Flood Forecasting using Hybrid Approach of Hydrological Model and Machine Learning

〇Sophal Try・Takahiro Sayama・Ty Sok・Sophea Rom Phy・Shunji Kotsuki・Kodai Yamamoto・Chantha Oeurng

Introduction and Objective

Flood is widely recognized as the most common and frequent natural phenomenon which currently threatens huge damage worldwide. Flood forecasting has been developed and implemented worldwide through out various methodologies, including physical-based hydrological model, empirical model, and direct analysis of climate variables. Nowadays, an accurate and timely prediction of flood forecasting is still a challenging task for forecasting system developers.

The Prek Thnot River (PTR) is one of the flood-prone areas where severe floods occur every year and causes the damages to residents in downstream in Phnom Penh, Kandal, and Kampong Speu provinces in Cambodia. The Kamo River (KR) is located in a mountainous region with steep topography in Kyoto Prefecture, Japan. Flood damages in both river basins are relatively large and affected the people and the economic activities in surrounding area. The characteristics of these two river basins are completely different. After heavy rainfall, the river flow could reach its peak in just around 4 hours in the KR, while it took generally approximately 2 days in the PTR.

In this study, a hybrid model which integrates the outputs from a physical-based hydrological model and machine learning approach is developed to predict the river flood in two case studies in the PTR in Cambodia and KR in Japan.

Methodology

This study used a fully distributed rainfall-runoff-inundation (RRI) model for river discharge and water level simulations (Sayama et al., 2012). The RRI model was calibrated and validated with gauged observed rainfall.

The forecasted rainfall from NICAM-LETKF numerical weather prediction (so-called GSMaPxNEXRA) dataset. GSMaPxNEXRA data is produced by Global Cloud Resolving Model with Data Assimilation with hourly 1° × 1° spatial resolution of a 5-day of the forecasting period. The GSMaPxNEXRA is used for the evaluation of forecasting of flood in the PTR.

Moreover, the composite rainfall product from gauged observation and radar from Japan Meteorological Agency (JMA) and its forecasting rainfall product was used as input to simulate and forecast flood in the KR river with 1km x 1km of spatial resolution and 6-hour lead time of forecast.

On the other hand, Support Vector Machine (SVM) a supervised learning model with associated learning algorithms, is used to improve the output result of the RRI model.

Two statistical indicators, the coefficient of extrapolation (CE) and the coefficient of persistence (CP), were used to evaluate the predictive capacity of forecasting performance.

$$\mathrm{CE} = 1 - \frac{\sum_{t=1}^{n}\left(Q_s^t - Q_o^t\right)^2}{\sum_{t=1}^{n}\left(Q_o^t - Q_1^t\right)^2}$$

$$\mathrm{CP} = 1 - \frac{\sum_{t=1}^{n}(Q_s^t - Q_o^t)^2}{\sum_{t=1}^{n}(Q_o^t - Q_o^{t-j})^2}$$

where j is the prediction lead. Q_1 is the forecast from the linear extrapolation of the two most recent measurements. Q_s^t and Q_o^t are simulated and observed datasets at time step t.

Results

First, the RRI model was calibrated and validated using gauged and radar precipitation for both study areas in the PTR and KR. Then, the same parameter setting of the model was used to simulate the extrapolation of flood forecasting by using real-time simulation as an initial condition.

The results of flood forecasting were compared and evaluated based on river discharge in the PTR and river water level in the KR. **Fig.1** compared the performance of forecasted discharge resulting from the RRI model before and after applying the machine learning. The accuracy of predictive capacity before applying the Machine Learning showed statistical indices of CE = -2.42–0.70 and CP = -2.56–0.07 for the forecasting periods of 1 to 5 days. After Machine Learning, these indices had significantly improved to CE = 0.71–0.91 and CP = 0.68–0.69.

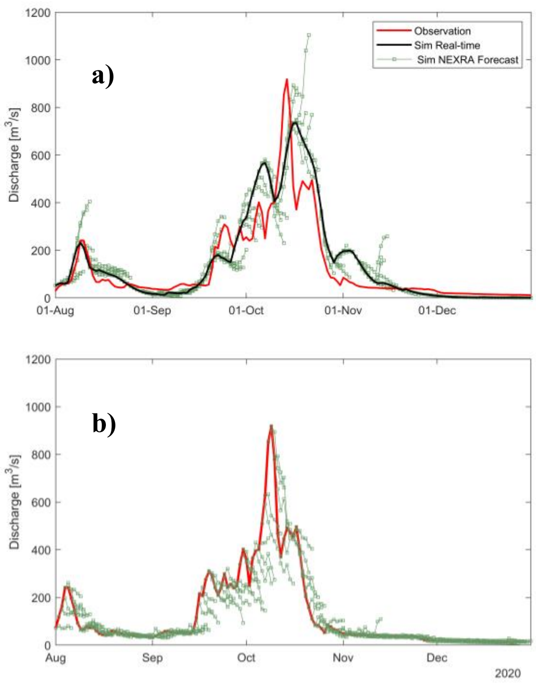

Fig.1 Performance comparison of flood forecast before (a) and after Machine Learning (b) for Prek Thnot River, Cambodia

In addition, the flood forecasting performance in the KR was also evaluated for the forecast lead of 6 hours (**Fig.2**). The primary results showed CE = 0.04–0.73 and CP = -0.23–0.55 for a range of 1-hour to 6-hour forecast lead. The Machine Learning could improve the accuracy to CE = 0.82–0.83 and CP = 0.63–0.78.

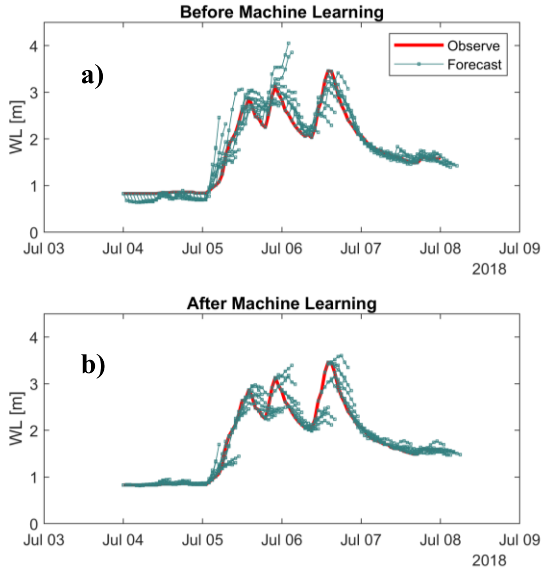

Fig.2 Performance comparison of flood forecast before (a) and after Machine Learning (b) for Kamo River, Japan

Conclusions

This study successfully evaluated the real-time flood forecasting in the PTR in Cambodia and KR in Kyoto Prefecture, Japan. The hybrid approach of physical-based hydrological model and Machine Learning could significantly improve the performance and accuracy of the predictive capacity of flood forecasting in the PTR and KR. Therefore, it would be beneficial to implement this approach to the operational real-time flood forecasting in these two river basins to increase the awareness of flood early warning and reduce the potential damages affected by severe flooding.

Acknowledgement

This research was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid: 21F21071 and the Cambodia Higher Education Improvement Project (Credit No. 6221-KH).

Reference

Sayama, T., Ozawa, G., Kawakami, T., Nabesaka, S., Fukami, K., 2012. Rainfall–runoff–inundation analysis of the 2010 Pakistan flood in the Kabul River basin. Hydrological Sciences Journal 57, 298–312.