Comparison of physics-based and data-driven models for streamflow simulation of the Mekong river

Giha Lee
Dept. of Construction & Disaster Prevention Eng.,
Kyungpook National University
Sangju, Korea
leegiha@knu.ac.kr

Yonghun Jung
Dept. of Construction & Disaster Prevention Eng.,
Kyungpook National University
Sangju, Korea
leegiha@knu.ac.kr

Daeup Lee
Dept. of Construction & Disaster Prevention Eng.,
Kyungpook National University
Sangju, Korea
leegiha@knu.ac.kr

Tae-Woong Kim
Dept. of Civil and Environmental Eng.,
Hanyang University
Ansan, Korea
twkim72@hanyang.ac.kr

***Abstract*—In recent, the hydrological regime of the Mekong River is changing drastically due to climate change and haphazard watershed development including dam construction. Information of hydrologic feature like streamflow of the Mekong River are required for water disaster prevention and sustainable water resources development in the river sharing countries. In this study, runoff simulations at the Kratie station of the lower Mekong River are performed using SWAT (Soil and Water Assessment Tool), a physics-based hydrologic model, and LSTM (Long Short-Term Memory), a data-driven deep learning algorithm. The SWAT model was set up based on globally-available database (topography: HydroSHED, landuse: GLCF-MODIS, soil: FAO-Soil map, rainfall: APHRODITE, etc) and then simulated daily discharge from 2003 to 2007. The LSTM was built using deep learning open-source library TensorFlow and the deep-layer neural networks of the LSTM were trained based merely on daily water level data of 10 upper stations of the Kratie during two periods: 2000~2002 and 2008~2014. Then, LSTM simulated daily discharge for 2003~2007 as in SWAT model. The simulation results show that Nash-Sutcliffe Efficiency (NSE) of each model were calculated at 0.9(SWAT) and 0.99(LSTM), respectively. In order to simply simulate hydrological time series of ungauged large watersheds, data-driven model like the LSTM method is more applicable than the physics-based hydrological model having complexity due to various database pressure because it is able to memorize the preceding time series sequences and reflect them to prediction**

Keywords—LSTM; Mekong River; SWAT

I. INTRODUCTION

In recent, the hydrological regime of the Mekong River is changing drastically due to climate change and haphazard watershed development including dam construction. Information of hydrologic feature like streamflow of the Mekong River are required for water disaster prevention and sustainable water resources development in the river sharing countries. In this study, runoff simulations at the Kratie station (Cambodia) of the lower Mekong River are performed using SWAT (Soil and Water Assessment Tool), a physics-based hydrologic model, and LSTM (Long Short-Term Memory), a data-driven deep learning algorithm.

II. METHDOLOGY

A. SWAT

The Soil & Water Assessment Tool (https://swat.tamu.edu/)is a small watershed to river basin-scale model used to simulate the quality and quantity of surface and ground water and predict the environmental impact of land use, land management practices, and climate change. SWAT is widely used in assessing soil erosion prevention and control, non-point source pollution control and regional management in watersheds. SWAT is a continuous time model that operates on a daily time step at basin scale. SWAT uses a two-level disaggregation scheme; a preliminary sub-basin identification is carried out based on topographic criteria, followed by further discretization using land use and soil type considerations. Areas with the same soil type and land use form a Hydrologic Response Unit (HRU), a basic computational unit assumed to be homogeneous in hydrologic response to land cover change. The SWAT is used as a basic tool for Mekong river hydrologic simulations in the Mekong River Commission [3], [4], [5].

B. LSTM

Long Short-Term Memory, an evolution of RNN, was introduced by Hochreiter and Schmidhuber [1] to address problems of the aforementioned drawbacks of the RNN vanishing gradient by adding additional interactions. LSTMs are a special kind of RNN, capable of learning long-term dependencies and remembering information for extended periods of time as a default. According to Olah [2], the LSTM model is organized in the form of a chain structure. However, the repeating module has a different structure. Instead of a single neural network like standard RNN, it hasfour interacting layers with a unique method of communication.

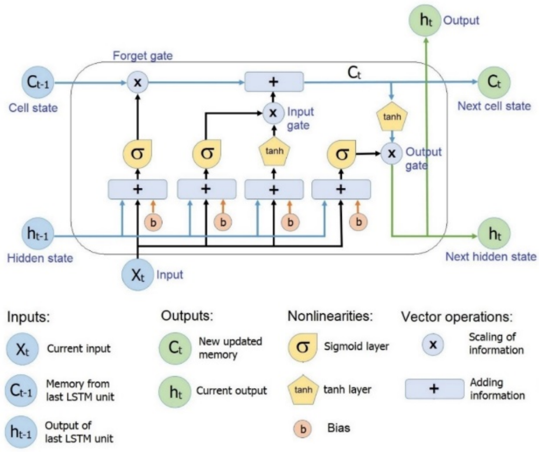

Fig. 1 The structure of LSTM neural network

The structure of the LSTM neural network is shown in Fig. 1.

A typical LSTM network is comprised of memory blocks called cells. Two states are being transferred to the next cell, the cell state, and the hidden state.The cell state is the main chain of data flow,which allows the data to flow forwardessentially unchanged.However, some linear transformations may occur. The data can be added to or removed from the cell state via sigmoid gates. A gate is similar to a layer or a series of matrix operations, which contain different individual weights. LSTMs are designed to avoid the long-term dependency problem because it uses gates to control the memorizing process.

C. Lower Mekong River

The Mekong River is a nationally shared river originating from the uppermost reaches of China and passing through the Mekong Delta of the most downstream Vietnam. The watershed covers about 10 times (795,000 km^2) of South Korea with an annual average flow of about 15,000 m^3/s. Fig. 2 shows the location of the main water level stations in the lower Mekong region (excluding upstream in China), and Table 1 shows the watershed area and runoff. As shown in Table 1, the annual average flow rate at the Kratie site is about 90% of the Mekong River main stream.

Table 1. Mean annual discharge at main water level stations

Station Name		Catchment area (km^2)	Mean annual discharge (m^3/s)	as % total Mekong
1	Chiang Saen	189,000	2,700	18
2	Luang Prabang	268,000	3,900	26
3	Chiang Khan	292,000	4,200	28
4	Vientiane	299,000	4,400	29
5	Nong Khai	302,000	4,500	30
6	Nakhon Phanom	373,000	7,100	47
7	Mukdahan	391,000	7,600	51
8	Pakse	545,000	9,700	65
9	Stung Treng	635,000	13,100	87
10	Kratie	646,000	13,200	88
Basin Total		795,000	15,000	100

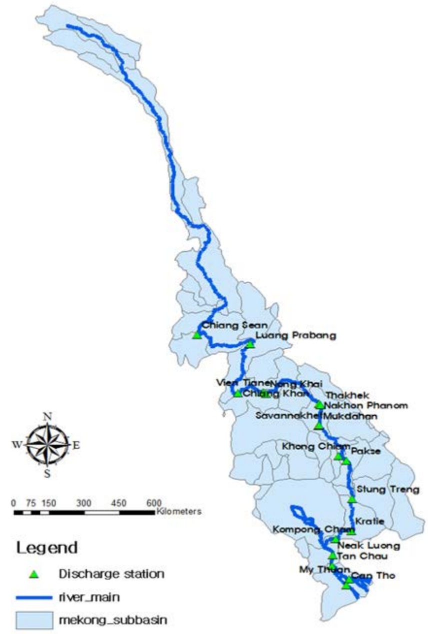

Fig. 2Main water stations of the Mekong River

III. APPLICATIONS

The SWATmodel was set up based on globally-available database(topography: HydroSHED, landuse: GLCF-MODIS, soil: FAO-Soil map, rainfall: APHRODITE, etc.) and then simulated daily discharge from 2003 to 2007. The data from 2000 to 2002 were used for model parameter optimization by SWAT-SUP, automatic calibration tool. The LSTM was built using deep learning open-source library TensorFlow (https://www.tensorflow.org) and the deep-layer neural networks of the LSTM were trained based merely on daily water level data of 10 upper stations of the Kratie during two periods: 2000 ~ 2002 and 2008 ~ 2014. Fig. 3 shows the water level time series of the stations for LSTM application. Then, LSTM simulated daily discharge for 2003 ~ 2007 as in SWAT model.

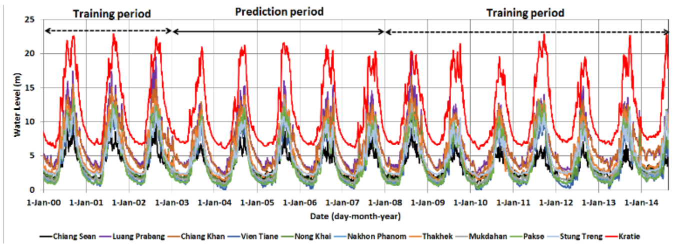

Fig. 3 Data sets for LSTM training and prediction (water level time series)

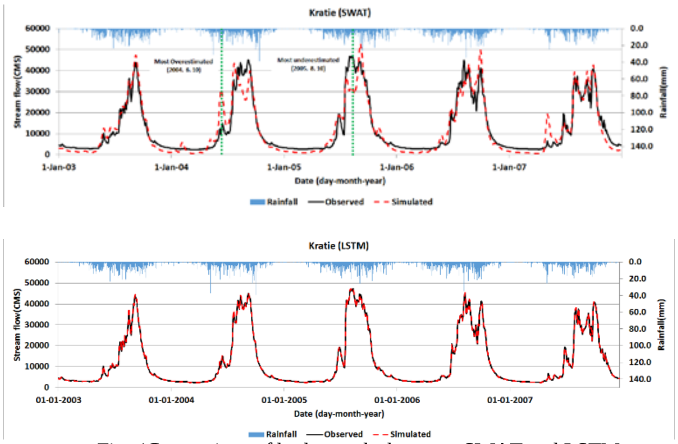

Fig. 4Comparison of hydrographs between SWAT and LSTM

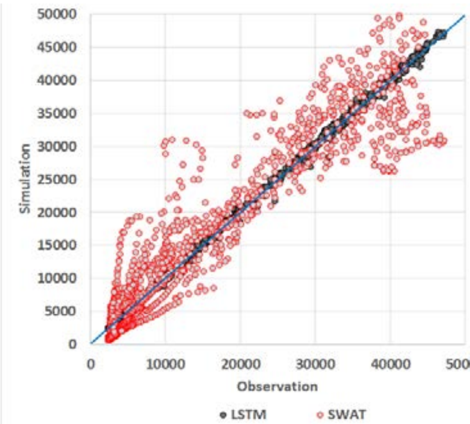

Fig. 5 Stream flow scatter plots of the two models

IV. RESULTS& DISCUSSION

To evaluate the performance of two models, NSE(Nash-Sutcliffle Efficiency) and RMSE(Root Mean Square Error)were used.

$$NSE = \left(1 - \frac{\sum_{i=1}^{n} \left(O_i - P_i \right)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}_i \right)^2} \right) * 100 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(O_i - P_i \right)^2} \quad (2)$$

Here, O_i and P_i are observed discharges and simulated discharge at time t, respectively; $\bar{O}$ is the mean of observed discharge; n is the total number of observations.

The results of the quantitative evaluation of the accuracy of the SWAT model were as follows: RMSE: 3941.71 m^3/s, NSE: 0.9. The SWAT simulation results were overestimated or underestimated for the simulation period. On the other hand, the LSTM model showed a very good agreement of hydrograph: NSE: 0.99 and RMSE: 330 m^3/s. The LSTM model provides very stable flow simulation results regardless of flow magnitude.

A variety of physical rainfall-runoff models have already been developed for the nonlinear nature of runoff analysis. When sufficient physical data and parameter correction are assured, these models have been proven to be very usefulfor simulation of rainfall-runoff and spatial hydrological variability However, the complexity of such a physical model may cause uncertainty problems in model structure, grid scale, and parameters, and may be limited by data construction and simulation time. Therefore, the LSTM model, which memorizes the preceding information in the time series prediction at a specific time steps and reflects it to the prediction,can be used as a complementary means.

ACKNOWLEDGMENT

This research was supported by International Hydrologic Program funded by Ministry of Environment in Korea.

REFERENCES

[1] Hochreiter, S., Schmidhuber, J. Long short-term memory. *Neural Comput.*, *9*, 1735-1780, **1997**.

[2] Olah, C. Understanding LSTM networks. In *http://colah.github.io/posts/2015-08-Understanding-LSTMs/*, **2015**.

[3] Shrestha, B., Babel, M. S., Maskey, S., Griensven, A. V., Uhlenbrook, S., Green, A., and Akkharath, I. Impact of climate change on sediment yield in the Mekong River basin: a case study of the Nam Ou basin, Lao PDR.*Hydrology and Earth System Sciences*, Vol. 17, No. 1, 1-20, **2013.**

[4] Vilaysane, B., Takara, K., Luo, P., Akkharath, I., and Duan, W. Hydrological stream flow modelling for calibration and uncertainty analysis using SWAT model in the Xedone river basin, Lao PDR. *Procedia Environmental Sciences*, Vol. 28, 380-390, **2015**.

[5] Vu, M. T., Raghavan, S. V., and Liong, S. Y. SWAT use of gridded observations for simulating runoff - A Vietnam river basin study. *Hydrology and Earth System Sciences*, Vol. 16, No. 8, 2801-2811, **2012**.