https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Discussions

Incremental learning for rainfall-runoff simulation on deep neural networks

Zeqiang Chen[1], Jiashun Li[1], Changjiang Xiao[2], Nengcheng Chen[1]

[1]National Engineering Research Center for Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China
[2]College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

Correspondence to: Changjiang Xiao (cjxiao@tongji.edu.cn)

Abstract. Rainfall-runoff simulation based on deep learning always costs plenty of time for training with large datasets. This may affect quick decision making in some flood emergency decision-making situations. To address this issue, this study proposes an incremental learning method to accelerate rainfall-runoff simulation with deep learning model. The method consists of two components, regular training and incremental operation. In regular training phase, the model is regularly trained using historical data. In the incremental operation phase, the method selects representative samples from historical data with distribution estimation metrics and time series similarity metrics, then updates the regularly trained model with the sampled data and recent data in case of emergency. The proposed method was tested using ten hydrological observation stations in the Yangtze River and Han River drainage basin, with three different modified Recurrent Neural Networks. The results show that the incremental learning method achieves a training efficiency acceleration of over 4 times, with only a little increase in percentage error and decrease in Nash-Sutcliffe efficiency coefficient. The results also illustrate the robustness of this method for different models in different places, as well as during continuous incremental scenarios. The findings indicate that the incremental learning method has great potential applications in rapid rainfall-runoff simulation for flood emergency decision-making.

1 Introduction

Rainfall-runoff simulation plays an important role in flood emergency decision-making. With the development of artificial intelligence and big data technologies, Deep Neural Networks (DNNs) have become prevalent in rainfall-runoff simulation for their powerful nonlinear simulation ability and high-precision simulation results. Recurrent Neural Networks (RNNs) as typical DNN sequence models are good at dealing with time series data and widely used in rainfall-runoff simulation (Chen et al., 2021; Hu et al., 2018; Li et al., 2021; Wang et al., 2020; Xu et al., 2022; Yin et al., 2021). When utilizing deep learning models to simulate rainfall-runoff relationships, the requirement for extensive training data often leads to a significant consumption of time, which may hinder swift decision-making in urgent flood emergency situations. For instance, a deep learning model employed to simulate the rainfall-runoff relationships of several key stations in the Yangtze River Basin in China, with decades of data, would take several hours to complete. Meanwhile, the government's flood control and disaster

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

reduction department of the Yangtze River necessitates the simulated results to conduct decision analysis and action on an hourly basis (Xie et al., 2018). Unfortunately, the sluggish pace of model simulation cannot meet the demands of real-time decision-making. This predicament is likely to become more pronounced as the utilization of training data becomes increasingly widespread. Therefore, a novel approach has emerged, which involves training the deep-learning model with an abundance of historical data during periods of low urgency and subsequently updating the trained model with a combination of historical sampled data and recent data in the event of an emergency. By doing so, the efficiency of rainfall-runoff simulation can be significantly improved to meet the requirements of timely decision-making. Incremental learning methods may offer a feasible solution to this problem.

Incremental learning, also called continual learning, refers to the ability of learning from stream of data which associate with different tasks or domains to solve the future problem with historical experience (Aljundi, 2019a). The main goal of incremental learning can be described as performing well both in historical tasks. Replay methods refer to re-introducing part of the samples from historical data to the learner when the new incremental samples come, the process of performing steps on the replayed samples is called rehearsal (Aljundi, 2019a). Hayes gives a detailed review of the derivation and idea of replay methods and illustrate how they are applied in deep learning (Hayes et al., 2021). Replay methods can be further classified into raw replay and generative relay methods referring to using the real historical samples and generating pseudo samples respectively. In raw replay methods, a buffer is usually set to store part of the historical data, which avoids frequent data selection when incremental data come while adds memory overhead. The Generative replay methods spring up as the generative adversarial network (GAN) plays an increasingly important role in deep learning and usually are applied in the situation that the previous data are difficult to acquire. The core problem of replay methods is the replayed data selection standards. Uniform random sampling, sampling with examples closet to class boundary, sampling with examples with highest entropy and sampling with the examples contributes to updating parameters the most are proposed standard and are verified with a good performance. The process of sampling can be described as an optimization problem on maximizing the chosen sample diversity with constrain (Aljundi, 2019b). The standards mentioned above can be seen as the objective of optimization. Similarity metrics also can be contributor to sampling. Overall sample selection methods for replay works well in some specific situation, while uniform random sampling still have a relatively stronger adaptivity to average problems. Apart from alleviating the catastrophic forgetting of the DNNs, replay methods can also facilitate better efficiency, which makes the network require fewer samples to learn new information. It is found that neural network required fewer training epochs to reach a target error on a new task after having learned other similar tasks. These findings between task similarity and network performance in terms of error and time have been studied (Davidson & Mozer, 2020), which can be further discussed considering inputs formulated as time series. Regardless of the specific form of replay, the replay methods usually combined with the regularization methods to improve the models' performance. Regularization methods refer to freezing parts of a model when training for successive incremental tasks, which can be interpreted as storing knowledge on how to solve different tasks in different parts of the model so that training on subsequent tasks does not interfere with this knowledge. Sometimes, the weights of the networks are not completely fixed but are normalized so that they do not change too much as the model is trained across

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Open Access
Discussions

different tasks. The elastic weight consolidation (EWC) uses this approach (Kirkpatrick et al., 2017). When the model is trained in a sequence of tasks, learning is slowed down by weights that are important to the previous task. Specifically, the learning of the important weights that are important to the previous tasks is slowed down. The reparameterization leads to a factorized rotation of the parameter space and makes the diagonal Fisher information matrix assumption more applicable. Moreover, other advanced methods are proposed on the calculation of the important weight. Memory Aware Synapses (MAS), structurally similar to EWC, adds an item which is obtained gradients of the squared L2-norm of the model's output in the loss function (Aljundi et al., 2018). MAS demonstrates a novel and more explanatory idea about the importance of parameters. Remanian Walk (RWalk) is a generalization of EWC and path integrals (Chaudhry et al., 2018), adopting a theoretical basis perspective based on KL divergence as well as several new metrics. A drawback of this method is the loss of the effective trainable ability of the model as increasing model parameters are regularized over time. SI (Zenke et al., 2017) shows state-of-the-art performance and comes closest to MAS. It estimates the importance weights in an online manner while training for a new task. Regularization method change to important parameters are penalized during training of incremental tasks. Above regularization methods concentrate on the penalty on loss function, given explanatory statistical standards into consideration, which perform relatively better when alleviating catastrophic forgetting. And it's common that the replay methods are usually combined with regularization methods to reach a more accurate in incremental learning tasks, among which ICARL is a typical method (Rebuffi et al., 2017). Besides, incremental learning methods usually focus on the picture data and are rarely applied in DNNs to analyze time series. Some researches use incremental learning on modified RNN, including constructing new hybrid incremental learning structure on LSTM (Sodhani et al., 2020) and enhancing short-term traffic prediction (Shao et al., 2021), which gives a diagram of incremental learning based on RNN. Yet, among the above incremental learning methods, partial data selection standards are not yet detailed discussed especially encountering time series data like rainfall-runoff data.

The theoretical description of time series similarity is proposed when addressing time series similarity/dissimilarity search (Agrawal et al., 1993), which is widely recognized in data mining field and rises different measurements coming into being. With different lengths and irregular sampling intervals, different distance measurement methods try to specify the similarity formulation by taking both time series representations and original raw time series into consideration. The prevalent methods include Euclidean distance, Longest Common Sub-Sequence (LCSS), Dynamic Time Warping (DTW) and so on (Paterson et al., 1994). Time series measurement methods can be roughly classified into methods based on time, methods based on shape and methods based on structure. The methods based on time consider similarity on each time step, and therefore it's proper to use the basic distance metrics like Euclidean distance measure. To reduce the calculation on the raw time series, Fourier transformer and Piecewise Aggregate Approximation (PAA) are usually used so that distance metrics are carried on transformed time series (Guo et al., 2010). The methods based on shape pay more attention on the series shaped similarity regardless of the points' features in time. Shape of the time series is a local characteristic of the complete time series, which tends to show the short-term variety rules of the time series. As for the methods based on structural similarity, modelling process such as Hidden Markov Models (HMM) (Smyth, 1996) are usually used so that measurement can be carried on the parameters of the model and time series. Structure of time series demonstrates the global characteristic, focusing on the long-

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

term variety rules of the time series. And the method based on dynamic programming perform the most effectively despite expensive time execution (the cost of comparing two time series is quadratic in the length of the time series) (Salvador & Chan, 2007), which construct a connection between the global and local features. Preliminary conclusion can be drawn from mentioned methods and related literatures is that the similarity/dissimilarity of time series depends on the target of utilizing the similarity, that so far most of the researches propose various measurement methods from time and global or local structural features based on relatively small dataset and that among the methods the most common methods such as Euclidean distance and DTW show high performance with relatively simple idea. Owing to the temporal characters of rainfall-runoff data, the similarity measurements for time series can be integrated to partial representative replayed data selection standards of the incremental learning method.

In this paper, we propose a novel incremental learning method to accelerate the training of Recurrent Neural Network (RNN) models for rainfall-runoff simulation. The method is designed to enable the model to learn from a large dataset while also adapting to changing conditions in real-time. The proposed method consists of two main components: regular training and emergency operation. During regular training, the model is trained on a diverse set of data samples, which enables it to learn the underlying patterns and relationships in the data. In emergency situations, the model can be quickly updated using a subset of representative data that captures the essence of the changing conditions. This approach allows the model to adapt to new data without requiring a complete retraining process. To ensure the efficiency and effectiveness of the incremental learning method, we introduce two key components: data distribution parameters and time series similarity metric. The data distribution parameters are used to select a representative subset of data that captures the underlying patterns and trends in the data. The time series similarity metric is used to measure the similarity between the new data and the existing data, allowing the model to adapt to changing conditions while still leveraging the knowledge gained from the regular training process. We evaluate the performance of the proposed incremental learning method using several modified RNNs and a study case of the middle reaches of the Yangtze River Basin and Han River Basin. The results demonstrate that the method is able to significantly accelerate the simulation speed of rainfall-runoff models while maintaining their accuracy and robustness. Our contribution is to propose a practical and robust incremental learning method that can accelerate the simulation speed of rainfall-runoff models, making them applicable to emergency flood management. The proposed method has important implications for improving the efficiency and effectiveness of flood management systems, particularly in situations where timely decisions are critical.

2 Materials and Methods

2.1 Study areas

The Yangtze River are prone to frequent flood disasters. This region experiences a high occurrence of flood and waterlogging incidents, which pose significant challenges to the local communities and infrastructure. These flood events result in widespread destruction, including damage to homes, agricultural lands, and transportation networks. Therefore, it is necessary to select the middle reaches region with significant ecological, economic, and social value as the study area. The Yangtze

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Discussions

River, the longest and largest river of China, is with length of 6300 km and a drainage basin area of 1.8 million km2 and is of great importance in ecology, economics and culture. As the longest tributary of the Yangtze River, with length of 1500 km and a drainage basin area of 1.8 million km2, the Han River flows into the artery of the Yangtze River in the middle reaches. The runoff changes have important impact on the spatial-temporal distribution of water resources in the whole river basin which means continuously frequent requirements on water resource analysis and management. Therefore, rainfall-runoff simulation is particularly significant for regional hydrological information patterns in the middle reaches of Yangtze River basin. The location and hydrological observation stations are shown in Figure. 1.

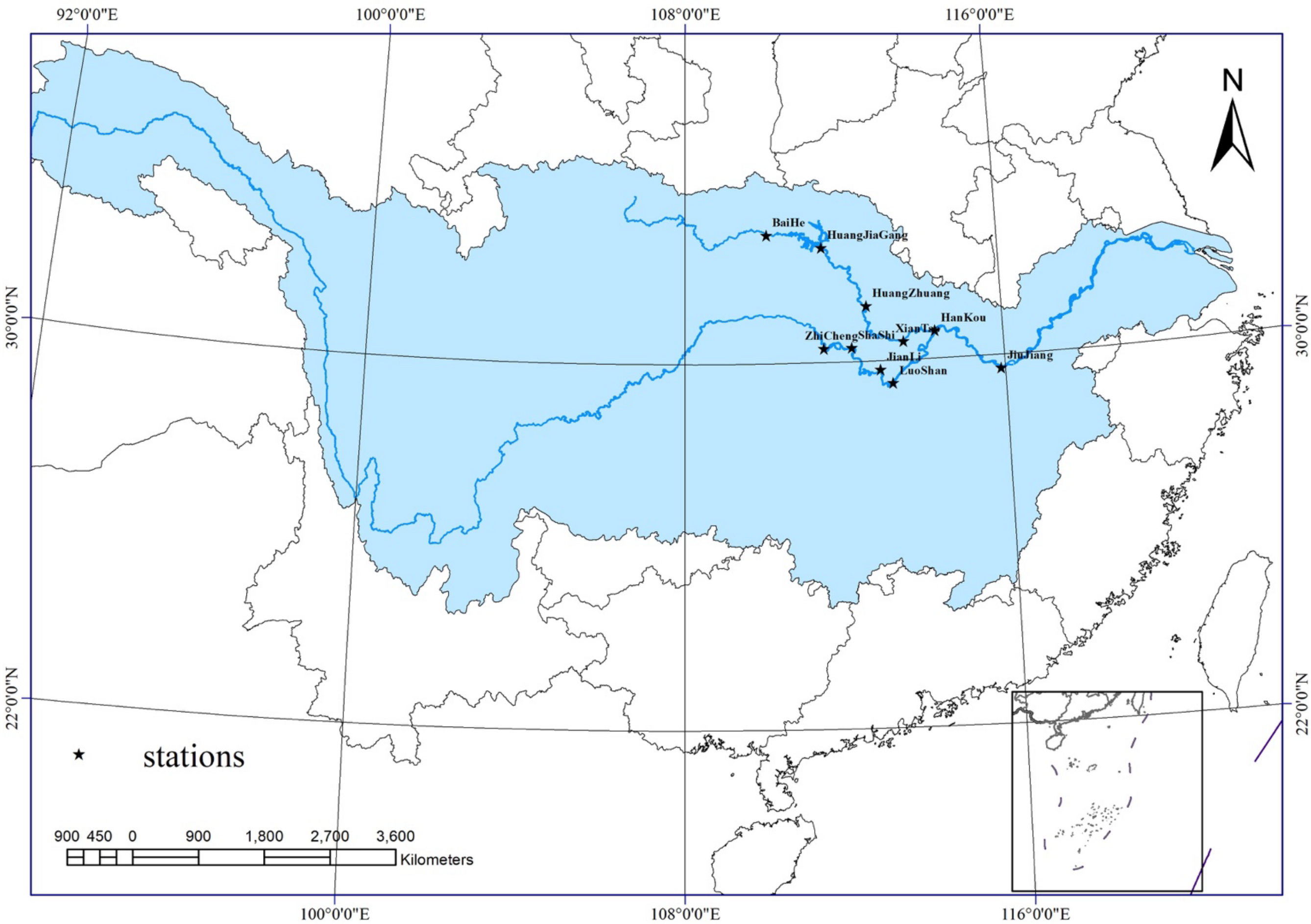


Figure 1: The location of the middle reaches of the Yangtze River and the Han River drainage basin and hydrological observation stations.

2.2 The Incremental Learning Method

Existing incremental learning methods for replay do not specifically target temporal feature data. We combine data distribution estimation, temporal similarity, and regularization methods to improve. We utilize partial representative data for incremental training, with a focus on time series similarity metrics that compare time series with the same length. Given the periodic

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Discussions

characteristic of rainfall-runoff series, we divide the complete time series into sub-time series of the same length, enabling the similarity between time series with different lengths to be transferred to the similarity among sub-time series with the same length.

We ensure that the data in each sub-dataset is similar in distribution and can be fit with a simple distribution, which can be estimated. We integrate similarity in both distribution and time series characteristics as partial representative data selection standards to ensure the representativeness of the selected data. As an additional penalty item on the loss function, parameter importance calculation is the core of regularization during incremental training.

Our method is based on regular network training, and as a result, the amount of calculation is significantly reduced, resulting in a notable acceleration of the training process. Moreover, owing to the representative partial data and regularization, the network model shows good performance on the incremental data. The structure of the incremental learning method can be elaborated as Figure. 2, the method can be divided into two parts: regular training for parameter initializing and incremental operation to deal with incremental training. Comprehensive consideration about data feature of both the historical and incremental data is used to produce partial representative data to reduce the magnitude of the input data. Parameter importance calculation as regularization constraint is added in incremental training to handle the error problem of the network when training real-time incremental data. Meanwhile when new incremental data is continuously input, the model may be trained at multiple times in a short period of time. The incremental learning method should also ensure the stability of the method effect under such conditions.

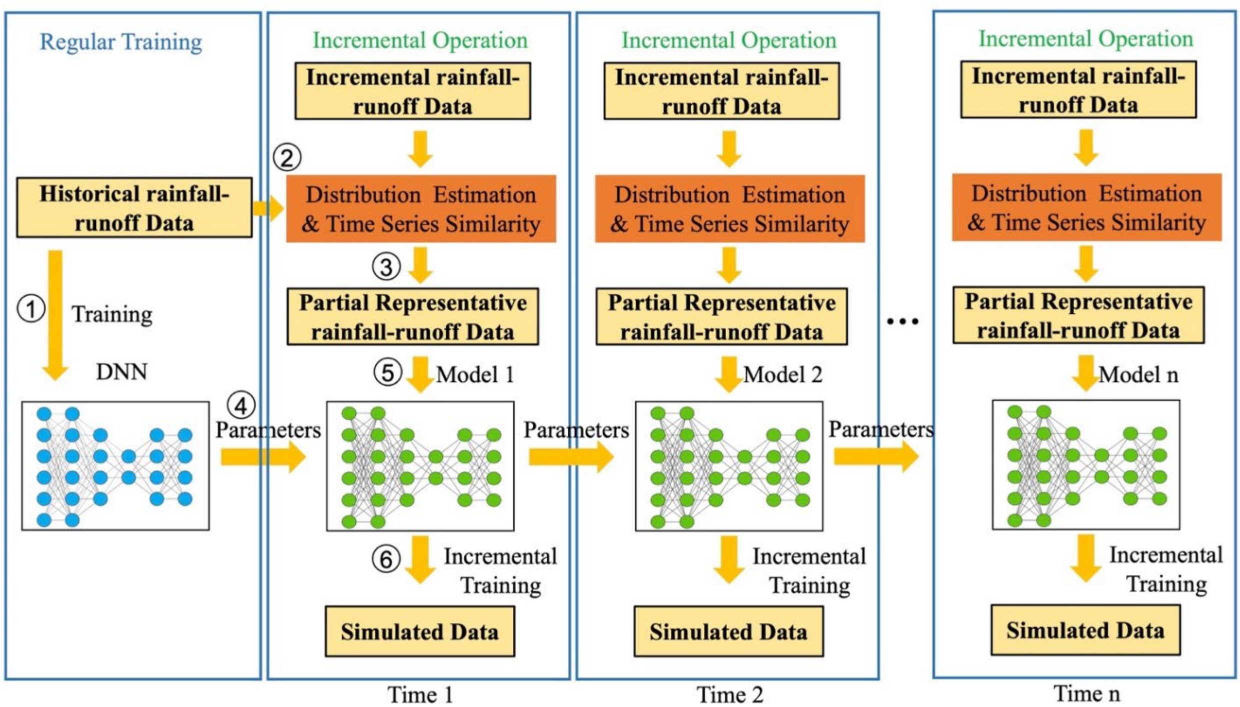


Figure 2: The structure of the incremental learning method.

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

The incremental operation part can be described as the following steps in detail. First, periodic analysis of time series is performed, and the combination of historical data and incremental data are sliced into multiple sub-time series. The distribution parameter calculation and timing similarity measurement calculation are performed for each sub-time series. By comparing the parameter difference between sub-time series and the overall time series and the timing similarity difference between sub-time series, the weights of the calculation result of the difference are assigned and the replay scores are obtained. The sub-time series are sorted according to the replay score, the number of sub-time series is determined according to the replay sample size level required by the incremental learning efficiency, and partial representative samples are selected for incremental training based on regularly trained network with the parameters initialized. Additionally, in the process of incremental training, parameter importance calculation selected as regularization constraints is imposed to the training loss of the model. L2 regularization normal form is introduced to impose penalties on the loss function of the deep learning model, and relevant parameters are adjusted. Eventually the training results are got. The process of the emergency operation part is shown in Figure. 3.

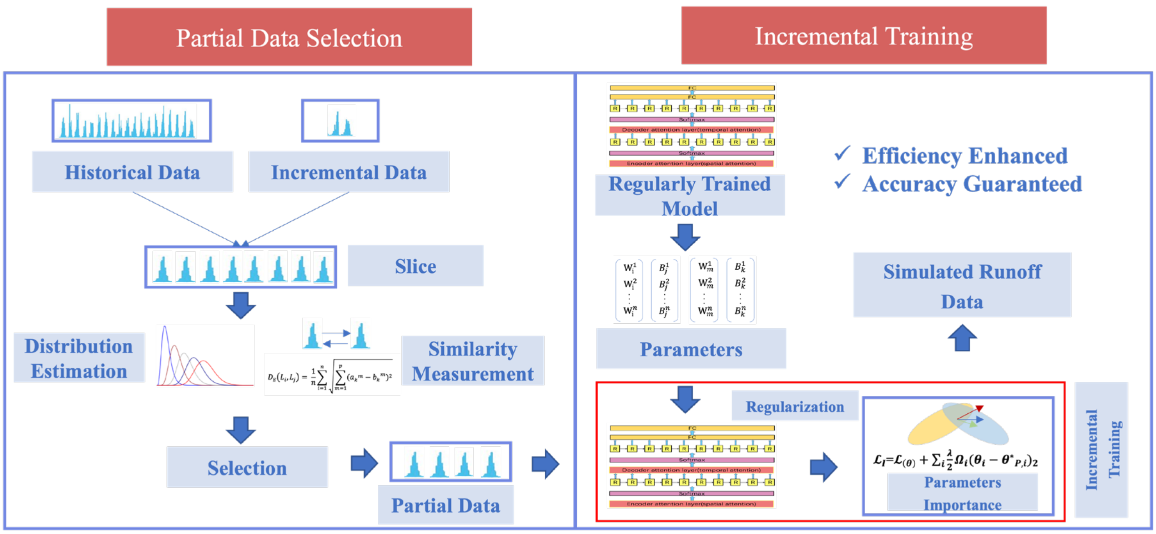


Figure 3: The process of the incremental operation part of the incremental learning method.

Formally, consider the moment t_i, the data that has been processed and trained in the deep learning model is called historical data, denoted as H_{t_i}, the incremental data arriving at this time is denoted as A_{t_i}, the deep learning model is denoted as $M_{t_i}(p_{t_i}^1, p_{t_i}^2 p_{t_i}^3, \ldots)$, $p_{t_i}^j$ is the jth parameter of the model at the moment. The historical data and incremental data become the historical data of the moment, and the depth model parameters after training at T_i become the input parameters of the moment model. The complete time series before periodic inspection is denoted as T_W and the sliced time series are T_{t_i}. Skewness and Kurtosis are selected as the distribution estimation metrics and standardized Euclidean distance works as the time series

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

similarity metric, the calculation process can be formulated as the following. $Skew$ is Skewness of T_{t_i}., $Kurt$ represents Kurtosis of T_{t_i}.SD is the standard deviation and $\bar{x}$ means the average of T_{t_i}.

$$Skew(X) = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^3}{SD^3} \quad (1)$$

$$Kurt(X) = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^4}{SD^4} - 3 \quad (2)$$

D_E refers to standardized Euclidean distance. α_s, α_k and α_D are the weights correspondent to the metrics to calculate replay score (S_{replay}). Replay score determines the probability that the sub-dataset will be chosen. Skewness and Kurtosis are to measure the distribution difference between $T_{t_i}$and T_W, standardized Euclidean distance is to measure the time series similarity among T_{t_i}. The replay score measures the representativeness of each sub-dataset. Generally, the magnitude of training data is positively correlative to training speed with the same parameters, therefore the incremental learning method can adjust the amount of representative data according to the anticipant speed that the incremental learning method needs to achieve. The number of selected sub-dataset is N_{replay} , and finally such many orders of magnitude sub-datasets with the highest replay scores are selected.

$$D_E(L_i, L_j) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{m=1}^{p}({a_k}^m - {b_k}^m)^2} \quad (3)$$

$$S_{replay} = \frac{\alpha_s}{\Delta Skew(X)} + \frac{\alpha_k}{\Delta Kurt(X)} + \frac{\alpha_D}{D_{\ E}}(L_i, L_j) \quad (4)$$

Then calculating the importance for each parameter in the network is attached to the loss function of the network, as regularization constraint. This can be described as the following formulations.

$$\mathcal{L}_I = \mathcal{L}_{(\theta)} + \sum_i \frac{\lambda}{2}\Omega_{ij}\left(\theta_i - {\theta^*}_{P,i}\right)_2 \quad (5)$$

$$\Omega_{ij} = \left\|\frac{\partial {l_2}^2\left(M(x;\theta)\right)}{\partial\theta_i}\right\| \quad (6)$$

$\mathcal{L}_I$ is the loss function of the model during incremental training, $\mathcal{L}_{(\theta)}$ is the loss function of the model during regular historical data training, $\sum_i \frac{\lambda}{2}\Omega_i\left(\theta_i - {\theta^*}_{P,i}\right)_2$ is the constraint item, θ_i is the parameter of incremental meta sample, ${\theta^*}_{P,i}$ is the standard to evaluate the parameter, which represents the difference between the previous and incremental meta sample, Ω_i is the l_2 regularization item, $M(x;\theta)$ is the output of the network, $\frac{\partial(M(x;\theta))}{\partial\theta_i}$ describes the gradient of the loss function of model with respect to parameter θ_i evaluated at the data point x . The importance of parameters can be described as the magnitude of the gradient. And λ can be adjusted with incremental data come.

Uniformly data of early years are set as historical data and data of lately years as incremental data. When the incremental data come at some time, both baseline and the incremental learning method are performed. The rainfall-runoff simulation on incremental data is defined as incremental tasks because the distribution of dataset has changed and three incremental tasks is set on each station. After selecting partial representative data, the incremental learning method uses the regularly trained attention-RNNs with learned parameters and finetuned model with a relatively lower learning rate and part of the changed hyperparameters. As for some of the model parameters, the weights and biases of the layers are updated when training by back-propagation approach. Iterations are performed with subsets of the training dataset which are called batches or a mini-batches.

Three modified RNNs are chosen as the networks. Besides, taken the rainfall and runoff of last station in the upstream as factors, attention mechanism (Vaswani et al., 2017) is applied in RNNs to capture the spatial relationship between the two stations. Inspired by a kind of dual stage attention (Qin et al., 2017), we construct a hybrid network called attention-RNN is shown as Figure. 3. attention-RNN consists of four neural network layers, including the encoder attention layer to capture spatial features, the decoder attention layer to capture temporal features, two single-track RNN layers and two full connected layers. RNN units can be regular RNN, LSTM and GRU. During the regular training step in the process of the incremental learning method, firstly expand forward calculation according to the time series, and then update the network parameters based on back propagation when three types of RNNs are picked up as the unit to make sure the incremental learning method can perform well in different RNNs. And the network can be specifically named as attention-RNN, attention-LSTM and attention-GRU respectively. Simplified illustration about the three RNNs can be described as following with several formulations.

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

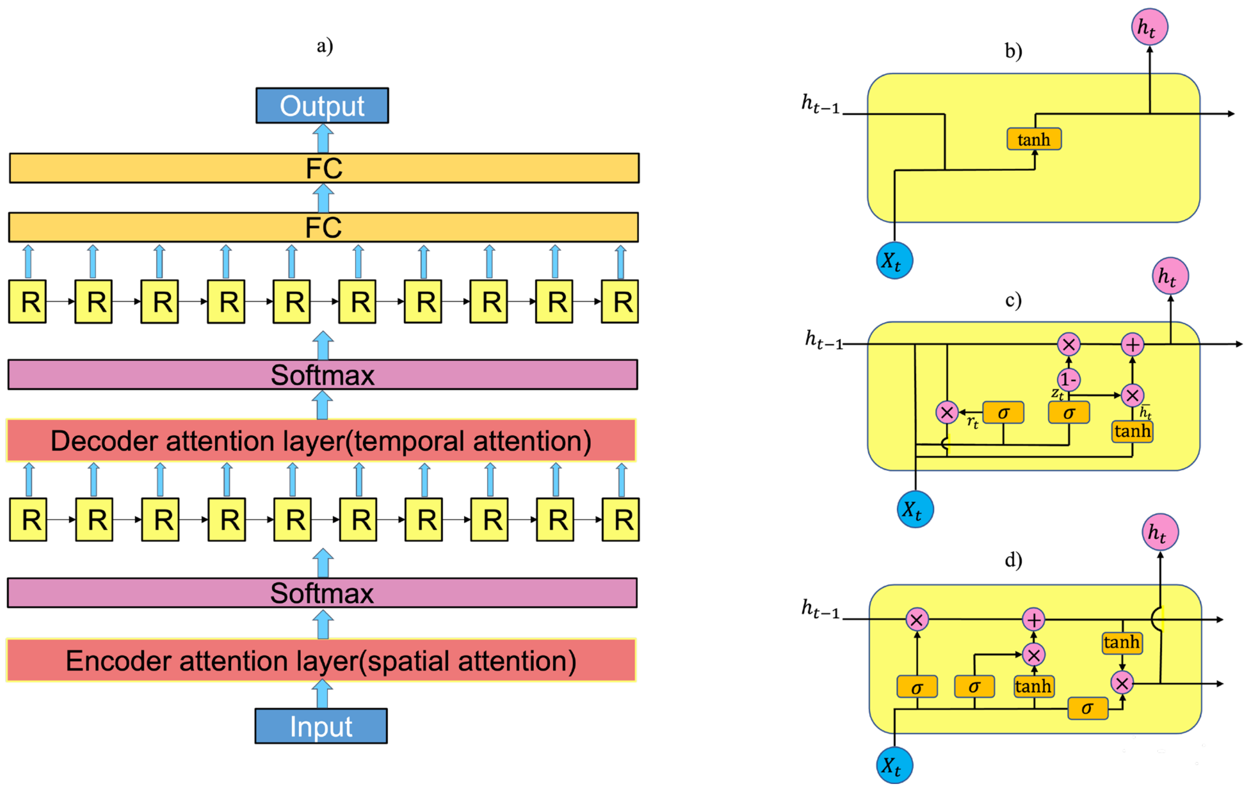


Figure 4: The structure of experiment models and inner structure of different RNNs. a) The structure of attention-RNNs (Qin et al., 2017). b) Inner structure of regular RNN. c) Inner structure of GRU. d) Inner structure of LSTM.

The first is regular RNN. W_{sh} is the weight matrix connecting input units with hidden units, W_{hh} is the connection weight matrix among the hidden units, W_{hy} is the weight matrix connecting hidden units with output units, b_h and b_y are the bias vectors. σ and $softmax$ are activate functions. The formulations clearly demonstrates that state at some moment rely on the past states and RNN can capture the relation among the states at different moment.

$$h_t = \sigma(W_{sh}x_t + W_{hh}h_{t-1} + b_h) \tag{7}$$

$$o_{t+1} = W_{hy}h_t + b_y \tag{8}$$

$$y_t = softmax(o_t) \tag{9}$$

LSTM (Hochreiter & Schmidhuber, 1997) is an advanced kind of RNN with gate control principle, which helps capturing the long short dependency relation among the time series by setting a memory cell. The gate control principle consists of three gates: input gate, output gate and forgetting gate. How LSTM works can be expressed as the following formulas. x_t and h_t correspence to input and output vector in hidden layer respectively, C_t is the memory cell, i_t is the input gate which controls the input data flowing into memory cells, f_t is the forgetting gate determining forgetting data in memory cells, o_t is the output gate, and tanh is also a kind of activation function.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{10}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{11}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{12}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{13}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{14}$$

$$h_t = o_t \times tanh(C_t) \tag{15}$$

GRU (Cho et al., 2014) takes similar gate principle to capture long time series dependency. But compared to LSTM, GRU only chooses two gates to control the data flows in RNN units, which are called reset gate and update gate respectively. The reset gate determines the combination of input and former memory and the update gate controls the amount of former memory data to be updated. The process can be formulated as following. r_t is the reset gate and the u_t is the update gate. Weight matrix and bias matrix are just like that in LSTM.

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{16}$$

$$u_t = \sigma(W_{xu}x_t + W_{hu}h_{t-1} + b_u) \tag{17}$$

$$\tilde{h}_t = tanh\,(W_{sh}x_t + W_{hh}(r_t \odot h_{t-1} + b_h) \tag{18}$$

$$h_t = u_t h_{t-1} + (1 - u_t) \odot \tilde{h}_t \tag{19}$$

2.3 Evaluation Metrics

To evaluate the performance of the proposed incremental learning approach, we use several regular evaluation indicators in regression (including Root Mean Square Error (RMSE), etc.) to show the error increase and computation time to evaluate the speed difference.

RMSE, referring to root mean square error, which represents the sample standard deviation of the difference between the predicted and observed values in order to account for the dispersion of the sample. The unit of RMSE in the experiments is cubic meters per second.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{20}$$

PE is a measurement metric used to quantify the difference between two values in terms of percentage error.

$$PE = \frac{|y_i - \hat{y}_i|}{\hat{y}_i} \times 100\% \tag{21}$$

Nash-Sutcliffe efficiency coefficient is generally used to verify the quality of hydrological model simulation results, and the changes of NSE can be significantly reflect on to what extend the method influence the quality of the baseline network.

$$NSE = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{22}$$

Equations for RMSE, PE and NSE are given in Eq. (20)-(22) respectively, where y_i is the predicted value, $\hat{y}_i$ is the observation value and m is the number of samples. It should be noted that the unit of the RMSE is cubic meters per second and the computational time is denoted as TIME with the unit of minutes.

3 Results and Discussion

3.1 Experiment Settings

The incremental learning method is performed on attention-RNN, attention-LSTM, and attention-GRU in dataset from stations. Meanwhile control group is set that the total data are used during network training, denoted as baseline. The experiments compare the performance of the baseline and incremental learning method with much concentration on to what extend does the proposed incremental learning method accelerate the training of the RNNs when guaranteeing acceptable error. All the experiments are performed on a computer with 128 GB RAM, Intel Silver 4210, and Quadro RTX A6000 48 GB GPU. The experiments are based on Python and deep learning framework Pytorch (Paszke et al., 2019).

By referring to the common rules of neural network training and repeated experiments, we choose to set the relative parameters as followings. The weights of distribution estimation metrics and time series metrics are both set to 1. The sample size of each batch or batch size is set to 128 (Gao et al., 2020). The mean square error (MSE) is selected as the original loss function for the gradient descent method to the baseline. And regularization constraint is added to MSE decrease in the incremental learning method. The optimization algorithm is Adam. The regularization parameter λ is set to 1 (Aljundi et al., 2018).

3.2 Incremental learning methods have improved the efficiency of rainfall-runoff simulation

The table 1 presents the results of a comparison between incremental learning and baseline methods on attention-LSTM in terms of root mean square error (RMSE), percentage of error (PE), Nash-Sutcliffe efficiency (NSE), and computation time (TIME) at different stations. The Incremental Tasks refer to simulation tasks of adding equal amounts of new real-time rainfall and runoff data sets divided in chronological order, represented by numbers 1, 2, and 3 respectively. The Method refer to whether using the incremental learning method, represented by baseline and increment respectively. The incremental learning method is seen to have a higher RMSE and lower NSE than the baseline method at all stations, indicating better performance

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

in forecasting. Specifically, the RMSE of the incremental learning method increases by 6.8% to 17.9% compared to the baseline method, with the rise of PE ascend from 1.34% to 2%, while the NSE is reduced vary from 0.02 to 0.06. Moreover, the computation time of the incremental learning method is shorter than that of the baseline method at all stations, with an average reduction to 22.8%. These results suggest that the incremental learning method is effective in improving the efficiency of hydrological forecasting, enabling prompt support for emergency flood prevention and mitigation decisions while maintaining an acceptable range of model training errors. Among the ten stations, the training time of the baseline method and the incremental learning method at the Baihe station is slightly shorter than that of the other stations due to the smaller number of data samples. However, the incremental learning method still shows a consistent pattern of improving training efficiency with the other stations, with an error range of around 1%.

Table 1: The comparison of performance between the incremental learning method and baseline conducted on attention-LSTM in different stations in Yangtze River basin.

Station	Incremental tasks	Method	RMSE (m^3/second)	PE	NSE	TIME (minute)
Zhicheng	1	baseline	1548.83	7.16%	0.92	153.60
		incre	1602.12	8.40%	0.86	34.18
	2	baseline	1574.65	7.82%	0.92	156.48
		incre	1639. 94	8.52%	0.87	34.17
	3	baseline	1618.65	7.94%	0.91	162.45
		incre	1696.47	8.36%	0.86	36.81
Shashi	1	baseline	1395.00	7.58%	0.92	156.5
		incre	1442.08	8.09%	0.86	36.15
	2	baseline	1464.53	7.56%	0.90	159.88
		incre	1479.54	8.20%	0.87	36.87
	3	baseline	1478.72	7.92%	0.91	165.12
		incre	1246.83	8.78%	0.86	38.80
Jianli	1	baseline	1347.80	7.76%	0.91	153.20
		incre	1432.12	8.23%	0.87	34.83
	2	baseline	1374.53	7.98%	0.90	155.40
		incre	1399.76	8.67%	0.86	34.45
	3	baseline	1409.76	8.02%	0.90	160.20
		incre	1446.34	8.90%	0.86	37.55
Luoshan	1	baseline	2204.96	7.15%	0.92	154.22
		incre	2332.47	7.97%	0.87	33.90

	2	baseline	1484.54	7.44%	0.92	159.43
		incre	1609.04	8.06%	0.86	35.80
	3	baseline	1505.28	7.38%	0.91	161.20
		incre	1546.53	8.03%	0.86	36.34
Hankou	1	baseline	2406.00	7.32%	0.91	154.60
		incre	2532.25	8.12%	0.86	29.18
	2	baseline	2474.66	7.82%	0.91	157.48
		incre	2639.12	8.36%	0.87	31.16
	3	baseline	2509.00	7.64%	0.90	159.46
		incre	2566.12	8.52%	0.86	32.73
Jiujiang	1	baseline	2476.83	7.33%	0.90	156.24
		incre	2532.09	8.47%	0.88	36.15
	2	baseline	2474.58	7.63%	0.90	159.27
		incre	2539.43	8.57%	0.88	36.88
	3	baseline	2498.65	7.72%	0.89	165.12
		incre	2546.61	8.41%	0.88	38.60

Table 2: The comparison of performance between the incremental learning method and baseline conducted on attention-LSTM in different stations in Han River basin.

Station	Incremental tasks	Method	RMSE (m^3/second)	PE	NSE	TIME (minute)
Baihe	1	baseline	202.03	11.64%	0.90	113.56
		incre	232.12	13.4%	0.86	29.19
	2	baseline	194.40	9.82%	0.90	119.47
		incre	239.95	13.2%	0.87	31.17
	3	baseline	208.65	9.64%	0.89	122.33
		incre	296.12	13.1%	0.86	32.74
Huangjiagang	1	baseline	160.65	10.04%	0.90	156.19
		incre	192.52	11.46%	0.86	36.16
	2	baseline	174.28	9.82%	0.90	159.28
		incre	209.34	11.60%	0.87	36.87
	3	baseline	181.45	9.64%	0.89	165.12
		incre	246.12	11.71%	0.86	38.78

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Huangzhuang	1	baseline	182.65	9.84%	0.90	153.17
		incre	242.77	11.2%	0.87	34.82
	2	baseline	191.37	10.36%	0.90	155.58
		incre	279.39	11.59%	0.86	34.42
	3	baseline	190.81	9.70%	0.90	160.20
		incre	226.04	11.54%	0.86	37.53
Xiantao	1	baseline	216.60	9.92%	0.89	154.25
		incre	252.10	11.97%	0.87	33.89
	2	baseline	224.65	10.17%	0.90	159.24
		incre	259.37	12.06%	0.86	35.79
	3	baseline	218.22	10.02%	0.90	161.22
		incre	296.12	12.20%	0.86	36.30

It can be obviously concluded from table 1 and table 2 that when the data size is at 20% of the entire dataset, if the model training time increases by more than 4 times, and the difference in error is less than 5%, and the difference in ratio-based metrics is less than 0.08. Consistent results across different attention-RNN models on the same dataset.

3.3 Good ability to continuous incremental learning

Over time, as incremental data keeps arriving, incremental learning methods should have good capabilities for handling continuous data additions. We divided the dataset used for training into three parts according to the time order and used them to simulate the scenario of continuously arriving new data. We conducted three incremental learning experiments with different models and different scenarios to compare the performance of the incremental learning method in different river basins. As for the continuous incremental scenarios, the performance of incremental learning method shows consistence that it performs the same in different incremental tasks. The greater difference between the distribution rules of incremental data and historical data is, the worse the effect of incremental learning is. When the different incremental task is close, the method shows the same performance. From incremental task one to three, the magnitude of replay data decreases, and correspondently the training efficiency generally ascends with error increase fluctuates in a relatively small range. Specifically, the run-time difference reaches over 4 times, the PE increase less than 3%, the NSE decease less than 0.05. However, the results show relatively weak self-adaptivity lower the ability of the online learning of the incremental learning method hard to handle the incremental data with rapidly changeable distribution. The incremental learning method shows stably consistent performance with relatively less self-adaptivity when the continuous incremental data trained. For example, in Jiujiang station, the PE increases 1.14%, 0.94%, 0.81% in the three incremental scenarios respectively, and NSE decreases 0.02, 0.04, 0.04 meanwhile

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

the efficiency improvement is 4.2, 4.2, 4.4. The Figure 5 shows the average PE, NSE and computation time difference of baseline and incremental learning method, which imply that when the incremental data are taken as continuously input, the incremental learning method gives the deep learning models the ability to continuous incremental learning.

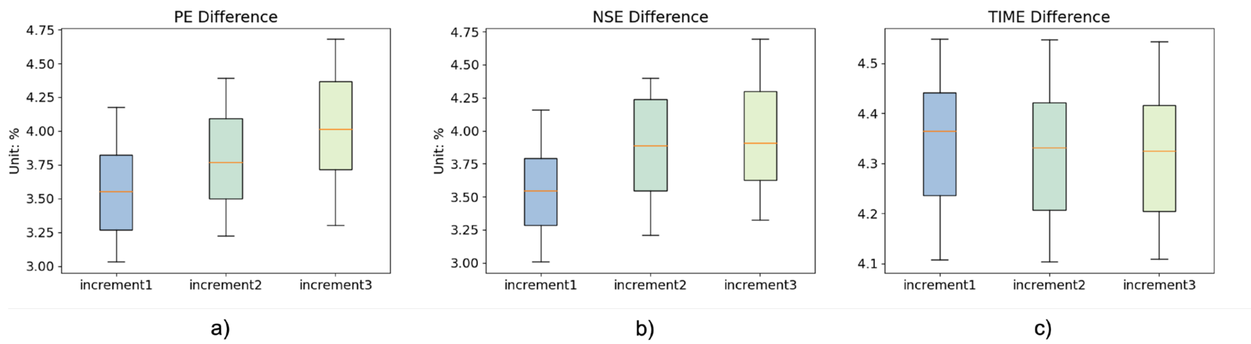
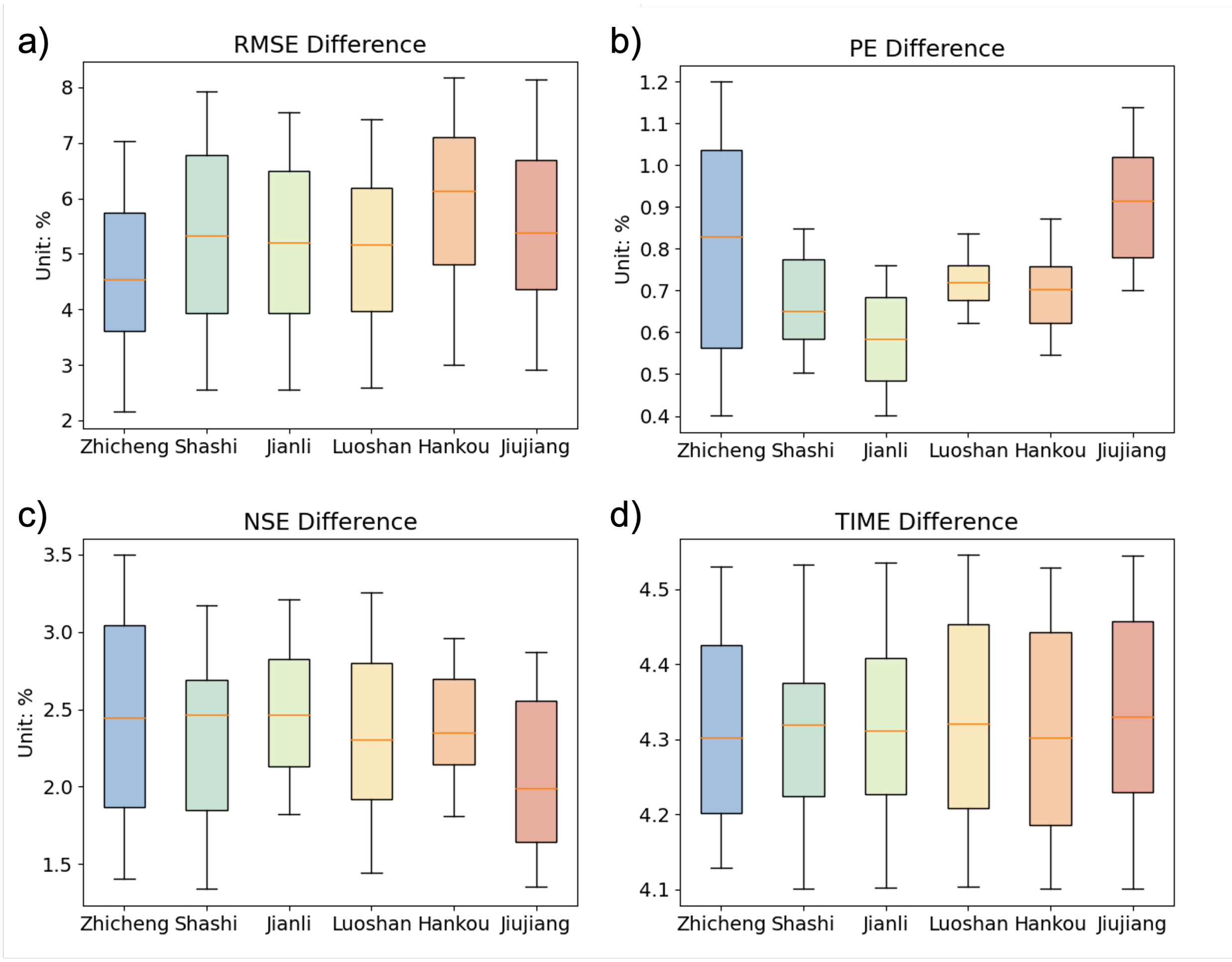


Figure. 5: The evaluation metrics difference between the baseline and the incremental learning method in the three incremental scenarios. a) the PE difference. b) the NSE difference. c) the multiplication factor difference for computation time.

3.4 Consistent Robustness Demonstrated in Different Networks and Locations

It can be summarized that the incremental learning method is well suitable for different river basins. As shown in Figure. 6 and Figure 7, by calculating the average results of different models of the same station, the conclusion can be drawn that under the condition of proper super parameter setting and good training, the incremental learning method has similar performance effect on different attention-RNNs, and the results are close. Take Hankou station as example, the PE increases 0.8%, 0.74%, 0.81% in the three incremental scenarios respectively, and NSE decreases 0.05, 0.04, 0.04 meanwhile the efficiency improvement is 4.2, 4.2, 4.1. Compared to Xiantao station, the PE increases 2.05%, 1.89%, 2.08% in the three incremental scenarios respectively, and NSE decreases 2.05, 1.89, 2.18 meanwhile the efficiency improvement is 4.3, 4.3, 4.2. The incremental learning method achieved similar effects at stations in the Yangtze River and Han River basins, rapidly increasing the training efficiency of the attention-RNNs model while maintaining a smaller error compared to the baseline model. However, it is notable that the baseline model and the incremental learning method had a higher error in the Han River basin than in the Yangtze River basin, likely due to the similar climatic conditions and rainfall patterns between the two regions. In this case, the smaller river is more prone to react to changes in rainfall, resulting in a more oscillatory hydrograph and a decreased fitting effect for the baseline model. Meanwhile, the incremental learning method amplifies the error caused by the model, making it more apparent.

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024



350 **Figure 6: The evaluation metrics difference between the baseline and the incremental learning method of four stations in Han River basin. a) the RMSE difference. b) the PE difference. c) the NSE difference. d) the multiplication factor difference for computation time.**

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

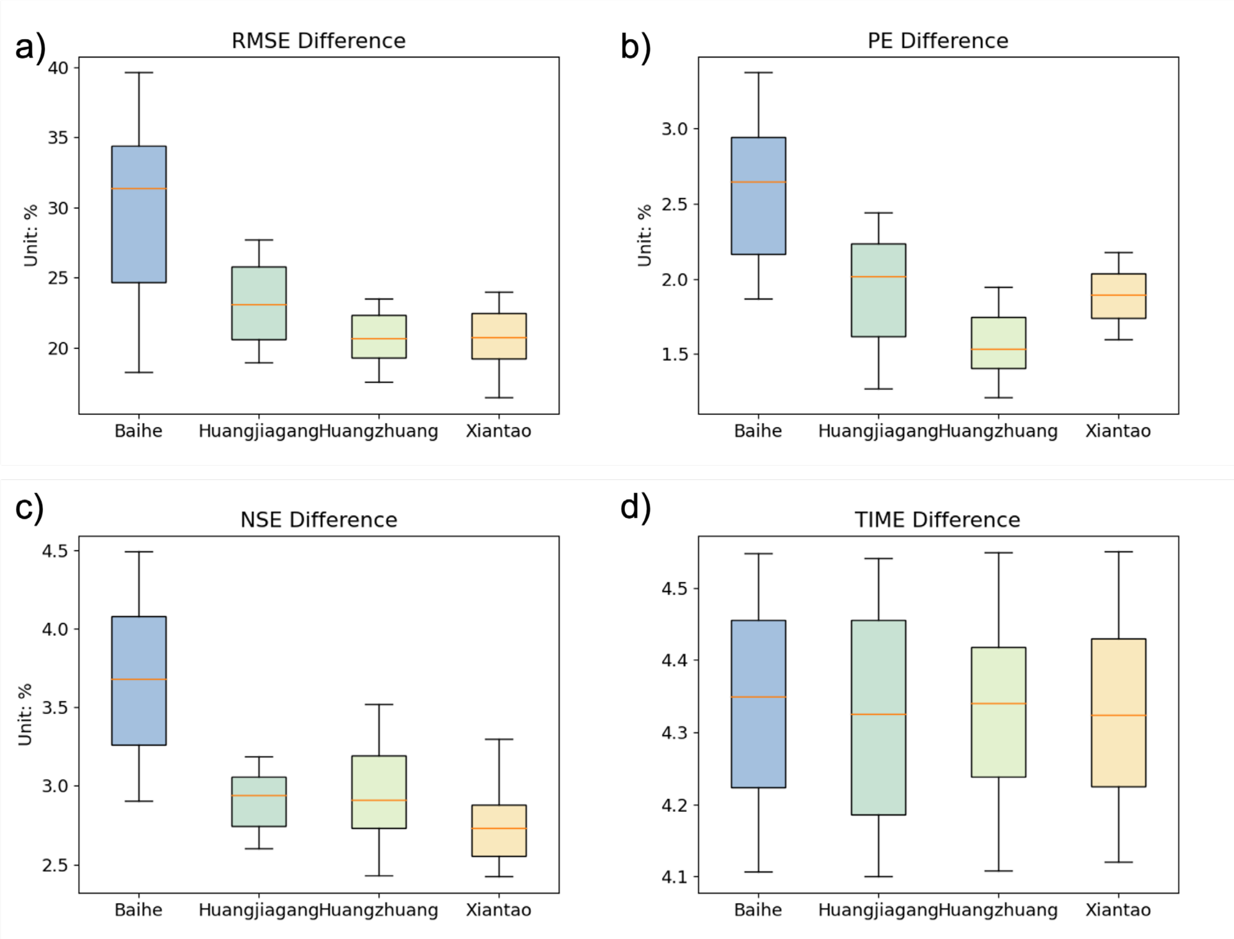


Figure 7: The evaluation metric differences between the baseline and the incremental learning method of four stations in Yangtze River basin. a) the RMSE difference. b) the PE difference. c) the NSE difference. d) the multiplication factor difference for computation time.

The incremental learning method is found to be suitable for various modified RNNs, as demonstrated by the comparable performance of different models under proper hyperparameter settings and adequate training. The attention-LSTM requires slightly more training time due to its larger number of model parameters, but the computation time improvement is similar to that of attention-GRU and attention-RNN. Besides, the similar increase intensity of evaluation metrics differences shows that the incremental learning method have the similar impact on the three modified RNNs. Hence, the incremental method is feasible for different RNNs, and the results are close.

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

4 Conclusions

This paper proposes an incremental learning method to accelerate the networks training. Several main conclusions can be summarized below:

1) The incremental learning method includes two components: regular training and emergency operation. Distribution evaluation metrics and time series similarity metrics are introduced as representative sample selection standards and parameter importance calculation is added into regularization under the situation of regular training networks in incremental operation.
2) The results turn out that the incremental learning method increases the training speed by over four times with around magnitude of all data and guarantee percentage error increase and NSE decrease less than 5%.
3) The incremental learning method shows suitability on different RNNs and the stability during continuous incremental scenarios. The results show the same tendency about efficiency improvement on average.

The incremental learning method has the potential to be applied in frequent rapid rainfall-runoff simulation, which can contribute to flood emergency decision making. To further enhance the applicability of the method, future research can focus on developing better interpretable selective sampling standards that can handle historical data with large extreme values and incremental data with dramatic changes in distribution. This will enable the incremental learning method to perform better in extreme precipitation situations, which are becoming increasingly common. Moreover, the use of novel Deep Neural Networks (DNNs) in rainfall-runoff modelling, as proposed by (Yin et al., 2022), can be explored for their suitability in the incremental learning method. By combining the strengths of DNNs and incremental learning, the accuracy and efficiency of rainfall-runoff simulation can be further improved, ultimately supporting more informed flood emergency decision making. In summary, the incremental learning method has the potential to be a valuable tool in flood emergency decision making, particularly in situations where timely and accurate rainfall-runoff simulation is critical. By refining the selective sampling standards and exploring the use of novel DNNs, the method's performance can be further enhanced, ultimately contributing to better flood risk management.

Author contribution

Zeqiang Chen: Conceptualization, Project administration, Writing-review. Jiashun Li: Methodology, Experiment, Writing-original draft. Changjiang Xiao: Writing-review. Nengcheng Chen: Funding, Writing-review.

Competing Interests

The authors declare that there is no competing financial or personal relationship interest influencing the works in this paper.

Code and data availability

The code is available upon request to the authors. The datasets are available at Xie Pingping., Joyce Robert., Wu Shaorong., Yoo S.-H., Yarosh Yelena., Sun Fengying., Lin Roger.,(2019), NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates. NOAA National Centers for Environmental Information. https://doi.org/10.25921/w9va-q159

Acknowledgments

This research is supported by the National Key Research and Development Program (2021YFF0704400), and the National Natural Science Foundation of China (41971351, 42001372)

References

Agrawal, R., Faloutsos, C., & Swami, A.: Efficient similarity search in sequence databases. International conference on foundations of data organization and algorithms, 69-84, doi:10.1007/3-540-57301-1_5,1993.

Aljundi, R.: Continual learning in neural networks. arXiv preprint arXiv:1910.02718, doi:10.48550/arXiv.1910.02718, 2019.

Aljundi, R., Massimo Caccia., Min Lin., Laurent Charlin., and Tinne Tuytelaars. (2019), Online continual learning with maximally interfered retrieval. arXiv preprint arXiv:1908.04742, 2. doi:10.48550/arXiv.1908.04742, 2019.

Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., & Tuytelaars, T.: Memory aware synapses: Learning what (not) to forget. Proceedings of the European Conference on Computer Vision (ECCV), 139-154, doi:10.1007/978-3-030-01219-9_9, 2018. Oiuy5tr4reedsk

Chaudhry, A., Dokania, P. K., Ajanthan, T., & Torr, P. H.: Riemannian walk for incremental learning: Understanding forgetting and intransigence. Proceedings of the European Conference on Computer Vision (ECCV), 532-547, doi:10.1007/978-3-030-01252-6_33, 2018.

Chen, Y.-C., Gao, J.-J., Bin, Z.-H., Qian, J.-Z., Pei, R.-L., & Zhu, H.: Application study of IFAS and LSTM models on runoff simulation and flood prediction in the Tokachi River basin. Journal of Hydroinformatics, 23(5), 1098-1111, doi:10.2166/hydro.2021.035, 2021.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, doi:10.48550/arXiv.1406.1078, 2014.

Davidson, G., & Mozer, M. C.: Sequential mastery of multiple visual tasks: Networks naturally learn to learn and forget to forget. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9282-9293, doi:10.48550/arXiv.1905.10837, 2020.

https://doi.org/10.5194/hess-2024-56
Preprint. Discussion started: 14 March 2024

Hydrology and Earth System Sciences
Open Access
Discussions

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q.: Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589, 125188, doi:10.1016/j.jhydrol.2020.125188, 2020.

Guo C, Li H, Pan D.: An improved piecewise aggregate approximation based on statistical features for time series mining[C]. Knowledge Science, Engineering and Management: 4th International Conference, KSEM 2010, Belfast, Northern Ireland, UK, September 1-3, Proceedings 4. Springer Berlin Heidelberg, 234-244, doi:10.1007/978-3-642-15280-1_23, 2010.

Hayes, T. L., Krishnan, G. P., Bazhenov, M., Siegelmann, H. T., Sejnowski, T. J., & Kanan, C.: Replay in deep learning: Current approaches and missing biological elements. Neural Computation, 33(11), 2908-2950, doi:10.1162/neco_a_01433, 2021.

Hochreiter, S., & Schmidhuber, J.: Long short-term memory. Neural Computation, 9(8), 1735-1780, doi:10.1162/neco.1997.9.8.1735, 1997.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z.: Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water, 10(11), 1543, doi:10.3390/w10111543, 2018.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., & Grabska-Barwinska, A.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 114(13), 3521-3526, doi:10.1073/pnas.1611835114s, 2017.

Li, W., Kiaghadi, A., & Dawson, C.: High temporal resolution rainfall–runoff modeling using long-short-term-memory (LSTM) networks. Neural Computing and Applications, 33(4), 1261-1278, doi:10.1007/s00521-020-05010-6, 2021.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32, 2019.

Paterson M, Dančík V.: Longest common subsequences. International symposium on mathematical foundations of computer science. Berlin, Heidelberg: Springer Berlin Heidelberg, 127-142, doi: 10.1109/SPIRE.2000.878178, 1994.

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G.: A dual-stage attention-based recurrent neural network for time series prediction. arXiv preprint arXiv:1704.02971, doi:10.48550/arXiv.1704.02971, 2017.

Rebuffi, S.-A., Kolesnikov, A., Sperl, G., & Lampert, C. H.: icarl: Incremental classifier and representation learning. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2001-2010, 2017.

Salvador, S., & Chan, P.: Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis, 11(5), 561-580, doi:10.3233/IDA-2007-11508, 2007.

Shao, Y., Zhao, Y., Yu, F., Zhu, H., & Fang, J.: The traffic flow prediction method using the incremental learning-based CNN-LTSM model: the solution of mobile application. Mobile Information Systems, doi:0.1155/2021/5579451, 2021.

Smyth, P.: Clustering sequences with hidden Markov models. Advances in neural information processing systems, 9, 1996.

Sodhani, S., Chandar, S., & Bengio, Y.: Toward training recurrent neural networks for lifelong learning. Neural Computation, 32(1), 1-35, doi:10.1162/neco_a_01246, 2020.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.: Attention is all you need. Advances in neural information processing systems, 30, 2017.

Wang, Q., Huang, J., Liu, R., Men, C., Guo, L., Miao, Y., Jiao, L., Wang, Y., Shoaib, M., & Xia, X.: Sequence-based statistical downscaling and its application to hydrologic simulations based on machine learning and big data. Journal of Hydrology, 586, 124875, doi.:10.1016/j.jhydrol.2020.124875, 2020.

Wu, Z., Xu, Z., Xiao, H., & Wu, H.: Spatio-Temporal Analysis of Drought Events in the Upper Reaches of Yangtze River Basin Based on Simulation of Soil Moisture (in Chinese). Resour. Environ. Yangtze Basin(27 (01)), 176–184, doi:10.11870/cjlyzyyhj201801020, 2018.

Xie Z, Du Y, Zeng Y.: Classification of yearly extreme precipitation events and associated flood risk in the Yangtze-Huaihe River Valley. Science China Earth Sciences, 61: 1341-1356, doi:10.1007/s11430-017-9212-8, 2018.

Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., & Wang, S.: Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. Journal of Hydrology, 608, 127553, doi:10.1016/j.jhydrol.2022.127553, 2022.

Yin, H., Zhang, X., Wang, F., Zhang, Y., Xia, R., & Jin, J.: Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model. Journal of Hydrology, 598, 126378, doi:10.1016/j.jhydrol.2022.127781, 2021.

Yin, H., Guo, Z., Zhang, X., Chen, J., & Zhang, Y.: RR-Former: Rainfall-runoff modeling based on Transformer. Journal of Hydrology, 609, 127781, doi:10.1016/j.jhydrol.2022.127781, 2022.

Zenke, F., Poole, B., & Ganguli, S.: Continual learning through synaptic intelligence. International Conference on Machine Learning, 70, 3987-3995, 2017