International Journal of Applied Science and Engineering

Streamflow prediction using machine learning approaches with different shared socioeconomic pathways (SSPs)

Ren Jie Chin*, **Ling Yong, Yuk Feng Huang, Lloyd Ling, Chai Hoon Koo, Ooi Kuan Tan**

Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang, Malaysia

ABSTRACT

Streamflow prediction is crucial for effective water resource management and flood prediction. Therefore, this study aims to predict streamflow within the Klang river catchment. Two machine learning approaches, namely artificial neural network (ANN) and support vector machine (SVM), were employed to forecast streamflow within the Klang river catchment. The performance of each model was evaluated using mean absolute error (MAE), root mean square error (RMSE), and percentage error. SVM outperformed ANN in streamflow prediction, achieving the lowest values of MAE, RMSE and percentage error, recorded as 7.23, 9.03 and 19.24, respectively. The model was then used to run the future scenarios, under two shared socioeconomic pathways (SSPs), which are SSP2-4.5 and SSP5-8.5, from coupled model intercomparison project phase 6 (CMIP6). SSP5-8.5 displays greater fluctuations than SSP2-4.5. This heightened variability evident in SSP5-8.5 can be attributed to its premise of rapid population expansion, significant technological advancements, and inadequate measures to address environmental issues. Consequently, these factors contribute to more frequent occurrences of extreme climate events.

Keywords: Artificial neural network, Climate model, CMIP6, Socioeconomic pathway, Streamflow prediction.

Received: December 27, 2024
Revised: March 07, 2025
Accepted: June 10, 2025

Corresponding Author:
Ren Jie Chin
chinrj@utar.edu.my

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

1.1 Background of the Study

Flooding is the most common, geographically widespread, and devastating natural disaster worldwide. In the twenty-first century, flooding is the most pressing issue due to climate change and growing urbanization. The urbanization process dramatically expands the impermeable paving area, thus increasing total runoff and flood risk (Danumah et al., 2016; Dang and Kumar, 2017). Besides, climate change will intensify the frequency and magnitude of rainfall increasing ocean temperatures can lead to increased water evaporation into the atmosphere, resulting in more moisture saturated air flowing over land and converging into storm systems. Consequently, heavy rainfall is the main factor in causing floods. Climate change, especially temperature and rainfall change, has a significant effect on flood and streamflow. Therefore, climate change projections are essential to assess the future variation in the hydrologic cycle (Teutschbein and Seibert, 2012; Chai et al., 2024). The climate model can be used to get the projection of future data like precipitation, humidity, temperature, wind speed, atmospheric pressure and others to predict the impact on streamflow patterns. Coupled model intercomparison project phase 6 (CMIP6) is one of the most common sources (Jha and Gassman, 2014; Mesgari et al., 2022).

On the other hand, using AI models has substantially improved accuracy and cost effectiveness in simulating streamflow and flooding (Ghose et al., 2022). AI models, such as artificial neural networks (ANN), support vector machines (SVM) and adaptive neuro-fuzzy inference systems (ANFIS), have demonstrated favorable outcomes in the streamflow prediction (Chin et al., 2019; Saraiva et al., 2021). These AI models require historical data to comprehend the interconnections between input factors like temperature, humidity, precipitation, wind speed and land use with streamflow. It should be trained with several percentages of data before it can reliably predict future streamflow. Hence, these AI models can manage nonlinear interaction with high accuracy. The occurrence and severity of flood events are significantly influenced by climatic changes, which are strongly related to the specific climate scenario (Arnell and Gosling, 2016; Deng et al., 2019).

1.2 Problem Statement

Despite the growing application of hydrological models and climate projections, predicting future streamflow under changing climate conditions remains a challenge. Conventional models rely on simplified assumptions and may fail to capture the non-linear relationships among meteorological parameters, land use changes, and hydrological responses (Soo et al., 2022; Chai et al., 2025). Additionally, climate change introduces uncertainties that complicate long term flood risk assessment (Deng et al., 2021; Liu et al., 2024) while, climate models like CMIP6 provide future climate projections, their direct application to streamflow prediction requires robust modelling techniques. Machine learning approaches, such as ANN and SVM, have demonstrated promising results in handling complex, nonlinear data relationships (Loh et al., 2021; Yao et al., 2021; Chin et al., 2023). However, a comprehensive assessment of their performance in streamflow prediction, especially under different climate change scenarios, is still lacking.

1.3 Research Motivation

Flooding is one of the most devastating natural disasters worldwide, with increasing frequency and intensity due to climate change and urbanization. The expansion of impervious surfaces in urban areas exacerbates flood risks by increasing surface runoff. Additionally, rising global temperatures intensify precipitation patterns, further impacting streamflow and flood occurrences. Traditional hydrological models often struggle to accurately capture the complex interactions between climate variables and streamflow patterns. However, the integration of advanced machine learning (ML) techniques with climate projections can enhance predictive accuracy (Soo et al., 2024). This study is motivated by the need to develop a reliable streamflow prediction model that can provide accurate forecasts under different climate change scenarios, thereby improving flood risk management and water resource planning.

1.4 Research Objective

This research addresses the research gap by integrating CMIP6 climate projections with machine learning models to improve streamflow prediction accuracy in the Klang river catchment. The research output is expected to provide essential data for understanding climate change, the likelihood of extreme weather events, and the adverse effects. By integrating machine learning techniques with future climate models, the study likely enhances the accuracy of streamflow predictions. The findings will aid in better flood risk mitigation strategies and sustainable water resource management.

1.5 Significance of Study

This study integrates CMIP6 climate projections with machine learning techniques to enhance the accuracy of streamflow predictions under different climate change scenarios. It evaluates the performance of various machine learning models, such as ANN and SVM, in capturing the nonlinear relationships between climate variables and streamflow. Additionally, the research investigates the impact of climate change, induced uncertainties on streamflow prediction, providing insights into model robustness and reliability. A case study is conducted on the Klang river catchment to demonstrate the practical application of the proposed methodology in flood risk assessment. By developing a data driven framework, this study supports policymakers and water resource managers in formulating effective flood mitigation and water resource planning strategies. Furthermore, the findings contribute to advancing climate resilient hydrological modelling, offering a novel approach that accounts for future climate variability and enhances urban resilience.

2. LITERATURE REVIEW

In recent decades, researchers have used various types of climate models to simulate both historical and future climate conditions. A climate model refers to a computational representation of the Earth's climate system, encompassing many components such as the atmosphere, ocean, land, and ice. Climate models are of the highest priority in comprehending the substantial effects of climate change on diverse social service aspects. They are utilized to simulate the historical and future changes in climate, encompassing both recent and distant periods, as well as aim to forecast the potential evolution of climate under several conceivable future human development scenarios and greenhouse gas emissions (Schoeman et al., 2023). The outputs generated by the regional climate models are used as direct inputs for hydrological models, enabling the simulation of the effect of climate on the water cycle at the catchment scale (Tootoonchi et al., 2023). Therefore, it has been widely used to extract meteorological data to simulate and predict streamflow. Climate models are broken into three categories, including global climate models (GCMs), regional climate models (RCMs) and earth system models (ESMs). GCMs are mathematical models that can accurately describe the physical processes occurring in the atmosphere and ocean. These models are utilized to simulate and predict the reaction of the global climate to the ongoing increase in greenhouse gas emissions (Saha and

Agrawal, 2020). RCM is a technique that combines regional characteristics into GCMs with lower resolution to predict large-scale variability (Saha and Agrawal, 2020). ESM has a broader range of components than GCM, as it includes physical, chemical and biological processes (Flato, 2011).

The coupled model intercomparison project (CMIP) is a global initiative coordinated by the working group on coupled modelling (WGCM) under the world climate research programme (WCRP). It is designed to enhance the understanding of past, present, and future climate variability and change by comparing multiple climate models from research institutions worldwide. The primary goal of CMIP is to assess the performance of global climate models by conducting standardized experiments that integrate atmospheric, oceanic, land surface, and sea ice components. This enables researchers to evaluate uncertainties in climate projections, improve model accuracy, and support climate impact studies across different regions and timescales (Ghose et al., 2022).

The latest phase of this initiative, CMIP6, represents a significant advancement over its predecessor, CMIP5, by incorporating new experimental designs, higher spatial resolution, and improved representation of dynamic climate processes. Unlike previous phases, CMIP6 adopts the shared socioeconomic pathway (SSP) framework rather than the representative concentration pathway (RCP) based emission scenarios. The SSP framework allows for a more comprehensive assessment of future climate change by integrating socioeconomic factors such as population growth, economic development, and technological advancements alongside greenhouse gas emissions. This approach facilitates a better understanding of how human activities influence climate change and provides more robust projections for policymakers and researchers (Chen et al., 2020).

Moreover, studies such as the Xin et al. (2020) and Ayugi et al. (2021) have demonstrated that CMIP6 outperforms CMIP5 in terms of model skill and predictive accuracy. The improved spatial resolution in CMIP6 models allows for more detailed simulations of regional climate patterns, extreme weather events, and ocean-atmosphere interactions. Additionally, the inclusion of enhanced physical processes, such as cloud dynamics, land-atmosphere feedback, and ocean circulation, contributes to more reliable climate projections. These advancements make CMIP6 a crucial tool for climate impact assessments, hydrological modeling, and the development of mitigation and adaptation strategies in response to global climate change.

In addition, machine learning techniques have become prevalent across various fields (Karim et al., 2017; Karim et al., 2019; Wang et al., 2022), with particular emphasis on the hydrological domain (Liu et al., 2023; Loh et al., 2024). Akbarian et al. (2023) applied multiple linear regression (MLR), eXtreme gradient boosting, (XGBoost), ANN, support vector regression (SVR) and random forest (RF) to forecast the streamflow using precipitation, runoff and temperature in Iran. Meanwhile, Achite et al. (2023) used the rainfall and runoff as the input while developing the ANFIS model for drought prediction in the Wadi Mina Basin. On the other hand, SVM is also one of the common models in soil moisture estimation. Ahmad et al. (2010), through integrating precipitation and normalized difference vegetation index, developed a SVM model to predict the soil moisture in the Colorado River Basin.

3. METHODOLOGY

3.1 Study Area

The focus of this study is centered on the Klang river catchment. The geographical location of the Klang river catchment includes Selangor and Kuala Lumpur. It has a length of 120 km and drains a basin of about 1288 km^2 (Kandari et al., 2018). Klang river catchment has 11 major tributaries, including Batu river, Gombak river, Kerayong river, Keruh river, Damansara river, Penchala river, Kuyoh river, and Ampang river. Eventually, it flows into the Straits of Malacca. Moreover, Klang Gates Dam and Batu Dam are two prominent dams located upstream of Klang river catchment, which supply water to the residents of Klang valley and serve as a means of flood control. The average amount of rain that falls in the area each year is between 1900 and 2600 mm.

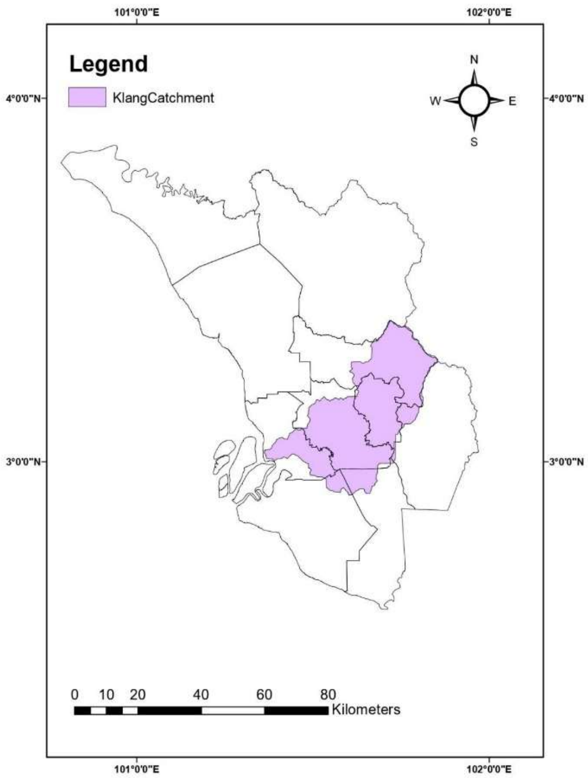


Fig. 1. Location of Klang river catchment in Selangor

Fig. 1 shows the location of the Klang river catchment at Selangor. Urbanization is taking place because of the rapid increase in population in Klang valley, and this is done to avert a lack of dwellings. Because of this, the riverbed of the Klang river is becoming shallower as a result of the fact that it already contains more than 50% urbanization (Hong and Hong, 2016). Hence, a flood occurs frequently downstream of the Klang river catchment, which is the city of Klang. Consequently, accurate streamflow predictions at the Klang river catchment are paramount in mitigating flood risks and effectively managing water distribution.

3.2 Data Acquisition

Historical rainfall data and streamflow data from 2011 to 2020 were acquired from the department of irrigation and drainage (DID), Malaysia based on the rainfall and streamflow stations, as shown in Fig. 2.

Weather parameters included in this study were air temperature, wind speed and humidity. Historical data such as air temperature, wind speed, and humidity were retrieved from the NASA Giovanni portal. The NASA Giovanni portal is a web-based platform that offers users access to an extensive array of Earth science data. It provides tools for visualizing, analyzing, and exploring diverse environmental parameters and phenomena. Users can access satellite, model, and observational data from NASA and other Earth-observing satellites and instruments through a user-friendly interface provided by the portal.

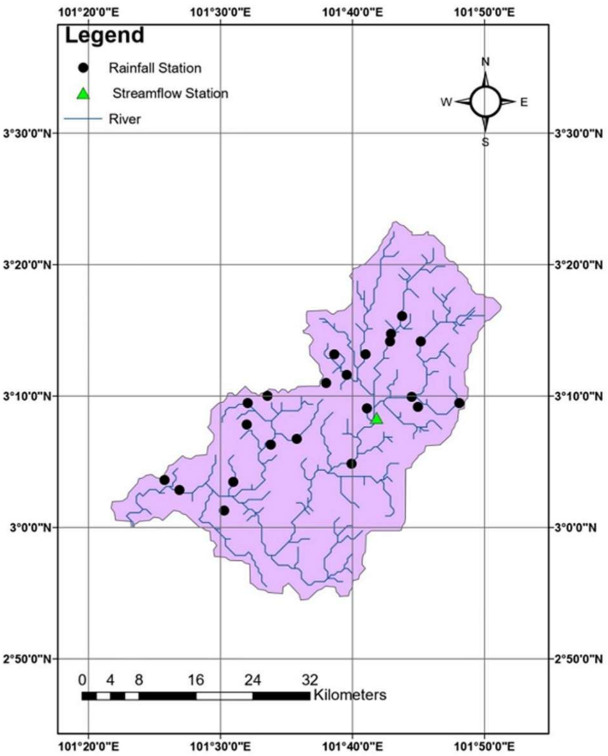


Fig. 2. Location of rainfall and streamflow stations

Apart from that, future data were extracted from the CMIP6 climate projections model at the copernicus climate change service portal. Additionally, CMIP6 is considering different scenarios of greenhouse gas emissions to forecast potential alterations in the Earth's climate in the coming years. Future weather parameters data were obtained under two scenarios, which are SSP2-4.5 and SSP5-8.5. SSP2-4.5 demonstrates the intermediate segment within the range of potential future forcing pathways. It has a moderate degree of global development and economic progress, characterized by certain advancements in living conditions and the elimination of poverty. Besides, the radiative forcing value of 4.5 W/m^2 signifies the radiative imbalance resulting from human activities, mainly the emission of greenhouse gases into the Earth's atmosphere. This particular level indicates a situation wherein the concentrations of greenhouse gases result in a rise in the average world temperatures by around 2.6 to 3.2 degrees. SSP5-8.5 assumes a hypothetical global setting whereby economic and population expansion persistently depend on fossil fuels, leading to a substantial increase in energy requirements and the consequent release of greenhouse gas emissions. Therefore, it results in significantly increased concentrations of greenhouse gases with the radiative forcing level of 8.5 W/m^2. This level refers to a potential situation where greenhouse gas emissions persistently escalate without substantial attempts to mitigate them, resulting in a considerable rise in world average temperatures of more than 4°C.

3.3 Data Pre-processing

Effective data pre-processing is essential to ensure accurate and meaningful results from machine learning models. Climate data from meteorological stations may contain missing values due to instrument malfunctions, power outages, or human errors during data collection. To maintain data quality, it is crucial to verify the completeness of the dataset before analysis. Therefore, rain gauge and streamflow data from DID were carefully screened. Rain gauge stations with more than 10% missing data during the study period were excluded from the analysis. However, if the missing data was less than 10% of the total dataset, the arithmetic mean method Equation (1) was used for imputation (Chai et al., 2024). This method is suitable when the rain gauges are evenly distributed, and individual measurements show minimal deviation from the average. Given the random nature of the missing data and its weak correlation with other variables, the arithmetic mean approach was deemed appropriate. In this process, missing values were replaced with the mean precipitation measurements from surrounding stations. The workflow for data screening and imputation is illustrated in Fig. 3.

$$P_t = \frac{\sum_{i=1}^{n} r_i}{n} \tag{1}$$

where P_t is the estimated value of the missing rainfall at the target station, n is the number of stations, i is the index of data and r_i is the observed rainfall neighboring station.

https://doi.org/10.6703/IJASE.202506_22(2).003

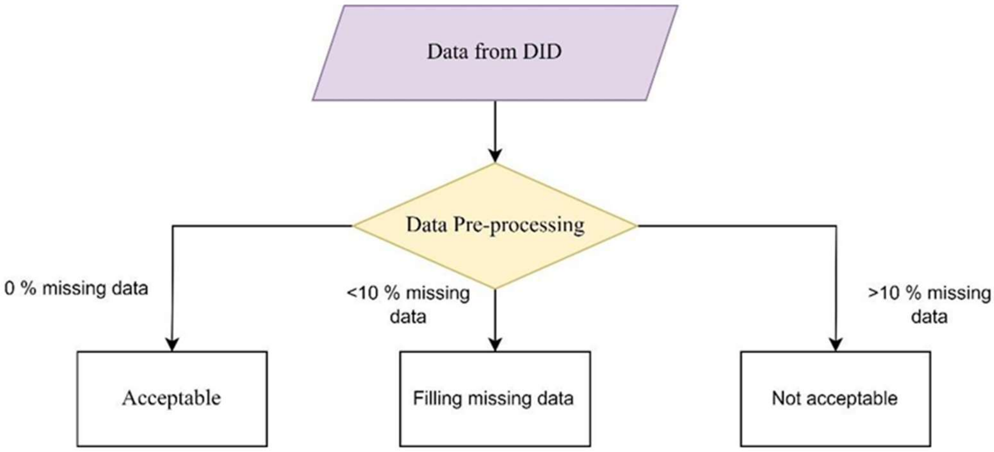


Fig. 3. Data screening and filling procedure

3.4 Model Development

The streamflow prediction model was developed using rainfall, air temperature, wind speed and humidity as input, while the streamflow acts as the output. The collected data was then divided into different training-to-testing ratio of 80:20, 70:30 and 60:40 for machine learning based model development. Two machine learning approaches were applied in this study, which are ANN and SVM.

ANN models are biologically inspired, and the creation of these models began with an interest in comprehending how the brain completes tasks. It consists of three different layers namely, input layer, hidden layer and output layer which are built with nodes. Feed forward neural network (FFNN) model is the most popular ANN (Adnan et al., 2017). The signal is transmitted sequentially via each layer, and if the output layer does not achieve the desired value, the error is propagated in the opposite direction. Based on the error, the model modifies the weights and thresholds in order to decrease the discrepancy between the predicted and observed values (Li et al., 2019). The fundamental aspect of training a neural network involves using backpropagation, a technique that adjusts the weights of the neural network based on the error rate observed in the previous iteration. The loss function is minimized based on the gradient descent method.

SVM is a feedforward networking method, much like ANN. Instead of creating a weight vector like the ANN, statistical learning with an SVM aims to measure the difference between a given target function f(x) and the output produced by the machine. SVM is frequently used for problems involving classification and regression. In this study, SVM for regression which is known as support vector regression (SVR) was used. SVR employs a kernel function to transform the input data into a higher-dimensional space, facilitating linear regression analysis. The choice of the kernel function is an important hyperparameter in SVR models that must be determined prior to their execution. The available kernel functions include the radial basis function (RBF), linear, polynomial, and sigmoid. In our study, RBF was chosen because it has high optimization efficiency and adaptability. The RBF kernel helps in capturing non-linear relationships between input features and target values. This kernel enables SVR to effectively model intricate patterns in the data that may not be linearly separable.

3.5 Performance Evaluation

Machine learning models were evaluated and compared using MAE, RMSE and percentage error, to identify the best-performing model. MAE quantifies the average absolute difference between predicted and actual values in a dataset. It treats all errors equally, regardless of magnitude. A lower MAE value signifies greater model accuracy.

$$\text{MAE} = \frac{\sum|y_i - x_i|}{n} \tag{2}$$

Meanwhile, RMSE evaluates the discrepancy between predicted and actual values, giving greater emphasis to larger errors by squaring the differences before averaging. A lower RMSE value indicates higher accuracy, as it brings the predicted values closer to the actual ones.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - x_i)^2} \tag{3}$$

Besides, percentage error quantifies the accuracy of a measurement, estimation, or prediction relative to the actual or expected value. An optimal percentage error of zero signifies perfect accuracy, while lower values indicate greater precision in predictions or estimations.

$$\text{Percentage error} = \frac{|True\ value - Predic\ \ \ \ value|}{True\ value} \times 100\% \tag{4}$$

where n is the number of data pairs, x is the observed variable and y is the predicted variable.

4. RESULTS AND DISCUSSION

The evaluation of training-to-testing data combinations was conducted and showcased through metrics such as RMSE, MAE, and percentage error. Lower error values indicate higher accuracy of the model. Table 1 shows the model comparison results of ANN and SVM with different combinations of training-to-testing ratios, which are 60:04, 70:30, and 80:20. It was found that Model IV yielded the most favorable results with the lowest MAE, RMSE, and percentage error, which are 7.23%, 9.03%, and 19.24% respectively. Based on the study by Nguyen et al. (2021), a training and testing dataset ratio of 70:30 was considered the most optimal for training and validating the models. Therefore, this finding aligns with the research of other scholars. The superiority of SVM over ANN in this case is consistent with studies such as Otchere et al. (2018), which showed that SVM often outperforms ANN in hydrological prediction tasks due to its capacity to handle non-linear relationships with fewer computational requirements.

The best-performing model (Model IV) was then used to forecast the streamflow under SSP2-4.5 and SSP5-8.5 scenarios. Future annual streamflow for SSP2-4.5 and SSP5-8.5 are displayed in Fig. 4.

Upon analyzing the overall trend, it is evident that SSP5-8.5 exhibits higher fluctuations compared to SSP2-4.5 over ten years. This trend aligns with global projections from the IPCC (2023), which predict increased climate extremes under SSP5-8.5 due to rapid economic growth, high fossil fuel dependency, and limited climate policies. The rapid expansion of the population could accelerate deforestation or urbanization rates, driven by the demand for increased housing construction. These factors lead to high greenhouse gas emissions and more severe climate change impacts. As a consequence of the extreme impacts of climate change, such as increased frequency and intensity of heat waves, more severe storms, and increased sea level rise, streamflow fluctuates dramatically. On the other hand, SSP2-4.5 exhibits a more consistent streamflow than SSP5-8.5 because it assumes a more sustainable and environmentally responsible path.

Moreover, the inconsistency of streamflow overtime may be due to the occurrence of extreme climate events, for example, El Niño, La Niña, and Indian Ocean Dipole (IOD), which can alter the global precipitation intensity and frequency, air temperature, and atmospheric circulation patterns. According to Generoso et al. (2020), both El Niño and La Niña cause global changes in temperature and rainfall approximately every two to seven years. Variations in temperature and rainfall patterns can impact streamflow, causing fluctuations and inconsistency over time.

In the case of SSP5-8.5, the peak streamflow occurred in 2023 at 537.35 m^3/s, while the lowest streamflow was recorded in 2021 at 517.64 m^3/s. The streamflow exhibits a vast increase from 2021 to 2023, followed by a steady decline leading up to 2026. Subsequently, there is a gradual rise until a second peak is observed in 2028, reaching a reading of 536.75 m^3/s, after which it declines once more.

In contrast with SSP5-8.5, SSP2-4.5 has the highest streamflow in the year 2021, with a value of 533.41 m^3/s. However, in the year 2027, the streamflow hit a low point measured at 525.60 m^3/s. The streamflow values are expected to decrease from the peak in 2021 to 2023, followed by a steady increase until 2025, before exhibiting a decreasing trend in 2027. Despite that, the streamflow from 2028 to 2030 shows relatively stable values. This pathway is associated with sustainable development and controlled emissions, echoing the findings of Khadka et al. (2023), who observed that less aggressive scenarios tend to produce less volatile hydrological behavior.

Upon analyzing the overall trend, it is evident that SSP5-8.5 exhibits higher fluctuations compared to SSP2-4.5 over ten years. The increased variability observed in SSP5-8.5 can be ascribed to its assumption of rapid population growth, substantial technological progress and limited efforts to mitigate environmental concerns. The rapid expansion of the population could accelerate deforestation or urbanization rates, driven by the demand for increased housing construction. These factors lead to high greenhouse gas emissions and more severe climate change impacts. As a consequence of the extreme impacts of climate change, such as increased frequency and intensity of heat waves, more severe storms, and increased sea level rise, streamflow fluctuates dramatically. On the other hand, SSP2-4.5 exhibits a more consistent streamflow than SSP5-8.5 because it assumes a more sustainable and environmentally responsible path.

Fig. 5 illustrates the monthly streamflow trends under the SSP2-4.5 scenario. The graph reveals a gradual increase in streamflow from January to May, followed by a steady decline until August, before experiencing a sharp rise towards the end of the year. The peak streamflow occurs in December, while the lowest point is observed in January.

Table 1. Transfer the sensor log data into training data

Model	Type	Training to testing ratio	MAE	RMSE	Percentage error (%)
I	ANN	60:40	15.18	18.39	39.25
II	SVM	60:40	8.40	10.41	19.55
III	ANN	70:30	18.00	20.98	48.59
IV	SVM	70:30	7.23	9.03	19.24
V	ANN	80:20	8.94	11.07	24.63
VI	SVM	80:20	8.51	10.17	23.44

Chin et al., International Journal of Applied Science and Engineering, 22(2), 2024431

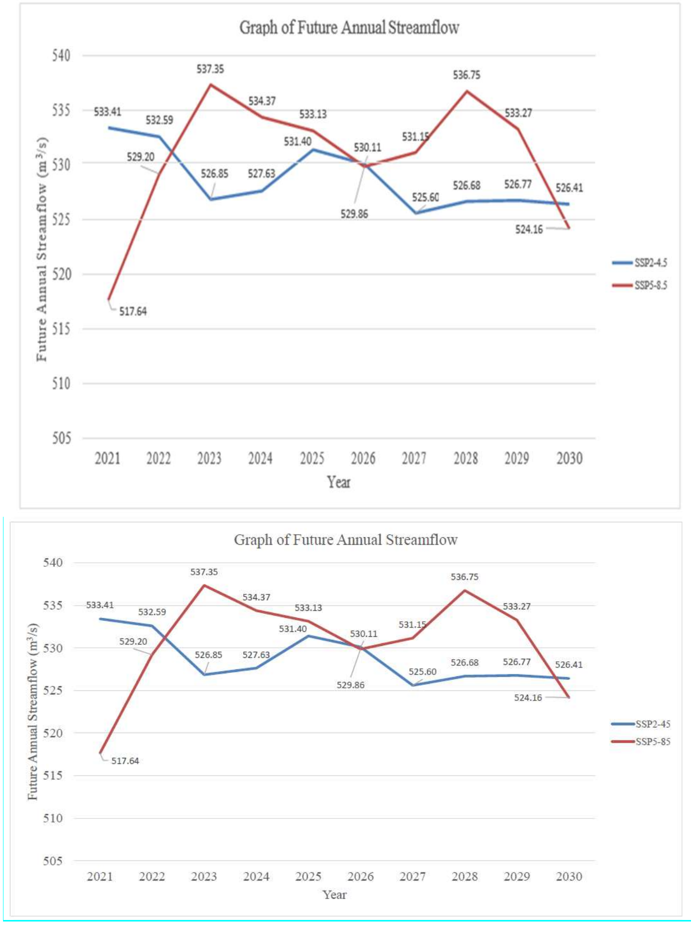
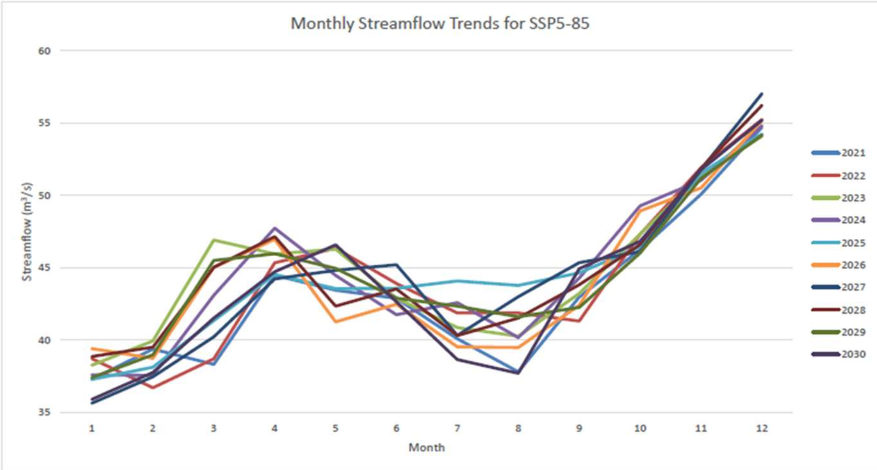


Fig. 4. Future annual streamflow under scenarios of SSP2-4.5 and SS5-8.5

This pattern aligns with Peninsular Malaysia's seasonal climate variations. From January to May, the Northeast Monsoon brings heavy rainfall, particularly to the eastern regions, including Selangor, leading to increased runoff within the Klang river catchment. However, as the monsoon subsides, precipitation levels decrease, temperatures rise, and evapotranspiration intensifies, resulting in a decline in streamflow between May and August. Towards the year's end, streamflow surges again with the onset of the rainy season. Under the SSP2-4.5 scenario, the highest predicted streamflow within the Klang River catchment is 55.04 m³/s in December 2024, while the lowest is 37.04 m³/s in January 2024. Additionally, a notable dip occurred in August 2029, with streamflow recorded at 40.83 m³/s. This decline could be attributed to extreme weather events, such as prolonged dry spells leading to reduced rainfall within the study area.

On the other hand, Fig. 6 presents the monthly streamflow trends under the SSP5-8.5 scenario, which follows a comparable pattern to SSP2-4.5. Streamflow rises from January to May, declines until August, and then increases again. However, fluctuations under SSP5-8.5 are more pronounced, with the highest predicted streamflow reaching 57.04 m³/s in December 2027 and the lowest recorded at 35.63 m³/s in January 2027. The heightened variability in SSP5-8.5 can be attributed to the scenario's assumption of more extreme climate change impacts, including intensified droughts, floods, and other hydrological extremes. Additionally, SSP5-8.5 accounts for increased deforestation and urbanization due to rapid population growth, further altering the hydrological cycle and contributing to greater fluctuations in streamflow.

The findings are in agreement with Tan et al. (2021), who demonstrated that high-emission scenarios lead to both higher peaks and lower troughs in streamflow due to intensified hydrological extremes. Moreover, the seasonal streamflow trends observed are also reflected in the work of Tiwari and Adamowski (2013), who emphasized the influence of local monsoon systems on streamflow behavior in Southeast Asia.



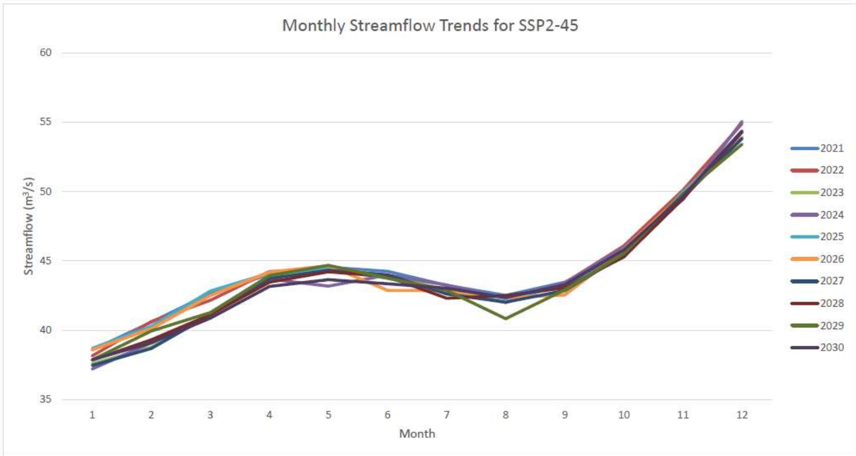
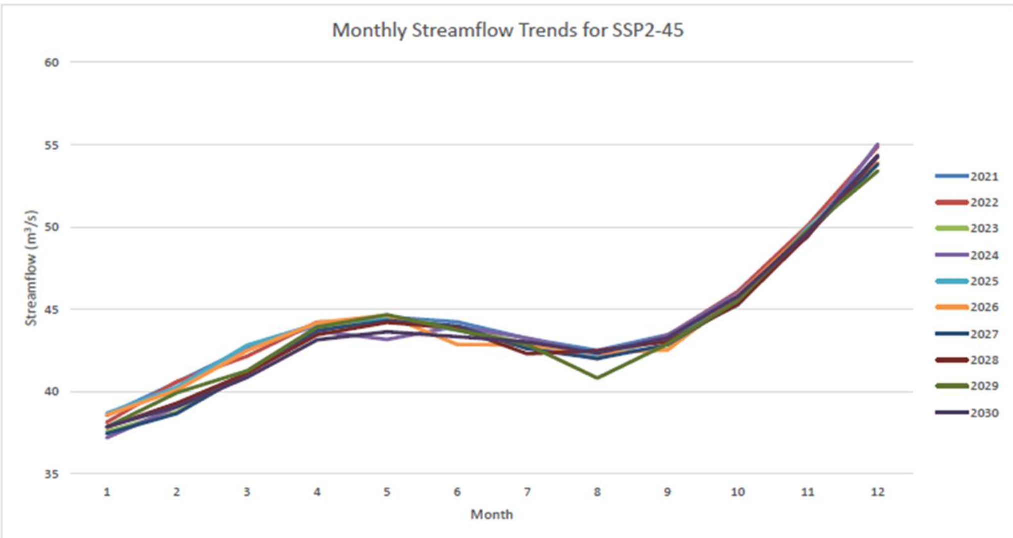


Fig. 5. Future monthly streamflow under scenarios of SSP2-4.5



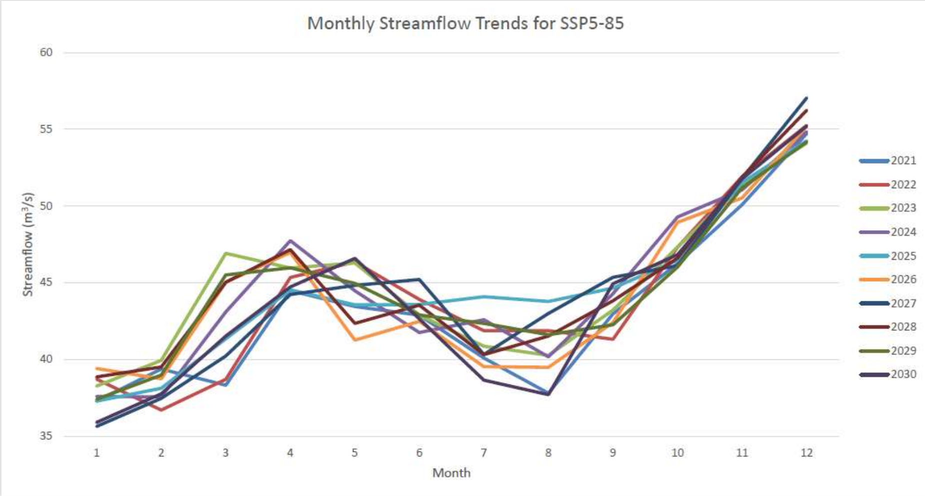


Fig. 6. Future monthly streamflow under scenarios of SS5-8.5

5. CONCLUSION

This study aims to develop a streamflow prediction model using various machine learning techniques, including SVM and ANN, with different training-to-testing ratios such as 60:40, 70:30, and 80:20. The model incorporates weather parameters such as air temperature, relative humidity, wind speed, and rainfall. Subsequently, performance evaluation of the models was conducted through statistical analyses involving MAE, RMSE, and percentage error. Among six models, SVM with a training-to-testing ratio of 70:30 yielded the best performance with the lowest value of MAE, RMSE, and percentage error recorded as 7.23%, 9.03%, and 19.23%, respectively. In the climate scenario assumption of SSP5-8.5, streamflow fluctuations are higher compared to SSP2-4.5. The maximum streamflow in SSP5-85 reaches 537.35 m^3/s in 2023, whereas in SSP2-4.5, it peaks at 533.41 m^3/s in 2021. Conversely, the lowest streamflow in SSP5-8.5 is observed in 2021, with a value of 517.64 m^3/s, while in SSP2-4.5, it occurs in 2027 at 525.60 m^3/s. Future studies should focus on enhancing machine learning models for streamflow prediction in the Klang river catchment by incorporating several improvements. Extending training datasets with long-term climate trends can improve model accuracy. Developing hybrid models that integrate machine learning with hydrological models, such as SWAT, can further refine predictions. Including additional weather parameters like atmospheric pressure and solar radiation may enhance model performance. Advanced techniques, such as deep learning models like convolutional neural networks (CNNs), should be explored to capture complex hydrological patterns. Expanding the range of CMIP6 climate scenarios, including SSP1, SSP3, and SSP4, will provide a more comprehensive assessment of climate impacts. Lastly, evaluating spatial and temporal variations in streamflow at different scales and time intervals will improve the reliability of predictions for effective water resource management.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme project (FRGS/1/2023/WAB02/UTAR/02/1) and was partly supported by Malaysia Toray Science Foundation (4417/0005). The authors wish to thank the Department of Irrigation and Drainage Malaysia for the data provided and Universiti Tunku Abdul Rahman for the facilities provided for the study.

REFERENCES

Achite, M., Gul, E., Elshaboury, N., Jehanzaib, M., Mohammadi, B., Mehr, A.D. 2023. An improved adaptive neuro-fuzzy inference system for hydrological drought prediction in Algeria. Physics and Chemistry of the Earth, 131, 103451.

Adnan, R.M., Yuan, X., Kisi, O., Yuan, Y. 2017. Streamflow forecasting using artificial neural network and support vector machine models. American Scientific Research Journal for Engineering, Technology, and Sciences, 29, 286–294.

Ahmad, S., Kalra, A., Stephen, H. 2010. Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33, 69–80.

Akbarian, M., Saghafian, B., Golian, S. 2023. Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. Journal of Hydrology, 620, 129480.

Arnell, N.W., Gosling, S.N. 2016. The impacts of climate change on river flood risk at the global scale. Climatic Change, 134, 387–401.

Ayugi, B., Zhihong, J., Zhu, H., Ngoma, H., Babaousmail, H., Rizwan, K., Dike, V. 2021. Comparison of CMIP6 and CMIP5 models in simulating mean and extreme precipitation over East Africa. International Journal of Climatology, 41, 6474–6496.

Chai, V.H., Chin, R.J., Ling, L., Huang, Y.F., Soo, E.Z.X. 2024. Regional and seasonal assessment of biases on high-resolution satellite precipitation estimations in peninsular Malaysia: 2011–2020. Journal of Hydroinformatics, 26, 1715–1734.

Chai, V.H., Chin, R.J., Ling, L., Huang, Y.F., Soo, E.Z.X. 2025. Performance evaluation and error decomposition of CMORPH satellite precipitation estimation for Klang Valley, Malaysia. 5th International Symposium on Civil and Environmental Engineering, Langkawi, Malaysia, 012050.

Chen, H., Sun, J., Lin, W., Xu, H. 2020. Comparison of CMIP6 and CMIP5 models in simulating climate extremes. Science Bulletin, 65, 1415–1418.

Chin, R.J., Lai, S.H., Loh, W.S., Ling, L., Soo, E.Z.X., Huang, Y.F., Yeo, Y.Q. 2023. Soft computing method for settling velocity prediction of fine sediment in retention structure. Proceedings of the 3rd International Civil Engineering and Architecture Conference, Kyoto, Japan, 687–693.

Chin, R.J., Lai, S.H., Ibrahim, S., Wan Jaafar, W.Z., Elshafie, A.H.K.A. 2019. New approach to mimic rheological actual shear rate under wall slips condition. Engineering with Computers, 35, 1409–1418.

Dang, A.T.N., Kumar, L. 2017. Application of remote sensing and GIS-based hydrological modelling for flood risk analysis: A case study of District 8, Ho Chi Minh City, Vietnam. Geomatics, Natural Hazards and Risk, 8, 1792–1811.

Danumah, J.H., Odai, S.N., Saley, B.M., Szarzynski, J., Thiel, M., Kwaku, A., Kouame, F.K., Akpa, L.Y. 2016. Flood risk assessment and mapping in Abidjan District using multi-criteria analysis (AHP) model and geoinformation techniques, (Cote D'ivoire). Geoenvironmental Disasters, 3, 10.

Deng, B., Lai, S.H., Jiang, C., Kumar, P., El-Shafie, A., Chin, R.J. 2021. Advanced water level prediction for a large-scale, river-lake system using hybrid soft computing approach: a case study in Dongting Lake, China. Earth Science Informatics, 14, 1987–2001.

Deng, B., Chin, R.J., Tang, Y., Jiang, C., Lai, S.H. 2019. New Approach to predict the motion characteristics of single bubbles in still water. Applied Sciences, 9, 3981.

Flato, G.M. 2011. Earth system models: An overview. Wiley Interdisciplinary Reviews Climate Change, 2, 783–800.

Generoso, R., Couharde, C., Damette, O., Mohaddes, K. 2020. The growth effects of El Niño and La Niña: Local weather conditions matter. Annals of Economics and Statistics, 140, 83–126.

Ghose, D.K., Tanaya, K., Sahoo, A., Kumar, U. 2022. Performance evaluation of hybrid ANFIS model for flood prediction. 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), 772–777.

Hong, J.L., Hong, K. 2016. Flood forecasting for Klang river at Kuala Lumpur using artificial neural networks. International Journal of Hybrid Information Technology, 9, 39–60.

IPCC, 2021. "Climate Change 2021-The Physical Science Basis". Cambridge University Press. Cambridge, United Kingdom.

Jha, M.K., Gassman, P.W. 2014. Changes in hydrology and streamflow as predicted by a modelling experiment forced with climate models. Hydrological Processes, 28, 2772–2781.

Kandari, P.N.A., Saudi, A.S.M., Chyang, P.J., Kamarudin, M.K.A., Saad, M.H.M., Azid, A., Saudi, N.S.M., Mahmud, M. 2018. Flood risk pattern recognition analysis in Klang river Basin. International Journal of Engineering and Technology, 7, 86–90.

Karim, S., Zhang, Y., Asif, M.R. 2017. Image processing based proposed drone for detecting and controlling street crimes. Proceeding of IEEE 17th International Conference on Communication Technology (ICCT), Chengdu, China, 1725–1730.

Karim, S., Zhang, Y., Yin, S., Brohi, A.A. 2019. Impact of compressed and down-scaled training images on vehicle detection in remote sensing imagery. Multimedia Tools and Applications, 78, 32565–32583.

Khadka, D., Babel, M.S., Kamalamma, A.G. 2023. Assessing the impact of climate and land-use changes on the hydrologic cycle using the SWAT Model in the Mun River Basin in Northeast Thailand. Water, 15, 3672.

Li, F.F., Wang, Z.Y., Qiu, J. 2019. Long-term streamflow forecasting using artificial neural network based on preprocessing technique. Journal of Forecasting, 38, 192–206.

Liu, Y., Wang, C., Jiang, C., Chin, R.J., Xiang, Z., Long, Y., Deng, B., Wu, Z. 2024. Water storage capacity estimation for a large complex lake system incorporating the water levels during flooding season. Journal of Hydrology: Regional Studies, 51, 101634.

Liu, Y., Yang, Y., Chin, R.J., Wang, C., Wang, C. 2023. Long short-term memory (LSTM) based model for flood forecasting in Xiangjiang river. KSCE Journal of Engineering, 27, 5030–5040.

Loh, W.S., Ling, L., Chin, R.J., Lai, S.H., Loo, K.K., Seah, C.S. 2024. A comparative analysis of missing data imputation techniques on sedimentation data. Ain Shams Engineering Journal, 15, 102717.

Loh, W.S., Chin, R.J., Ling, L., Lai, S.H., Soo, E.Z.X. 2021. Application of machine learning model for the prediction of settling velocity of fine sediments. Mathematics, 9, 3141.

Mesgari, E., Hosseini, S.A., Hemmesy, M.S., Houshyar, M., Partoo, L.G. 2022. Assessment of CMIP6 models' performances and projection of precipitation based on SSP scenarios over the Menap region. Journal of Water and Climate Change, 13, 3607–3619.

Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Le, H.V., Tran, V.Q., Prakash, I., Pham, B.T. 2021. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Mathematical Problems in Engineering, 2021, 1–15.

Otchere, D.A., Ganat, T.O.A., Gholami, R., Ridha, S. 2021. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. Journal of Petroleum Science and Engineering, 200, 108182.

Saha, A.K., Agrawal, S. 2020. Mapping and assessment of flood risk in Prayagraj District, India: A GIS and remote sensing study. Nanotechnology for Environmental Engineering, 5, 11.

Saraiva, S.V., Carvalho, F.D.O., Santos, C.A.G., Barreto, L.C., Freire, P.K.D.M.M. 2021. Daily streamflow forecasting in Sobradinho reservoir using machine learning models coupled with wavelet transform and bootstrapping. Applied Soft Computing, 102, 107081.

Schoeman, D.S., Gupta, A.S., Harrison, C.S., Everett, J.D., Brito-Morales, I., Hannah, L., Bopp, L., Roehrdanz, P.R., Richardson, A.J. 2023. Demystifying global climate models for use in the life sciences. Trends in Ecology and Evolution, 38, 843–858.

Soo, E.Z.X., Chin, R.J., Ling, L., Huang, Y.F., Lee, J.L., Lee, F.W. 2024. Streamflow simulation and forecasting using remote sensing and machine learning techniques. Ain Shams Engineering Journal, 15, 103099.

Soo, E.Z.X., Wan Jaafar, W.Z., Chin, R.J., Ling, L., Ng, C.Y., Prashant, S. 2022. Streamflow evaluation using IHACRES model in Kelantan river basin, Malaysia. 2nd International Conference on Civil and Environmental Engineering (ICCEE 2022), Kajang, Malaysia, 04008.

Tan, M.L., Liang, J., Samat, N., Chan, N.W., Haywood, J.M., Hodges, K. 2021. Hydrological extremes and responses to climate change in the Kelantan River Basin, Malaysia, based on the CMIP6 HighResMIP experiments. Water, 13, 1472.

Teutschbein, C., Seibert, J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology, 456, 12–29.

Tiwari, M.K., Adamowski, J. 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resources Research, 49, 6486–6507.

Tootoonchi, F., Todorović, A., Grabs, T., Teutschbein, C. 2023. Uni- and multivariate bias adjustment of climate model simulations in Nordic catchments: Effects on hydrological signatures relevant for water resources management in a changing climate. Journal of Hydrology, 623, 129807.

Wang, L., Yin, S., Alyami, H., Rashid, M., Almotiri, J., Alyamani, H.J., Alturise, F. 2022. A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images. Geoscience Data Journal, 11, 237–251.

Xin, X., Wu, T., Zhang, J., Yao, J., Fang, Y. 2020. Comparison of CMIP6 and CMIP5 simulations of precipitation in China and the East Asian summer monsoon. International Journal of Climatology, 40, 6423–6440.

Yao, Y., Yang, X., Lai, S.H., Chin, R.J. 2021. Predicting tsunami-like solitary wave run-up over fringing reefs using the multi-layer perceptron neural network. Natural Hazards, 107, 601–616.