TA-138L

PREDICTING THE RESERVOIR INFLOW OF BHUMIBOL DAM USING XGBOOST MACHINE LEARNING ALGORITHM

Pheeranat Dornpunya[1], Areeya Rittima[2*], Yutthana Phankamolsil[3], Allan Sriratana Tabucanon[4], Wudhichart Sawangphol[5], Jidapa Kraisangka[6], Yutthana Talaluxmana[7], Varawoot Vudhivanich[8], and Assada Kijpayung[9]

[1, 2*]Department of Civil and Environmental Engineering, Faculty of Engineering, Mahidol University, Thailand

[3]Environmental Engineering and Disaster Management Program, Mahidol University, Kanchanaburi Campus, Thailand

[4]Faculty of Environment and Resource Studies, Mahidol University, Thailand

[5, 6]Faculty of Information and Communication Technology, Mahidol University, Thailand

[7]Department of Water Resources Engineering, Faculty of Engineering, Kasetsart University, Thailand

[8]Department of Irrigation Engineering, Faculty of Engineering at Kamphaengsaen, Kasetsart University, Thailand

[9]Royal Irrigation Department, Thailand

*Corresponding author's e–mail: areeya.rit@mahidol.ac.th

Keywords— Bhumibol Dam, Extreme Gradient Boosting (XGBoost), Reservoir Inflow Prediction

ABSTRACT

XGBoost, a tree-based ensemble machine learning algorithm, was used to predict the daily and monthly reservoir inflows of Bhumibol Dam in Thailand. The prediction models were developed using the observed inflow and climate data from 2000 to 2020 as major features. The structures of prediction model were determined by defining the future reservoir inflow at the predictive lead time steps as a function of the past inflow, precipitation, and humidity at time step t, as well as average inflow at the delayed time steps. Consequently, 54 XGBoost scenarios of the daily and monthly inflow prediction models were trained and validated by altering the model parameters namely; training–testing dataset ratio, learning rates, maximum number of iterations, and early stopping rounds. The statistical performance metrics namely; RMSE, MSE, R^2, R, and NSE were employed to evaluate the model performance. It can be drawn from the validation results that the XGBoost model can deliver reliable and robust prediction outcomes. In addition, the XGBoost model is capable of predicting the complete performance of the daily reservoir inflow with higher accuracy than the monthly inflow.

INTRODUCTION

A more frequent occurrence of flood and drought in Thailand particularly in the past decade reflects the uncertainty of hydrological data due to changes in the regional climate and economic growth of the country. This has significant implications for the operational actors to revise the strategic plan based upon the data-driven decision-support tools to reduce disaster risks and losses. The accurate and reliable hydrological prediction plays vital role in the decision-making process specifically for real time operation of dam-reservoir system. Machine Learning (ML) which is the advanced area of Artificial Intelligence (AI), has been extensively used to improve predictive accuracy and understand hydrological uncertainty and provide the multiple lead times. It has proved a great success in predicting hydrological data such as rainfall, reservoir inflow, and river flow. [1]. Therefore, this study aims at evaluating the predictability of machine learning-based prediction models for reservoir inflow prediction. The extreme gradient boosting (XGBoost) algorithm with R programming language was employed to develop the daily and monthly prediction models of Bhumibol Dam where high variability of reservoir inflow has apparently found and probabilistic forecast has become increasingly important.

METHODOLOGY

Setting up the prediction model structures were performed by specifying the highly-correlated predictor variables as the model features including number of average inflow at the delayed time steps, climate data at time step t, and observed inflow data at time step t. Three datasets of training–testing ratio namely; 60:40, 70:30, and 80:20 and learning rates of 0.1, 0.01, and 0.001 were specified. Accordingly, 54 XGBoost scenarios of daily and monthly models were trained and validated to produce good predictive results. In the predictive modelling process, predictor variables were firstly imported into the prediction models. Secondly, the time series of selected variables were divided into training and testing datasets according to the designated ratio. Thirdly, development of XGBoost model was controlled by the hyperparameter setting such as number of iterations (nrounds), learning rate (Eta), and early stopping rounds parameters. Accordingly, the maximum number of iterations was 10,000. The learning rate allows model to achieve faster convergence of training dataset. So, the learning rates of 0.1, 0.01, and 0.001 were determined in this study. The early stopping rounds are generally used to stop training procedures when the loss on training dataset starts increasing. Lastly, the level of agreement between the predicted values and observed values were evaluated by the statistical methods namely; RMSE, MSE, R^2, R, and NSE.

RESULTS AND DISCUSSIONS

It is appeared that the best daily reservoir inflow prediction model can be made by specifying the reservoir inflow at lead time t+1 as a function of the observed inflow at time step t, average inflow at the delayed time steps t–1 to t–3 with learning rate of 0.1. The best input structures for monthly prediction model are the observed inflow at time step t, average inflow at the delayed time steps t–1 to t–7, precipitation and humidity at time step t with learning rate of 0.001. Moreover, splitting the training and testing datasets using 60:40 and 80:20 ratio gave the robust performance for the daily model and monthly model, respectively. The predictive performance for the daily model reached high with R^2 of 0.8854 and NSE of 0.8619 after the validation process was completely done. However, it is found that the predictive performance was lower for the monthly model with R^2 of 0.6788 and NSE of 0.6746. Fig.1 depicts the qualitative performance of the best daily and monthly prediction models for reservoir inflow of Bhumibol Dam. It was likely similar in terms of the inflow pattern between the observed and predicted inflows during 2000–2020. The average daily predicted inflow performed by the testing dataset of prediction model was really closed to the observed average values with small percentage difference of +0.27% and -2.85% for the daily and monthly predicted inflows, respectively. However, under-estimated predictive results were found for the daily and monthly prediction models when the peak inflows were considerably investigated.

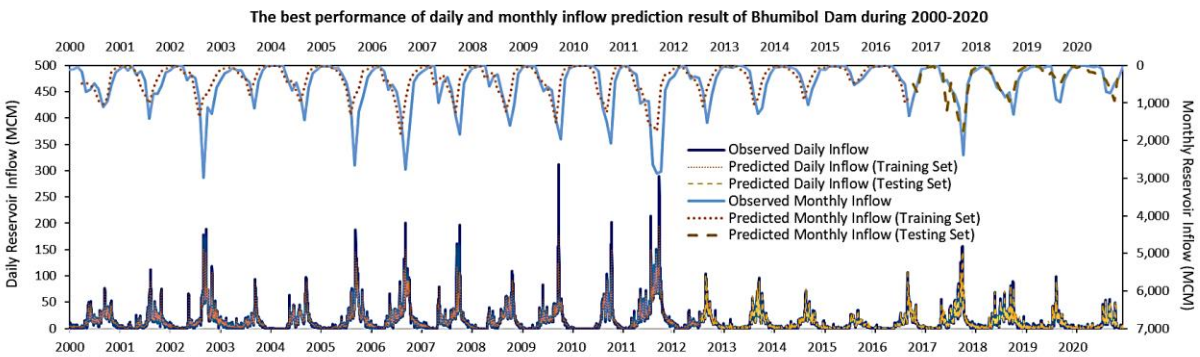

Fig.1 The qualitative comparison between observed and predicted inflows of the best daily and monthly prediction model of the Bhumibol Dam

SUMMARY AND ACKNOWLEDGMENTS

The daily and monthly reservoir inflows of the Bhumibol Dam were predicted using XGBoost algorithm. It is exhibited that the XGBoost model provided consistent and robust prediction results particularly for the daily prediction model with the greatest values of R^2, R, NSE, and the lowest values of RMSE and MSE. The XGBoost model is capable of predicting the complete performance of the daily reservoir inflow with higher accuracy than the monthly inflow.

This work was financially supported by the Agricultural Research Development Agency (Public Organization).

REFERENCES

[1] Chai, S., (2004). Artificial intelligence and social theory: a one–way street. Perspectives, 27(4):11–12.