A hybrid modelling approach for forecasting fluvial flood occurrences in the Ratnapura area, Sri Lanka

K. S. Saubhagya[1], C. D. Tilakaratne[1], G. P. Lakraj[1], M. A. Mammadov[2]

[1]*Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka*
[2]*School of Info Technology, Faculty of Science, Geelong Warun Ponds Campus, Deakin University, Australia.*

Development of an effective flood forecasting and early warning system is not straightforward due to many reasons (e.g., extreme events, missing values). Ratnapura, located in the Kalu river basin is affected by frequent flooding, resulting in fatalities and damage to property. Due to the existing shortfall in a robust flood mitigation system in the area, this study proposes a novel approach to predict flood occurrences 3 days ahead. This approach consists of 3 stages: first, impute non-random missing values in the rainfall data; second, forecast three-days-ahead rainfall with high accuracy capturing its extreme and heavy intensities; and third, produce robust flood forecasts based on river level forecasts up to 3 days in advance. The study used daily climatic and hydrological data in both numeric and image forms for the period 2014-2020. A novel optimal weight allocation method was introduced to predict missing rainfall values of neighbouring stations, combining the predictions generated by Spatial Kriging and Multilayer Perceptron Neural Network. The proposed hybrid model resulted in lower RMSE and MAE values than the individual models. The selection of inputs and their past lags for the rainfall and water level forecasting models was done using Random Forest, Granger Causality test and Cross correlation analysis. The best Bi-LSTM model which provided the lowest RMSE and MAE, and the highest accuracy in forecasting extreme events was used to obtain three-days-ahead rainfall forecasts. This model was trained with the past fourteen days values of selected input variables (e.g., relative humidity, dew point, clouds). Three-days-ahead forecasts of the river level in Ratnapura were produced using a separate Bi-LSTM model trained with ten-days-back water level of the river at identified locations and actual rainfall at the target location. Then, rainfall and river level forecasts were used as the independent variables of a Time Series Regression model to further enhance the accuracy of the river level forecasts. Numerical experiments show that the proposed hybrid model is capable of forecasting actionable flood events 1, 2 and 3 days-ahead with accuracies of 80%, 80% and 100%, respectively.

Keywords: *Flood forecasting, Kalu river, Bi-LSTM, Missing values, Feature selection*

Acknowledgements: We acknowledge financial support from the University of Colombo, Sri Lanka.