Impact of Spatial Distribution Information of Rainfall in Runoff Simulation Using Deep-Learning Method

Yang Wang[1], Hassan A. Karimi[1]

[1]Geoinformatics Laboratory, School of Computing and Information, University of Pittsburgh, 135 N Bellefield Ave, Pittsburgh, PA 15213, USA

Correspondence to: Yang Wang (yaw70@pitt.edu)

Abstract. Rainfall-runoff modelling is of great importance for flood forecast and water management. Hydrological modelling is the traditional and commonly used approach for rainfall-runoff modelling. In recent years, with the development of artificial intelligence technology, deep learning models, such as the long short-term memory (LSTM) model, are increasingly applied to rainfall-runoff modelling. However, current works do not consider the effect of rainfall spatial distribution information on the results. Focusing on ten catchments from the CAMELS dataset, this study compared the performance of LSTM with different look-back windows (7, 15, 30, 180, 365 days) for future one-day discharges and for future multi-day simulations (7,15 days). Secondly, the differences between LSTMs as individual models trained independently in each catchment and LSTMs as regional models were also compared across ten catchments. All models are driven by catchment mean rainfall data and spatially distributed rainfall data, respectively. The results demonstrate that regardless of whether LSTMs are trained independently in each catchment or trained as regional models, rainfall data with spatial information improves the performance of LSTMs compared to models driven by mean rainfall data. The LSTM as a regional model did not obtain better results than LSTM as individual model in our study. However, we found that using spatially distributed rainfall data can reduce the difference between LSTM as a regional model and LSTM as an individual model. In summary (a) adding information about the spatial distribution of the data is another way to improve the performance of LSTM where long-term rainfall records are absent and (b) understanding and utilizing the spatial distribution information can help improve the performance of deep learning models in runoff simulations.

Deleted: average

1 Introduction

Rainfall-runoff simulations are vital for watershed water resources management and risk analysis (Montanari, 2005; Neitsch et al., 2011). In addition, rainfall-runoff simulation plays an increasingly important role as a technical basis for hydrological forecasting due to the frequent occurrence of extreme hydrological events caused by climate change (Grayman, 2011; Panagoulia and Dimou, 1997). As the most widespread and essential tool for water science research, hydrological model plays a pivotal role in the rainfall-runoff simulation (Krause et al., 2005; Sood and Smakhtin, 2015). The development of hydrological models cannot be separated from the continuous research on hydrological processes. It is on the basis of the continuous understanding of hydrological processes that hydrological researchers have enough theoretical basis for building models that describe the interrelationship between the various hydrological elements and can simulate the overall hydrological cycle. The development of hydrological models has gone through two main stages, namely, lumped hydrological models and distributed hydrological models (Devia et al., 2015). For example, the Stanford model is the first lumped hydrological model with a solid theoretical basis (CRAWFORD and H., 1966). In 1977, British, Danish and French researchers jointly proposed the SHE (Systeme Hydrologique Europeen) hydrological model, which is the first generation of distributed hydrological models (Sahoo et al., 2006). The Variable Infiltration Capacity (VIC) is a large-scale distributed hydrological model developed by the University of Washington, the University of California at Berkeley, and Princeton University (Liang et al., 1996). The distributed VIC model is based on the idea of gridding to achieve distributed simulation of watersheds.

However, the fact that we cannot accurately describe every process of the hydrologic cycle leads to the necessary simplifications in the hydrologic model calculation process, which is one of the contributing factors to simulation errors. Since models based on physical mechanisms cannot fully describe the physical processes of the hydrologic cycle, researchers started to explore data-driven models for hydrologic modelling (Solomatine and Ostfeld, 2008). For example, Support Vector Machines (SVMs) are often used to manage the processing of hydrological model input data or to perform hydrological simulations directly due to their advantages in processing nonlinear problems (Ahmad et al., 2010; Sivapragasam et al., 2001). Artificial neural networks (ANNs) are a type of machine learning method that have been used for hydrological modelling since the 1990s. In the following years, more research has demonstrated that ANN models can achieve comparable results to physical models while requiring less data (Chang et al., 2015; Ömer Faruk, 2010). Although the robustness of ANN models needs to be further investigated, the ability of ANNs to capture the nonlinearity associated with hydrologic applications has led to its widespread use (Ghumman et al., 2011).

In recent years, with the development of deep learning techniques, LSTM (Long Short-Term Memory), as one type of RNN (Recurrent Neural Network) structure, has gained much attention in processing time series data. Compared with the traditional version of RNN, LSTM can solve its inherent problem of gradient disappearance or explosion (Greff et al., 2017; Hochreiter and Schmidhuber, 1997). LSTM has been used in many fields, including hydrological, and has achieved better results than traditional RNNs. A For example, Hu et al. (2018) compared the difference between ANN and LSTM in simulation of flood events, and the results show that LSTM models perform significantly better than ANN models. Kratzert et al. (2018) trained

Deleted: SHE

Deleted: n recent years, due to the development of deep learning techniques, such as LSTM in natural language processing and time series data, these techniques have also been widely used in simulation of rainfall-runoff. Among these, LSTM has garnered more attention of researchers due to its suitability for processing and predicting events with very long intervals and delays in time series

Deleted: (

Deleted: ,

Field Code Changed

Deleted: (

Deleted: ,

LSTM models with rainfall-runoff data from several watersheds, demonstrating the potential of LSTM as a regional hydrological model, one of which can predict flows in various watersheds. A LSTM model was also used in combination with Sequence-to-Sequence to simulate the discharge for the next few hours (Xiang et al., 2020). Gauch et al. (2020) 's study illustrated that LSTM can process different input variables at different time scales. Gauch et al (2021) used LSTM as a regional model and studied the relationship between LSTM and training period length, number of training basins. Gao et al. (2020) compared RNN, LSTM, and the Gated Recurrent Unit (GRU) network. Their results show that accuracy of LSTM and GRU models gradually improves and stabilizes with the increase of time step.

Analysing the current research applying LSTM to rainfall-runoff simulation, we find that the spatial information of the input is not fully utilized. The input features involved in the rainfall-runoff simulation, for example, rainfall and temperature, are not spatially distributed. Current research mostly uses an aggregated value, for example, surface-mean value, to drive LSTM models, which to some extent loses the spatial distribution information of the features. The uneven spatial distribution of these factors has a significant impact on the formation of runoff, especially the formation of peak discharge.

The aim of this study is to explore the potential impact of spatial distribution information in rainfall-runoff simulation using LSTM. Considering that rainfall is the most direct and influential factor on rainfall-runoff simulation, the main objective of this study is to compare the difference between the results obtained using the LSTM model driven by rainfall data with spatial distribution information and the LSTM model driven by basin mean rainfall data. The comparison includes the following differences:

(1) Look-back windows. Analyze how spatial information affects the results of LSTM models under different input sequence length.

(2) Look-forward windows. With the fixed input sequence length, use the LSTM with 'many-to-one' structure to simulate next day discharge and use the 'many-to-many' structure of the LSTM to simulate next multi-day discharge. The effect of spatial information under different look-forward windows is also analyzed.

(3) Training settings. Analyze the effect of spatial information on regional LSTM. LSTM is often used as a regional model, combining data from catchments within the region to train the model. The regional setting is of particular interest because it allows the model to encapsulate different hydrological processes by learning from more data and situations. The effect of spatial information on the results when training LSTM using data from each catchment separately is also considered.

The paper is structured is as follows. Section 2 describes the data, the model structure, and the experimental design. Section 3 analyses and discusses results. Section 4 provides concluding remarks and discusses future research.

2 Methods and Dataset

2.1 The CAMELS Dataset

In this study, we use the CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) data set from the National Center for Atmospheric Research (NCAR) (Addor et al., 2017; Newman et al., 2015). The dataset contains lumped

Deleted: (

Deleted: ,

Deleted: (

Deleted: ,

Deleted: (

Deleted: ,

Deleted: T

Deleted: A

Deleted: In summary, the goal of this study was: Firstly, the deep learning model represented by LSTM for rainfall-runoff simulation does not focus on the spatial distribution information of rainfall. It is known that rainfall is the most direct and influential factor on the formation of runoff. Current research mostly uses surface-mean rainfall to drive LSTM models, which to some extent loses the spatial distribution information of rainfall, which in a physical sense has a very important impact on the formation of runoff, especially the formation of peak discharge.

Second, there are two main ways to use the LSTM for runoff simulation. First, the LSTM is trained separately using data from each catchment. Secondly, the LSTM is used as a regional model, combining data from the catchments in the region while training the regional model. The latter has received a lot of attention because it can increase the amount of training data. However, current studies do not compare LSTMs as individual models trained independently and LSTMs as regional models driven by different types of rainfall data.

Third, there are different types of LSTM models such as 'many to one', which is to predict the value of the next time step using the data of the past *n* time steps and 'many to many', which is to predict the value of the future multiple steps using the data of the past *n* time steps. The current research on rainfall-runoff simulation by LSTM mainly uses the 'many to one' type. Analysing the performances of 'many to many' type of LSTM can help better apply the LSTM model to rainfall-runoff simulation

The main objective of this study is to explore the difference between the results obtained using the LSTM model driven by rainfall data with spatial distribution information and the LSTM model driven by basin mean rainfall data. We focus not only on the simulation of the next one-time step, but also on the simulation of multiple future time steps. We also compare the differences between LSTM as an individual model and as a regional model.

Deleted: HYDROMETEOROLOGICAL TIME SERIES meteorological forcing data and observed discharges on a daily time scale starting in 1980 for most basins. Lumped meteorological forcing data were mainly calculated from three grid data sources, namely Daymet (Thornton et al., 2014), Maurer (Livneh et al., 2013), and NLDAS (Xia et al., 2012). We used Daymet data in this study since it has a better resolution of 1 km than the other two data sources. CAMELS contains a total of 671 catchments with minimal anthropogenic disturbance in the contiguous United States (CONUS). All catchments are divided into 18 hydrologic units (HUCs) according to the U.S. Geological Survey's HUC map. In this study, we selected 10 catchments, 5 from the Ohio region and 5 from the Pacific Northwest. The Ohio region is located in the east and the Pacific Northwest is located in the west, which can better describe the different hydrological conditions. For each catchment, CAMELS has the basin mean forcing (lump) dataset, which includes the driving data when using the lumped hydrologic model. These are: (i) daily cumulative rainfall, (ii) daily minimum air temperature, (iii) daily maximum air temperature, (iv) mean short-wave radiation, and (v) vapor pressure. Here the daily cumulative rainfall is treated as the basin mean rainfall data without spatial distribution information. For each catchment, CAMELS also includes the hydrologic response units it contains. As can be seen in Figure 1, instead of using the catchment mean rainfall data (see the top of Figure 1b), we extract the rainfall of all hydrologic response units in the catchment to form a vector. The bottom of Figure1b shows that the catchment has 8 hydrologic response units from which we extract the corresponding 8 rainfall data to form a vector of size 8. Since each value in this vector represents the rainfall at different locations in the catchment, we assume that the vector can describe the rainfall at different locations, which means it has spatial distribution information. We extracted the rainfall data of each hydrologic response unit and created a dataset for the corresponding catchment, and regarded it as rainfall with spatial distribution information. The locations of the ten catchments are shown in Fig. 1a. Table 1 shows the basic information on each catchment and the size of the corresponding spatially distributed rainfall data. In addition, CAMELS data include simulation results from the hydrologic model, which is the Snow-17 models coupled with the Sacramento Soil Moisture Accounting Model(Newman et al., 2015). In this study we use the results of this model as benchmark to compare with the results of LSTM in Experiment 1.

Deleted: We used the Daymet data in this study since it has a resolution of 1 km, which is better than the other two

Deleted: Since the values in this vector represent rainfall information at different locations in the catchment, our assumption is that the vector is rainfall data with spatial distribution information.

Table 1. Overview of the selected catchments; for precipitation and temperature, mean and standard deviation is reported.

ID	Region name	Code	Area (km^2)	Mean precipitation (mm day−1)	Daily minimum air temperature (C)	No. of HRU
1	Ohio	03164000	46.15	3.66±8.07	4.20±8.52	64
2		03069500	58.41	4.00±6.69	2.07±0.24	32
3		03070500	64.72	3.67±6.58	3.84±9.30	8
4		03213700	59.10	3.47±6.43	5.54±8.91	41
5		03281500	69.22	3.76±7.52	6.26±9.15	27
6	Pacific Northwest	13340000	73.87	2.96±4.32	-1.44±6.88	193
7		12025000	33.43	4.58±7.21	4.82±4.86	12
8		12358500	81.11	3.36±5.18	-2.38±8.14	36
9		13337000	89.69	3.61±5.54	-1.56±6.99	34
10		13338500	73.75	2.37±3.75	-1.34±6.95	41

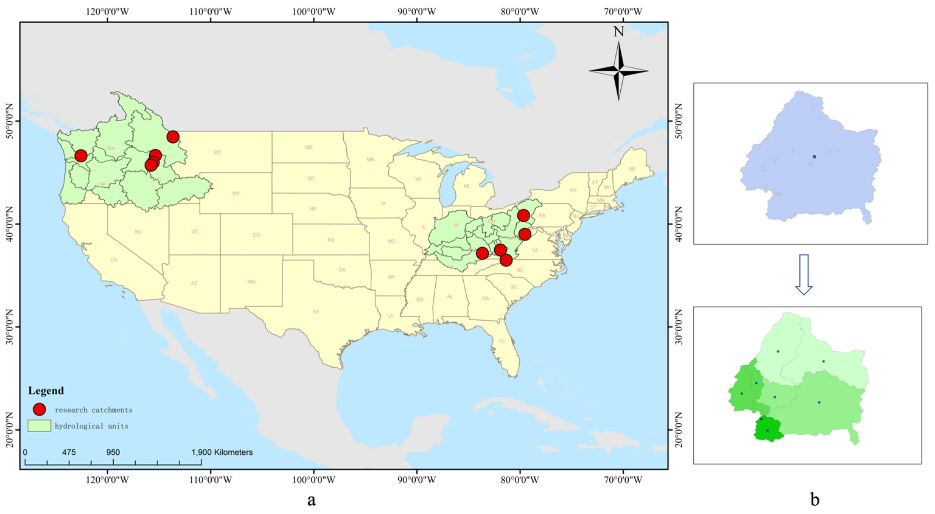


Figure1. a: Ten catchments and their locations in the State; b: Examples of spatially distributed rainfall data in this study

2.2 Long-short term memory network

RNNs are one of the most frequently used deep learning models to deal with sequential data, which are a superset of feedforward neural networks, augmented by the inclusion of recurrent edges that span adjacent time steps, introducing a notion of time to the models (Lipton et al., 2015). The main problem with RNN models is the occurrence of long-term dependencies, which arises when the nodes of a neural network have gone through many time steps of computation and the features from a relatively long time ago have been covered by the latest features.(Sherstinsky, 2020). Bengio et al (1994). The main problem with RNN models is the occurrence of long-term dependencies, which arises when the nodes of a neural network have gone through many time steps of computation and the features from a relatively long time ago have been covered by the latest features. The motivation for a LSTM model is to solve the problem mentioned above. As the name implies, Long Short Term Memory is a neural network with the ability to remember both long and short-term information. LSTM was first proposed by Hochreiter and Schmidhuber (1997) in 1997, and it has gone through several generations, resulting in a more systematic and complete LSTM framework that has been widely used in many fields. The reason why LSTM can solve the long-term dependency problem of RNN is that LSTM introduces the gate mechanism for controlling the delivery and loss of features. The basic structure of LSTM is shown in Fig. 2. In equations below, Ws are the weight matrices for different gates (W_f for

Deleted: In addition, CAMELS data include simulation results from the hydrologic model, which is the Snow-17 models coupled with the Sacramento Soil Moisture Accounting Model. For each catchment, 10 models were calibrated by rRMSE (oot mean squared error) as the objective function, and the model with the lowest RMSE was selected for validation. In this study we use the results of this model as benchmark to compare with the results of LSTM in Experiment 1.¶

Deleted: models when

Deleted: (

Deleted: ,

Deleted: RNN is one of the most frequently used models when using deep learning to deal with temporal problems, and the reason why RNN performs well on temporal data is that the input to RNN is the hidden node at time as the current time and the previous information at time

Deleted: (Sherstinsky, 2020)

Deleted: (

Deleted: ,

Deleted: and due to the rise of deep learning in 2012, LSTM

Deleted:

forget gate, W_i for input gate, W_c for output gate, and W_o for gate unit). bs are the bias vectors for different gates (b_f for forget gate, b_i for input gate, b_c for output gate, and b_o for gate unit). $tanh$ is hyperbolic tangent activation function, and σ is sigmoid activation function.

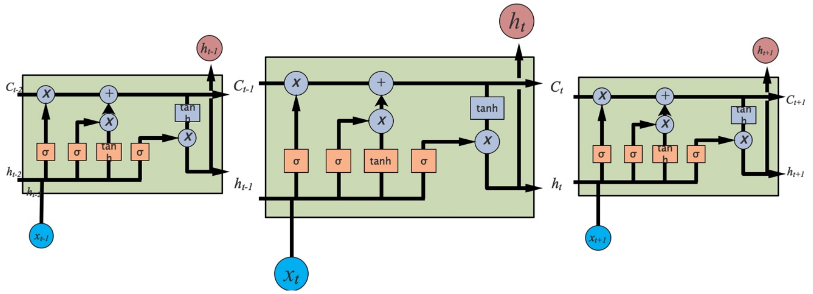


Figure 2. Basic LSTM layer structure with a detailed calculation illustration shown in the LSTM cell at time step *t*

Whenever information passes through a LSTM cell, there are actions that determine what old information is discarded and what new information is added. The structure that controls the addition and subtraction of information to and from the cell state is called gates. There are three such gates in a LSTM cell, namely forget gate, input gate, and output gate.

Deleted: a flow

The forget gate determines which information needs to be noted and which can be ignored. The information from the current input x_t and the hidden state h_{t-1} is passed through the sigmoid function. Sigmoid generates a value between 0 and 1, which can be used to describe whether a part of the old output is necessary (by bringing the output closer to 1). This value of f_t is the output of forget gate.

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

The input gate performs two steps to update the cell state. First, the current state x_t and the previously hidden state h_{t-1} are passed to a second sigmoid function. Next, the same information about the hidden state and the current state is passed through the tanh function. To regulate the network, the tanh operator creates a vector $c_t$where all possible values are between -1 and 1.

$$i_t = \sigma \cdot \left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$\tilde{c}_t = tanh \cdot \left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{3}$$

The next step is to decide and store the information from the new state in the cell state c_t. The previous cell state $c_{t-1}$is multiplied by the forget vector f_t . If the result is 0, the information is removed from the cell state. Next, the network takes the output value of the input vector i_t, which updates the cell state and thus provides the network with a new cell statec_t.

$$c_t = c_{t-1} \odot f_t + \tilde{c}_t \odot i_t \tag{4}$$

The output gate will determine the value of the next hidden state, which contains information about the previous input. First, the model passes the current state and the value of the previous hidden state to a third sigmoid function. The resulting new cell state is then passed through the tanh function. Based on this output value, the network decides what information the hidden state should have. This hidden state is used for output. The new cell state and the new hidden state are transferred to the next time step.

$$o_t = \sigma \cdot (W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

In summary, the forget gate determines what relevant information from previous steps is needed. The input gate determines what relevant information can be added to the current step, and the output gate ultimately determines the next hidden state.

2.3 Performance Evaluation Criteria

In this study, the performance of each model is evaluated by statistical error measurements and characteristics of discharge process error including Nash-Sutcliffe efficiency coefficient and root mean square error.

The Nash-Sutcliffe efficiency coefficient (NSE) is often used to verify the goodness of the hydrological model simulation results. NSE is calculated as follows:

$$NSE = 1 - \frac{\sum_1^n (p^t - q^t)^2}{\sum_1^n (p^t - \bar{q}^t)^2} \tag{12}$$

where p is the model simulation discharge at time t, q^t is the observed discharge at time t, and $\bar{q}^t$ is the mean of observed discharge. NSE takes the value of negative infinity to 1. NSE close to 1 means that the model quality is good and credible; NSE close to 0 means that the simulation results are close to the mean level of the observed values, i.e., the overall results are credible, but the simulation error is large; if NSE is much less than 0, the model is not credible.

The RMSE assesses how well the predictions match the observations. Depending on the relative range of the data, values can range from 0 (perfect fit) to +∞ (no fit). RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_1^n (p^t - q^t)^2}{n}} \tag{13}$$

where p^t is the model's simulation discharge at time t, q^t is the observed discharge at time t. n is the length of the sequence.

We also used the error of peak discharge (EPD) to measure the ability of the model to simulate peak discharge. Since there are multiple peak discharges in the sequence, we use the mean of all peak discharge EPDs as an indicator. EPD can be calculated as follows:

$$EPD = \frac{1}{n} \cdot \sum_1^n \frac{|p^t{}_p - q^t{}_p|}{q^t{}_p} \cdot 100\% \tag{14}$$

Deleted: In the formula above, Ws are the weight vectors for different gates (W_f for forget gate, W_i for input gate, W_c for output gate, and W_o for gate unit). bs are the bias vectors for different gates (b_f for forget gate, b_i for input gate, b_c for output gate, and b_o for gate unit). $tanh$ is hyperbolic tangent activation function, and σ is sigmoid activation function.¶

Deleted: generally

Deleted: f^t

where $q^{t}_{p}$is the observed peak discharge at time *t*, p^{t}_{p} is the modelled peak discharge at time *t*. n is the number of peak discharges in the dataset.

2.5 Experimental Setup

Considering the start and end times of rainfall data for all stations in the two catchments, for each catchment, we have a total of 11680 daily data, which is from Jan. 1, 1980 to Dec. 23, 2011. All model calibration and training were performed using data from Jan. 1, 1980 to Oct. 11, 2008. All model evaluation was done using data from Oct. 12, 2008 to Nov. 23, 2011. All data are normalized before being imported into the model. In this study, we set up three experiments. Experiment 1 and Experiment 2 are used to study the performance of LSTM for 'one time step output'. In Experiment 1, LSTM was trained as an individual model in each catchment separately. For each catchment, we used catchment mean rainfall data and spatially distributed rainfall data as input rainfall data, respectively. We considered different lengths of the input sequence, mainly 7 days, 15 days, 30 days, 180 days, and 365 days. In Experiment 2, LSTM is used as regional model. Specifically, training data from catchment 1-5 are combined to train regional model 1, and training data from catchment 6-10 are combined to train regional model 2. The trained regional model is used for the corresponding catchment in order to test the effect of the model. For each regional model, we use different types of rainfall data separately. It is noted that the vector length used to describe the rainfall spatial distribution information is not consistent for each catchment due to the inconsistent number of HRUs contained in each catchment. When using the regional LSTM, we need to keep the vector lengths of the corresponding catchments the same, so that we can gather the data from different catchments together to train the model. We standardize the length of spatially distributed rainfall data for each catchment to 20 when using the regional LSTM. For the catchment whose length is greater than 20, we fuse some of the hydrologic response units and take the average value as the rainfall of the fused units. For the catchment whose length is less than 20, we add enough 0 elements to the vector to make the length to be 20. Experiment 3 was designed to examine the performance of LSTM for 'n time step output'. In Experiment 3, the look-back window of the LSTM is set to 365 days based on the results of the first two experiments. We examined the model for the next 7 and 15 days and considered the difference between LSTM as an individual model for each catchment and a regional model. Each model is also driven separately using different rainfall data. We use M for meteorological data including daily minimum air temperature, daily maximum air temperature, mean short-wave radiation, and vapor pressure; D for discharge data, and P for rainfall data. The input data and output data for the three experiments (Experiments 1-3) are shown in Table 1. We tested two-layer LSTM with hidden states of 64 and 128, and batch sizes of 64 and 128. Finally, in all experiments, we used a two-layer LSTM structure with a cell/hidden state of 128 for each layer. The dropout rate is set at 0.2 in the experiment, and the batch size is 128. For each training procedure in the three experiments, the number of epochs is 200. We repeated each training procedure 10 times and selected the best performing model parameters by validation data for the future test.

Deleted: These data are divided into three parts, 70% of the data are used for model training, 20% for model validation, and 10% for model testing

Deleted:

Deleted:

Deleted: I

Table 2. Input data, output data, and training process for three experiments

ID	Input data	Type of rainfall	Training process	Output
Exp. 1	M, P	1. basin mean rainfall data 2. spatially distributed rainfall	1. LSTM as individual model for each catchment	D for the next one day
Exp. 2	M, P	1. basin mean rainfall data 2. spatially distributed rainfall	2. LSTM as regional model	D for the next one day
Exp. 3	M, P	1. basin mean rainfall data 2. spatially distributed rainfall	1. LSTM as individual model for each catchment/ 2. LSTM as regional model	D for the next few days (7/15 days)

3 Results and Discussion

3.1 Comparison of the results from different types of rainfall driven data for 'one time step output' simulation using LSTM as individual model (Experiment 1)

In Experiment 1, each catchment is trained separately. We compared the model results for different look-back windows driven by different types of rainfall data. The simulation results of the hydrological model are also placed in each table for comparison. Table 3 shows the performance of Experiment 1 driven by catchment mean rainfall data. From the table, we can see that there is a gradual improvement in the performance of the simulation as the length of look-back windows increase. Except for catchment 7 where the best model occurs at look-back windows of 30 and 180, the best results for all other catchments take place at look-back windows of 180 and 365 days. The catchment with the largest improvement in RMSE is catchment 3, where the RMSE is 1.92 with a look-back window of 7 days and 1.45 with a look-back window of 365 days. The catchment with the largest improvement in NSE is catchment 4, where the NSE is 0.56 with a look-back window of 7 days and 0.81 with a look-back window of 365 days. Comparing the results of the LSTM model with the benchmark, we can see that the results of the LSTM model are overall better than the benchmark. When the look-back window is 7 days and 15 days, the results of some catchments are slightly worse than the benchmark, such as catchment 4 and catchment 1. However, when the look-back window is larger than 15 days, the results of LSTM outperform the benchmark.

Table 4 shows the performance of Experiment 1 driven by spatially distributed rainfall data. We can see that for the LSTM driven by spatially distributed rainfall data, the results are better than the shorter look-back windows when the look-back window is 180 or 365 days. For example, for catchment 2, the RMSE for look-back windows of 7 days and 365 days are1.78 and 1.30, respectively, with an improvement of 0.48. The largest improvement in NSE is with catchment 4, with an NSE of 0.56 when the look-back window is 7 and 0.81 when the look-back window is 365. We also compared the results of LSTM with the benchmark. The results are similar to those driven by catchment mean rainfall data. The results of the LSTM model are generally better than the benchmark. Based on the results in Table 3 and Table 4, we can conclude that for runoff simulation, increasing the look-back window can improve the simulation performance of the LSTM. In our experiments, regardless of the type of rainfall data used to drive the LSTM, the simulations with look-back windows of 180 and 365 days outperform the models with 7, 15 and 30 days. Compared with RNN, LSTM can learn long-term dependence. The long look-back window can provide more information for establishing the relationship between the input and output data, which can improve the performance of the model.

Formatted: Centered

Deleted: ¶

Table 3 Performance of Experiment 1 driven by catchment mean rainfall data

	7 days		15 days		30 days		180 days		365 days		Benchmark	
ID	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE
1	0.71	0.74	0.61	0.80	0.60	0.81	0.62	0.80	0.59	0.82	0.67	0.65
2	1.82	0.70	1.52	0.79	1.42	0.81	1.29	0.85	1.40	0.82	1.97	0.78
3	1.92	0.59	1.77	0.65	1.76	0.65	1.44	0.77	1.45	0.76	2.08	0.64
4	1.14	0.56	1.03	0.65	0.87	0.74	0.73	0.82	0.76	0.81	0.92	0.60
5	1.60	0.71	1.53	0.74	1.60	0.71	1.16	0.85	1.21	0.83	1.77	0.78
6	0.96	0.80	0.98	0.79	0.89	0.83	0.59	0.92	0.58	0.92	0.91	0.84
7	1.70	0.86	1.75	0.85	1.51	0.89	1.51	0.89	1.65	0.87	1.92	0.84
8	1.25	0.80	1.30	0.86	1.24	0.87	0.97	0.84	1.17	0.83	1.71	0.82
9	1.39	0.76	1.38	0.80	1.34	0.81	1.12	0.87	1.06	0.85	1.75	0.81
10	0.60	0.75	0.61	0.78	0.57	0.80	0.37	0.89	0.37	0.90	0.65	0.73

Table 4 Performance of Experiment 1 driven by spatially distributed rainfall data

	7 days		15 days		30 days		180 days		365 days		Benchmark	
ID	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE
1	0.67	0.76	0.61	0.81	0.54	0.85	0.54	0.85	0.50	0.87	0.67	0.65
2	1.78	0.71	1.56	0.78	1.22	0.86	1.21	0.87	1.30	0.85	1.97	0.78
3	1.94	0.58	1.66	0.69	1.56	0.72	1.43	0.77	1.43	0.77	2.08	0.64
4	1.08	0.61	1.04	0.63	0.86	0.75	0.69	0.84	0.70	0.83	0.92	0.60
5	1.56	0.72	1.39	0.78	1.37	0.79	1.04	0.88	1.02	0.88	1.77	0.78
6	0.87	0.84	0.85	0.85	0.80	0.86	0.53	0.94	0.53	0.94	0.91	0.84
7	1.73	0.85	1.68	0.86	1.42	0.90	1.53	0.88	1.51	0.89	1.92	0.84
8	1.34	0.85	1.32	0.85	1.31	0.85	1.47	0.82	1.30	0.86	1.71	0.82
9	1.63	0.72	1.47	0.78	1.36	0.81	1.08	0.88	1.08	0.88	1.75	0.81
10	0.64	0.75	0.62	0.77	0.51	0.84	0.38	0.91	0.39	0.91	0.65	0.73

Figure 3 shows how we compare the differences in the results obtained by driving LSTM with different types of rainfall data. The comparison for RMSE is shown in the left panel. Positive values indicate that the results driven by spatially distributed rainfall data are better than those driven by mean rainfall data. The right panel shows the comparison of NSE. Negative values indicate that the results driven by spatially distributed rainfall data outperform mean rainfall data-driven results. We find that the results driven by spatially distributed rainfall data are better than those driven by mean rainfall data. In particular, when the look-back windows are 180 and 365 days, which represent the better models for each catchment, the results driven by spatially distributed rainfall data are generally better than the results driven by mean rainfall data. For example, for NSE, when the look-back window is 365 days, the results obtained from spatially distributed rainfall data are better than those obtained from mean rainfall data. However, we find that for catchment 8, the RMSE obtained for the mean rainfall data with a look-back window of 180 is half smaller than that the one obtained for the corresponding spatially distributed rainfall data, which is the largest difference.

Table 5 compares the error of peak discharge obtained from different types of driver data. As can be seen from the table, the simulation of peak discharge is better in the results obtained using spatially distributed rainfall data. Except for catchment 4 where the best simulation results occur in the mean rainfall data, the best results for the other nine catchments occur in the results of spatially distributed rainfall data. Figure 4 compares the catchment 10 discharge process using different types of rainfall data. We can see that the discharge process obtained by spatially distributed rainfall is closer to the actual one. The results obtained by spatially distributed rainfall are also better in the simulation of flood peaks. Coupling NSE with RMSE, we can see that good performance can be achieved by using LSTM for runoff simulation. The results of LSTM using longer look-back windows are generally better than those of the benchmark and shorter look-back windows. The spatially distributed rainfall data can provide more information to the input data, which helps the LSTM to better identify the relationship between the input and output data, thus build more accurate model.

Table 5 Comparison of EPD of Experiment 1 using different types of rainfall data

		1	2	3	4	5	6	7	8	9	10
180 days	(1)	0.24	0.24	0.25	0.27	0.26	0.16	0.16	0.17	0.22	0.22
	(2)	0.17	0.25	0.25	0.28	0.21	0.14	0.16	0.15	0.23	0.22
360 days	(1)	0.20	0.28	0.25	0.31	0.24	0.17	0.18	0.20	0.22	0.21
	(2)	0.15	0.23	0.26	0.29	0.22	0.14	0.18	0.20	0.21	0.20

(1) driven by catchment mean rainfall data; (2) driven by spatially distributed rainfall data

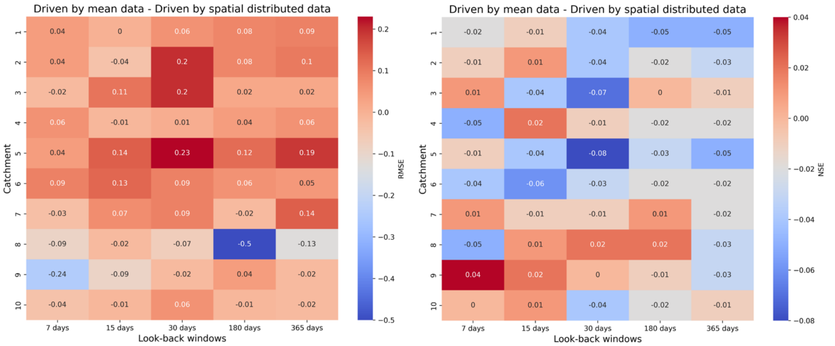


Figure 3 Comparison of performance of Experiment 1 using different types of rainfall data

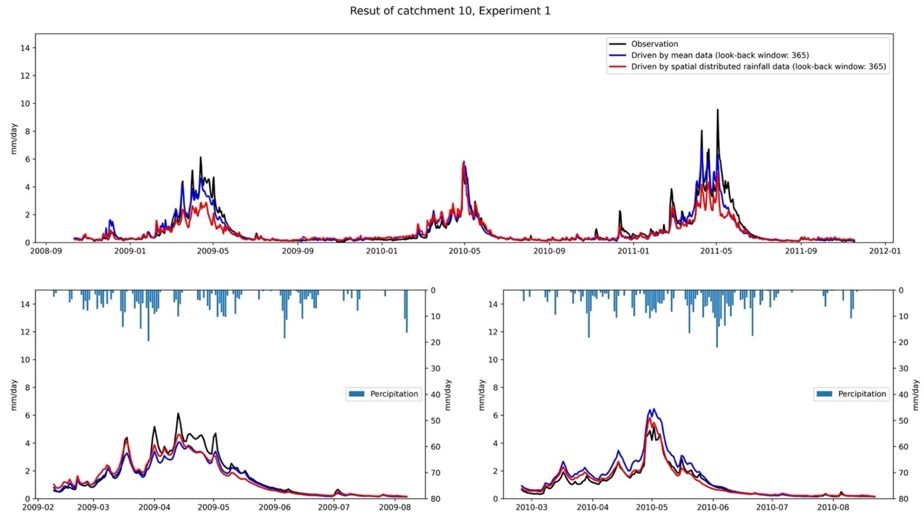
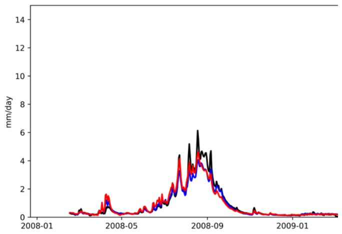


Figure 4 Comparison of performance of Experiment 1 using different types of rainfall data



Deleted:

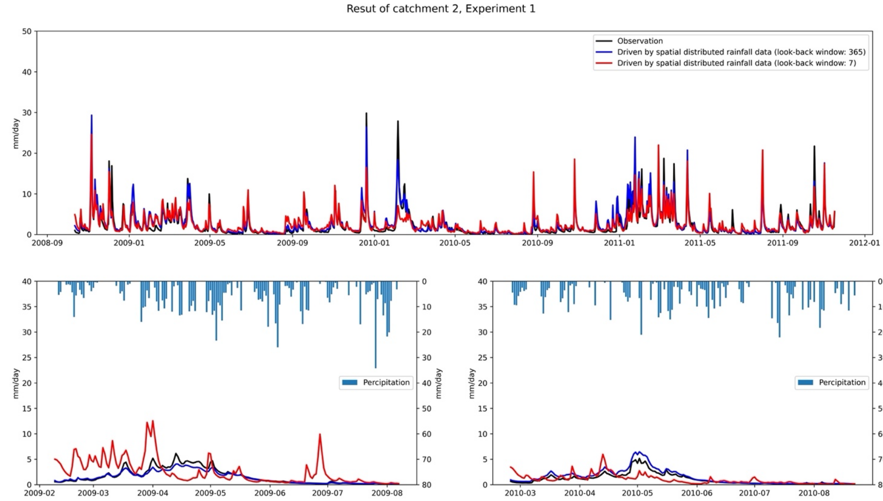
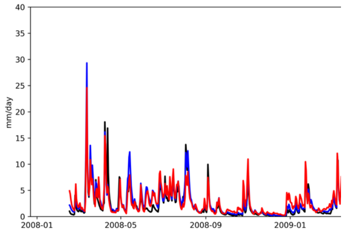


Figure 5 Comparison of performance of Experiment 1 using different look-back windows

3.2 Comparison of the results from different types of rainfall driven data for 'one time step output' simulation using LSTM as regional model (Experiment 2)

In Experiment 2, we examined the effect of different types of rainfall data on the model when the LSTM is used as a regional model. The data from catchment 1-5 were used together to train regional model HUC 1, and the data from catchment 6-10 were used together to train regional model HUC 2. We applied the trained model to each catchment separately and compared the performance.

Table 6 and Table 7 show the results obtained by training the regional model with different types of rainfall data. Firstly, combining the results of the two regional models, we can see the same trend as in Experiment 1, that is, for each model, the optimal performance occurs when the look-back windows are 180 and 365 days. This also proves that increasing the look-back windows can improve the model's performance. For HUC 1, we found that spatially distributed rainfall data in all catchments achieved better results except for catchment 2 where mean rainfall data achieved slightly better simulation results. Similarly, we found that in HUC 2, except for catchment 8 where the mean rainfall data obtained slightly better results than the spatially distributed rainfall data, the spatially distributed rainfall data also obtained better results in the other catchments. Figure 6 shows the EPDs of HUC1 and HUC2, where we only count the results for the look-back windows of 180 and 365



Deleted:

Deleted: . It is noted that when training and testing the model using spatially distributed rainfall data, we need to keep the length of spatially distributed rainfall data consistent in order to use data from different catchments. We standardize the length of spatially distributed rainfall data for each catchment to 20. For the catchment whose length is greater than 20, we fuse some of the hydrologic response units and take the average value as the rainfall of the fused units. For the catchment whose length is less than 20, we add 0 to change the length to 20 where the models are more effective. From the figure, we can see that for both HUCs, the EPDs obtained by training the models with spatially distributed rainfall data are generally smaller than those obtained by training with catchment mean rainfall data. This illustrates that adding information on the spatial distribution of rainfall can also improve the simulation of the model when the LSTM is used as the regional model.

Table 6 Comparison of performance of regional model (HUC 1) using different types of rainfall data

	Driven by catchment mean rainfall data									
	7 days		15 days		30 days		180 days		365 days	
ID	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE
1	1.40	0.27	1.79	0.22	1.38	0.21	1.00	0.48	0.99	0.48
2	1.90	0.67	1.74	0.72	1.40	0.82	1.27	0.85	1.30	0.84
3	1.90	0.59	1.69	0.68	1.54	0.73	1.41	0.77	1.54	0.73
4	1.13	0.57	1.35	0.38	1.19	0.52	1.31	0.43	1.04	0.64
5	1.68	0.68	1.59	0.71	1.70	0.67	1.36	0.79	1.43	0.77
	Driven by spatially distributed rainfall data									
1	1.14	0.42	1.26	0.32	1.37	0.21	0.91	0.57	0.82	0.65
2	1.94	0.65	1.62	0.76	1.43	0.81	1.43	0.81	1.30	0.84
3	1.79	0.64	1.63	0.70	1.48	0.75	1.36	0.79	1.38	0.79
4	1.08	0.61	1.31	0.42	1.33	0.40	0.87	0.75	1.00	0.66
5	1.72	0.67	1.59	0.71	1.60	0.71	1.27	0.82	1.11	0.86

Table 7 Comparison of performance of regional model (HUC 2) using different types of rainfall data

	Driven by catchment mean rainfall data									
	7 days		15 days		30 days		180 days		365 days	
ID	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE	RMSE (mm/d)	NSE
6	0.83	0.85	0.92	0.82	0.91	0.82	1.05	0.75	0.93	0.77
7	1.80	0.84	1.82	0.84	1.83	0.84	1.80	0.84	1.59	0.88
8	1.35	0.84	1.37	0.84	1.50	0.81	0.95	0.92	1.02	0.91
9	1.36	0.81	1.40	0.80	1.37	0.81	0.87	0.92	0.82	0.93
10	0.96	0.44	0.83	0.58	0.88	0.53	0.55	0.82	0.57	0.80
	Driven by spatially distributed rainfall data									
6	1.07	0.75	1.05	0.76	1.03	0.77	0.77	0.87	0.78	0.87
7	1.60	0.87	1.46	0.90	1.56	0.88	1.32	0.91	1.33	0.91
8	1.63	0.77	1.60	0.78	1.54	0.80	1.17	0.88	0.97	0.92
9	1.59	0.74	1.56	0.75	1.37	0.80	0.90	0.92	0.79	0.92
10	0.77	0.64	0.80	0.61	0.99	0.40	0.51	0.84	0.38	0.91

The LSTM as a regional model is a widely used method for runoff simulation. One of the main reasons is that sufficient training data is a prerequisite for a deep learning model to achieve good results, and using data from different catchments of the same hydrological unit can increase the amount of training data. Here we compare the results obtained by using LSTM as a regional model with those obtained by using LSTM as an individual model for each catchment. Figure 8 shows the differences between the three metrics. In the figure, a positive value of RMSE means that the regional model is worse than the individual

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

Formatted: Centered

model; a positive value of NSE means that the regional model is better than the individual model; a positive value of EPD means that the regional model is worse than the individual model. From the figure, we did not observe the general phenomenon that LSTM as a region model achieves better results than an individual model. For example, for catchments 1, 4, 5, 6, and 10, the RMSE and NSE using the LSTM as an individual model for each catchment are better than the LSTM as a regional model. This result is consistent for two different types of derived data. One possible reason is that although LSTM as a regional model can be learned with more training data, data from different catchments increase the possibility of inconsistency in the data. Similar discharges may correspond to different input data in different catchments, and similar inputs may correspond to different discharges in different catchments. This may have a negative impact on the learning process of LSTM. However, when comparing the difference of LSTM as a regional model and LSTM as an individual model from different types of data, we find that using spatially distributed rainfall data can reduce the difference between LSTM as a regional model and LSTM as an individual model. We counted the absolute values of different metrics in Figure 8. The RMSE, NSE and EPD between LSTM as a regional model and LSTM as individual model are 0.2±0.15, 0.11±0.11, and 0.06±0.05, respectively when using mean rainfall data to drive the model. When a spatially distributed rainfall data-driven model is used, the RMSE, NSE and EPD between LSTM as a regional model and LSTM as an individual model are 0.19±0.10, 0.07±0.07, and 0.04±0.03, respectively.

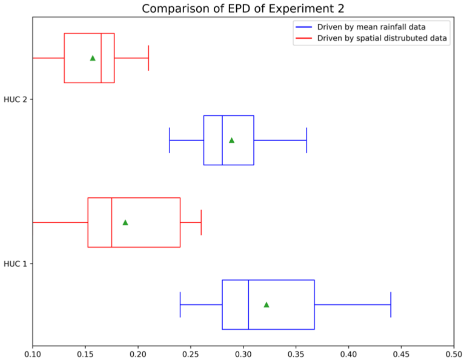


Figure 6 Comparison of EPD of Experiment 2 using different types of rainfall data

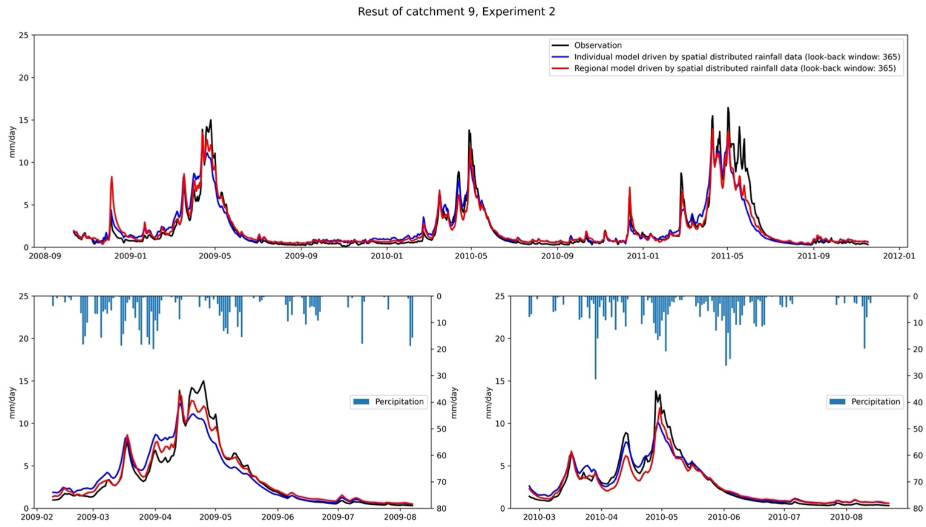
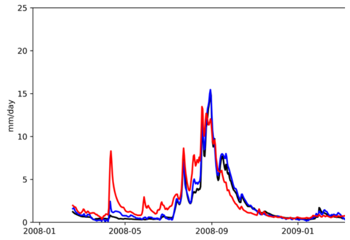


Figure 7 Comparison of performance of Experiment 2 using LSTM as regional model and individual model

Deleted:



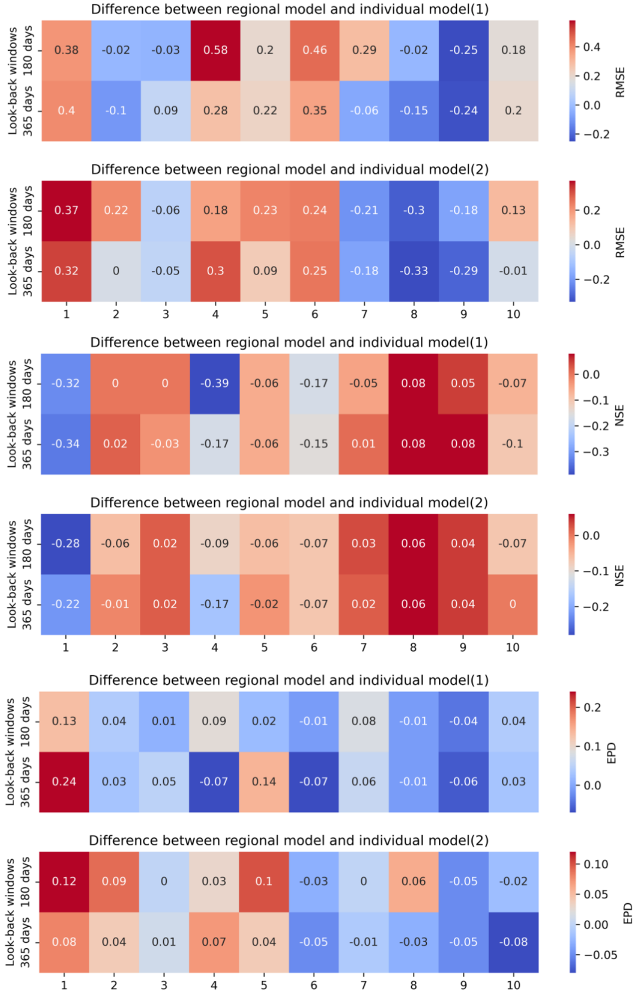


Figure 8 Comparison of the performance between using LSTM as a regional model and an individual model: (1) driven by catchment mean rainfall data; (2) driven by spatially distributed rainfall data.

Deleted:

3.3 Comparison of the results from different types of rainfall driven data for ‘n time step output’ simulation (Experiment 3)

In Experiment 3 we tested the simulation ability of LSTM for n time steps output. Based on the results of Experiment 1 and Experiment 2, we found that longer look-back windows can achieve better simulation results. For the future multi-day simulation, we used a look-back window of 365. Our goal is to simulate the future 7-days and 15-days discharges. The results of using LSTM as an individual model are shown in Table 8. We can see that the error obtained by simulating the discharge for the next 7 days is smaller than the error obtained by predicting the discharge for the next 15 days. Prediction for multiple days in the future is a much more difficult task. Comparing the simulation results of mean rainfall data and spatially distributed rainfall data, we find that the results obtained by using spatially distributed rainfall data are better than those obtained by mean rainfall data. For the next 7 days of simulations, catchment 8 and 10 have the same results for different types of driven data. For the next 15 days of simulations, the results obtained for spatially distributed rainfall data in all catchments are significantly better than those obtained for mean rainfall data.

Table 8 Comparison of performance using LSTM as individual model for ‘n time step output’

		1	2	3	4	5	6	7	8	9	10
7day (1)	RMSE (mm/d)	0.62	1.2	1.5	0.68	1.28	0.61	1.49	1.28	1.14	0.39
	NSE	0.80	0.87	0.75	0.85	0.81	0.92	0.88	0.86	0.87	0.91
	EPD	0.23	0.28	0.23	0.29	0.28	0.14	0.18	0.24	0.21	0.21
7day (2)	RMSE (mm/d)	0.57	1.14	1.49	0.66	1.05	0.42	1.47	1.28	1.03	0.39
	NSE	0.83	0.88	0.75	0.85	0.88	0.96	0.88	0.86	0.89	0.91
	EPD	0.13	0.25	0.30	0.27	0.24	0.09	0.20	0.24	0.18	0.21
15day (1)	RMSE (mm/d)	0.67	1.39	1.64	0.73	1.28	0.58	1.74	1.36	1.19	0.43
	NSE	0.77	0.82	0.70	0.82	0.81	0.93	0.85	0.84	0.85	0.89
	EPD	0.21	0.24	0.29	0.25	0.26	0.16	0.20	0.26	0.29	0.27
15day (2)	RMSE (mm/d)	0.57	1.23	1.51	0.58	1.10	0.51	1.60	1.34	1.11	0.34
	NSE	0.83	0.86	0.74	0.89	0.86	0.94	0.88	0.85	0.87	0.93
	EPD	0.15	0.24	0.28	0.23	0.21	0.10	0.18	0.25	0.19	0.16

(1) driven by catchment mean rainfall data; (2) driven by spatially distributed rainfall data

We also examined the simulation results for future multiple days when LSTM is used as a regional model. From Table 9, we can see that, in general, the regional model obtained using spatially distributed rainfall data has better simulation results. Except for catchment 6, the best models for the next 7 days are spatially distributed rainfall data driven models. The spatially distributed rainfall data-driven model has better results for all catchments for the next 15 days. The results for multi-day simulations are the same as those of Experiment 1 and Experiment 2. We can conclude that the rainfall data with spatial distribution information can improve the rainfall simulation results of LSTM. In particular, for the future multi-day simulations, the addition of rainfall data with spatial analysis information gives a significant advantage over the LSTM driven by mean rainfall data. By comparing different types of regional models, we also find that rainfall data with spatial analysis information can also improve the simulation results of LSTM as a regional model.

Table 9. Comparison of performance using LSTM as regional model for 'n time step output'

		1	2	3	4	5	6	7	8	9	10
7day (1)	RMSE (mm/d)	1.79	1.51	1.59	1.25	1.36	0.58	1.33	1.24	1.14	0.76
	NSE	0.66	0.79	0.71	0.48	0.79	0.93	0.91	0.87	0.87	0.65
	EPD	0.89	0.35	0.30	0.40	0.29	0.13	0.19	0.23	0.23	0.40
7day (2)	RMSE (mm/d)	0.73	1.43	1.57	0.93	1.04	0.77	1.32	1.01	0.87	0.37
	NSE	0.73	0.81	0.72	0.71	0.88	0.87	0.92	0.91	0.92	0.92
	EPD	0.30	0.30	0.33	0.27	0.24	0.10	0.16	0.15	0.14	0.15
15day (1)	RMSE (mm/d)	1.34	1.60	1.88	1.22	1.36	0.88	1.65	1.27	0.90	0.76
	NSE	0.70	0.77	0.60	0.50	0.79	0.90	0.87	0.88	0.92	0.65
	EPD	0.58	0.32	0.29	0.37	0.30	0.14	0.28	0.18	0.15	0.38
15day (2)	RMSE (mm/d)	0.98	1.26	1.40	1.14	1.17	0.86	1.50	1.04	0.81	0.43
	NSE	0.50	0.85	0.78	0.57	0.84	0.84	0.89	0.91	0.93	0.89
	EPD	0.31	0.25	0.28	0.31	0.25	0.13	0.16	0.16	0.14	0.18

(1) driven by catchment mean rainfall data; (2) driven by spatially distributed rainfall data

Figure 6 shows the comparison of LSTM as an individual model for each catchment and as a regional model for future multi-day simulations. As in Experiment 2, we did not observe a significant advantage of LSTM as a regional model. In general, the regional model is better than the individual model for catchments 7, 8, and 9, which means that the regional model is slightly better than the results of LSTM as individual model for HUC 2. For catchments 1, 2, 3, 4, 5, the results of the individual model are generally better than those of the regional model. When comparing the effect of different types of driving data on the differences, we also find that the differences between individual and regional models driven by spatially distributed rainfall data are relatively small.

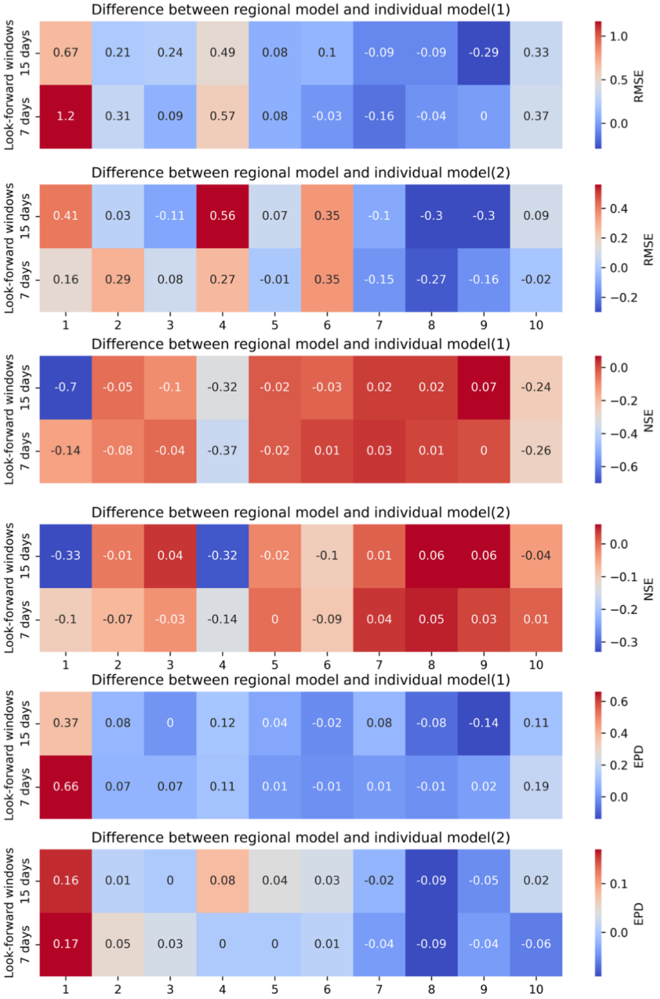


Figure 9 Comparison of the experiment 3 between using LSTM as regional model and individual model

4 Conclusions and Future Research

Deep learning models, especially LSTMs, have received increasing attention in rainfall-runoff simulation studies. The current LSTM-based studies are still mainly from a data-driven perspective and few studies have investigated the different simulation results from different types of meteorological data or construction of models based on the physical relationships of rainfall and runoff. In this study, rainfall, which has the greatest influence on runoff, is used as the object of study. The basin mean rainfall data is used as the rainfall data without spatial distribution information, and the vector composed of rainfall on hydrologic response units in the catchment is used as the rainfall data with spatial distribution information. The impact of the two types of rainfall data on the performance of the deep learning model is compared and analysed.

Based on the results of Experiment1 and Experiment 3, we conclude that the LSTM has a good performance compared to the benchmark when performing runoff simulations. In our experiments, the model performs better with look-back windows of 180 days and 365 days than with look-back windows of 7 days, 15 days and 30 days. The trend holds for both one-day and multi-day simulations. The long look-back window can provide more information for establishing the relationship between the input and output data, which can improve the performance of the model. The trend is not affected by the type of rainfall data.

We used two approaches to train the LSTM model. One is to treat the LSTM as an individual model and train it independently in each catchment. The second way is to use the LSTM as a regional model, using data from all catchments in the region for training. Based on the results of Experiment 2 and Experiment 3, we found that regardless of the approach, rainfall data with spatial information can improve the model's performance when compared with the model driven by mean rainfall data. In particular, the spatially distributed rainfall data improves the simulation results more when simulating the next multi-day discharges. The spatially distributed rainfall data can provide more information to the input data, which helps the LSTM to better identify the relationship between input and output and thus build a more robust model. Our findings suggest that increasing the spatial distribution information of the input data can improve the performance of the model, whether the LSTM is used as an individual model or as a regional model for runoff simulation.

Deleted: more significantly

Deleted: show

We also compared the difference between LSTM as an individual model and as a regional model. According to the results of our experiments, we did not observe that LSTM as a regional model achieved better results than LSTM as an individual model. In some catchments the regional model gives better results, while in others the individual model gives better results. This conclusion applies to both one-day and multi-day simulations. However, we found that using spatially distributed rainfall data can reduce the difference between LSTM as a regional model and LSTM as an individual model. Although LSTM as a regional model can be learned with more training data, data from different catchments increase the possibility of inconsistency in the data. Similar discharges may correspond to different input data in different catchments, and similar inputs may correspond to different discharges in different catchments. This may have a negative impact on the learning process of LSTM.

When we compared the results of spatially distributed rainfall data in Experiment 1, which mean increasing the spatial distribution information of the input data, with the results of mean rainfall data in Experiment 2, which mean increasing the size of the training data, we found that the results of the two are comparable. The variables related to runoff generation are characterized by uneven spatial distribution, such as rainfall, temperature, humidity, etc. Understanding and utilizing the spatial distribution information of these variables can help improve the performance of deep learning models in runoff simulations. This is especially true for those regions where data are scarce, since raster rainfall data with spatial distribution information are currently available from many sources. Adding information about the spatial distribution of the data is another way to improve the performance of deep learning models.

There are some gaps that can be continued to be investigated in the future. For example, in this study, the rainfall of the hydrological response unit of catchment is used to represent the spatial distribution of rainfall information. We can obtain raster-type rainfall data from satellite data, climate models, and other sources, which may be able to better represent the spatial distribution of rainfall. We only consider comparing the basin mean rainfall and spatially distributed rainfall, other driving data, such as temperature and a pressure, also have spatial distribution characteristics. How to increase the spatial distribution information of all features on the basis of the uniform resolution of different features and compare the influence of the input conditions on the model results is also a research direction worth conducting in the future.

Code and data availability. The CAMELS input data are freely available at the homepage of the NCAR (https://ral.ucar.edu/solutions/products/camels). Model outputs as well as code may be made available by request to the corresponding author.

Author contributions. YW and HK jointly developed the project idea and performed research. YW proposed the LSTM architecture and conducted all the experiments and analyzed the results. HK supervised the manuscript from the machine-learning perspective. All authors read and approved the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Not applicable.

References

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: Catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21(10), 5293–5313, doi:10.5194/hess-21-5293-2017, 2017.

Ahmad, S., Kalra, A. and Stephen, H.: Estimating soil moisture using remote sensing data: A machine learning approach, Adv. Water Resour., 33(1), 69–80, doi:10.1016/j.advwatres.2009.10.008, 2010.

Bengio, Y., Simard, P. and Frasconi, P.: Learning Long-Term Dependencies with Gradient Descent is Difficult, IEEE Trans. Neural Networks, 5(2), 157–166, doi:10.1109/72.279181, 1994.

Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A. and Vachaud, G.: Modeling water quality in an urban river using hydrological factors - Data driven approaches, J. Environ. Manage., 151, 87–96, doi:10.1016/j.jenvman.2014.12.014, 2015.

CRAWFORD and H., N.: Digital Simulation in Hydrology : Stanford Watershed Model IV., Stanford Univ. Tech. Report., 39 [online] Available from: http://ci.nii.ac.jp/naid/10007403485/en/ (Accessed 12 July 2021), 1966.

Devia, G. K., Ganasri, B. P. and Dwarakish, G. S.: A Review on Hydrological Models, Aquat. Procedia, 4(Icwrcoe), 1001–1007, doi:10.1016/j.aqpro.2015.02.126, 2015.

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

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J. and Hochreiter, S.: Rainfall–runoff prediction at multiple timescales with a single long short-term memory network, arXiv, 2045–2062, doi:10.5194/hess-2020-540, 2020.

Gauch, M., Mai, J. and Lin, J.: The proper care and feeding of CAMELS: How limited training data affects streamflow prediction, Environ. Model. Softw., 135, 0–2, doi:10.1016/j.envsoft.2020.104926, 2021.

Ghumman, A. R., Ghazaw, Y. M., Sohail, A. R. and Watanabe, K.: Runoff forecasting by artificial neural network and conventional model, Alexandria Eng. J., 50(4), 345–350, doi:10.1016/j.aej.2012.01.005, 2011.

Grayman, W. M.: Water-related disasters: A review and commentary, Front. Earth Sci., 5(4), 371–377, doi:10.1007/s11707-011-0205-y, 2011.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R. and Schmidhuber, J.: LSTM: A Search Space Odyssey, IEEE Trans. Neural Networks Learn. Syst., 28(10), 2222–2232, doi:10.1109/TNNLS.2016.2582924, 2017.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9(8), 1735–1780, doi:10.1162/neco.1997.9.8.1735, 1997.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N. and Lou, Z.: Deep learning with a long short-term memory networks approach for rainfall-runoff simulation, Water (Switzerland), 10(11), 1–16, doi:10.3390/w10111543, 2018.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M.: Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22(11), 6005–6022, doi:10.5194/hess-22-6005-2018, 2018.

Krause, P., Boyle, D. P. and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Adv.

Geosci., 5, 89–97, doi:10.5194/adgeo-5-89-2005, 2005.

Liang, X., Wood, E. F. and Lettenmaier, D. P.: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, Glob. Planet. Change, 13(1–4), 195–206, doi:10.1016/0921-8181(95)00046-1, 1996.

Lipton, Z. C., Berkowitz, J. and Elkan, C.: A Critical Review of Recurrent Neural Networks for Sequence Learning, , 1–38 [online] Available from: http://arxiv.org/abs/1506.00019, 2015.

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P. and Lettenmaier, D. P.: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions, J. Clim., 26(23), 9384–9392, doi:10.1175/JCLI-D-12-00508.1, 2013.

Montanari, A.: Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, Water Resour. Res., 41(8), 1–13, doi:10.1029/2004WR003826, 2005.

Neitsch, S. ., Arnold, J. ., Kiniry, J. . and Williams, J. .: Soil & Water Assessment Tool Theoretical Documentation Version 2009, Texas Water Resour. Inst., 1–647, doi:10.1016/j.scitotenv.2015.11.063, 2011.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19(1), 209–223, doi:10.5194/hess-19-209-2015, 2015.

Ömer Faruk, D.: A hybrid neural network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell., 23(4), 586–594, doi:10.1016/j.engappai.2009.09.015, 2010.

Panagoulia, D. and Dimou, G.: Sensitivity of flood events to global climate change, J. Hydrol., 191(1–4), 208–222, doi:10.1016/S0022-1694(96)03056-9, 1997.

Sahoo, G. B., Ray, C. and De Carlo, E. H.: Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream, J. Hydrol., 327(1–2), 94–109, doi:10.1016/j.jhydrol.2005.11.012, 2006.

Sherstinsky, A.: Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, Phys. D Nonlinear Phenom., 404, 132306, doi:10.1016/j.physd.2019.132306, 2020.

Sivapragasam, C., Liong, S. Y. and Pasha, M. F. K.: Rainfall and runoff forecasting with SSA-SVM approach, J. Hydroinformatics, 3(3), 141–152, doi:10.2166/hydro.2001.0014, 2001.

Solomatine, D. P. and Ostfeld, A.: Data-driven modelling: Some past experiences and new approaches, J. Hydroinformatics, 10(1), 3–22, doi:10.2166/hydro.2008.015, 2008.

Sood, A. and Smakhtin, V.: Revue des modèles hydrologiques globaux, Hydrol. Sci. J., 60(4), 549–565, doi:10.1080/02626667.2014.950580, 2015.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R. and Cook, R. B.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Data set., Oak Ridge Natl. Lab. Distrib. Act. Arch. Center, Oak Ridge, Tennessee, USA., 2014.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y. and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, J. Geophys. Res. Atmos., 117(3), doi:10.1029/2011JD016048, 2012.

Xiang, Z., Yan, J. and Demir, I.: A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning, Water Resour. Res., 56(1), doi:10.1029/2019WR025326, 2020.