MCQRNN	XAJ-MCQRNN
Training	*t*+1	91	71	68	3.22	3.91	3.91	1.75	1.87	1.94
	t+2		89	84		3.51	3.51		1.79	1.90
	t+3		115	106		3.00	3.22		1.68	1.85
	t+4		137	124		2.53	2.81		1.54	1.78
Validation	*t*+1	102	77	73	3.00	3.51	3.51	1.72	1.85	1.93
	t+2		94	88		3.22	3.51		1.78	1.91
	t+3		121	111		2.66	3.00		1.69	1.84
	t+4		146	130		2.41	2.81		1.60	1.77
Testing	*t*+1	84	90	80 (4.7)∗	3.22	3.22	3.51 (9.0)	1.85	1.85	1.93 (4.3)
	t+2	96	103	91 (5.2)	3.00	3.00	3.51 (17.0)	1.79	1.74	1.90 (6.1)
	t+3	122	131	114 (6.6)	2.41	2.30	2.81 (22.2)	1.70	1.67	1.85 (8.8)
	t+4	149	168	137 (8.1)	2.12	1.97	2.53 (28.4)	1.61	1.56	1.79 (11.2)

∗ The number in the bracket is the improvement rate of the XAJ-MCQRNN model over the XAJ model.

To further evaluate the forecast accuracy of MCQRNN models for making median forecasts of floods, the regression analysis of the observed and forecast streamflow data is presented in Fig. 4. It is too hard to judge the difference in model quality when the correlation between observations and forecasts is high. However, this is different for the case of low correlation, where the gap between the regression and 45-degree lines is large. A large gap denotes a significant biased-prediction. From the standpoint of correlation (Fig. 4), the XAJ-MCQRNN model can more adequately alleviate the biased-prediction phenomenon (under-prediction in our case) because of the larger values of the coefficient of determination R^2 as well as the smaller gaps between the regression and 45-degree lines, compared to the MCQRNN model. As expected, there are higher correlations between observations and forecasts at horizons *t*+1 and *t*+2, whereas larger improvements of the correlation between observations and forecasts occur at horizons *t*+3 and *t*+4.

The hybrid model still consistently underestimates most flood processes (Fig. 4).

The back-fitting algorithm with the autoregressive strategy (Zhang et al., 2018) was performed to correct the systematic bias of point forecasts of the hybrid model. After implementing bias-correction, Table 3 summarizes the Kolmogorov-Smirnov (KS) test results from the standpoint of density distribution. Due to the high variation of flood datasets (streamflow value$\geqslant$1000 m^3/s), the observations at each horizon are divided into three parts (Part I: 1000 $m^3/s \leq$ value < 3000 m^3/s; Part II: 3000 $m^3/s \leq$ value < 9000 m^3/s; and Part III: value$\geqslant$9000 m^3/s) to effectively examine the predictability of the proposed model at various flood magnitudes. The larger KS test indicator values suggest smaller difference between observed and forecast distributions and better model performance. Interestingly, the XAJ-MCQRNN model creates a pretty good similarity between observed and forecast distributions of the three parts at all horizons whereas the MCQRNN model only performs well when flood magnitudes are small (Part I). From the table, we can see that the probability density distribution of the forecasts acquired from the XAJ-MCQRNN model is more similar in shape to that of observations at each horizon. In other words, the values of the forecasts obtained from the XAJ-MCQRNN model in three parts are closer to those of observations, in comparison to the forecasts obtained from the MCQRNN model. The results demonstrate that the XAJ-MCQRNN model can not only sufficiently forecast the probability density distribution of flood data but also accurately forecast different flood magnitudes at four horizons.

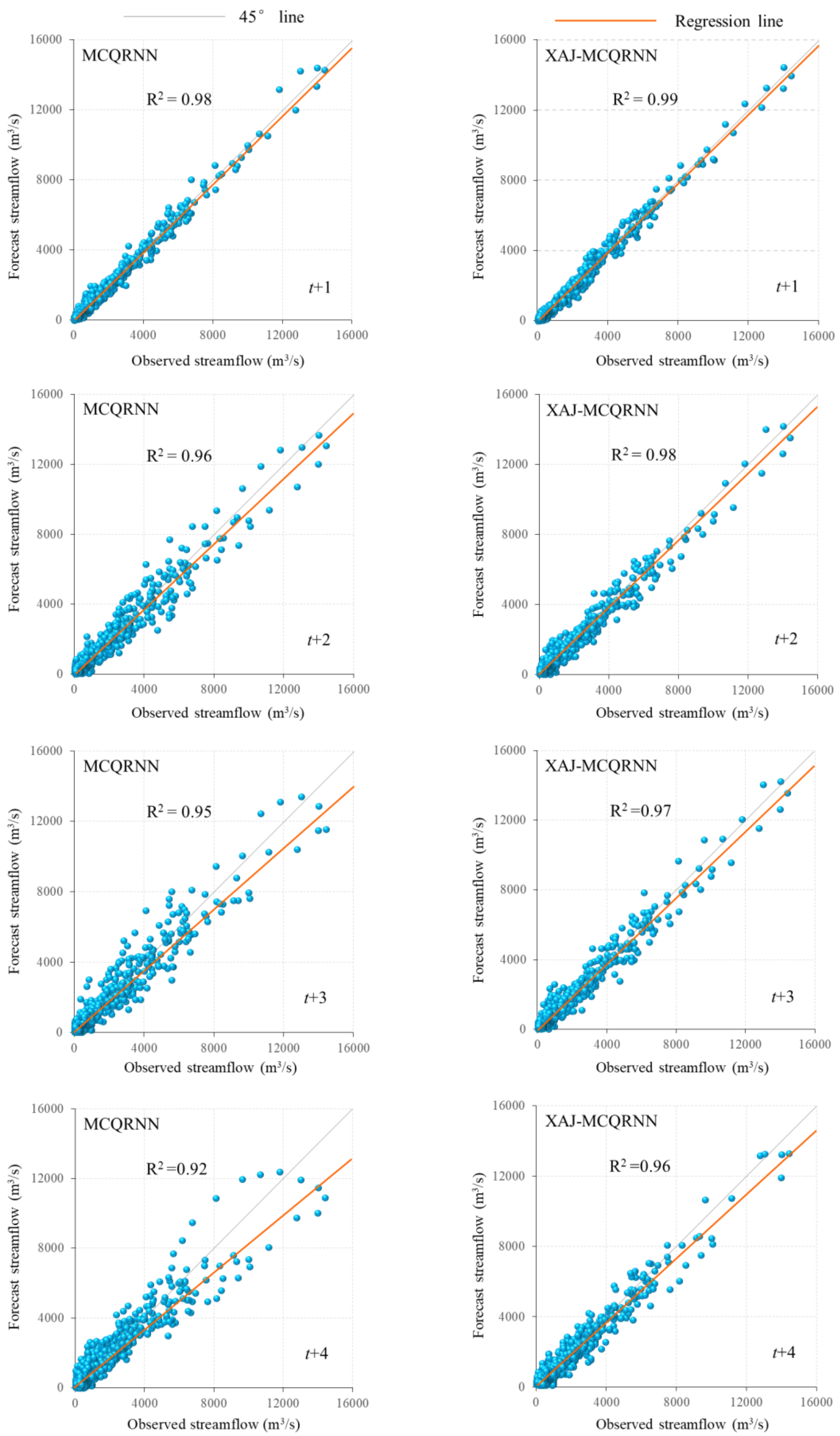

Fig. 4 Regression analysis between the observed streamflow data and the median forecasts of streamflow data in the testing stage at four horizons.

Table 3 Kolmogorov Smirnov (KS) test results of the two distributions of observations and forecasts in three parts for each given horizon at the testing stage

Model	Horizon	KS test indicator*		
		Part I: 1000 $m^3/s \leqslant$ value<3000 m^3/s	Part II: 3000 $m^3/s \leqslant$ value<9000 m^3/s	Part III: value $\geqslant$ 9000 m^3/s
MCQRNN	*t*+1	0.03	0.06	0.08
	t+2	0.03	0.07	0.10
	t+3	0.04	0.10	0.13
	t+4	0.05	0.13	0.15
XAJ-MCQRNN	*t*+1	0.01	0.03	0.06
	t+2	0.01	0.04	0.06
	t+3	0.02	0.06	0.08
	t+4	0.03	0.07	0.09

*****The KS test is performed at a significance level of 0.05. The null hypothesis states that the probability density distribution of forecasts is similar in shape to that of observations. If the value of the KS test indicator is smaller than the value of D(n, 0.05) (=0.05, 0.09 and 0.11 for the Part I, II and III, respectively), the null hypothesis would not be rejected.

In brief, the point (i.e., median) forecasts of floods made by the XAJ-MCQRNN model offered solid evidence of good model performance and favorable stability in multi-step-ahead flood forecasting. The next subsection is concerned with the comparison of the MCQRNN and XAJ-MCQRNN models for making interval forecasts of floods.

4.2 Reliability of MCQRNN models on interval forecasts of flood processes

The interval forecasts of floods for the case of the quantiles ($0.01 \leq \tau_k \leq 0.99$, with an increment of 0.01) were employed to test the reliability of the two models (MCQRNN and XAJ-MCQRNN). The back-fitting algorithm with the autoregressive strategy (Zhang et al., 2018) was also executed to correct the systematic bias of interval forecasts of the hybrid model. Table 4 provides the RB and CR values of the MCQRNN and XAJ-MCQRNN models in three stages at four horizons. The results indicate that the XAJ-MCQRNN model acquires pretty good forecast accuracy at all the horizons whereas the MCQRNN model only performs well at shorter horizons of *t*+1 and *t*+2 (CR is higher than 92%, but RB is lower than 12%). For 12-hour forecasts, the XAJ-MCQRNN model can enhance the CR value by 1.5% up to 5.9% and decrease the RB values by 8.8% up to 25.7% in the testing stages, as compared with the MCQRNN model. This means the XAJ-MCQRNN model can not only significantly improve probabilistic forecast accuracy in terms of a narrow prediction interval (as represented by CR values) but also adequately eliminate the impacts of flood magnitudes on the bandwidth of the prediction intervals (as represented by RB values). The 12-hour forecast accuracy of the XAJ-MCQRNN model is superior to that of the MCQRNN model since the XAJ-MCQRNN model utilizes the streamflow forecasts made by the XAJ model to reduce the uncertainties of the rainfall forecasts from the ECMWF, whereas the MCQRNN model adopts the rainfall forecasts of the ECMWF to make multi-step-ahead flood forecasts in testing stages.

Table 4 Results of Coverage Ratio (CR) and Relative Bandwidth (RB) values acquired from the MCQRNN and XAJ-MCQRNN models in training, validation and testing stages at four horizons

Stage	Horizon	RB (%)		CR (%)	
		MCQRNN	XAJ-MCQRNN	MCQRNN	XAJ-MCQRNN
Training	*t*+1	4.91	4.49 (8.5)*	95.17	96.98 (1.9)
	t+2	8.62	7.77 (9.9)	93.83	95.99 (2.3)
	t+3	12.56	11.20 (10.8)	90.60	94.04 (3.8)
	t+4	17.24	15.05 (12.7)	87.72	91.67 (4.5)
Validation	*t*+1	6.55	6.02 (8.1)	94.93	96.45 (1.6)
	t+2	10.10	9.15 (9.4)	93.26	95.31 (2.2)
	t+3	14.81	13.34 (9.9)	90.09	93.33 (3.6)
	t+4	19.37	16.62 (14.2)	86.36	90.07 (4.3)
Testing	*t*+1	8.16	7.44 (8.8)	94.32	95.73 (1.5)
	t+2	11.65	10.01 (14.1)	92.79	94.74 (2.1)

	t+3	16.00	13.07 (18.3)	89.56	92.87 (3.7)
	t+4	23.33	17.33 (25.7)	84.43	89.41 (5.9)

∗ The number in the bracket is the improvement rate of the XAJ-MCQRNN model over the MCQRNN model.

Moreover, to clearly determine the predictabilities of the MCQRNN and XAJ-MCQRNN models, the three flood events: (a) high flood magnitude with total precipitation of 134 mm and maximal streamflow of 14400 m^3/s (maximum streamflow with a return period of 45a); (b) medium flood magnitude with total rainfall of 60 mm and maximal streamflow of 10100 m^3/s (the return period of this peak is 20a); and (c) low flood magnitude with total rainfall of 26 mm and maximal streamflow of 5200 m^3/s (the return period of this peak is 10a), were selected to examine both models by assessing whether the observed streamflow data fall within the 90% prediction interval, i.e. [5% quantile, 95% quantile], at horizon t+4 in testing stages, as illustrated in Fig. 5. Three interesting results are acquired. First, the 90% prediction intervals made by the XAJ-MCQRNN model can cover the majority of the observed streamflow data, in comparison to those of the MCQRNN model. Second, the XAJ-MCQRNN model can create a predictive distribution narrower than that of the MCQRNN model in each flood event. Third, the XAJ-MCQRNN model can largely mitigate the biased-prediction phenomenon (under-prediction in our case). The prediction interval aims to promote the sharpness of the predictive distributions as much as possible, where sharpness is used to represent the concentration of the predictive distributions. Concerning short-term forecasts, the XAJ-MCQRNN model can raise the concentration of the predictive distributions while obtaining high coverage of the observed streamflow data in all cases.

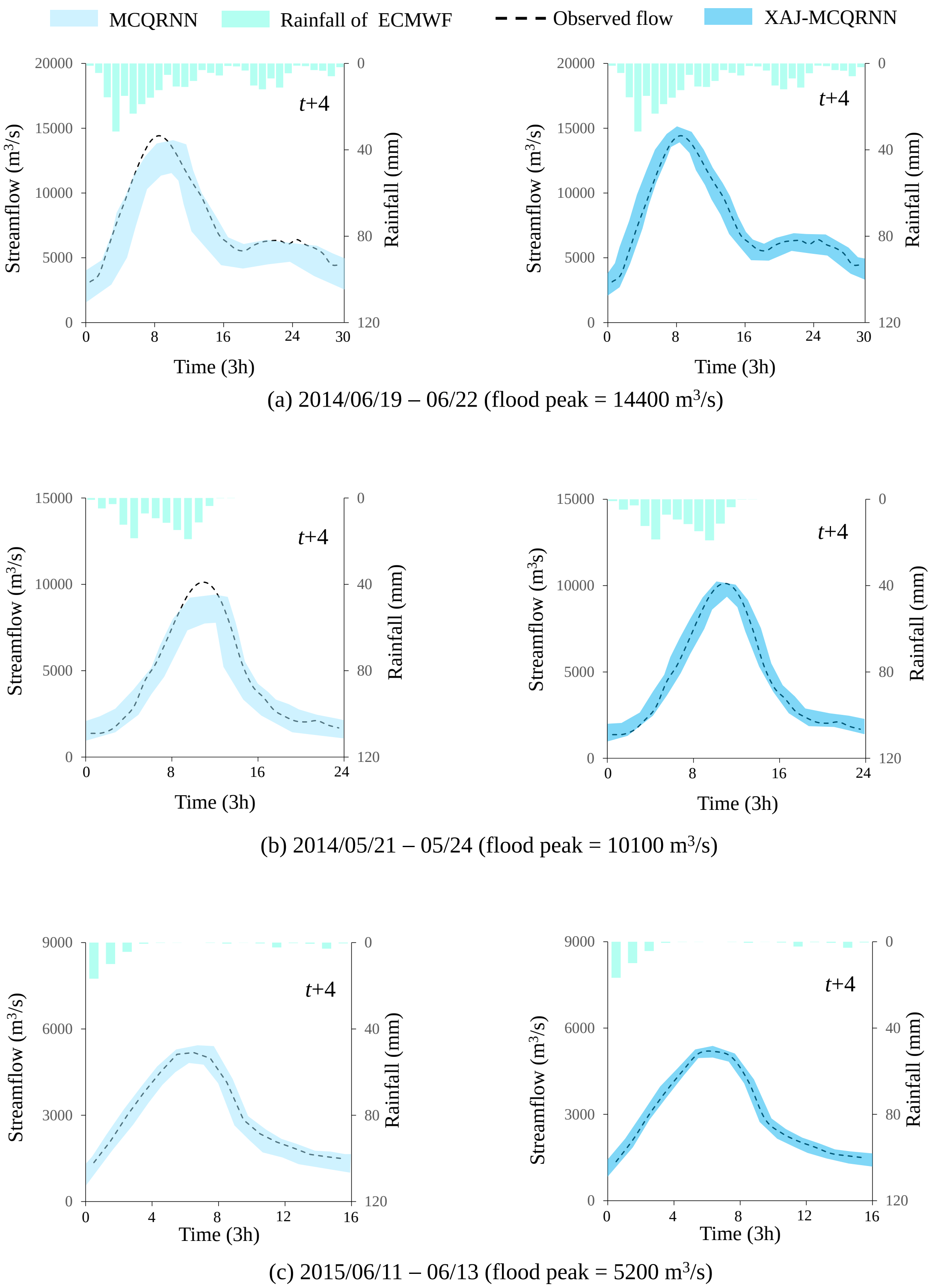

Fig. 5 Interval forecast results of three flood events at horizon *t*+4 in the testing stages, given a 90% prediction interval of [5% quantile, 95% quantile]. The rainfall forecasts during horizons *t*+1 up to *t*+4 were collected from the ERA-Interim datasets of the ECMWF. Left panel for MCQRNN, and right for XAJ-MCQRNN.

The density distribution (i.e., violin plot) is also adopted to evaluate the reliability of interval forecasts. Fig. 6 illustrates the predictive violin-plots of the flood peaks of the above-specified flood events at horizon t+4 in testing stages. In each violin plot, the probability density distribution of the forecasts made by the MCQRNN (XAJ-MCQRNN) model is located at the left (right) part, while the statistical boxplot of all forecasts is located at the central part. The violin plots point out that the difference between the forecasts acquired from MCQRNN and XAJ-MCQRNN models displays a trend climbing from the low, the medium to the high values of flood peaks at horizon t+4. The bandwidths of the prediction intervals are notably larger in the medium-high flood peaks than in the low flood peak due to the larger fluctuations of flood magnitudes in the former. Interestingly, the range of kernel density functions corresponding to the XAJ-MCQRNN model is significantly narrower than that of the MCQRNN model while the central locations (orange dashed lines) of kernel density functions related to the XAJ-MCQRNN model are closer to the observed flood peaks (purple points) in all cases. That means the XAJ-MCQRNN model produces higher sharpness and predictions with smaller bias than the MCQRNN model. Briefly, the XAJ-MCQRNN model not only can produce more accurate and reliable probability density forecasts but also can alleviate the biased-prediction phenomenon associated with the extreme flood events through integrating the data-driven and conceptual-based mechanisms for rainfall-runoff modelling.

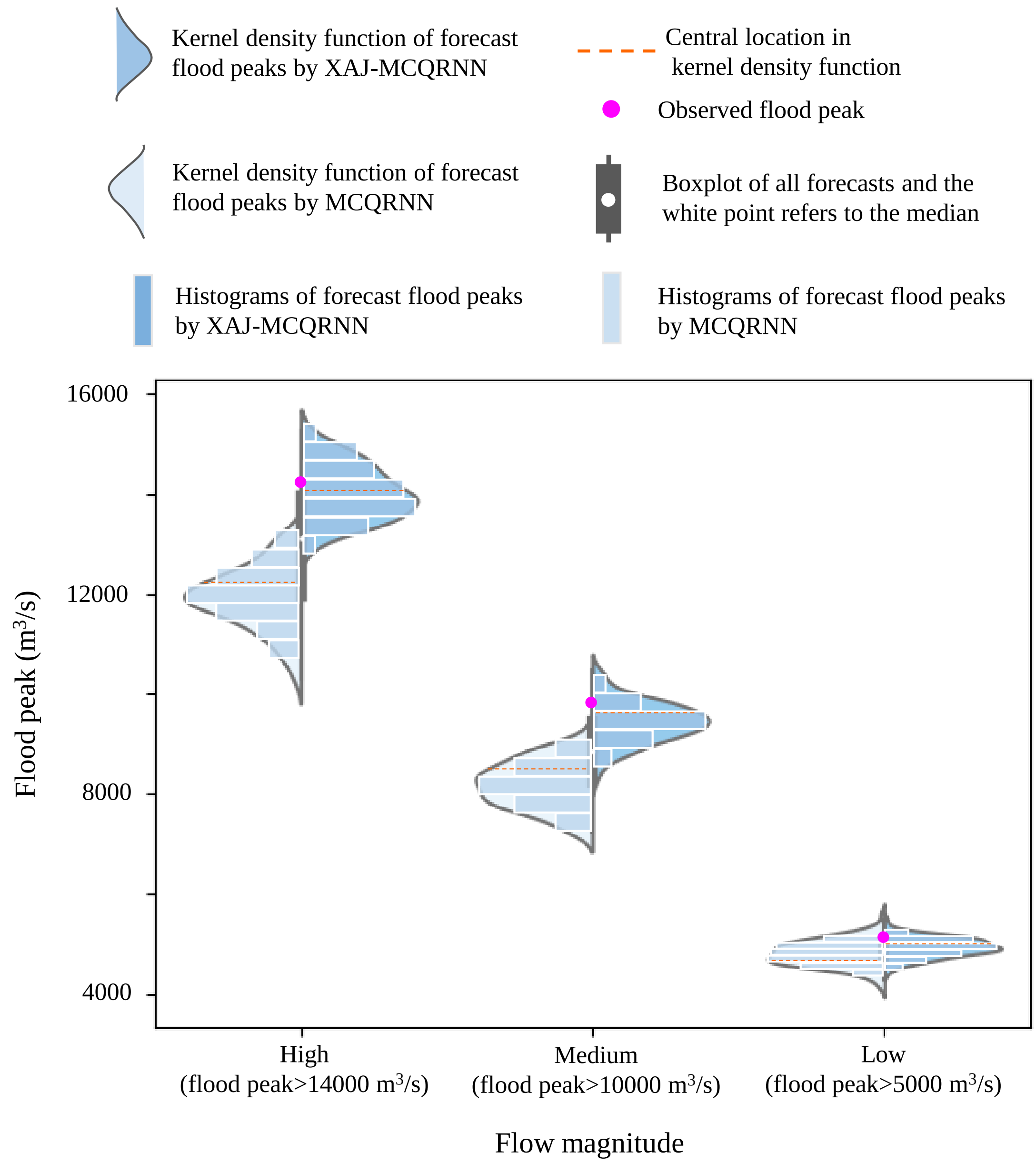

Fig. 6 Violin plots for evaluating the uncertainty of MCQRNN and XAJ-MCQRNN model forecasts on flood peaks of high, medium and low flood magnitudes at horizon *t*+4 in the testing stages.

5 Conclusions and discussion

The demand for the hybridization of machine learning and conceptual models is motivated by real-world applications in the best interests of enhancing the accuracy and reliability of short-term FPD forecasting. This study proposed a hybrid rainfall-runoff model (i.e., XAJ-MCQRNN) that integrated the Xinanjiang (XAJ) conceptual model and the MCQRNN to make short-term FPD forecasts. Its capability of efficient learning and accurate forecasting is tested and verified with the long-term (2009-2015) flood events at the Jianxi River catchment in China. The MCQRNN model is constructed and used for comparison purpose.

The results of median point forecasts on flood processes reveal that the proposed XAJ-MCQRNN model performs distinguishably better than the MCQRNN model in short-term (12 hours) flood forecasting with higher NSE and lower RMSE values as well as a better goodness-of-fit to observed flood processes. The results of interval forecasts of flood processes further demonstrate that the XAJ-MCQRNN model is superior to the MCQRNN model at four horizons. The XAJ-MCQRNN model has smaller RB, larger CR values, and higher sharpness of the predictive distribution. These achievements provide solid evidence that the XAJ-MCQRNN model can make considerably more reliable and accurate point and interval forecasts of flood processes at long forecast horizons and can more effectively overcome the phenomena of overfitting and time-shift than the MCQRNN model. Moreover, the reason that the XAJ-MCQRNN model is successful in obtaining superb probability density forecast results is attributed to the pivotal strategy that the XAJ conceptual model supplements the point forecasts of flood processes through training the MCQRNN model to alleviate predictive uncertainty.

Therefore, the XAJ-MCQRNN by virtue of data-driven and conceptual-based hydrological models can provide forecasting and early warning of flood processes, thus reducing flood risks associated with the extreme rainstorm phenomenon. This study is a good demonstration of the hybrid approach which unites two different groups of scientists including data scientists and disciplinary scientists (e.g., hydrologists). To use the strength of various point forecasts meanwhile alleviating their weaknesses, model combinations in the form of hybridization (e.g., Zhang et al., 2018; Lin et al., 2020; Hosseiny et al. 2020; Konapala et al., 2020; Kao et al., 2021) have been applied to making short-term flood point forecasts for basins with humid subtropical climate. This is the first study to apply conceptual models with artificial neural networks for flood probability density (FDP) forecasting for such basins. Notwithstanding these promising achievements, further research can be conducted to investigate the far-reaching effects of short-term and even long-term flood probability forecasting and warning on flood prevention and water resources management in humid, semi-humid and arid areas by combining more hybrid modelling techniques with medium-long term weather forecast products.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2021YFC3200303 and 2018YFC0407904), the National Natural Science Foundation of China (No. 51538173 and U20A20317), and the Research Council of Norway (FRINATEK Project 274310). The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that are greatly contributive to enriching the manuscript.

References

Bandai, T., & Ghezzehei, T. A. (2020). Physics-informed neural networks with monotonicity constraints for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements. Water Resources Research.

Berrisford, P., Kållberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A. J., Sato, H. (2011). Atmospheric conservation properties in ERA-Interim. Quarterly Journal of the Royal Meteorological Society, 137(659), 1381-1399.

Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. Stochastic Environmental Research and Risk Assessment, 32(11), 3207–3225.

Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Computers & Geosciences, 37(9), 1277–1284.

Chadalawada, J., Herath, H. M. V. V., & Babovic, V. (2020). Hydrologically informed machine learning for rainfall-runoff modeling: A Genetic Programming - Based Toolkit for automatic model induction. Water Resources Research, 56(4).

Chen, C. (2007). A finite smoothing algorithm for quantile regression. Journal of Computational and Graphical Statistics, 16(1), 136–164.

Cui, Z., Zhou, Y., Guo, S., Wang, J., Ba, H., & He, S. (2021). A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting. Hydrology Research. DOI: 10.2166/NH.2021.016.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology, 158(3), 265–284.

Farfán, J. F., Palacios, K., Ulloa, J., & Avilés, A. (2020). A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds. Journal of Hydrology: Regional Studies, 27, 100652.

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. Water Resources Research, 56(9).

Ghaseminejad, A., & Uddameri, V. (2020). Physics-inspired integrated space-time Artificial Neural Networks for regional groundwater flow modeling. Hydrology and Earth System Sciences, 24(12), 5759–5779.

Giuliani, M., Zaniolo, M., Castelletti, A., Davoli, G., & Block, P. (2019). Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations. Water Resources Research, 55(11), 9133–9147.

Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in Genetic Algorithms. In Foundations of Genetic Algorithms, 1: 69–93.

Gu, H., Xu, Y.-P., Ma, D., Xie, J., Liu, L., & Bai, Z. (2020). A surrogate model for the Variable Infiltration Capacity model using deep learning artificial neural network.

Journal of Hydrology, 588, 125019.

Hitokoto, M., & Sakuraba, M. (2020). Hybrid deep neural network and distributed rainfall-runoff model for real-time river stage prediction. Journal of JSCE, 8(1), 46–58.

Hosseiny, H., Nazari, F., Smith, V., & Nataraj, C. (2020). A framework for modeling flood depth using a hybrid of hydraulics and machine learning. Scientific Reports, 10(1), 8222.

Huber, P. J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics, 35(1), 492–518.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., & Maier, H. R. (2016). A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. Journal of Hydrology, 540, 623–640.

Hunter, J. M., Maier, H. R., Gibbs, M. S., Foale, E. R., Grosvenor, N. A., Harders, N. P., & Kikuchi-Miller, T. C. (2018). Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems. Hydrology and Earth System Sciences, 22(5), 2987–3006.

Kan, G., Liang, K., Yu, H., Sun, B., Ding, L., Li, J., Shen, C. (2020). Hybrid machine learning hydrological model for flood forecast purpose. Open Geosciences, 12(1), 813–820.

Kao, I.-F., Zhou, Y., Chang, L.-C., & Chang, F.-J. (2020). Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. Journal of Hydrology, 583, 124631.

Kao, I.-F., Liou, J.-Y., Lee, M.-H., & Chang, F.-J. (2021). Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts. Journal of Hydrology, 126371.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In ICLR 2015 : International Conference on Learning Representations 2015.

Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environmental Research Letters, 15(10), 104022.

Kumanlioglu, A. A., & Fistikoglu, O. (2019). Performance enhancement of a conceptual hydrological model by integrating artificial intelligence. Journal of Hydrologic Engineering, 24(11), 4019047.

Kurian, C., Sudheer, K. P., Vema, V. K., & Sahoo, D. (2020). Effective flood forecasting at higher lead times through hybrid modelling framework. Journal of Hydrology, 587, 124945.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness - of - fit” Measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233–241.

Li, H., Gao, H., Zhou, Y., Xu, C.-Y., Ortega, Z. M. R., & Sælthun, N. R. (2020). Usage of SIMWE model to model urban overland flood: a case study in Oslo. Hydrology Research, 51(2), 366–380.

Li, J., Yuan, D., Liu, J., Jiang, Y., Chen, Y., Hsu, K. L., & Sorooshian, S. (2019). Predicting floods in a large karst river basin by coupling PERSIANN-CCS QPEs with a physically based distributed hydrological model. Hydrology and Earth System Sciences, 23(3), 1505–1532.

Li, W., Kiaghadi, A., & Dawson, C. (2020). Exploring the best sequence LSTM modeling architecture for flood prediction. Neural Computing and Applications, 1–10.

Li, X., & Willems, P. (2020). A hybrid model for fast and probabilistic urban pluvial flood prediction. Water Resources Research, 56(6).

Lin, K., Sheng, S., Zhou, Y., Liu, F., Li, Z., Chen, H., Guo, S. (2020). The exploration of a Temporal Convolutional Network combined with Encoder-Decoder framework for runoff forecasting. Hydrology Research, 51(5), 1136–1149.

Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y., & Song, T. (2018). A review of advances in China's flash flood early-warning system. Natural Hazards, 92(2), 619–634.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282 - 290.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2020). What role does hydrological science play in the age of machine learning. Water Resources Research.

Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. Water Resources Research, 56(5).

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27 (3), 379–423.

Sharma, A., Mehrotra, R., Li, J., & Jha, S. (2016). A programming tool for nonparametric system prediction using Partial Informational Correlation and Partial Weights. Environmental Modelling and Software, 83, 271–275.

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

Sood, A., & Smakhtin, V. U. (2015). Global hydrological models: a review. Hydrological Sciences Journal, 60(4), 549–565.

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: Can we learn from mismatch? Water Resources Research, 55(2), 1179–1195.

Takeuchi, K., Chavoshian, A., & Simonovic, S. P. (2018). Floods: From Risk to Opportunity. Journal of Flood Risk Management, 11(4).

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. Hydrology and Earth System Sciences, 24(5), 2527-2544.

Thiessen, A. H. (1911). Precipitation averages for large areas. Monthly Weather Review, 39(7), 1082–1089.

Tian, Y., Xu, Y.-P., Yang, Z., Wang, G., & Zhu, Q. (2018). Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. Water, 10(11), 1655.

Taylor, J. W. (2000). A quantile regression neural network approach to estimating the

conditional density of multiperiod returns. Journal of Forecasting, 19(4), 299–311.
Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resources Research, 56(1).
Xie, S., Wu, W., Mooser, S., Wang, Q. J., Nathan, R., & Huang, Y. (2021). Artificial neural network based hybrid modeling approach for flood inundation modeling. Journal of Hydrology, 592, 125605.
Xiong, L., & O'Connor, K. M. (2008). An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modeling. Journal of Hydrology, 349(1), 115–124.
Xu, Q., Deng, K., Jiang, C., Sun, F., & Huang, X. (2017). Composite quantile regression neural network with applications. Expert Systems With Applications, 76, 129–139.
Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., et al. (2019). Evaluation and machine learning improvement of global hydrological model-based flood simulations. Environmental Research Letters, 14(11), 114027.
Yoganath Adikari & Junichi Yoshitani (2009). Global Trends in Water-Related Disasters: An Insight for Policymakers. International Centre for Water Hazard and Risk Management (UNESCO), Tsukuba, Japan.
Young, C.-C., Liu, W.-C., & Wu, M.-C. (2017). A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Applied Soft Computing, 53, 205–216.
Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community. Water Resources Research, 56(10).
Zhang, X., Liu, P., Cheng, L., Liu, Z., Zhao, Y. (2018). A back-fitting algorithm to improve real-time flood forecasting. Journal of Hydrology 562, 140-150.
Zhao, R. (1992). The Xinanjiang model applied in China. Journal of Hydrology, 135(1), 371–381.
Zhou, Y., Guo, S., Xu, C.-Y., Chang, F.-J., & Yin, J. (2020). Improving the reliability of probabilistic multi-step-ahead flood forecasting by fusing unscented Kalman filter with recurrent neural network. Water, 12(2), 578.
Zhou, Y., Guo, S., & Chang, F.-J. (2019). Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. Journal of Hydrology, 570, 343–355.

Abstract

Making accurate and reliable probability density forecasts of flood processes is fundamentally challenging for machine learning techniques, especially when prediction targets are outside the range of training data. Conceptual hydrological models can reduce rainfall-runoff modelling errors with efficient quasi-physical mechanisms. The Monotone Composite Quantile Regression Neural Network (MCQRNN) is used for the first time to make probability density forecasts of flood processes and serves as a benchmark model, whereas it confronts the drawbacks of overfitting and biased-prediction. Here we propose an integrated model (i.e. XAJ-MCQRNN) that incorporates Xinanjiang conceptual model (XAJ) and MCQRNN to overcome the phenomena of error propagation and accumulation encountered in multi-step-ahead flood probability density forecasts. We consider flood forecasts as a function of rainfall factors and runoff data. The models are evaluated by long-term (2009-2015) 3-hour streamflow series of the Jianxi River catchment in China and rainfall products of the European Centre for Medium-Range Weather Forecasts. Results demonstrated that the proposed XAJ-MCQRNN model can not only outperform the MCQRNN model but also prominently enhance the accuracy and reliability of multi-step-ahead probability density forecasts of flood process. Regarding short-term forecasts in testing stages at four horizons, the XAJ-MCQRNN model achieved higher Nash-Sutcliffe Efficiency but lower Root Mean Square Error values, while improving Coverage Ratio and Relative Bandwidth values in comparison to the MCQRNN model. Consequently, the improvement can benefit the mitigation of the impacts associated with uncertainties of extreme flood and rainfall events as well as promote the accuracy and reliability of flood forecasting and early warning.

Highlights (3 to 5 bullet points (maximum 85 characters including spaces per bullet point)

Highlights

- Machine learning assists hybrid model to promote flood forecasting and early warning
- Hybridizing MCQRNN with XAJ model for flood probability density forecasting
- XAJ-MCQRNN conquers overfitting and biased-prediction bottlenecks
- XAJ-MCQRNN improves accuracy and reliability of flood probability density forecasts

Credit Author Statement

Conceptualization: Yanlai Zhou; Shenglian Guo; Chong-Yu Xu

Data curation: Zhen Cui; Kangling Lin; Sheng Sheng

Formal analysis: Yanlai Zhou; Kangling Lin; Sheng Sheng

Funding acquisition: Hua Chen; Shenglian Guo; Chong-Yu Xu

Investigation: Zhen Cui; Kangling Lin; Sheng Sheng

Methodology: Yanlai Zhou; Shenglian Guo; Chong-Yu Xu

Project administration: Shenglian Guo; Chong-Yu Xu

Resources: Yanlai Zhou; Shenglian Guo; Chong-Yu Xu

Software: Yanlai Zhou; Zhen Cui; Hua Chen

Validation: Zhen Cui; Kangling Lin; Hua Chen

Visualization: Zhen Cui; Kangling Lin

Writing - original draft: Yanlai Zhou; Chong-Yu Xu

Writing & editing: Shenglian Guo; Chong-Yu Xu

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: