An optimized LSTM-based approach applied to early warning and forecasting of ponding in the urban drainage system

Wen Zhu[1], Tao Tao[1], Hexiang Yan[1], Jieru Yan[1], Jiaying Wang[1], Shuping Li[1], Kunlun Xin[1]

[1]College of Environmental Science and Engineering, Tongji University, Shanghai, 200000, China

Correspondence to: Tao Tao (taotao@tongji.edu.cn), Jieru Yan (yan_jieru@tongji.edu.cn)

Abstract. In this study we propose aAn optimized LSTM-based approach which is applied to early warning, and forecasting of ponding in the urban drainage system is proposed in this study. This approach can quickly identify and locate pondinglocations and the ponding process with relatively high accuracy. Based on the approach, a model is developed, which The model is constructed with by two tandem processes and utilizesintroduces a multi-task learning mechanism. The superiorityThe results of the developed model was demonstrated by comparingare compared with those of two widely used neural networks (LSTM, CNN) to validate their advantages. Then, the model is was further revised with available monitoring data in the study area to achieve higher accuracy. We also discussed hHow the number of selected monitoring points influences influenced the performance of the corrected model's performance is also discussed. In this study, oOver 15000 designed rainfall events are were used for model training, covering various extreme weather conditions.

1 Introduction

The intensity and frequency of urban floods are growing as a result of the increased frequency of extreme weather, rapid urbanization, and climate change (Hossain Anni et al., 2020; Guo et al., 2021; Huong and Pathirana, 2013). It is becoming increasingly clear that urban floods significantly impact city management and endanger the safety of people's life and property. The ability to reliably characterize and forecast urban floods and generate high-precision flood risk maps has become critical in flood mitigation and decision-making.

The most common approach to simulating urban floods is to develop a hydrodynamic model (i.e. storm-inundation simulation). It, which utilizesuses the collected topographic map, information on the pipe network, historical rainfall data, monitoring data, and other information in the study area (Jamali et al., 2018; Aryal et al., 2020; Balstrøm and Crawford, 2018; Tian et al., 2019). However, a realistic hydrodynamic model for continuous simulation requires vast data. S, such as comprehensive information on topography, infiltration conditions, and sewage system data, (including exact locations, depths, and diameters of sewage pipes)., However, the above dataall of which are difficult to obtain, especially in metropolitan areas (Rahman et al., 2002; Kuczera et al., 2006). Furthermore, calculation in storm-inundation simulation is sophisticated, and, often computationally intensive, whichand takes a long time to execute. The most detailed representation of the storm-inundation simulation is the 1D-2D model (Djordjević et al., 1999; Djordjević et al., 2005). It which

summarizes the dynamic interaction between the flow that enters the underground drainage network and the overloaded flow that spreads to the surface flow network during high-intensity rainfall. ~~Some of the available models that realize storm inundation simulation~~Representatives of such model include XPSWMM, TUFLOW, and MIKE FLOOD (Leandro and Martins, 2016; Teng et al., 2017; Zhang and Pan, 2014).

The lack of underlying information has~~The above problems have -~~ hampered the continuous development of hydrodynamic models in urban flood forecasting. As a result, deep learning has emerged as another viable forecasting tool. Deep learning is a form of training that involves creating a unique dataset, and then ~~analyzing~~analysing and forecasting urban flooding using algorithms (Mudashiru et al., 2021; Sit et al., 2020; Shen, 2018; Moy De Vitry et al., 2019). It can compensate for ~~the impact of actual~~ data scarcity by training on a large designed data set. And unlike traditional hydrodynamic models, deep learning~~It~~ does not require any assumptions on the physical processes behind it ~~as traditional hydrodynamic models require~~.

However, some factors need to be improved in order to promote the application of deep learning in urban flood forecasting. Firstly, the data-set for training needs to be ~~revised~~ enriched to reflect the superiority of the approach. There are studies utilizing deep learning algorithms for urban flood forecasting, but the developed model is trained on a small number of samples. For example, Cai and Yu (2022) used ~~only~~ 25 historical floods for forecasting ~~calculations, and.~~ Abou Rjeily et al. (2017) used only 10~~ten~~ rainfall events for training and verification, which was insufficient ~~need to be revised~~ to reflect the characteristics of rainfall distribution. Secondly, monitoring equipment is expensive and thus not frequently available. Therefore, researchers have to rely on simulations produced from hydrodynamic models, ~~without regard for the accuracy of the models.~~however, often without considering the accuracy of the models. For example, Chiang et al. (2010) used synthetic data from the SWMM model as the target values to train the recurrent neural network (RNN),~~.~~ a~~A~~nd ~~they~~ compared the predictions with simulation results to evaluate the model~~'s~~ accuracy in estimating water levels at ungauged locations. Thirdly, some studies have focused on building more complex deep-learning architectures to improve ~~the~~ model~~'s~~ performance. Such as the automatic encoder (Bai et al., 2019), encoder-decoder (Kao et al., 2020c), and customized layers based on Long Short-Term Memory (LSTM) (Sit et al., 2020; Kratzert et al., 2019; Kratzert et al., 2019). For example, an encoder-decoder LSTM has been proposed for runoff forecasting up to 6 hours and 24 hours ahead (Xiang et al., 2020b; Kao et al., 2020c). Nevertheless, the urban flooding forecasting tasks with multiple time steps are mainly based on the precipitation forecast hours in advance, which is not available in this paper. With the real-time rain data in a short duration, getting enough data like the continuing runoff data to support the hours ahead prediction is yet to be available.

In this study, we propose ~~This study proposes~~ an optimized LSTM-based approach, which is applied to early warning and forecasting of ponding in ~~the~~ urban drainage system. This approach can quickly identify and locate ponding ~~locations and the ponding process~~ with relatively high accuracy. The model ~~was~~ is constructed by~~with~~ two tandem processes and introduce~~d~~ a multi-task learning mechanism. The~~Its~~ evaluation results of the model were compared with those of two widely used neural networks (LSTM, CNN) ~~to validate its advantages~~. The model was further revised with monitoring data in the study area to achieve higher accuracy. ~~This paper~~ We also discussed the influence of the number of monitoring points selected on the

model's performance in the model updating procedure. Over 15000 designed rainfall events were used for model training, covering various extreme weather conditions.

The rest of the paper is organized as follows: Section 2 introduces the methodology used to develop, including the proposed LSTM-based modelingmodelling framework, as well as; this part also explains the experimental setup and application of the model;. Sections 3 presents the results of the model. and Section 4 presents discussion and discuss the results, respectively, and; and finally, Section 5 concludes this paper by drawings brief conclusions.

2 Methodology

2.1 LSTM-based model

When existing hydrodynamic models work, different mechanisms are used to calculate surface runoff and pipeline confluence. Like a hydrodynamic model, which is generally composed of two processes: runoff process and flow confluence process, tThe LSTM-based model proposed in this study is also constructed with two stages, the runoff process and the flow confluence process, to reduce the computational burden of this data-driven model. Fig. 1 illustratespresents the model architecture from input (i.e. rain intensity) to output (i.e. ponding volume of each node) for runoff and flow confluence processes.

The two processes are in tandem: the inputs of the flow confluence process are inherited and concatenated from the outputs of all nodes in the runoff process. However, during the training process, the two processes are trained separately and without mutual interference when as the inputs and outputs of both processes are produced from a hydrodynamic modelthe simulated results.

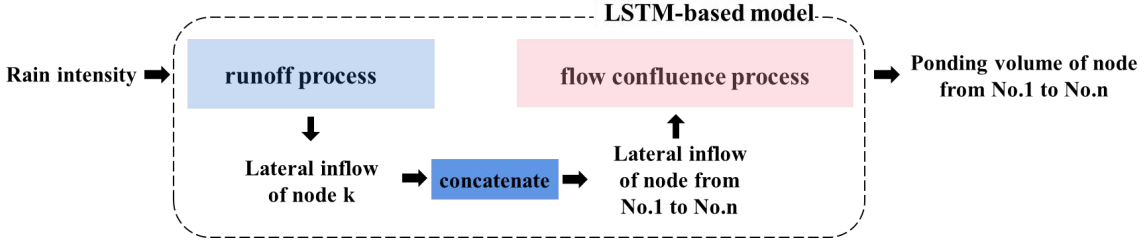


Figure 1: The architecture of the LSTM-based model.

2.1.1 Runoff process

With a general understanding of the a hydrodynamic models, the runoff process involves surface runoff and infiltration,. while tThe most important influential factor is rainfall. As aThe mass rainfall curve can reflects the characteristics of a specific rainfall process. Thus, it can be directly used as the input of a neural network. The output of the neural network (i.e. lateral inflows at each node) reflects the hydraulic state in the runoff process. Meanwhile, lateral inflows at each node as the outputs reflect the hydraulic state in the runoff process.

Fig. 2 illustratesdetails the runoff process's training, validationerification, and testing processesprocedures in the runoff process. As illustrated shown in Fig.2, a training set with two time-series data is fed into the neural network: , i.e., rainfall intensity and lateral inflows that enterat each nodech node. At each epoch, use four indicators are used to evaluate the consistency between the predicted lateral inflows and the simulation from the hydrodynamic model. If the model converges, evaluate the network is further evaluated on the test set. Otherwise, the next training epoch is startedgo to the next epoch for training.

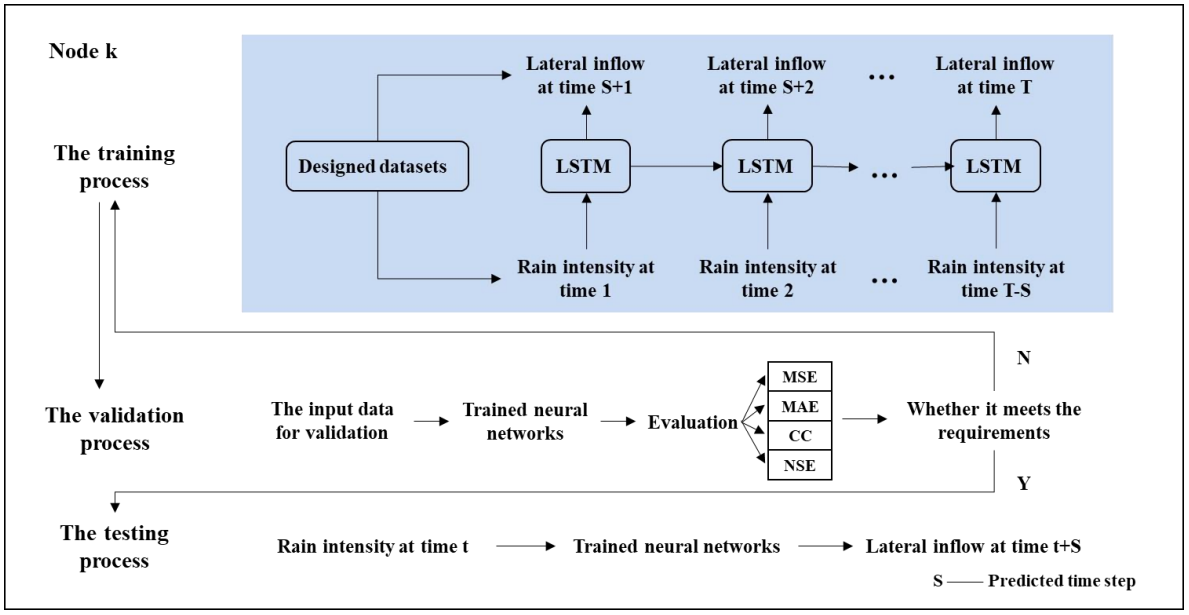


Figure 2: The training, verificationvalidation, and testing procedures in used when developing the LSTM-based runoff emulator.

In the figure, (MAE is short for-- Mean Absolute Error, MSE is -- Mean Squared Error, CC is the-- Correlation coefficient, and NSE --is the Nash-Sutcliffe efficiency coefficient). S represents the advanced time step for prediction fixed at 5min in this paper because of the interval for the measured data obtained.

2.1.2 Flow confluence process

The flow confluence process is set up in the same manner as with the understanding of the simulation process of a hydrodynamic model (e.g., the SWMM model). If we compare the urban drainage system to is regarded as a black box, onlymainly the lateral inflows at each node and outflows from the outlets enter and leave the system, respectively (Archetti et al., 2011). Therefore, the flow confluence process inputs are the lateral inflows at each node and the outflows. If the a free outflow condition is consideredsare set free, namely the hydraulic state behind the outlets has little influence on the interior of the ease areasystem, then the inputs of the flow confluence process are only the lateral inflows at each node. Thus the inputs could remove them.

Fig. 3 illustrates the details of the network architecture in the flow confluence process.

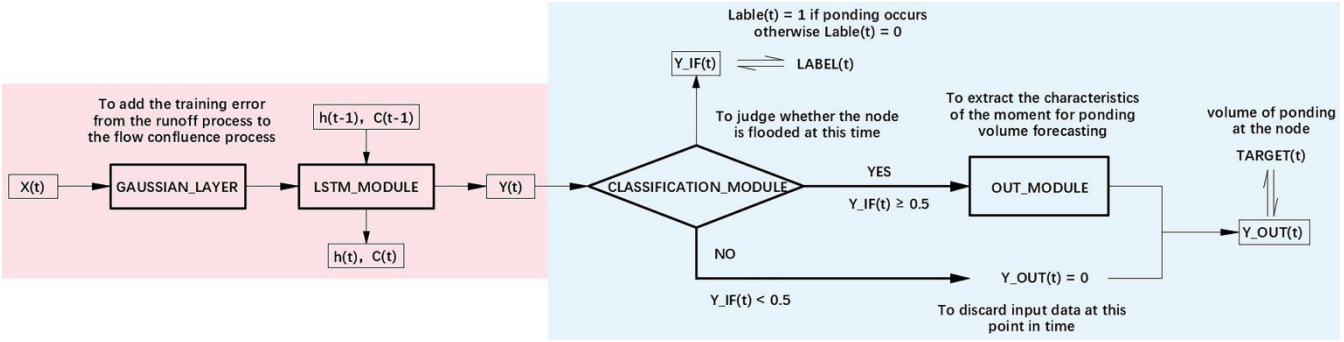
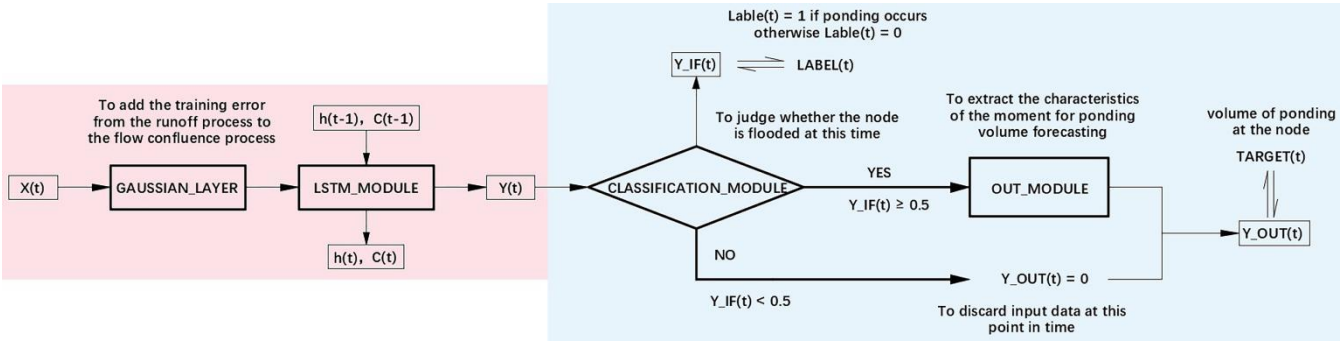


Figure 3: The network structure of the flow confluence process (for a single node).

As illustrated in the pink block in Figure 3, ~~First,~~ a Gaussian layer is added after the input layer in the flow confluence process during training. The gaussian layer serves as a filter to compensate for the inaccuracy of the prediction (by the hydrodynamic model) in the runoff process. The model is trained to minimize the differences between the predictions (from the neural network, i.e. the output from the runoff process) and the simulations (from the hydrodynamic model). ~~is used to avoid interference caused by the predictions in the runoff process. The flow confluence process is trained using simulations, however, the inputs will be the predictions by the runoff process in application. The difference between the predictions and simulations needed to be compensated during training. Sect. 2.2 introduces the calculation about how the error propagates in detail.~~ Then (as illustrated in the blue block in Figure 3), a classification layer ~~task is introduced as an auxiliary task to ponding volume forecasting. The classification module~~ is added after the outputs of ~~each time step in~~ the LSTM module to judge whether ponding occurs at the time step. Only ~~if~~ when ponding ~~has occurred~~occurs at the time step, can the output of the LSTM module ~~at the time step~~ enters the 'OUT_MODULE' ~~(as shown in Fig. 3)~~ to continue with the learning. Otherwise, ~~discard~~ the output of the LSTM module at this time step is discarded. In this way, the interference of the time points without ponding on the ponding volume forecasting is eliminated to the ~~greatest~~ a great extent. The higher the classification accuracy is, the more accurate the prediction of ponding volume will be. Moreover, the multi-task learning has a hard parameter sharing mechanism ~~of parameters~~, which ~~. It could also~~ effectively alleviate the over-fitting of the model. The parameters ~~layers~~ in the ~~whole~~ 'LSTM_MODULE' ~~module~~ (~~,~~ including the parameters of the LSTM layers, batch normalization~~BN~~ layers, activation functions, ~~and~~ etc.), are shared by the 'CLASSIFICATION_MODULE' ~~classification~~ and 'OUT_MODULE' ~~out modules~~.



~~**Figure 3: The network structure in of the flow confluence process (for a single node).**~~

2.2 Error transmission

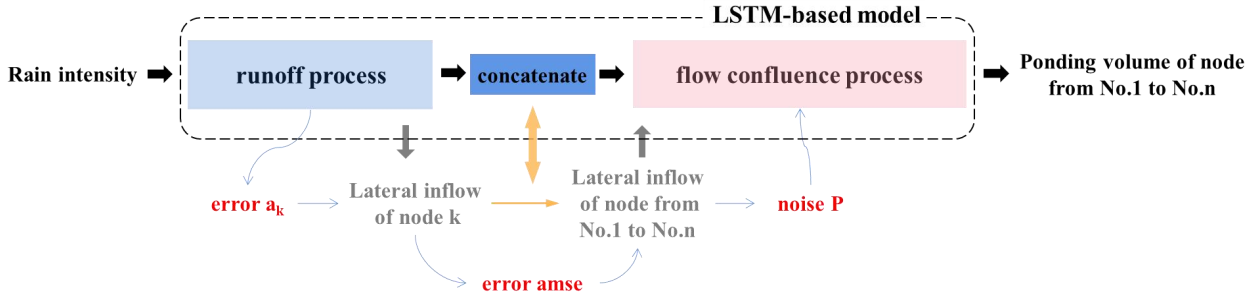


Figure 4: The error transmission during training from the runoff process to the flow confluence process.

Fig. 4 illustrates how the training error of ~~from~~ the neural network propagates from~~in~~ the runoff process ~~propagates~~ to the flow confluence process during training. Noise P ($P \sim N(0, p^2)$) is added to the lateral inflows before feeding the data into the neural network in the flow confluence process~~. It will~~ in order to avoid the interference caused by the training error in the runoff process and also ~~effectively~~ to alleviate the over-fitting of the neural network. The magnitude of noise P can be determined as follows~~:~~.

1 The mean squared error (MSE) is used to characterize the training error in the runoff process~~s, where~~s~~. Compute~~ the error at node K can be computed by~~as follows:~~

$$a_k = \frac{\sum_{i=1}^{T}\sum_{j=1}^{S}\left(\hat{X}_{ij} - X_{ij}\right)^2}{T \cdot S} \quad (1)$$

where T represents the duration of event j in min, S represents the number of events in the training data, $\hat{X}_{ij}$ represents the simulated lateral inflow at node K ~~of~~ at the i-th time step in the j-th rainfall event~~s~~ in $L \cdot s^{-1}$~~L/s~~, X_{ij}~~X$_{ij}$~~ represents the output of the runoff process at node K ~~of~~ at the i-th time step in the j-th sample event in $L \cdot s^{-1}$~~L/s~~.

2 Then compute the average mean squared error of all nodes ~~as follows:~~by

$$amse = \frac{\sum_{k=1}^{N} a_k^2}{N} \quad (2)$$

where N represents the number of nodes.

3 Then convert amse into noise percentage ε with the mean value of the predicted lateral inflows at all nodes in the training set by. ~~Eq. (3) is the definition of noise percentage(ε) and the relationship between amse and the noise percentage:~~

$$\varepsilon = \sqrt{\frac{P_N}{P_S}} = \sqrt{\frac{\sum_{k=1}^{N}\sum_{i=1}^{T}\sum_{j=1}^{S}(\hat{X}_{kij}-X_{kij})^2}{\sum_{k=1}^{N}\sum_{i=1}^{T}\sum_{j=1}^{S}(X_{kij})^2}} \le \sqrt{\frac{\sum_{k=1}^{N}\sum_{i=1}^{T}\sum_{j=1}^{S}(\hat{X}_{kij}-X_{kij})^2}{\frac{1}{N\cdot T\cdot S}(\sum_{k=1}^{N}\sum_{i=1}^{T}\sum_{j=1}^{S}X_{kij})^2}} = \frac{\sqrt{amse}}{\frac{1}{N\cdot T\cdot S}\sum_{k=1}^{N}\sum_{i=1}^{T}\sum_{j=1}^{S}X_{kij}} \quad (3)$$

~~Where~~ where P_S represents signal power, and P_N represents noise power.

4 ~~Add~~ Finally add noise P to the inputs (X) in the flow confluence process during the training process. Namely~~Firstly~~, generate a set of random numbers G with the length of X ~~by~~ using Pseudorandom Number Generator, where G~~which~~ obeys a normal distribution (G~N(0,1)), i.e., $P = p \cdot G$, where p is computed by~~.~~

$$p = \varepsilon \cdot \sqrt{\frac{1}{T}\sum_{i=1}^{T}(X_i)^2} \quad (4)$$

2.3 Model correction system

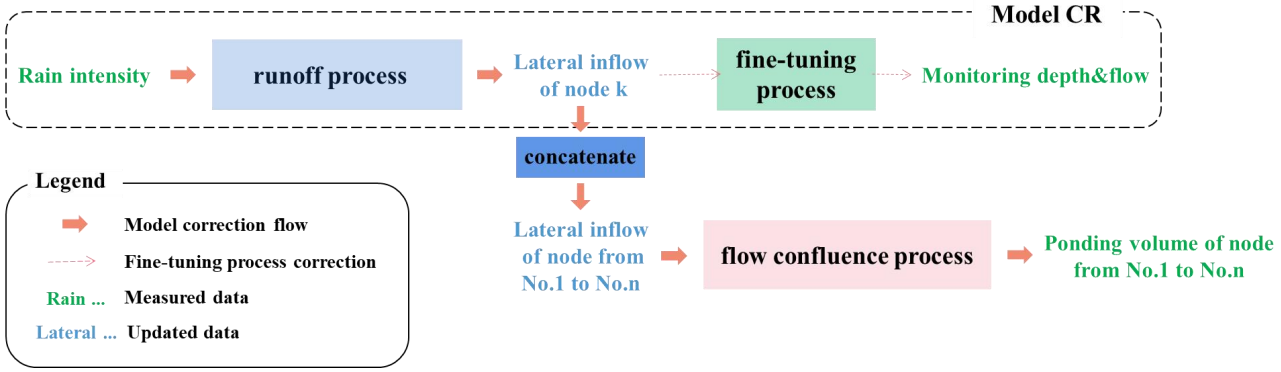


Figure 5: Model correction system. CR is abbreviated for correction of the runoff process.

The LSTM-based model is built ~~constructed~~ based on the simulation results of a relatively accurate hydrodynamic model. However, the ~~the~~ differences between the simulation from the hydrodynamic model at the monitoring points and the obtained ~~actual measured~~ monitoring data always exist ~~persists~~ during the operation of the pipe network ~~operation. It,~~ which leads to a discrepancy between the predicted results from the proposed LSTM-based model and the actual situation. Thus it is necessary to c~~C~~orrect~~ing~~ the model using~~with~~ the measured level and flow data at the monitoring points ~~will alleviate this problem~~, and how to revise the model properly using the available data is also ~~. Revising the model with the measured data is~~ one of the focuses of this study. Fig. 5 ~~indicates~~ describes the model correction process using the measured rainfall data, ~~monitoring~~ depths and flows at~~obtained from~~ the monitoring points, and ponding data at any node. Specifically, t~~T~~he LSTM-based model is corrected ~~with~~ using the following two steps:

1. The runoff process is corrected with the measured rain, level, and flow data referring to the transfer learning (PAN, S J, et al., 2010). Transfer learning is mainly used to transfer the knowledge of one domain (source domain) to another domain (target domain) sucho that the target domain can achieve better learning effects.

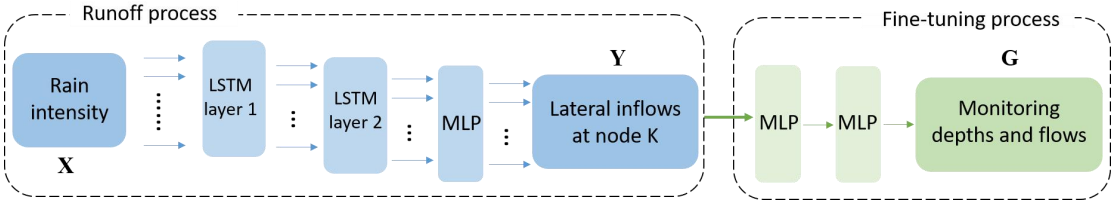


Figure 6: The architecture of Model CR. (MLP -- Multi-Layer Perception)

Fig. 6 shows the schematic of Mthe model CR. It migrates the network structure inof the runoff process from rain data (X) to lateral inflows (Y) to the input-output connection between X and monitoring data (G). And add multiple fully connected layers after the output layer of Y. Model CR is designed to updatebuilt for updating the runoff process in the primary LSTM-based model. The correction has two stepsis divided into two processes: training and updating. Firstly, the model CR is trained based on a pre-trained mapping between from X toand Y (as shown in Fig.6). Then, it is updated on pairs of measured rain data, and measured monitored water depths and flows.

2. The flow confluence process is corrected usingwith the updated lateral inflows of all concatenated nodes and the measured ponding volume.

2.4 Case study

2.4.1 Study area

The LSTM-based model trains nodes in the pipe network one by one. Namely N sub-models. with the same architecture are generated, where N is the number of nodes in the system. Both in the runoff process and flow confluence process, these sub-models are trained separately.the training process of different nodes will not interfere with each other. Due to thisese structural characteristics, the size of the case area doeswill not limit the model's performance. Regarding the model structure, the output of the runoff process's output is the lateral inflow at a single node's lateral inflow. Likewise, The flow confluence process is similar to the runoff process, where the output of the flow confluence process is the the ponding volume at a single node. Regardless of the size of the pipe networkcase area, the output of the model is in at each nodes. The size of the case area directly affects the network scale, i.e., the number of nodes. However, aA large-scale pipe networkcase area with lots ofmore nodes will significantly increase the time spent training the model and also require extra processing power. However, it will maintain its accuracy in ponding forecasting.

To verify the feasibility of the modelling method framework above, a small-scale case area, JD,, a residential district in S city, is selected as the study area. Fig. 7(a) shows the elevation map of the study area. There are 32 residential buildings in the district, with a total area of 6.128hm^2. The study area is separated from the municipal roads by walls with three entrances

on the community's north, east, and west sides. Rain pipes in the study area are circular pipes with 200mm, 300mm, 400mm, 500mm, orand 600mm, most 300mm in diameter (mostly 300mm). The total length of this pipe network is 5.5 km. The network contains 336 nodes and 340 pipes, and is connected to the municipal pipe networks through 4 outlets, as denoted by the green trianglein green in Fig. 7(b). There are 15 level gauges and three flowmeters in the current pipe network. The layout of monitoring points is also shownsn in Fig. 7(b).

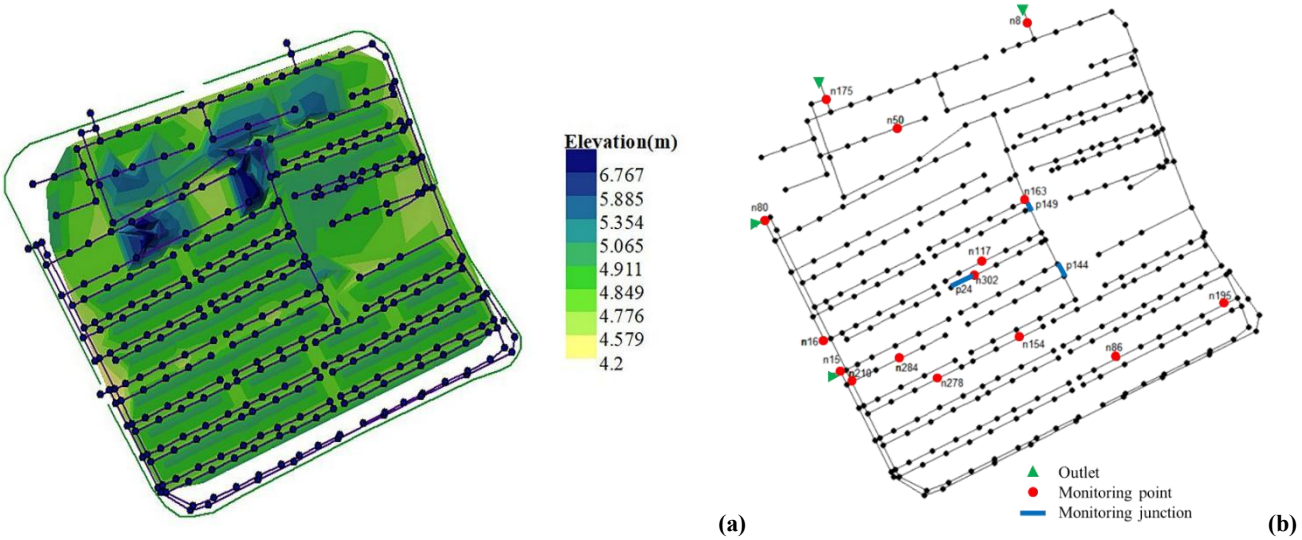


Figure 7: Study area - JD residential district. (a) The elevation map and stormwater system in the case area. (b) Locations The layout of the monitoring points in the case area.

2.4.2 Rainfall data

Eq. (5) computed Tthe rainstorm intensity for S city is designed using Eq. (5), which. iIt was obtained according to a universal design storm pattern proposed by Keifer&Chu. The storm pattern is broadly used both at home and abroad proposed by Keifer&Chu, and. The storms the generated storms are usuallywere extreme enough to reflect the state of the pipe networks under the most unfavorable conditions (Skougaard Kaspersen et al., 2017)..

$$q = \frac{167A(1+C\log P)}{(t+b)^n} = \frac{1600(1+0.846\log P)}{(t+7.0)^{0.656}} \quad (5)$$

where q is the rainstorm intensity in $L \cdot s^{-1} \cdot hm^{-2}$ P is the reappearing rainfall period in a, t is the duration of rainfall in min, and A, C, b, and n are parameters of the rainstorm intensity design formula.

The rainstorm intensity before or after the peak is determined using Eq. (6).

As shown in Eq. (6), tThe rainstorm intensity before or after the peak shown in Eq. (6) was determined by referring to the local rainstorm intensity formula, design storm pattern, and historical rain data, respectively.

$$\begin{cases} I(t_b) = \dfrac{A(1 + C\lg P)\left[\dfrac{(1-n)}{r}t_b + b\right]}{\left(\dfrac{t_b}{r} + b\right)^{n+1}} \\ I(t_a) = \dfrac{A(1 + C\lg P)\left[\dfrac{(1-n)}{1-r}t_a + b\right]}{\left(\dfrac{t_a}{1-r} + b\right)^{n+1}} \end{cases} \quad (6)$$

where t_b and t_a are the time before and after the peak in min, respectively, and r is the rainfall peak coefficient.

Then single-peak rainfall scenarios were~~were~~ constructed unevenly by using different rainfall reappearing periods (P) ranging from 0.5a to 100a, peak coefficient (r) ranging from 0.1 to 0.9, and duration (T) ranging from 60 to 360 min.

$$q = \frac{167A(1 + C\log P)}{(t+b)^n} = \frac{1600(1 + 0.846\log P)}{(t+7.0)^{0.656}} \quad (5)$$

~~where q is the rainstorm intensity in $L \cdot s^{-1} \cdot hm^{-2}$, P is the reappearing rainfall period in a, t is the duration of rainfall in min, and A, C, b, and n are parameters in of the rainstorm intensity design formula.~~

$$\begin{cases} I(t_b) = \dfrac{A(1 + C\lg P)\left[\dfrac{(1-n)}{r}t_b + b\right]}{\left(\dfrac{t_b}{r} + b\right)^{n+1}} \\ I(t_a) = \dfrac{A(1 + C\lg P)\left[\dfrac{(1-n)}{1-r}t_a + b\right]}{\left(\dfrac{t_a}{1-r} + b\right)^{n+1}} \end{cases} \quad (6)$$

~~where t_b and t_a are the time before and after the peak in min, respectively, and r is the rainfall peak coefficient.~~

In addition to single-peak rainfall scenarios, we also considered bimodal rainfall scenarios. A~~a~~ccording to the historical bimodal rainfall data in S city, the rainfall peaks corresponding to the bimodal design storm pattern ~~in~~ with the duration ~~of rainfall~~ from 60 to 360 min could be computed by Pilgrim & Cordery. Pilgrim & Cordery ~~was~~ is a method to count the historical rainfall data and deduce the rainstorm pattern from it (Pilgrim & Cordery, 1975). ~~This method divided the duration into several periods. For each rainfall event, determine the sequence number of each period according to the rainfall in each period from large to small, where significant rainfall corresponded to a small sequence number. Then, average the serial numbers of each period, calculate the percentage of each rain to the total rainfall in each period, and take the average rate in each period.~~

~~Then, construct double-peak rainfall scenarios where the reappearing period was from 0.5a to 100a.~~

Table 1: The bimodal design storm pattern.

(a) ~~Duration is~~ 60 min~~utes duration~~.

T(5min)	P/Pmax(%)
1	3.19

T(5min)	P/Pmax(%)
2	16.71
3	8.74
4	1.52
5	2.27
6	4.02
7	5.89
8	22.68
9	10.51
10	12.76
11	7.05
12	4.68

(b) ~~Duration is~~ 120 min duration~~.~~

T(5min)	P/Pmax(%)
1	0.74
2	1.76
3	6.42
4	3.75
5	2.05
6	1.52
7	2.73
8	4.43
9	9.23
10	11.17
11	17.57
12	13.81
13	7.76
14	5.3
15	2.38
16	1.13
17	3.17
18	1.32
19	0.98

T(5min)	P/Pmax(%)
20	0.84
21	0.54
22	0.64
23	0.44
24	0.33

Table 1 shows the bimodal design storm patterns with ~~respectively in~~ 60 min and 120 min duration time, respectively~~.~~, where P/Pmax ~~in the table~~ represent~~sed~~ the distribution~~s~~ of ~~the~~ rainfall intensity over time (with 5 min~~. The~~ unit period~~) was 5 minutes~~. Then, double-peak rainfall scenarios were constructed according to Table 1 using reappearing periods ranging from 0.5a to 100a.

The produced s~~S~~ingle-/~~ and~~ double-peak rainfall data ~~produced~~ were then added with Gaussian white noise (produced according to the procedures described in Section 2.2) to ~~. It~~ ensured that the obtained datasets contains~~ed~~ enough~~a majority of~~ extreme ~~rainfall event~~conditions. Take the rainfall ~~in~~ with a~~the~~ return period of 5a as an example. Fig. 8 shows the ~~process~~ effect of adding noise~~.~~, where the subfigure ~~Fig. 8~~(1) shows~~ed~~ the randomly generated Gaussian white noise ~~generated randomly~~ over the duration, the subfigure ~~Fig. 8~~(2) shows ~~was~~ the distribution of reordered white noise, and the subfigure ~~Fig. 8~~(3) zooms in on the part circled in (2). The subfigures~~showed the details around the rainfall peak in Fig. 8(2), Figs.~~ 8(4) - (6) show~~were~~ the design rainfalls after adding 30%, 50%, and 70% white noise, respectively. Specifically, we have limited the noises near the rainfall peak, i.e., only negative noises are allowed there.~~In Fig. 8(2), adjust the noise near the peak artificially to be harmful to smooth the extreme maximum rainfall, so the rainfall was closer to the actual situation.~~

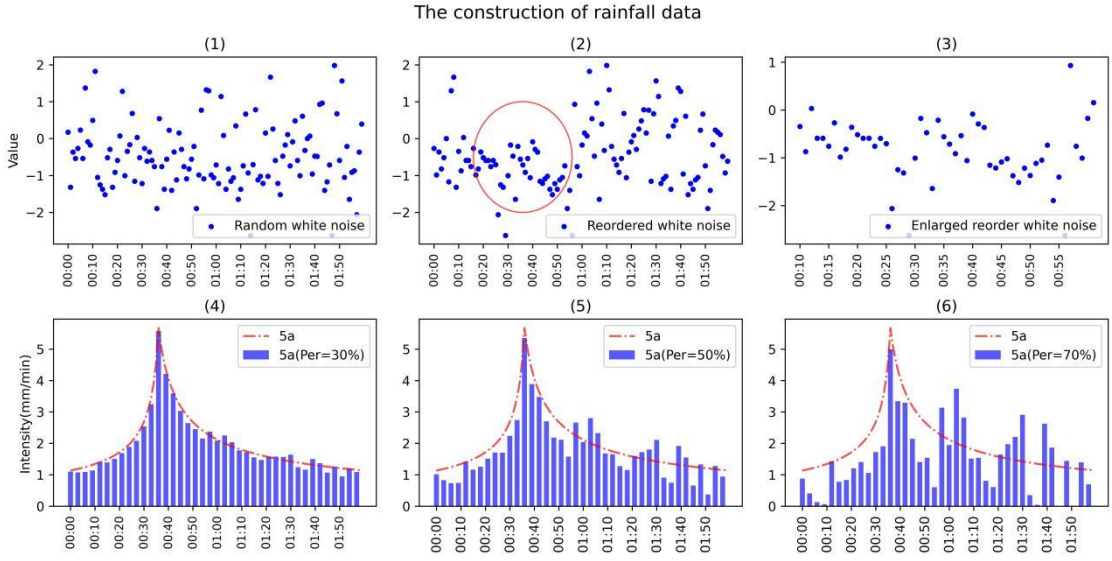


Figure 8: **A demonstrative example to show the effect of adding white noise.~~The construction of the design rainfall data.~~**

In this study, tThe noise percentages went from 0 to 100% withat an incrementinterval of 10% to blur the characteristics of the design storm pattern and intensify the extreme conditionsbring the rainfall closer to reality. The synthetic dataset contained a total of 16960 rainfall events, and . Furthermore, the ratios of the training, validation, and test sets were 80%, 10%, and 10%, respectively.

In general, a small training set will normally leads to a poor approximation effect. Thus. aA sort of convergence test waisperformed to evaluate how much data is was sufficient required for the proposed LSTM-based model trainingto obtain the desired approximation effect. In Fig. 9, Tthe model performances using different sizes amount of training data wereare compared, as shown in Fig. 9. When the data size is was reduced to 2/3 of the origin volume, the model performance reaches fell down to 90% of the original.. And if the data size wasis reduced to halfhalved, less than 80% of the origin model performance remainsed.

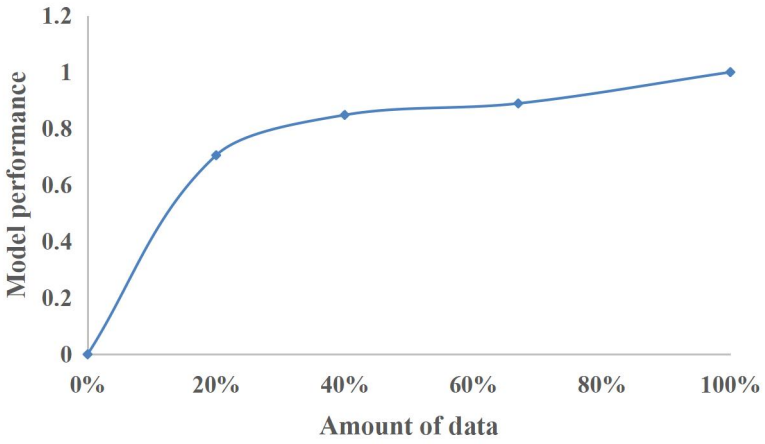


Figure 9: The learning curve which describesto desor ibe the relationship between model performance and data volume.

2.4.3 Simulated and measured data

A hydrodynamic model was established for the case pipe network. The simulation results (i.e., the lateral inflows and the volume of ponding at each node, as well as the level and flow data at the monitoring points) were obtained using the constructed rainfall events described in Sect. 2.4.2. The constructed rainfall events were used in the training, validation, and testing processes. In the simulation process, we considered a uniform rainfall distribution in space. A simplified representation of the sewer system and a constant, uniform infiltration rate in the green area were also considered for runoff computation (Roland L¨owe et al.,2021). Meanwhile, this paper does notwe did not consider the two-dimensional surface overflow process.

Besides, the measuredactual rain data and the monitoring data (measured water depthh and flow) data at the monitoring pointsof in the past five historical rainfall events were used to verify the performance of the corrected model. The uncertainty of the measurements was not considered (Huong and Pathirana, 2013) — for example, errors in data transmission of data monitoring equipment.. In this study, weThis paper considered the simulatiothe simulationn results of the verified hydraulic model as the ground truth.hcredible. Thus, in the process of model correction, we only fine-tuned the trained weight parameters without requiring too much measured data.

In the model correction process, the uncertainty of the measurement (i.e., rainfall data, water depth, and flow data at monitoring points) was not considered (Huong and Pathirana, 2013) — for example, errors in data transmission of data monitoring equipment.

Table 2 shows the measurements of the five historicalmeasured rainfall events used in the process of model correction system. Among the five events, tThree were used to correct Mthe model CR and the flow confluence process, and the other twos were used to evaluate the reliability of the approach's reliability to update the model.

Table 2: 5 historical rainfall data used to correct the LSTM-based model.

Dataset	Rainfall event	Rainfall (mm)	Max. rain intensity (mm/min)	Duration (min)
Training	No.1	494.50	8.13	180
	No.2	146.63	2.61	90
	No.3	254.51	3.61	240
Testing	No.4	442.61	7.26	150
	No.5	254.41	4.97	120

2.4.4 Model Construction

Table 3: Hyper-parameters configuration in the model setup and correction processes.

	Hyper-parameters	Model CR		Flow confluence process
		Runoff process	Fine-tuning process	
Model setup	Normalization	Z-score	Z-score	Min-Max
	Batch size	150	150	150
	Epoch	300	300	300
	Learning rate	1e-2	5e-3	1e-2
	Optimizer	Adam	SGD	SGD
	LSTM hidden layer neurons	16	-	256
	MLP hidden layer neurons	16	1536/3072	256/128*

	Hyper-parameters	Model CR		Flow confluence process
		Runoff process	Fine-tuning process	
Model correction	LSTM layers	2	-	4
	MLP layers	1	2	2*
	Learning rate	1e-4		5e-5
	Optimizer	Adam		SGD

Note: * ~~represents~~ means to set~~ting~~ the hyperparameters of 'CLASSIFICATION_MODULE' and 'OUT_MODULE'~~the classification and out modules~~ in the flow confluence process to the same values.

~~Most of t~~The hyper-parameters used in this paper were mainly determined by Hyperopt (BERGSTRA, J et al., 2013). Hyperopt is a Python library for hyper-parameter optimization that adjusts parameters using~~by~~ Bayesian optimization~~,~~ ~~allowing obtaining optimal parameters for a given model~~. ~~The search space of hyper-parameters was set~~ determined ~~by trial and error first. Then, it received a random combination of hyper-parameters from the search space to do the training and returned the training loss. It used the built-in search algorithms like the tree of Parzen estimators (TPE) to determine the following hyper-parameters combination. Thus, these iterations gradually build an optimal hyper-parameters combination to get the lowest training loss.~~

Table 3 shows the hyper-parameters in the learning process of the model setup and model correction obtained by Hyperopt.~~Hyper-parameters in the learning process of the model setup and Table 3 show the model correction obtained by Hyperopt.~~

2.4.5 Performance evaluation

Mean Absolute Error (MAE), Mean Squared Error (MSE), Correlation coefficient (CC) and Nash-Sutcliffe efficiency coefficient (NSE)~~MAE, MSE, NSE, and CC~~ are broadly ~~always~~ used indicators to assess the performance of a data-driven model~~s~~. In this study, ~~This paper~~we used MAE and MSE to quantify the size of the errors, i.e., difference at each node between the predict~~ioned results~~ by the proposed LSTM-based model~~neural network~~ and simulation from the hydrodynamic model ~~at each node~~. Moreover, NSE and CC ~~here~~ were also used to evaluate the level of agreement at all nodes. Eq~~.~~uations (7)-(10) ~~described~~ list the formulas of these 4 indicators~~criteria~~.~~.~~

$$MAE = \frac{1}{DT}\sum_{s=1}^{D}\sum_{t=1}^{T}\left|Y_{st} - \hat{Y}_{st}\right| \quad (7)$$

$$MSE = \frac{1}{DT}\sum_{s=1}^{D}\sum_{t=1}^{T}\left(Y_{st} - \hat{Y}_{st}\right)^2 \quad (8)$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(\frac{1}{D}\sum_{s=1}^{D}Y_{st} - \frac{1}{D}\sum_{s=1}^{D}\hat{Y}_{st}\right)^2}{\sum_{t=1}^{T}\left(\frac{1}{D}\sum_{s=1}^{D}\hat{Y}_{st} - \frac{1}{DT}\sum_{t=1}^{T}\sum_{s=1}^{D}\hat{Y}_{st}\right)^2} \quad (9)$$

$$CC = \frac{\sqrt{\sum_{t=1}^{T}\left(\frac{1}{D}\sum_{s=1}^{D}Y_{st} - \frac{1}{DT}\sum_{t=1}^{T}\sum_{s=1}^{D}Y_{st}\right)\left(\frac{1}{D}\sum_{s=1}^{D}\hat{Y}_{st} - \frac{1}{DT}\sum_{t=1}^{T}\sum_{s=1}^{D}\hat{Y}_{st}\right)}}{\sqrt{\sum_{t=1}^{T}\left(\frac{1}{D}\sum_{s=1}^{D}Y_{st} - \frac{1}{DT}\sum_{t=1}^{T}\sum_{s=1}^{D}Y_{st}\right)^2}\sqrt{\sum_{t=1}^{T}\left(\frac{1}{D}\sum_{s=1}^{D}\hat{Y}_{st} - \frac{1}{DT}\sum_{t=1}^{T}\sum_{s=1}^{D}\hat{Y}_{st}\right)^2}} \quad (10)$$

where D is the number of events in the test set, T is the ~~total~~ number time steps of~~of~~ the relevant rainfall event, Yst is the prediction given ~~value predicted~~ by the neural network at the t-th time step in the s-th event, and $\hat{Y}_{st}$ is the simulation given by~~result from~~ the hydrodynamic model.

To evaluate the accuracy of the proposed model in predicting ponding ~~judgment of whether a node was flooding as predicted by the model,~~, we have introduced 56 indicators (as shown in Table 4): ~~showed some other criteria.~~ Accuracy (ACC), Precision (PPV), and False ~~omission~~ Omission ~~rate~~ Rate (FOR) ~~were used~~ to ~~verify~~ evaluate the ~~~~ model accuracy in predicting the occurrence of ponding at a single node~~of the time of occurrence and duration when ponding occurred in a rainfall event;~~. $S-PPV$ and $S-FOR$ ~~were used~~ to evaluate the model accuracy in predicting the occurrence of ponding for a single ~~the~~ event ~~where ponding occurred~~.

Table 4: ~~Score values~~Indicators used to evaluate accuracy of the proposed model in predicting ponding for a single node and for a single event.~~for measuring the level of agreement between the judgment of whether ponding occurred by the neural network and simulation from the hydrodynamic model at one node.~~

Score	Purpose	Equation	Range	Best Value
ACC	Mean accuracy for time points classified correctly	$ACC = \frac{1}{D}\sum_{s=1}^{D}\frac{TP_s + TN_s}{TP_s + TN_s + FP_s + FN_s}$	$0-1$	1
PPV	Mean precision for well-judged time point	$PPV = \frac{1}{D}\sum_{s=1}^{D}\frac{TP_s}{TP_s + FP_s}$	$0-1$	1
FOR	Mean proportion of omission on the timeline	$FOR = \frac{1}{D}\sum_{s=1}^{D}\frac{FN_s}{TN_s + FN_s}$	$0-1$	0
$S-PPV$	Percentage~~Precision~~ for well-judged samples	$S-PPV = \frac{TP}{TP + FP}$	$0-1$	1
$S-FOR$	Percentage of samples with false negatives	$S-FOR = \frac{FN}{TN + FN}$	$0-1$	0

where TP and TN denote the number of occurrences when~~events where~~ a ponding case and a normal case (no ponding occurs) are correctly identified, respectively, ~~occurs and does not occur that are correctly classified, and~~ FP is~~and FN are~~ the number of occurrence~~event~~s when~~re~~ a normal case is incorrectly identified as a ponding case, ~~ponding~~ and FN is the number

of occurrences when a ponding case is ignored by the model.~~occurs and does not occur incorrectly classified, respectively.~~ And the subscript "s" denotes ~~Besides, TPs and TNs are~~ the number of time steps in the s-th event. ~~when ponding occurs and does not occur correctly classified. FPs and FNs are the number of time steps that are incorrectly classified.~~

3 Results

3.1 Model setup

~~t~~The LSTM-based model was trained by the ~~Designed~~designed rainfall data and simulation produced ~~trained the LSTM-based model~~ from the hydrodynamic model. ~~following the procedure described in Sect. 2.4.~~ According to the procedures described~~steps~~ in Sect. 2.2, the noise (ε) transmitted from the runoff process to the flow confluence process was ~~calculated as~~equal to 1.9412% in the case pipe network. For the sake of convenience ~~of calculations~~, the noise was set to 2%. ~~Fig. 10 shows the performance evaluation results of the trained model's performance.~~

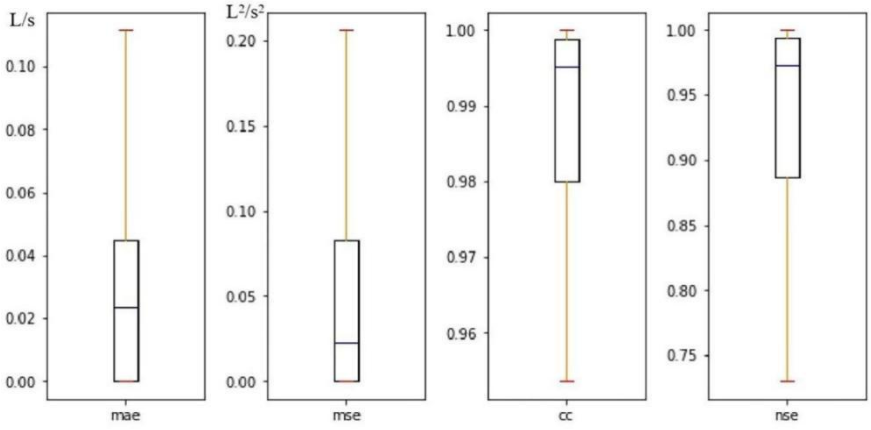


Figure 10: Box plots of score values for comprehensive evaluation of all nodes in the case area in the model development procedure.

Figure~~.~~ 10 described ~~the model's~~ the overall performance of the model using~~with~~ four box plots of mean scores (of all nodes) on the test set, ~~of all nodes in the study area~~ with the outliers removed. As shown in the figure, t~~T~~he median values of MAE and MSE were much smaller~~far less~~ than 0.1, indicating that the model ~~the training~~ has~~d~~ converged at all nodes. The median value of CC was close to 1~~.~~ , and ~~In contrast,~~ even the minimum value ~~-~~ was higher ~~more significant~~ than 0.95. The median value of NSE was higher ~~more significant~~ than 0.95, yet ~~while the~~the minimum value was ~~was~~ about~~round~~ 0.75, which~~. It~~ indicated~~d~~ that although the model's performance at each node was slightly different, the overall prediction ~~results~~ was generally~~of the model were~~ reliable.

Due to the limited space, we~~This article~~ only listed the evaluation results of 6 representative nodes ~~due to the limited space~~. The six nodes (as shown in Fig. 11) were selected ~~according to~~because of the ~~their~~ severity of consequence once ponding occurred, and also because ~~the spatial distribution~~ they were relatively uniformly distributed in the pipe network.~~to show the~~

~~difference in model performance between nodes.~~ Moreover, t~~T~~hree of them (Nodes 2, 238, ~~??,~~ and ~~??~~313) were chosen because the positive~~their percentages of~~ samples (where ponding occurred) accounted for less than 50% ~~where ponding occurred of~~~~in~~ the training set ~~were less than 50%~~, and the other ~~three~~s were in the opposite case. For example, at Node 238, the positive samples accounted for 18.33% of the training set ~~percentage of Node 238 was 18.33%~~, while at Node 95, up to~~that of Node 95 was~~ 98.6% samples were positive. ~~Moreover, Fig. 11 marks their locations.~~

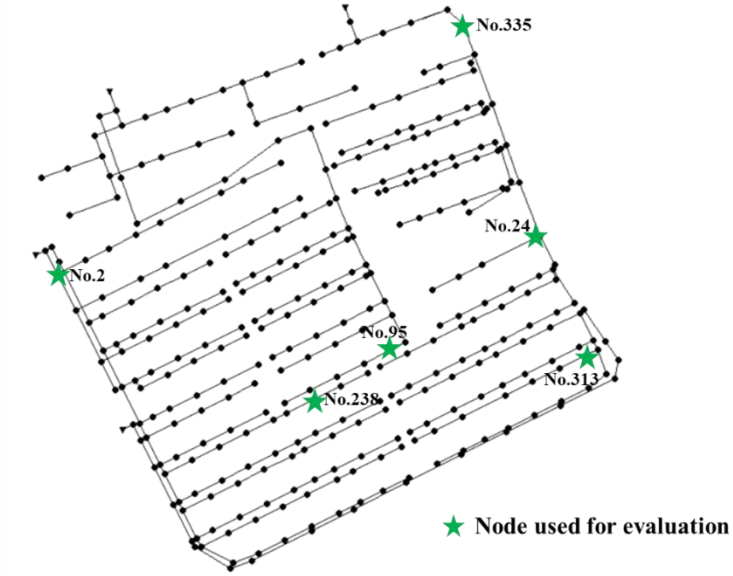


Figure 11: The locations of the six selected nodes used for evaluation.

~~On the premise that the model performed differently from node to node, Table 5 listed the score values of the six selected nodes on the judgment of ponding, and Table 6 presented the scores on the volume forecasting.~~

Table 5: The s~~S~~core ~~values~~s at~~of~~ the six selected nodes ~~for~~for evaluating the model ~~accuracy~~ performance in ponding occurrence prediction. ~~of the judgment of whether ponding occurred at some time step and in a rainfall event predicted by the model.~~

Node No.	ACC	PPV	FOR	S − PPV	S − FOR
2	99.90%	95.11%	0.04%	98.00%	0.00%
24	99.56%	95.21%	0.27%	99.33%	0.00%
95	98.66%	93.47%	0.98%	100.00%	0.00%
238	99.81%	88.33%	0.06%	94.00%	0.00%
313	99.67%	95.75%	0.19%	99.33%	0.00%
335	99.56%	95.29%	0.23%	100.00%	0.00%

~~-~~Table 5 lists the scores at the six selected nodes for evaluating the model performance in ponding occurrence prediction. In the t~~T~~able ~~5~~, Column ACC~~the second~~ to FOR ~~fourth columns~~ reflect~~ed~~ the accuracy of ponding occurrence prediction in the sense of time (i.e. averaged in time). ~~model's sensitivity in judging the ponding's starting and ending time.~~ The mean ACC

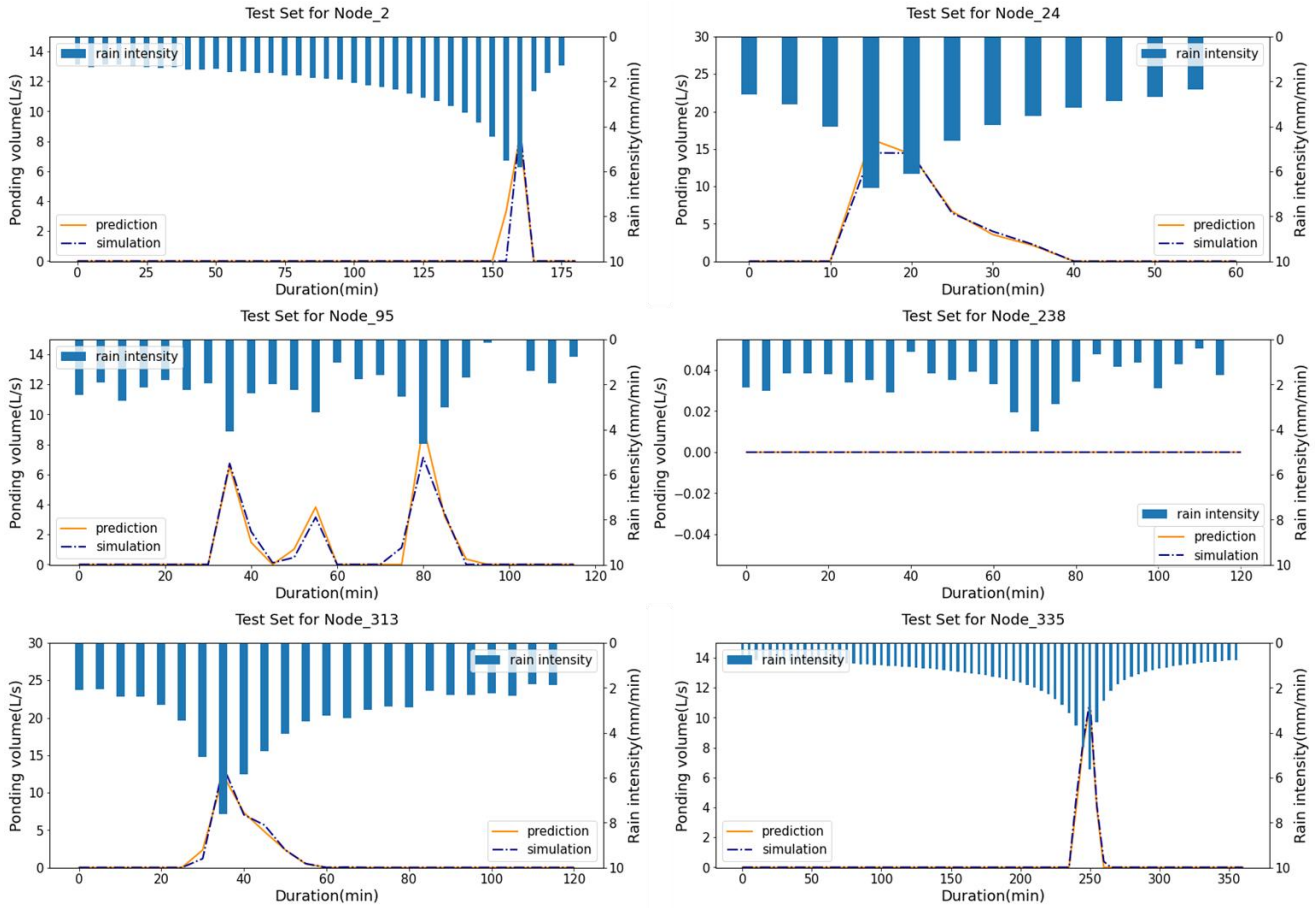
values ~~accuracy~~ (Accuracy) for all the 6 nodes were higher than 98.5% ~~of judging whether ponding occurred at some time step (ACC) of each node was above 98.5%~~. Compared to ACC, ~~Although~~ the mean PPV values (Precision~~PPV~~) ~~of each node~~ were~~as~~ slightly lower ~~than ACC,~~ with the minimum value about 88% at Node 238~~it was higher than 88%~~, which indicate~~d~~d that a ponding case had at least 88% chance to be correctly identified~~about 12% of the time points were judged incorrectly~~.
The mean FOR value (False Omission Rate) of e~~E~~ach node's ~~FOR~~ was generally lower than 1%, and among them the worst performance occurred at~~that of~~ Node 95 (FOR=0.98%)~~was slightly higher~~, which~~. It~~ indicated that the model had a relatively small chance to ignored ponding~~. at some point. However, its proportion was far less than the false discovery rate.~~ The last two columns ~~of~~ in Table 5 reflect the accuracy of the model in predicting ponding occurrence for a single rainfall event. For example, falsely reported events took up 6% of the testing events at Node 238 (S-PPV=94%), which was already the worst performance among the 6 selected nodes. While the S-FOR values at all these 6 nodes were zero, which ~~respectively showed that the model performed well at each node in judging whether ponding occurred in the rainfall event. Node 238 falsely reported whether ponding occurred in 6% of the testing events, resulting in the lowest $S-PPV$. $S-FOR$ in Table 5~~ indicate~~ded~~d that the model did not miss any ponding incidents ~~during~~ in the testing set~~the rainfall events~~.
The scores for evaluating the model performance in ponding volume prediction are listed in Table 6. As shown in the table, the MAE and MSE scores were generally small, with the highest MAE score (0.0770$L \cdot s^{-1}$~~L/s~~) occurred at Node 95 and the highest MSE score (0.3788$L^2 \cdot s^{-2}$~~L^2/s^2~~) occurred ~~had the highest MAE score, 0.0770L/s, while the maximum MSE score was 0.3788L^2/s^2~~ at Node 2. Compared to the MAE and MSE scores, the variability of CC scores was much smaller. All of them were very ~~varied slightly between nodes which were~~ close to 1. As for the NSE scores, the lowest score (NSE=0.8195 at Node 238) was above 0.8. ~~However, Node 238 had the lowest NSE score, which remained higher than 0.8.~~ The results shown in Table 6 indicated that the proposed model had a relatively good performance in ponding volume prediction.
~~the predicted volume of ponding at each node was reliable.~~

Table 6: The s~~S~~core ~~values~~ at~~of~~ the six selected nodes for evaluating the model performance in~~on the~~ ponding volume prediction.~~forecasting.~~

Node No.	MAE($L \cdot s^{-1}$ ~~L/s~~)	MSE($L^2 \cdot s^{-2}$ ~~L^2/s^2~~)	CC	NSE
2	0.0170	0.3788	0.9997	0.9811
24	0.0414	0.1876	0.9941	0.9754
95	0.0770	0.2740	0.9860	0.9599
238	0.0073	0.0260	0.9999	0.8195
313	0.0183	0.0505	0.9974	0.9825
335	0.0349	0.0826	0.9968	0.9882

Furthermore, in the above analysis, mean score values on the test set were used for evaluation, and the variability ~~individual differences among the test set were~~was ignored. Figure~~.~~ 12 shows the predicted ponding volume ~~of ponding~~ at the selected nodes compared ~~to~~ with the simulation results in ~~several~~six testing rainfall events. As shown in ~~Fig. 12~~Subfigure (1),~~ a~~ the

predicted start time of ponding was 5 minutes earlier than the simulation ~~misjudgment occurred at the beginning of the ponding, which was 5 minutes ahead~~ at Node 2. As shown in Subfigure (2),~~Three~~ three peaks appeared in the ponding process of Node 95, and the model has identified each of them~~one~~. No ponding occurred at node 238 given the testing precipitation, as shown in Subfigure (5), ~~with the testing precipitation,~~ and the prediction of the model was in consistent with ~~predicted~~ it ~~correctly~~. Overall, ~~the model~~ the prediction of the model was relatively accurate ~~at each node~~.



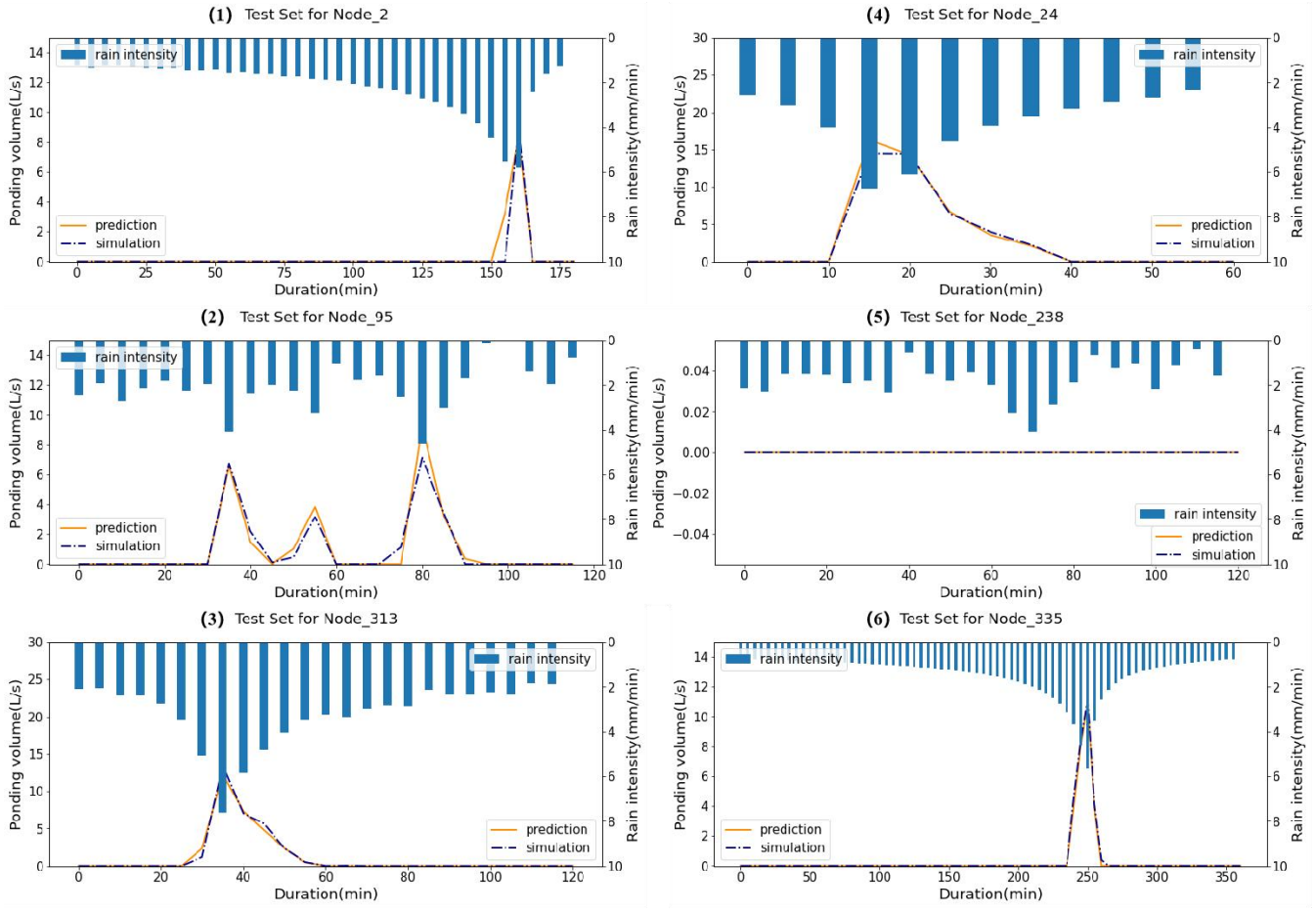


Figure 12: Comparison between the predicted ponding volume and simulation from the hydrodynamic model at the selected nodes in ~~several~~ 6 testing rainfall events that were chosen randomly ~~chosen~~.

~~The charts and tables above show the comparison of the simulation from the hydrodynamic model and the predicted resultsion of from the neural networks. From them, it was reliable to judge the occurrence and indicate the location and process of ponding using the LSTM-based model trained by the simulation from the hydrodynamic model.~~

3.2 Model correction

~~This paper~~In this study, ~~trained the~~ the model was trained based on the simulation results from a~~the~~ hydrodynamic model. Though the hydrodynamic model has been verified~~However~~, the differences between the simulation (from the hydrodynamic model) at the monitoring points and the ~~actual measure~~monitoring~~d~~ data persisted during the essential operation of the pipe network, which~~. It~~ inevitably degraded the accuracy of the LSTM-based model in~~on~~ ponding forecast~~ing~~. Thus, it is necessary to correct the model ~~the model must be corrected~~ using the measured rainfall data, level or flow data at the monitoring points, and ponding data.

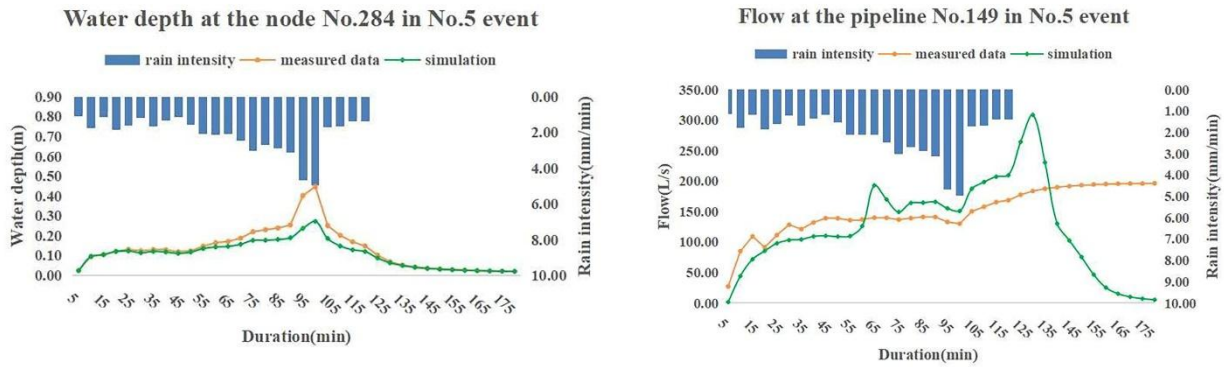


Figure 13: Comparison between the measured data and simulation from the hydrodynamic model at Node 284 and Pipeline 149 in No.5 event.

The discrepancy between the measurements and simulation from the hydrodynamic model can be exemplified by Fig. 13. As shown in the figure~~Fig. 13~~, rainfall event No.5 was one of the measured precipitation event where the maximum precipitation intensity reached 4.97mm/min. ~~In~~ For this event, the measured water depth and flow~~flow or level~~ data were compared with the simulation from the hydrodynamic model, as shown on the left and right panels, respectively, in Fig. 13. ~~at the monitoring points and the simulations using the measured precipitation were compared to present the discrepancy between the measured and simulated results. For example, Figure 13 shows the disparity at monitoring node 284 and pipeline 149.~~

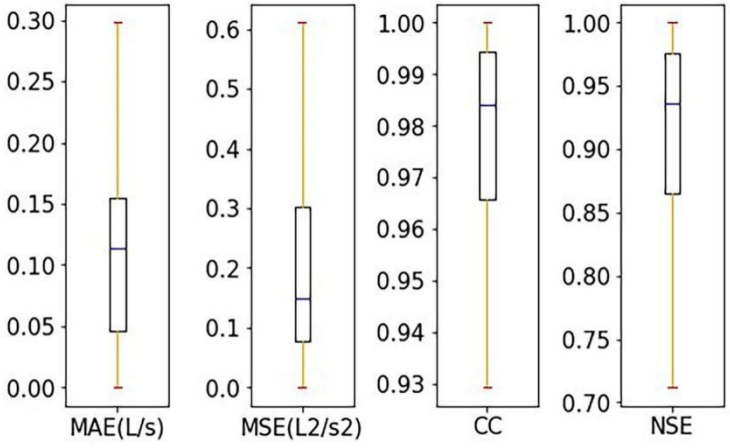


Figure 14: Box plots of mean score values on the test set for all nodes in the model updating procedure.

The ponding process predicted by the corrected model was compared with the ~~actual~~ monitor~~ing~~ed ponding data to evaluate the model~~'s~~ performance. ~~The approach to fixing the model was reliable if the two agreed.~~ Figure~~.~~ 14 illustrates the overall performance of the corrected model ~~of~~ using 4 indicators as described in Section 2.4.5. ???~~all nodes on the ponding volume forecasting. The approach to fixing the model was reliable if the two agreed.~~

In specific, the 4 box plots~~Fig. 14~~ shows the range of the mean score values on the test set ~~of~~ at all nodes ~~after being updated~~. As shown ~~by~~ in the figure, the median values of CC and NSE scores maintained above 0.98 and 0.9, respectively. In contrast, the maximum values of MAE and MSE scores remained lower than 0.30$L \cdot s^{-1}$~~L/s~~, and ~~that of MSE was lower than~~ 0.6$L^2 \cdot s^{-2}$~~L^2/s^2~~, respectively.~~.~~

Specifically, the mean score values at the six selected nodes obtained using the corrected model are summarized in Table 7. As shown in the t~~T~~able ~~7~~, the MAE and MSE scores were generally small, the NSE score at each node was stably above 0.9~~.~~, and t~~T~~he CC scores w~~a~~eres all above 0.95.~~,~~ The results shown in Table 7~~which~~ suggested that the corrected model performed well ~~on the test set. The updated model performed well~~ at differenteach nodes. ~~and was reliable.~~

~~Table 8 summarizes the mean scores of all nodes for five measured rainfall events on the model performance before and after modification.~~

Table 7: Mean score values ~~on~~ at the six selected nodes obtained~~' test set on predicting the ponding volume in the model updating procedur~~ using the corrected modele.

Node No.	MAE($L \cdot s^{-1}$~~L/s~~)	MSE($L^2 \cdot s^{-2}$~~L^2/s^2~~)	CC	NSE
2	0.0912	0.1604	0.9774	0.9545
24	0.1359	0.2426	0.9863	0.9586
95	0.2916	0.7378	0.9583	0.9071
238	0.0855	0.1842	0.9773	0.9302
313	0.0642	0.0571	0.9896	0.9727
335	0.0943	0.1135	0.9942	0.9683

To test further the capability of the corrected model, the mean scores of all nodes for five measured rainfall events are summarized in Table 8, where the results from the model without correction are also listed as a comparison. As shown in Table 8, a~~A~~ll of the four indicators suggested that the corrected model performed much better than the model without correction.~~from the corrected model were much better than those from the uncorrected model in Table 8.~~ Specifically, the NSE score obtained from the model without correction~~before the correction~~ was less than 0, while this score rose up to~~compared to~~ 0.8316 after applying the model correction procedure~~the modification. It~~, which indicated t~~proved t~~he necessity of~~for~~ the ~~model~~ correction~~updating~~.

Table 8: Mean score values of all nodes for five measured rainfall events, obtained by using ~~from~~ the model with/without correction.~~ed model versus those from the uncorrected model.~~

	MAE($L \cdot s^{-1}$ ~~L/s~~)	MSE($L^2 \cdot s^{-2}$ ~~L^2/s^2~~)	CC	NSE
Model without correction~~Before~~	0.5719	4.5045	0.1139	< 0

	MAE($L \cdot s^{-1}$ ~~L/s~~)	MSE($L^2 \cdot s^{-2}$ ~~L^2/s^2~~)	CC	NSE
~~updating~~ Model with correction~~After updating~~	0.1504	0.5919	0.9309	0.8316

To further demonstrate the effect of model correction procedure, we have shown the predicted ponding process at the 6 selected nodes for rainfall event No.5, obtained by using the model with and without correction, as shown in Fig. 15. As shown in the figure, the corrected model ~~Fig. 15 plotted the specific ponding process at the selected nodes in the measured rainfall event No.5 predicted by the corrected model compared to those anticipated before correction. Fig. 15 showed that the revised model~~ performed better at all the selected nodes, e.g., more accurate prediction of start/end time of ponding, more accurate ponding curves (more similar to the measure ones). ~~in terms of not only starting and ending time but also the process of ponding than the uncorrected model. The prediction by the corrected model had a high consistency with the measured ponding volume at each node in the figure, which proved the reliability of the approach to introducing the measured monitoring data to correct this LSTM-based model as well.~~

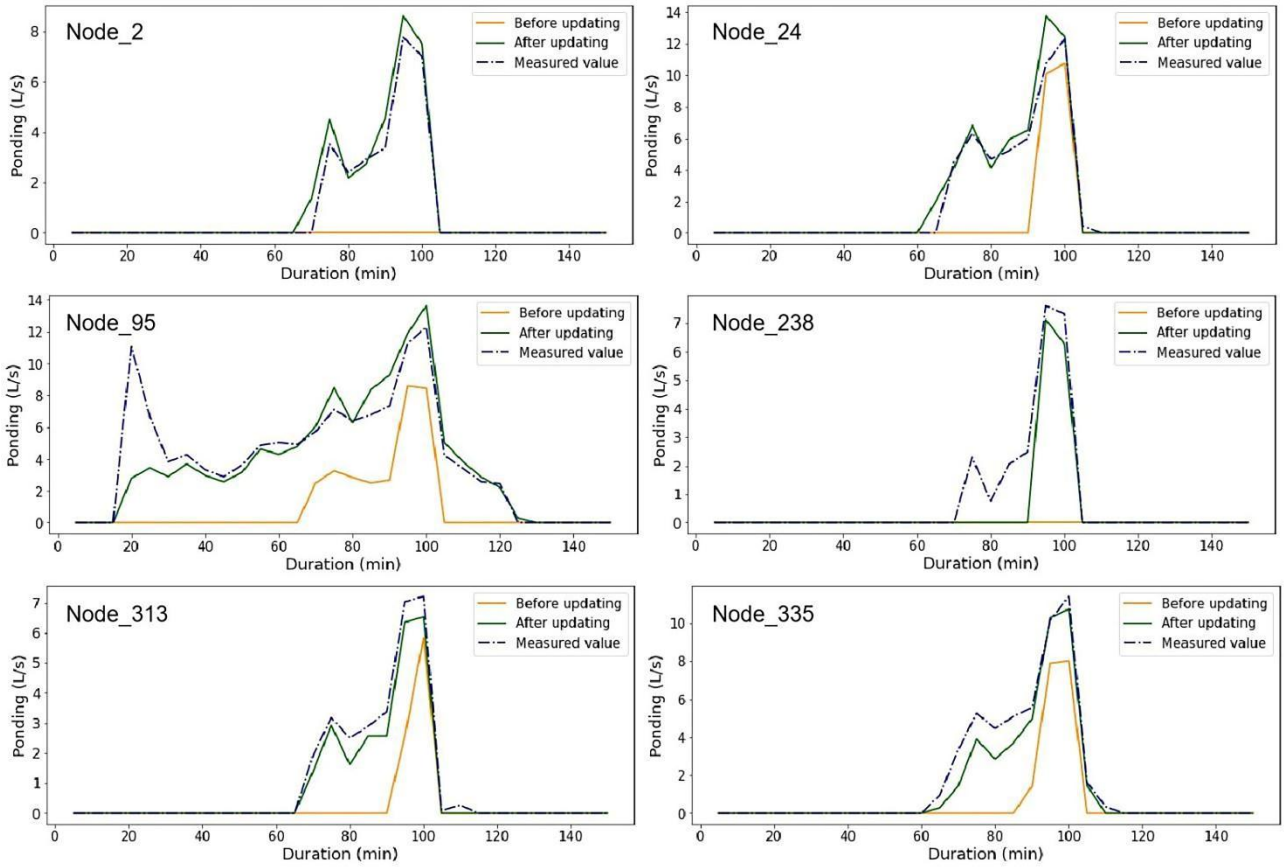


Figure 15: Comparison between the predicted ponding volume and measured values at the selected nodes in rainfall event No.5.

All the results shown above demonstrate the superiority of the corrected model compared to the original one, where the monitoring data were introduced in the model correction procedure. ~~By the comparison between the prediction by the model before and after correction, it proved the reliability of the approach to introducing the measured monitoring data to correct the model.~~

4 Discussion

4.1 Comparison of neural network structures

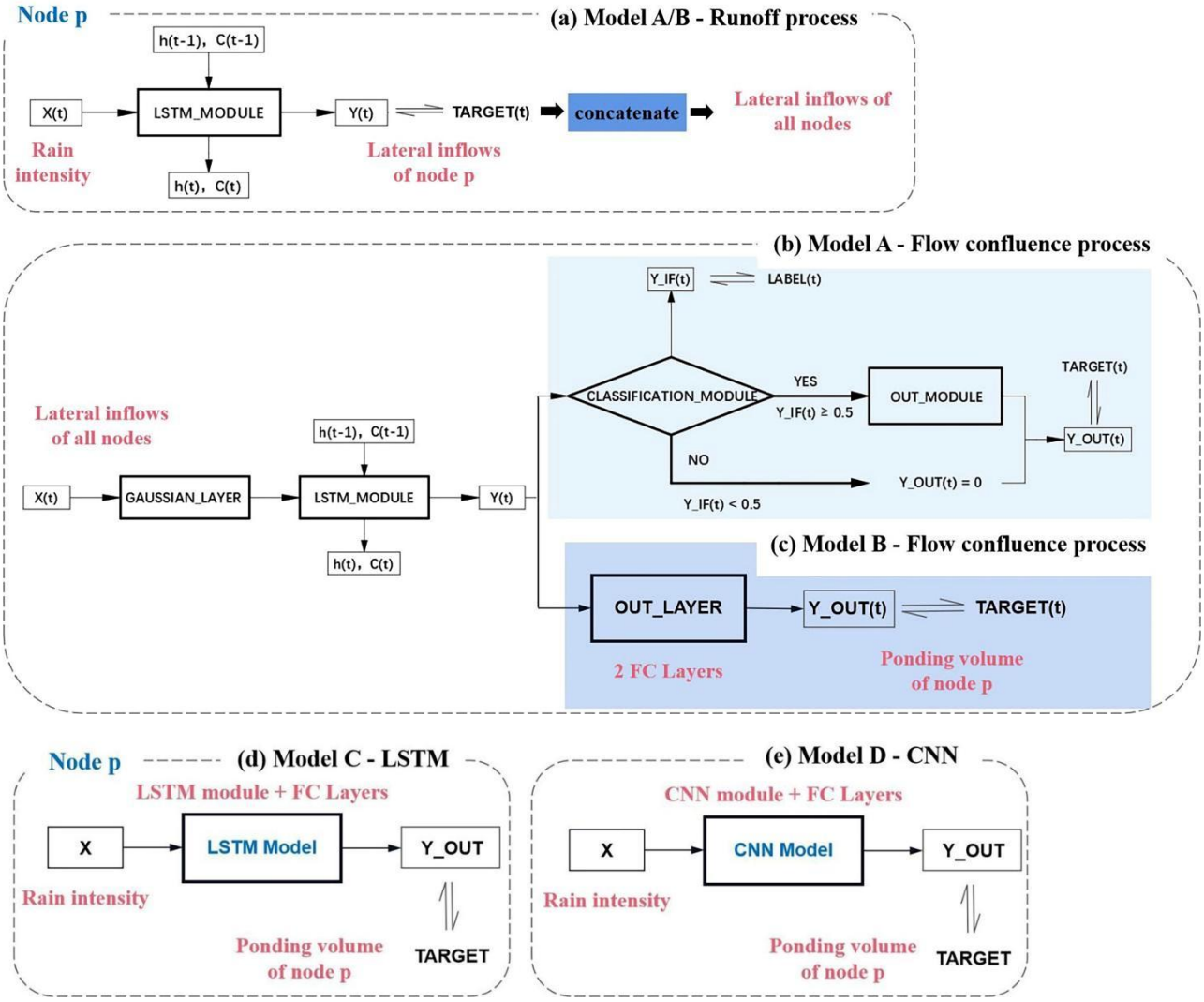
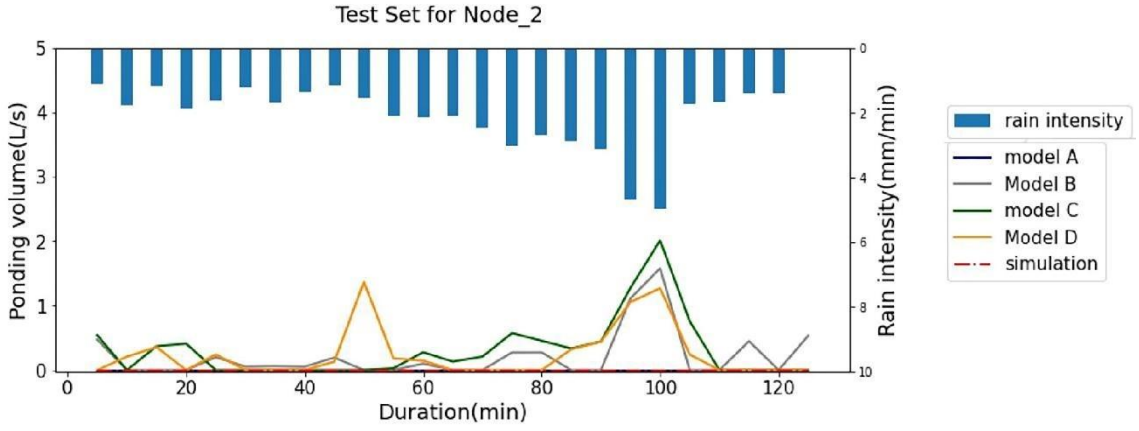


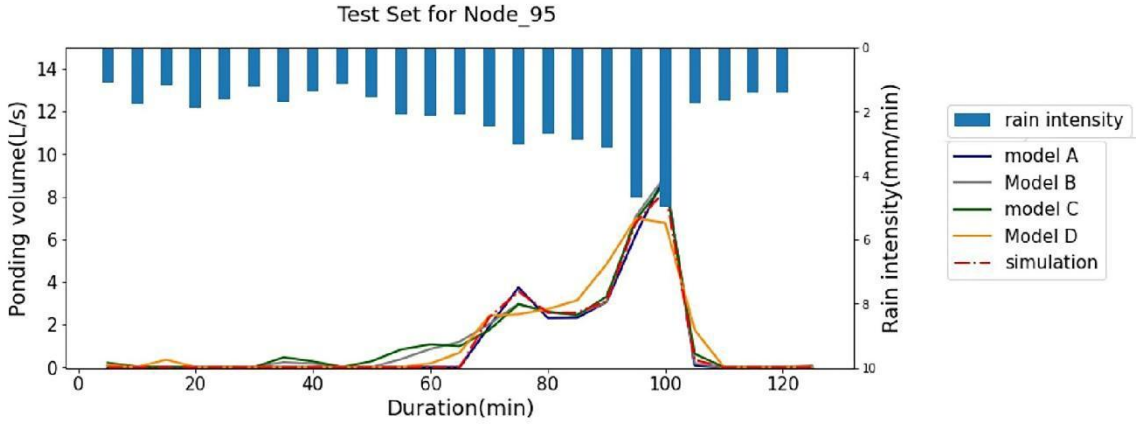
Figure 16: Schematic diagrams of different network structures for comparison. (a) The same runoff process in model A and B. (b) The multi-target learning in the flow confluence process of model A, marked in light blue. (c) The flow confluence process in model B marked in dark blue. (d) The LSTM structure in model C. (e) The CNN structure in model D.

The proposed model (termed Model A) was compared with the conventional LSTM structure (termed model B) ~~in Fig. 16(a ~ c)~~ to show the superiority of the variant of the LSTM structure in the flow confluence process. The schematic diagrams of the two models are shown in Fig. 16 (a ~ c). As shown in the subfigures, Model B has exactly the same structure as model A in the runoff process. The only difference of the two models lies in the flow confluence process, where a multi-task learning mechanism is introduced in the learning process of Model A.

Furthermore, Model A proposed in this paper was compared with two other models (Models C and D) to illustrate the necessity of two processes in tandem, i.e. the runoff and flow confluence processes. The network structures of Models C and D are shown in Fig. 16 (d) and (e), respectively, where the ponding information was obtained directly from rainfall data without extracting the characteristics of lateral inflows.



(a) Case a where ponding did not occur at Node 2.



(b) Case b, where ponding occurred at Node 95.

Figure 17: A comparison between the predicted ponding volume by the LSTM-based model (Model A) and by Models B, C, and D for a particular rainfall event.

Figure 17 shows two examples. In the first example, as shown in the subfigure (a), ponding did not occur at Node 2. However, about 2-4$L \cdot s^{-1}$ ponding volume was falsely reported by the three alternative models (Models B, C, and D), while Model A predicted no ponding at this Node, which was consistent with the simulation (considered as the ground truth). In the second example, as shown in the subfigure (b), where ponding occurred and lasted for about 40 minutes, Model A predicted a more accurate ponding curve than the other three alternative models. ~~Model B hasd the same structure as model A with in the runoff and flow confluence processes. The network architecture in model B's runoff process was the same as in model A, presented in Fig. 16(a). However, it was different in model B's flow confluence process (Fig. 16 b ~ c). There was no multi-task learning mechanism in its learning process.~~

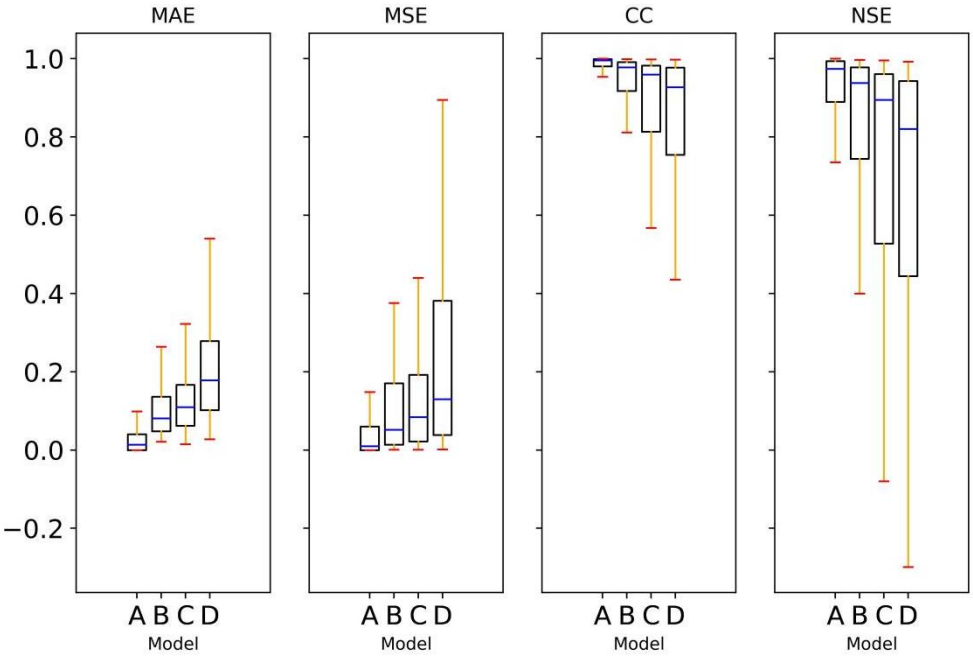


Figure 178: Comparison of model performance on the ponding volume forecasting. The results of the proposed model A are compared to those obtained from models B, C, and D.

~~Fig. 17 presented the range of mean score values on the test set of all nodes on model A, B, C, and D performance on the ponding volume forecasting. As shown in Fig. 17, the range of model A's MAE and MSE scores was half that of model B. CC scores of model A were close to 1, while those of model B varied from about 0.8 to 1. NSE scores of model A were higher than 0.7, while model B's were higher than about 0.4. Obviously, the structure in Fig. 16(b) was superior to that in Fig. 16(c) on forecasting ponding volume.~~

~~Furthermore, model Model A proposed in this paper was compared with two other models (Models C and D) to illustrate the necessity of two processes in tandem, i.e. , the runoff and flow confluence processes. As shown in Figures 16(d) and 16(e),~~

~~The network structures of Mmodels C and D are shown in Fig. 16 (d) and (e), respectively, where the obtained ponding information was obtained directly from rainfall data without extracting the characteristics of lateral inflows' characteristics.~~

Fig~~.~~ure 1~~7~~8 presente~~d~~s the range of mean score values on the test set for~~of~~ all nodes, obtained by using ~~on~~ M~~m~~odels A~~, B, C, and D performance on the ponding volume forecasting~~. As shown in Fig. 17the figure, the range of model A's~~ of the MAE or~~and~~ MSE scores from Model A was~~was~~ half that of model B. The CC scores from~~of~~ model A were very close to 1, while the CC scores from~~those of~~ model B varied from about 0.8 to 1. The NSE scoress from~~of~~ M~~m~~odel A were generally higher than 0.7, while the NSE scores from model B's were unstable and generally lower than those from Model A~~higher than about 0.4~~. Obviously, Model A performed much better than Model B in ponding volume prediction, as indicated by all of the four indicators~~. the structure in Fig. 16(b) was superior to that in Fig. 16(c) on forecasting ponding volume.~~

Also as shown in Fig. 1~~7~~8, the obvious superiority of Model A (or B) over Models C and D demonstrates the necessity of having two processes in tandem. Besides, it is also shown in the figure ~~Fig. 17 showed~~ that the range of all these four indicators expanded gradually from Model A to D, which indicated a decreased steadiness. ~~the MAE and MSE scores of models A, C, and D almost expanded gradually, while the trend in the CC and NSE scores was the opposite. It showed that the structure of two processes in tandem in model A had an advantage over model C. The LSTM structure was better than CNN in dealing with these time series data.~~

Table 9: Mean score values of all nodes obtained from models for predicting the volume of ponding.

Model	MAE($L \cdot s^{-1}$~~L/s~~)	MSE($L^2 \cdot s^{-2}$~~L^2/ s^2~~)	CC	NSE
A	0.0309	0.1624	0.9960	0.9462
B	0.0622	0.1815	0.9578	0.8552
C	0.0849	0.2584	0.8823	0.7424
D	0.1358	0.3480	0.9257	0.7391

Table 9 ~~also supported this. It showeds the~~ the mean score values at~~of~~ all nodes (on the test set) obtained by using~~of~~ the four models ~~on the test set~~. According to the ~~table~~results, the performance ranking of the four models was ~~m~~Model A > ~~m~~Model B > ~~m~~Model C > ~~m~~Model D.

The comparative ~~analysis~~ analyses above indicated that the LSTM-based model proposed in this paper had remarkable superiority over the other three alternatives in ponding volume forecasting. There are two reasons behind this. First, the proposed model had two processes in tandem: the runoff and flow confluence processes. The second is due to ~~because of the following two reasons. One of them was the two processes in tandem: the runoff process and the flow confluence process. The other one was~~ the auxiliary classification task introduced~~variant of the LSTM structure~~ in the flow confluence process. The two tandem processes~~first one~~ reduced the computational burden of this data-driven ~~model~~ approach and avoided

interference with each other duringwhen training separately. While the classification task introduced facilitated the capability of the model to identify ponding.

~~The second one introduced an auxiliary task to judge whether ponding occurred and tried to eliminate the interference of the time points without ponding on the ponding volume forecasting.~~

4.2 The influence of the number of monitoring points on ~~the~~ model ~~modification~~correction

It was easy to spot from the trial that the performance of the corrected ~~modified~~ model~~'s performance~~ depended on whether the layout of the monitoring points~~' current layout~~ could reflect~~ed~~ the ~~the pipe network's~~ hydraulic conditions of the pipe network. An unreasonable design of the monitoring equipment might lead to a failure ~~of~~ in model ~~modification~~correction. There were 15 level gauges and three flowmeters in the case~~current~~ pipe network, as shown in~~.~~ Fig. 7(b). To analyse how the number of monitoring points impacted the performance of the revised model, ~~shows the layout of the monitoring points. d~~Different numbers of monitoring points were randomly selected as a quantitative control group ~~to analyze whether and how the number of monitoring points impacted the performance of the revised model~~. ~~Scores in~~ Figure~~.~~ 19 presented the evaluation results of the revised model on ponding volume forecasting obtained by using ~~when correcting the model with~~ different numbers of monitoring point~~site~~s.

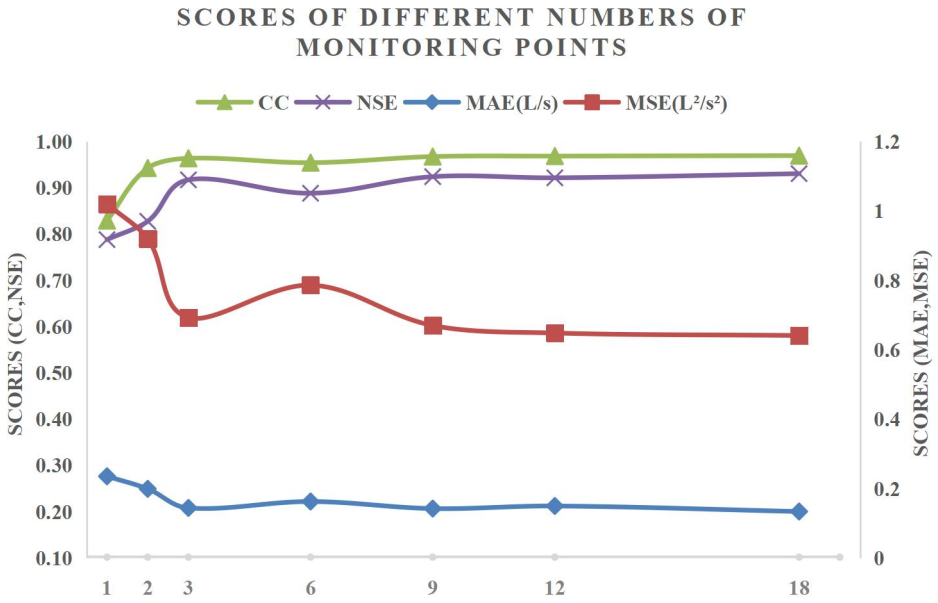


Figure 19: Score values obtained by using ~~of~~ different numbers of monitoring points ~~on the performance of~~in the model ~~processed model~~correction.

As shown in the figure, the NSE scores stayed around 0.9 when the number of monitoring points exceeded ~~six~~6, the CC scores showed similar trends as the NSE scores. ~~a~~And the other scores showed the opposite trends~~showed similar trends as~~

~~the NSE scores.~~ ~~were almost consistent with NSE.~~ It turned out that, when the number of monitoring points was over 1 per hectare, increasing the number of monitoring points further had ~~little~~ limited effect~~influence~~ on improving the ~~the~~ accuracy of the corrected model. ~~when the number of monitoring sites was over one per hectare.~~ However, when the number of monitoring points was~~went~~ below~~down to about~~ 0.5 per hectare (~~,~~ i.e., the number of monitoring points was less than 3), it is highly effective to increase the number of monitoring points in the pipe network. ~~the results of the revised model could have been better.~~ For example, the NSE score was lower than 0.8 when the number of monitoring points was only 1.

In summary, one monitoring point per hectare is the critical point. If~~Suppose~~ the number of monitoring point~~site~~s was less than ~~one per hectare~~this limit~~. In that case~~, the performance of the revised model could not be guaranteed. ~~Otherwise, if the number of monitoring sites was one per hectare or even more, the accuracy of the revised model on the ponding volume forecasting could be much higher.~~

5 Conclusions

In this study, a~~A~~n optimized LSTM-based model is proposed, which is applied to early warning~~,~~ and forecasting of ponding in ~~the~~ urban drainage system ~~was proposed in this paper~~. According to the research results, the main conclusions of this study are summarized as follows:

1 The LSTM-based model is constructed based on a relatively accurate hydrodynamic model. However, the difference between the simulation from the hydrodynamic model and the measured monitoring data persists~~s~~, which leads to a discrepancy between the predicted results from the proposed LSTM-based model and the actual situation (reflected by the monitoring data). Therefore, by calibration using real-life data, the corrected ~~LSTM-based~~ model can closely reflect the actual condition of the pipe network, and is capable of identifying and locating ponding with relatively high accuracy and efficiency. ~~. The revised model realizes the real-time prediction of ponding with sure accuracy in the case pipe network. The predicted information includes the ponding location and ponding process at each node.~~

2 The model is split into two tandem processes: runoff process and flow confluence process. In t~~T~~he runoff~~first~~ process,~~is to simulate the process of~~ the lateral ~~inflows~~ inflows entering ~~the~~ individual nodes are simulated, and in the flow confluence process,~~the second is to predict the~~ ponding volume~~s of ponding~~ at individual ~~each~~ node~~s are calculated in the pipe network~~. Moreover, the model introduces a multi-task learning mechanism to optimize the network structure in the flow confluence process. The above~~se~~ two characteristics ~~structures~~ significantly improve the performance of the LSTM-based model. The ~~advantage~~ superiority of the proposed ~~variant of LSTM~~model (~~model~~ Model A, a LSTM variant) in handling the task of ponding volume forecasting is demonstrated by a comparison with the conventional LSTM structure (~~model~~ Model B, also with two tandem processes as Model A). Then, the superiority of having two tandem processes is proved by comparing Model A (or B) with Models C and D, i.e., LSTM and CNN structures with a single process.~~compared to the models based on LSTM and CNN structures (models C and D), the construction with two processes in tandem based on LSTM is reliable.~~

3 The performance of the corrected model on ponding volume ~~forecasting~~ prediction is reliable if the number of monitoring ~~sites~~ points is over one per hectare. Increasing the number of monitoring points further has ~~little~~ limited ~~impact~~ effect on improving the model~~'s~~ performance.

Overall, the proposed LSTM-based approach provides a new possibility for early warning and forecasting of ponding in ~~the~~ urban drainage system~~ and has a remarkable improvement after being corrected by the measured monitoring data~~. In this study, all operations were conducted ~~This paper carried out all operations on model correction~~ in an offline mode. In the future study, we will~~Future studies should~~ explore the capability of the proposed model in real-time event analysis. And also, we will optimize the model by considering the influence of two-dimensional overland flow in ponding volume prediction~~online correction. Besides, considering the variation of ponding areas with the two-dimensional groundwater overflow process will further optimize the model~~.

Code availability. The pieces of code used for all analyses are available from the authors upon request.

Data availability. All data used in this study are available from the authors upon request.

Competing interests. The contact author has declared that neither they nor their co-authors have competing interests.

Acknowledgment. The authors gratefully acknowledge the financial supports by the National Natural Science Foundation of China under Grant numbers 51978493, and ~~The authors~~also ~~would like to~~ thank all the team members for their insightful comments and constructive suggestions to polish this paper in high quality.

References

Abou Rjeily, Y., Abbas, O., Sadek, M., Shahrour, I. and Hage Chehade, F.: Flood forecasting within urban drainage systems using NARX neural network, Water Science and Technology, 76, 2401-2412, https://doi.org/10.2166/wst.2017.409, 2017.

Archetti, R., Bolognesi, A., Casadio, A. and Maglionico, M.: Development of flood probability charts for urban drainage network in coastal areas through a simplified joint assessment approach, Hydrology and Earth System Sciences, 15, 3115-3122, http://dx.doi.org/10.5194/hess-15-3115-2011, 2011.

Aryal, D. et al.: A Model-Based Flood Hazard Mapping on the Southern Slope of Himalaya, Water, 12, 540, https://doi.org/10.3390/w12020540, 2020.

Bai, Y., Bezak, N., Sapač, K., Klun, M. and Zhang, J.: Short-Term Streamflow Forecasting Using the Feature-Enhanced Regression Model, Water Resources Management, 33, 4783-4797, https://doi.org/10.1007/s11269-019-02399-1, 2019.

Balstrøm, T. and Crawford, D.: Arc-Malstrøm: A 1D hydrologic screening method for stormwater assessments based on geometric networks, Computers & Geosciences, 116, 64-73, https://doi.org/10.1016/j.cageo.2018.04.010, 2018.

Bergstra, J., Yamins, D., Cox, D.D., 2013. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures.

Cai, B. and Yu, Y.: Flood forecasting in urban reservoir using hybrid recurrent neural network, Urban Climate, 42, 101086, https://doi.org/10.1016/j.uclim.2022.101086, 2022.

Chiang, Y., Li-Chiu, C., Meng-Jung, T., Yi-Fung, W. and C., F.: Dynamic neural networks for real-time water level predictions of sewerage systems-covering gauged and ungauged sites, Hydrology and Earth System Sciences, 14, 1309-1319, https://doi.org/10.5194/hess-14-1309-2010, 2010.

Djordjević, S., Prodanović, D. and Maksimović, Č.: An approach to simulation of dual drainage, Water Science and Technology, 39, 95-103, ~~https://doi.org/~~https://doi.org/10.1016/S0273-1223(99)00221-8, 1999.

Djordjević, S., Prodanović, D., Maksimović, Č., Ivetić, M. and Savić, D.: SIPSON – Simulation of Interaction between Pipe flow and Surface Overland flow in Networks, Water Science and Technology, 52, 275-283, https://doi.org/10.2166/wst.2005.0143, 2005.

Guo, K., Guan, M. and Yu, D.: Urban surface water flood modelling-a comprehensive review of current models and future challenges, Hydrology and Earth System Sciences, 25, 2843-2860, http://dx.doi.org/10.5194/hess-25-2843-2021, 2021.

Hossain Anni, A., Cohen, S. and Praskievicz, S.: Sensitivity of urban flood simulations to stormwater infrastructure and soil infiltration, Journal of Hydrology, 588, 125028, https://doi.org/10.1016/j.jhydrol.2020.125028, 2020.

Huong, H.T.L. and Pathirana, A.: Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam, Hydrology and Earth System Sciences, 17, 379-394, http://dx.doi.org/10.5194/hess-17-379-2013, 2013.

Jamali, B. et al.: A rapid urban flood inundation and damage assessment model, Journal of Hydrology, 564, 1085-1098, https://doi.org/10.1016/j.jhydrol.2018.07.064, 2018.

Kao, I., Zhou, Y., Chang, L. and Chang, F.: Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting, Journal of Hydrology, 583, 124631, https://doi.org/10.1016/j.jhydrol.2020.124631, 2020.

Kratzert, F. et al.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resources Research, 55, https://doi.org/10.1029/2019WR026065, 2019.

Kratzert, F. et al.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089-5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.

Kuczera, G. et al.: Joint probability and design storms at the crossroads, Australian journal of water resources, 10, 63-79, https://doi.org/10.1080/13241583.2006.11465282, 2006.

Leandro, J. and Martins, R.: A methodology for linking 2D overland flow models with the sewer network model SWMM 5.1 based on dynamic link libraries, Water Science and Technology, 73, 3017-3026, https://doi.org/10.2166/wst.2016.171, 2016.

Moy De Vitry, M., Kramer, S., Dirk Wegner, J. and Leitao, J.P.: Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network, Hydrology and Earth System Sciences, 23, 4621-4634, http://dx.doi.org/10.5194/hess-23-4621-2019, 2019.

Mudashiru, R.B., Sabtu, N., Abustan, I. and Balogun, W.: Flood hazard mapping methods: A review, Journal of hydrology (Amsterdam), 603, 126846, https://doi.org/10.1016/j.jhydrol.2021.126846, 2021.

Pan, S.J. and Yang, Q.: A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering, 22, 1345-1359, https://doi.org/10.1109/TKDE.2009.191, 2010.

Pilgrim, D.H. and Cordery, I.: Rainfall Temporal Patterns for Design Floods, Journal of the Hydraulics Division, 101, 81-95, https://doi.org/10.1061/JYCEAJ.0004197, 1975.

Rahman, A., Weinmann, P.E., Hoang, T.M.T. and Laurenson, E.M.: Monte Carlo simulation of flood frequency curves from rainfall, Journal of hydrology, https://doi.org/10.1016/S0022-1694(01)00533-9, 2002.

Skougaard Kaspersen, P., Hoegh Ravn, N., Arnbjerg-Nielsen, K., Madsen, H. and Drews, M.: Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding, Hydrology and Earth System Sciences, 21, 4131-4147, http://dx.doi.org/10.5194/hess-21-4131-2017, 2017.

Yang, T., Hwang, G., Tsai, C. and Ho, J.: Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts, Hydrology and Earth System Sciences, 20, 4731-4745, http://dx.doi.org/10.5194/hess-20-4731-2016, 2016.