

# LSTM Based Hybrid Method for Basin Water Level Prediction by Using Precipitation Data

Liu Shuofeng<sup>1</sup>, Lei Puwen<sup>2</sup>, Koji Koyamada<sup>1</sup>

<sup>1</sup>Academic Center for Computing and Media Studies, Kyoto University, Japan

<sup>2</sup>Graduate school of engineering, Kyoto University, Japan

`liu.shuofeng.65n@st.kyoto-u.ac.jp`

Received: December 25, 2019; Accepted: January 8, 2021; Published: January 21, 2021

**Abstract.** Water level prediction is becoming increasingly important. However, physical models tend to become difficult to apply when it comes to some small rivers which have insufficient hydrological data. To address it, nowadays, deep learning methods are increasingly being applied to climate prediction analysis as an alternative to computationally expensive physical models for its features of flexible data-driven learning and universality. In our paper, we focus on the precipitation-only water level forecasting problem by using long-short-term memory (LSTM) based hybrid model, and try predicting the future water level of all the rivers in Japan by using simulated precipitation data from the database for Policy Decision making for Future climate change (d4PDF).

**Keywords:** Water level prediction, Deep machine learning, LSTM, Hybrid model

## 1. Introduction

Recent years, flood disaster has already become a heat topic with the increasing of the rainfall, which is getting more and more attention all over the world. The extreme destructive power and lethality tend to cause massive loss. In 20 July 2018, 225 people were confirmed dead across 15 prefectures with a further 13 people reported missing attribute to the extremely severe flood disaster in west Japan. Therefore, it's of great significance to predict the water level in advance to prevent flood disasters. In response to this situation, many corresponding countermeasures have been formulated. Water level prediction is a greatly important method to prevent the flood disaster and to save a quantity of people as well as families. Some good results have been achieved in the prediction of rainfall water levels for large rivers, since they have complete data and numerous observation stations, which means complicated prediction models can be applied to these kinds of large rivers. However, in some small river areas which don't have many observation stations and data, serious flood disasters can also occur. In 2008, 4

people were washed away and killed by a small river called toga river when the flood disaster came suddenly in Kobe. It is obvious that many small watersheds are also quite dangerous during flooding, but because there are no complete and sufficient data, many complex prediction models cannot be applied to these watersheds. Therefore, these small rivers' potential danger may create huge hidden disaster to life and property. The purpose of our study is to propose a universal and simple prediction model based on precipitation data only called LSTM based hybrid model and discuss its accuracy.

Long Short Term Memory (LSTM) is a time-recurrent neural network (RNN), which is suitable for processing and predicting important events with very long intervals and delays in time series. In terms of flood disaster prevention, the earlier a flood can be predicted, the more time people will have for the sake of making the countermeasures. Therefore, only predicting water level after an hour or two hours seems not enough, which means we need to predict the occurrence of flood disaster as early as possible. In this situation, there is a large time gap between prediction and real time, so LSTM is an available alternative method due to its idiographic memory characteristic. In this paper, we predict future water level in the next 6 hours at the Hiwatashi location in the Oyodo river.

## **2. Related works**

At the present stage, many water level prediction methods have already been proposed and applied to practice. Different forecasting methods are selected for different river basins. According to different algorithms, the prediction models can be generally divided into three big different groups.

### **2.1. Outflow model**

The first group is outflow model. This kind of model expresses the behavior of precipitation as closely as possible to a real phenomenon and progress or illustrate the phenomenon in a simulated way [1]. Outflow model can be applied to mountains, forests, arable land, grassland and wasteland in some parts of Japan and overseas [2]. It mainly includes physical model and conceptual model. Distributed model is a typical representative of the physical model, which is mainly used in many first-class rivers nationwide [3]. Distributed model can reflect the distribution of terrain and precipitation and show the detailed relationship of precipitation-outflow. However, the complicated model makes it impossible to be widely applied to practice, because on the one hand, the distribution of terrain is not the same in different rivers, on the other hand, constructing distributed model is going to cost too much time and energy. What's more, implementing parameter adjustments is another challenge which cannot be ignored for physical model [4]. Conceptual model which can be applied to various of rivers contains tank model and storage function method in general. Tank model is relatively easy to calculate and able to

indicate various outflow waveforms. When it turns to snowpack collection in the mountain watersheds, tank model can perform well [5]. Storage function model is easy to calculate as well and applied to the first-class large rivers in Japan at present. Development and evaluation of storage function method capable of long-term outflow calculation is still ongoing these years [6]. Moreover, not only in the prediction of river levels, some storage function method for converging sewers are also being developed [7]. Even though conceptual models have advantages in terms of calculation, the physically of the parameter is a little bit weak. Therefore, although the accuracy can be accepted, it is not fairly satisfactory due to all kinds of existing errors [8].

## **2.2. Direct water level prediction model**

The second group is the statistical based water level prediction model. It mainly contains time series analysis and machine learning model. Time series analysis is an ancient method that is easy to build models; it's not necessary to import physical models. Such as AR, ARMA, ARIMA, etc. However, it uses previous time series data only to forecast future water level, so only the linear relationship can be presented [9], which means the accuracy is not high. In addition, it's impossible to predict long-time water level due to the linear characteristic, since precipitation changes with time dramatically and it is hard to be a linear relationship. Machine learning model is another statistical based water level prediction model mainly using neural network posed recently, such as SVM, ANN, linear regression model. It can predict future water level by learning from previous flood patterns. Obtaining the results by calculating the statistical analysis and showing the non-linear relationship are also possible [10]. Moreover, even if the quantity of data is necessary and calculation progress is complicated, because the accuracy performs well, machine learning methods are still used more and more to practice.

## **2.3. Hybrid model**

There is another group called hybrid model. This kind of model combines physical model with statistical model to make predictions. Because of the hybrid nature, the accuracy is improved in some cases. However, most of the hybrid models require plenty of data, which makes them hardly use in small rivers that have insufficient data. To address this conundrum, we propose a new method aimed at using hybrid model based on LSTM for prediction in small rivers with precipitation data only.

# **3. Proposed Model**

## **3.1. Background: Long Short Term Memory (LSTM) Model**

Before introducing the LSTM, we would like to have a brief introduction of Recurrent Neural Network (RNN). RNN is a kind of neural network used for sequence data processing, especially

for time series data because of its feature of recursive data processing[11]. Compared with other kind of neural model, RNN can form a directed circle inside every unit. As shown in Fig. 1, the change of hidden state  $h_t$  can be passed through by using the update equations (1). This internal self-looped structure can help RNN learn to "remember" its previous states to capture the dynamic temporal characteristics in time series data. RNN has proven to be effective for sequence processing model like Natural Language Processing (NLP) and other time series data predictions.

$$h_t = \tanh(b_h + Wh_{t-1} + Ux_t) \quad (1)$$

However, the shortcomings of basic RNN are also obvious. Basic RNN is difficult to capture long-term time correlation because it cannot handle the problem of long-term recursion, weight exponential explosion or gradient disappearance, which is also called vanishing gradient problem[12]. These problems may make RNN difficult to learn long-term dependencies in the sequence data, and to solve this problem, LSTMs are proposed.

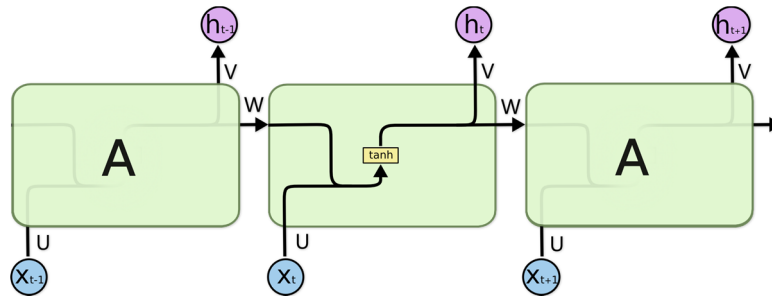


Figure 1: the basic structure of RNN

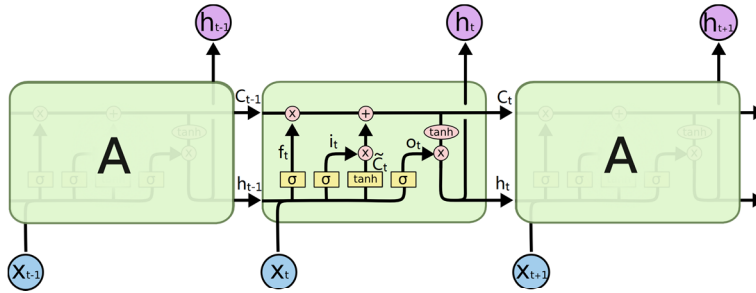


Figure 2: the basic structure of LSTM

Long Short-Term Memory networks, usually called LSTMs, are a special kind of RNN to solve the problem of learning long-term dependencies. LSTMs were initially introduced by Hochreiter and Schmidhuber in 1997, and were refined and applied into various domains by other researchers after that[13]. It has been proven that LSTMs can work tremendously well on a large variety of problems, and they are still with high popularity now. The basic structure of LSTMs is shown in Fig.2. Like RNNs, LSTMs also have a chain-like structure, but the

repeating module is more complicated. Instead of having a single neural network layer, there are four with different modules, interacting in a very special way.

As shown in Fig. 2, a green block is the constant error carousel cell, which will pass straight down the whole recursive process without activation function so that the problem of vanishing gradient will not happen. That is why LSTMs can handle longer-term dependency of sequence data than RNNs. Moreover, there are three special units called input, forget, and output gates respectively, together with an activation function. As their names have shown, these gates focus on the different information and decide to what extent these information should be used or abandoned by the cell.

Similar to RNNs, LSTMs derive the output sequence by calculating the network unit activation as follow:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4)$$

$$C_t = i_t \otimes \tanh(W_c x_t + U_c h_{t-1} + b_h) + f_t \otimes C_{t-1} \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

where  $W_i$ ,  $W_f$ , and  $W_o$  are the matrix of weights corresponding to the input, forget, and output gates.  $U_i$ ,  $U_f$ , and  $U_o$  are the weight matrix for the hidden states.  $b_i$ ,  $b_f$ ,  $b_o$  are the bias vectors of the respective gates.  $\sigma$  means the activation function, where the logistic sigmoid function tends to be applied.  $i_t$ ,  $f_t$ ,  $o_t$  and  $C_t$  are the state output of corresponding gates and the cell at time step  $t$ .

## 3.2. Proposed Model

### 3.2.1. Input Data Processing

Although deep learning based statistical model like LSTMs is powerful when dealing with time series forecasting, it is caught into a dilemma when comes to the problem of water level prediction because the input data we would like to use is the precipitation data only. As we know, rainy days only account for a small part of the whole year. Moreover, the precipitation of a basin tends to come almost simultaneously and will not last for too long. As the result, there is a wide range of zero input in our precipitation data, which will result in a terrible effect on the prediction performance, as shown as Fig. 3. Since the basic principle of deep learning model is similar to linear regression, the sudden disappearance of input will inevitably result in the rapid decrease of the output. In this situation, although the model can capture the rise of water level, its dissatisfaction of flow conservation makes its prediction less reliable.

Since the performance of deep learning model is excellent when both the precipitation data and water level data serve as the input, we hold the view that the water level related input is the key point to keep the change of water level reasonable.

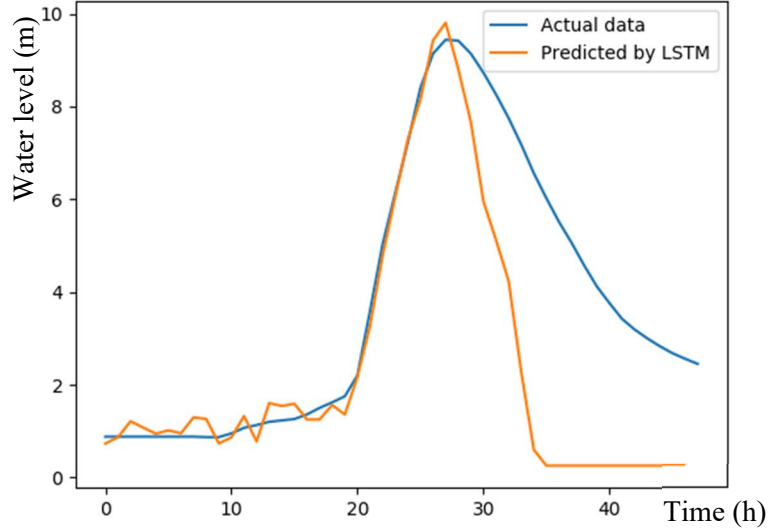


Figure 3: terrible effect on the prediction performance of LSTM

To resolve this, for details, a water level related input needs to be generated and added in every timestep of prediction to ensure that the model will not meet the problem of input disappearance. It is one of the biggest advantages of physical models to maintain flow conservation. One of the physical models that has been widely used for water level prediction in Japan is called Tank model based on Storage Function Method. To calculate the outflow of the river, according to the model, the amount of water flowing into the river basin does not completely equal to the rainfall, but the amount of water stored in the basin. The relationship among precipitation, outflow and stored water can be described using the following equations[3]:

$$S = kq^p \quad (7)$$

$$\frac{dS}{dt} = R_{t-T_L} - q_t \quad (8)$$

Here,  $S$  is the amount of storage,  $R$  is the amount of rainfall,  $k$  and  $p$  are the outflow parameters,  $q$  is the amount of outflow, and  $T_L$  is the time step. According to (7), the amount of outflow  $q$  depends on the storage volume  $S$ . Therefore, the flow change  $\Delta q$  is almost equal to the storage volume change  $\Delta S$ . In addition, changes in the water level can be considered to correspond to changes in the discharge. Based on the above considerations, it is considered that the change in storage volume  $\Delta S = R_{t-T_L} - q_t$  is more suitable for the calculation rather than the rainfall  $R$ . Therefore, the proposed hybrid model uses the change in storage capacity  $\Delta S$

and the change in water level as input data to improve the prediction accuracy. What's more, parameter  $k$  and  $p$  can be determined as follows. However, since it is not the core aspect in our paper, we just explain the theory readily. First, the value of  $R_{t-T_L}$  and  $q_t$  can be obtained at any time period, so the amount of storage  $S$  can be calculated at every moment. Second, time step (time lag)  $T_L$  we used in our prediction is from one to six hours, so according to  $T_L$ , we have six different  $k$  and  $p$  respectively. Third, in order to determine the value  $k$  and  $p$ , we take the logarithm operation for eq. (7) and we get eq. (9):

$$\ln S = \ln k + p \cdot \ln q_c \quad (9)$$

Finally, according to  $T_L$ ,  $k$  and  $p$  can be determined by linear regression on logarithmic axis. Parameter  $k$  and  $p$  both meant to calculate the value  $S$ , and we only do the calculation by using eq. (7) and (8). Of course, we can use  $k$  and  $p$  to calculate many attributes of the river basin, but in our research, we only need to get the value  $S$  as our input. To calculate  $\Delta S$ , it is needed to know the change of water level. To calculate it, we change the structure of data input and recurrently concatenate the prediction output  $O_t$  with the next input  $I_{t+1}$ , as shown in Fig. 4.

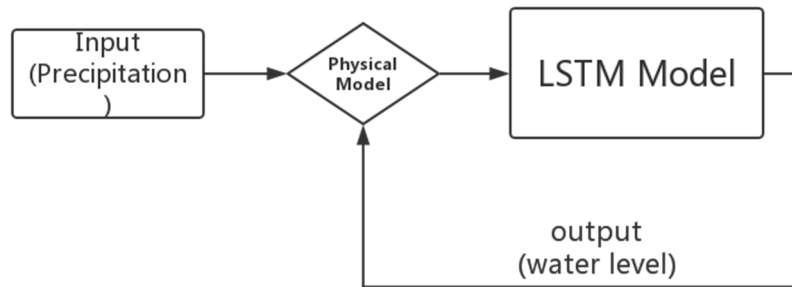


Figure 4: The flow of model training

Actually, we use precipitation as our input, but some positions' water level has a huge influence on Hiwatashi where we meant to make a prediction. What's more, as we mentioned we need to calculate the change in storage volume  $\Delta S$ , so 5 vital positions' water level have also been considered as our input to make the prediction have a better performance. What we need to do is to put all the precipitation data from 15 positions and water level data from 5 necessary positions simultaneously, and the calculation progress will implement inside our proposed Hybrid model. Therefore, there seems no feedback showing upon the model structure, but the feedback has already been inside the model.

### 3.2.2. Loss Function

The key point for deep learning model to fit the data is to minimize the value of the loss function. In order to meet the flow conservation, it will be convenient to revise the loss function to make it satisfy the restriction of flow conservation. The basic loss function, we use here is MSE (Mean Squared Error), which is calculated in the following way:

$$MSE = \frac{1}{n} \sum_{i=1}^N (O_i - \hat{O}_i)^2 \quad (10)$$

To keep the balance of flow conservation, the change of the storage volume  $\Delta S$  should keep the balance with the inflow, which can be represented by the precipitation  $q$  somehow in this situation, so that the revised loss function can be defined as follow:

$$Loss = MSE + \Delta S - kq \quad (11)$$

### 3.2.3. Model Setting

In our research a hybrid model is proposed, which is the combination of LSTM and Tank model. The structure of the proposed model and the parameters setting are shown below as Table 1 and Fig. 5:

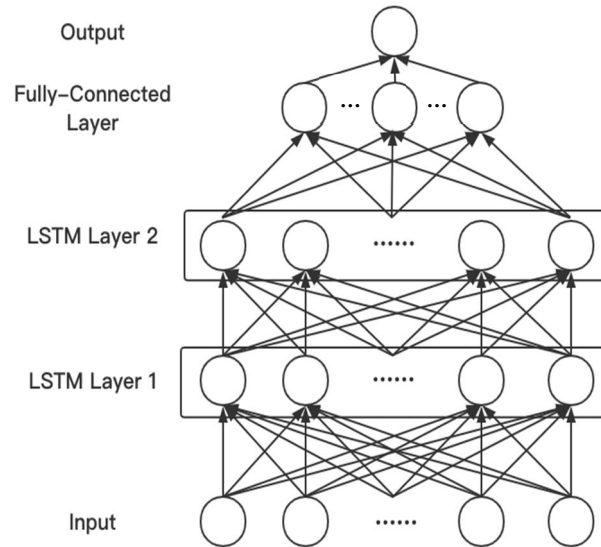


Figure 5: Model structure



Table 1: Model setting

Input	Precipitation
Output	Water level
Model structure	120-80-40-1
Input time range	6hrs
Forecasting time	1hr - 6hr
Optimizer	Adam
Dropout	0.5
Batch size	200
outflow rate	0.8
inflow rate	0.4

As we mentioned before, we tend to use the precipitation of 15 different areas around Hiwatashi to predict its water level. Thus, it is obvious that the number of input layer is 15, corresponding to 15 different locations which are in proximity of our target position, Hiwatashi. In addition, for sake of predicting the water level in Hiwatashi from next 1 to 6 hours, we cannot use only one time period as our input data, because we found it is extremely hard to obtain a decent prediction result. After we decided to use 6 as our time step, we got a brilliant result (the specific forecast results will be given in the following chart). Since we have plenty of precipitation data of each area, we are likely to input 200 data into LSTM for training at a time, which is reflected in batch size, and set Dropout to 0.5, in order to meet our forecast accuracy requirements. Moreover, we use Adam, which is a first-order optimization algorithm that can replace the traditional stochastic gradient descent process, as our Optimizer. Adam is able to iteratively update neural network weights based on training data, so it seems like a perfect method for our hybrid model.

It is very hard to deny that using precipitation data only to make prediction has some flaws, because never should we ignore the influence of physical factors, such as topography, runoff, penetration and so forth, on the inflow of the rain. This means if we only use rainfall data as input, the predicted result we get will be larger than the actual result, due to the fact that it is impossible for all the rainfall to flow into the river basin. Therefore, we use two parameters, inflow rate and outflow rate, to simulate the average situation in order to compensate for the inaccuracy of using rainfall data only. According to our specific calculation, we set outflow rate and inflow rate to 0.8 and 0.4 respectively so as to restore the real situation. Finally, we got our hybrid model and made water level prediction.

## 4. Comparison

### 4.1. Comparison of overall prediction trend

In order to prove our proposed model has advantages when using precipitation data only to make predictions, we compared ours with another two different kinds of statistical models. We meant to make comparison with outflow models at the beginning as well. However, after we collected information about outflow models, we found that it is impossible to make predictions by using precipitation data only such as our proposed model, since all kinds of out flow models need many data related to the shape of areas, such as surface runoff, soil permeability and so on. Our proposed model aims at predicting water level all over the country, but it is also impossible for outflow models to be applied to different terrains. Form previous paper, it is proved

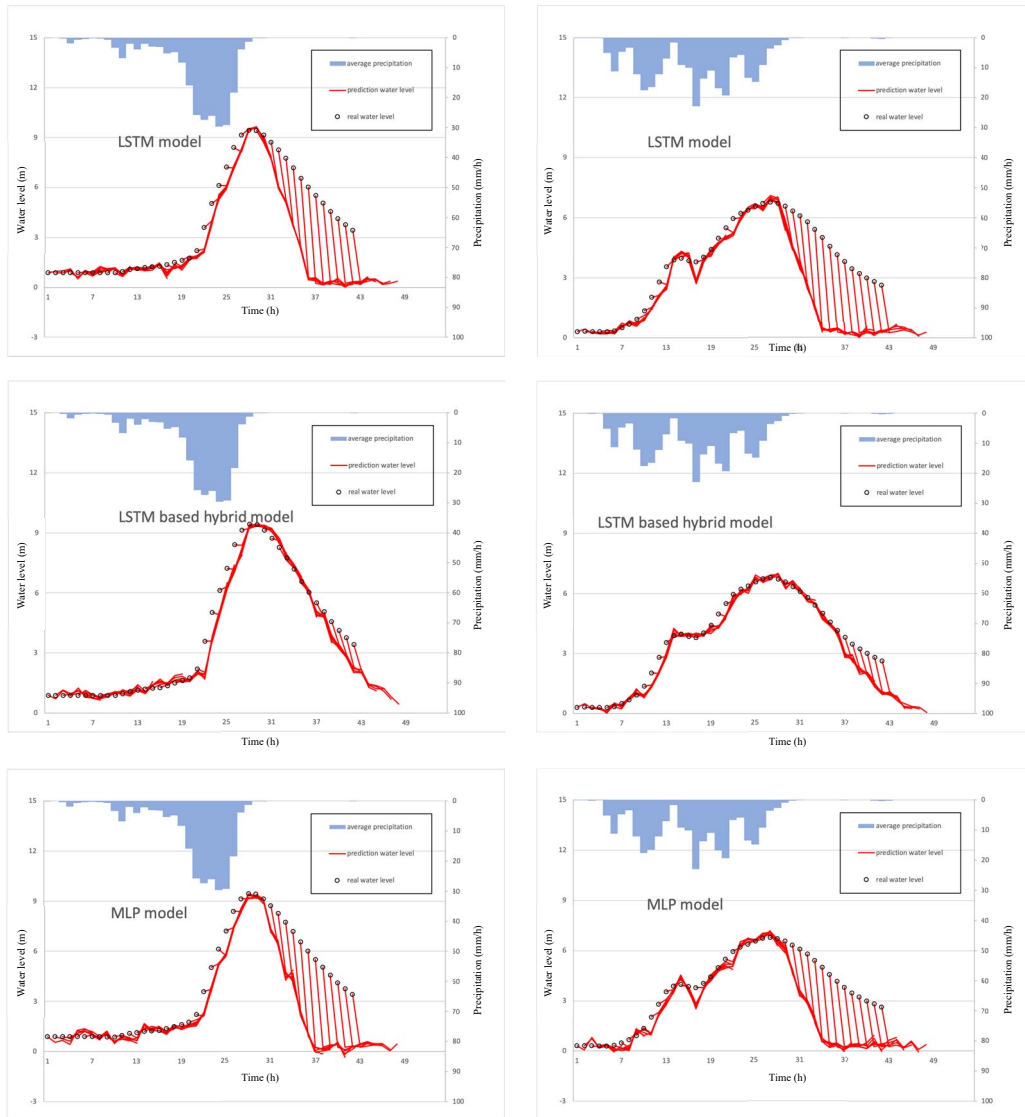


Figure 6: Prediction result (left 9/28/1990, right 7/3/2008)

that when it comes to water level prediction, statistical models can perform better than outflow models [5]. Therefore, we give up comparing outflow by quantitative analysis ultimately. We chose Hiwatashi in the Oyodo river in Kyushu as our prediction area and select representative two time periods, 9/28/90 and 7/3/2008, to make predictions in 42 hours. We use the real water level for the first 6 hours before the prediction moment as input and predict the water level for next 1 to 6 hours. The results are presented below as Fig. 6.

From overall view, according to the result predicting by three different methods, our proposed model, LSTM based hybrid model, performed best.

It is obvious from the average precipitation that the pattern of precipitation is very different in 1990 and 2008. In 1990, the precipitation is relatively concentrated, which means the storm came very suddenly. However, in 2008, almost every used time period had precipitation, which means the precipitation is relatively steady. No matter in which kind of precipitation patterns, our proposed model's prediction result is much closer to the actual water level. The red line also had a better fit, especially at the end of the precipitation which problem we meant to overcome. We noticed that large errors appeared in both LSTM and MLP models around 28<sup>th</sup> time point, but in LSTM based hybrid model, when it comes to 37<sup>th</sup> time point only some small errors gradually appeared. Therefore, our proposed model solved problems with poor forecasting accuracy when it comes to the end of rainfall effectively by using precipitation data only in some degree.

LSTM based hybrid model performed better not only at the end of prediction, but also at the beginning of prediction. Either LSTM or MLP model had sharp turning point according to the red line, especially on 7/3/2008. However, our proposed model's prediction red line is smoother than those two. In addition, the flutter of the red line is feeble at the beginning of prediction compared with the other two methods, which also means LSTM based hybrid model's prediction result is closer to the actual water level.

#### **4.2. Comparison of 1-hour prediction to 6-hour prediction**

In order to more intuitively reflect the prediction accuracy of each model in one to six hours, we use RMSE (Root Mean Square Error) to analyze the prediction results from 1-hour to 6-hour respectively to calculate each model's RMSE. The results are shown below as Table 2.

Table 2 RMSE result

	1 hr	2 hr	3 hr	4 hr	5 hr	6 hr
Hybrid	0.64	0.67	0.72	0.81	0.82	0.86
LSTM	1.43	1.51	1.54	1.58	1.62	1.73
MLP	1.46	1.56	1.63	1.66	1.72	1.79

From Table 6 we can notice that RMSE became higher as the forecast time increases in every model, and our proposed hybrid model's growth value is approximately 0.2. However, LSTM and MLP models' growth value is around 0.3, which means LSTM based hybrid model has more reliable prediction results in the long-term prediction. What's more, it's obvious that from 1-hour prediction to 6-hour prediction, LSTM based hybrid model's RMSE is lower than LSTM and MLP in every prediction time point, even 6-hour prediction result's RMSE of our proposed model is lower than 1-hour's RMSE of the other two.

## 5. Conclusion

Under the premise of using precipitation data only as input, we created LSTM based hybrid model to make prediction. According the prediction results, we are able to say our proposed model performs best among three different models. One of the reasons is it can make prediction in any pattern of precipitation and especially, when the rain stops, the output can also make sense. Another reason that cannot be ignored is compared with statistical models, the accuracy of LSTM based hybrid model has been greatly improved due to the restriction of the tank model we have added. No matter short-term prediction or long-term prediction, our proposed model has available performance. Moreover, as it is not limited by terrain, we can imagine LSTM based hybrid model can be applied to either first-class river basin or small river basin all over Japan by using precipitation data.

## References

- [1] Shuichi Kure and Tadashi Yamada: Theoretical derivation of the conceptual rain-fall-runoff models, *Japan Soc Hydro. and Water Resour.*, 22:5 (2009), 386-400.
- [2] Hideshi Ikeda, Takashi Wakamatsu, Ko Nakaya and Seiya ABE: A review of sediment yields observed in mountainous forest areas and its prediction method, *Japan Soc Hydro. and Water Resour.*, 25:6 (2012), 396-409.
- [3] Masayuki Hitokoto and Masaaki Sakuraba: Hybrid deep neural net-work and distributed rainfall-runoff model for the real-time river stage prediction, *Proceeding of JSCE.*, 73:1 (2017), 22-33.
- [4] Gen Nagatani, Naoya Mizuno, Hiroya Ishida, Kazuya Ozawa and Kaoru Takara: Application of distributed rainfall runoff model to slope failure simulation, *Proceeding of JSCE F5.*, 68:1 (2012), 16-26.
- [5] Noto Fumikazu, Maruyama Toshisuke, Hayase Yoshio, Takimoto Hiroshi and Nakamura Kimihito: An evaluation of snow storage depth in the tedoru river basin using tank model, *Irrigation, Drainage and Rural Engineering Journal.*, 78:4 (2010), 265-271.
- [6] Yamamoto Takahiro and Lu Minjiao: The improvement of storage function method for continuous simulation, in *Japan Society of Hydrology and Water Resources Research and Publish Conference*, 2008.

- [7] Tadakatus Takasaki, Akira Kawamura and Hideo Amaguchi: Development of the storage function model for urban flood considering the outflow through combined sewer system, *Japan Soc Hydro. and Water Resour.*, 21:3 (2008), 228-241.
- [8] Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P. : Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrology and Earth System Sciences.*, 16:10 (2012), 3863-3887.
- [9] Zhang, Jinxin, et al.: *Time Series Analysis*. Handbook of Medical Statistics 269, 2017.
- [10] Faruk, Durdu Ömer. "A hybrid neural network and ARIMA model for water quality time series prediction." *Engineering Applications of Artificial Intelligence.*, 23:4 (2010), 586-594.
- [11] Koutnik, Jan, et al.: *A clockwork rnn*. arXiv preprint arXiv:1402.3511, 2014.
- [12] Zhu X, Sobihani P, Guo H. Long short-term memory over recursive structures, in *International Conference on Machine Learning*, 2015, 1604-1612.
- [13] Ordóñez F, Roggen D.: Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors.*, 16:1 (2016), 115.