Borneo Journal of Sciences & Technology, 4(1): 44-50
DOI: http://doi.org/10.3570/bjost.2022.4.1-07
e-ISSN: 2672-7439

Submitted: 16th November 2021 Accepted: 20th December 2021 Published: 31st January 2022

A Statistical Prediction Model of River Flood

Siew-Ping Yiiong and Nancy Bundan

School of Computing and Creative Media (SCM), University of Technology Sarawak, 96000 Sibu, Sarawak, Malaysia

Abstract: Floods are the most common and widespread among the weather-related natural hazards. Floods can affect thousands of people each year and have the potential to cause catastrophic loss of life and property. Due to the severe weather threat, these days, a lot of efforts have been put into developing flood monitoring and warning systems to minimize the impacts of flooding. This paper presented a statistical model for the prediction of the river flood using fast calculations and provides real-time results with the purpose of saving lives and reducing the damages caused by floods in the studied area. An easy and speed efficient regression method, robust linear regression, had been applied in the prediction model as it can provide reliable real-time results with low resource utilization. In addition, quadratic fit function had also been applied to approximate the rising water level, in which it can determine if the water level may exceed the flood line in the near future. A regression equation which represented the relationship between the rainfall and water level and a high R-squared value were respectively obtained. It was also found that the developed prediction model based on the robust linear regression was efficient and accurate, and the prediction of the water levels in the near future were quite in line with the observed values.

Keywords: Flood prediction, river flood, robust fit function, linear regression, quadratic fit function.

INTRODUCTION

Nowadays, disasters from nature are a global phenomenon and various communities cooperate in addressing these natural disasters. One of the most common forms of natural disasters is floods. It is the most common natural disaster experienced in Malaysia, for example, flash flood, seasonal flood, river flood, coastal flood, and others. Floods always cause a huge impact on people, the country's economy, and also the environment. The number of damage can be extensive, such as poor harvests. Hence, flood loss prevention and mitigation such as structural or non-structural flood protection measures play a vital role in addressing the flood problems. The structural flood prediction measures like building dams are always effective in addressing the mentioned problem, however, it is not available in most areas especially the areas with high population with limited space for the construction of dams. Therefore, non-structural measures, such as the application of the flood prediction model, are truly important in reducing the damage caused by flood as it can provide an early warning of the disaster and then extend the period for the implementation of the evacuation plan. Generally, flood prediction models can be developed using both numerical models and statistical models. Numerical models, usually, could be categorized into conceptual models and empirical models, in which empirical model is always adopted by the researchers in comparison with conceptual model due to its simplicity and easier to develop [1]. On the other hand, a statistical model always represents the mathematical relationship between one or various variables, in which numerous types of statistical methods could be considered such as regression, clustering, logistic regression and etc. [2]. In this study, a statistical model on the river flood at Bawang Assan area (located in Sibu) was developed and the data was collected through the flood monitoring devices, such as rain sensors and ultrasonic sensors. The motivation for this study originates from the frequent occurrences of the river flood which always causes inconveniences to the community at the mentioned area. River flood generally occurs over days or months as it occurs in a large basin that includes major tributaries and is generally the result of numerous rains over many days. There are many factors that could cause a river flood and torrential rains are the most common cause of river flooding. The incoming precipitation is diverted to the main arm of the river and the water level of the river gradually rises. Large-scale river floods always cause great loss of property and human life. Therefore, an accurate prediction model that can monitor the behaviour of the river is indeed desired in order to

Corresponding Author: Siew-Ping Yiiong. University of Technology Sarawak (UTS), email: siew.ping@uts.edu.my

predict the possible river flood that will affect the community in that area, and thus proper action could be taken by the community and the damages caused by river floods could then be reduced.

LITERATURE REVIEW

Basically, a river flood occurs when a river fills with water beyond its capacity and then the water overflows the river bank and runs into the low-lying lands which are very near to the river. The most common factors that contribute to river floods are heavy rainfall and high tides. As a result of the occurrences of river floods, the affected communities always suffered from the enormous destruction of property. Fortunately, these losses could be reduced by appropriate structural control measures and non-structural measures. In fact, flood forecasting is one of the flood controls measures that can be used to reduce losses. In general, there are two types of considerations that are included in flood forecasting, namely, meteorological and hydrological. Meteorological considerations concern with the phenomena of the atmosphere, including rainfall, snowmelt, temperature and others. Meanwhile, hydrological considerations, include the water levels in rivers, river discharge, groundwater level and others [3]. These two considerations can be used as the input parameters for flood prediction model.

These days, a substantial amount of statistical and numerical works had been carried out in flood predicting. [4] established a research work to develop a technology for flash flood warning systems based on a predictive model, multiple regression analysis. The regression equation was developed while training data from seven trials was inserted into the regression model. The data parameters collected were water level and water velocity. Meanwhile, [5] also used the same predictive model in developing an intelligent disaster forecasting system to monitor flood risk in the flood-prone area. In the forecasting model proposed by [5], knowledge parameters of water level, water pressure and precipitation were collected instead. Besides that, multiple simple regression analysis was used by [6] to urge the equation within the development of flood prediction. There were two methods for analysing water levels, inclusion the residual and the non-inclusion residual, and this was also used in the predictive models of [4] and [5]. According to [6], it was found that the residual non-inclusion method had a more robust performance in predicting water levels if compared to the residual inclusion method. Therefore, the non-inclusion residuals methodology was believed to be more suitable for the flood forecasting model. Meanwhile, [7] used a linear regression model to predict river floods. This model was used to predict the increase in runoff water and the water level of the river. It was found that the linear regression model was more suitable for the primary approximation for large river discharge predictions. However, this model was no longer suitable for predicting more than two days, as precipitation forecasts must be included retrospectively. Therefore, another model was needed to continue predicting the flood of the river. In addition, two types of regression methods had been applied in the work of [8] which were linear and symbolic regressions; linear regression method was applied for investigating the relationship between a dependent variable and various explanatory variables, and the least squares method was generally used to fit the linear regression data for the purpose of reducing the sum of squared residuals. Symbolic regression, on the other hand, was applied in order to perform the mathematical expressions in symbolic form that best fit the regression data. In the work of [8], it was noticed that a linear model was not feasible in the flood scenario especially if the water level was high. Therefore, the symbolic regression model was suggested to be applied in order to achieve the high precision of the river flood prediction. Furthermore, in order to improve the accuracy of the flood prediction, an easy flow predicting strategy namely base difference model was applied in the work of [9]. The base difference model was found to be less complicated in comparison with both linear regression model and multiple linear regression model as the calculation of the regression coefficient was not required in the base difference model. In an earlier work of [10], a predictive model that using robust linear regression with multiple variables had been proposed. In this predictive model, the amount of data parameters can be added or removed if necessary. In other words, the model was independent of the amount of data parameters since it can be adjusted as necessary; thus, the predictive model of [10] was more preferable if compared with the those of [4] and [5] due to the flexible feature of the model Several functions had also been applied in the model of [10], such as a quadratic water level fit function, a robust fit function, and a time multiplier function for different purposes. It is noteworthy that the time multiplier function was applied with the aim to decide the time interval between two consecutive readings that had been taken; here, it played a role as a worst-case predictor, in which the fastest or slowest time of the next reading to be taken will be informed. And, the other two functions in the work of [10] were used to approximate the trend of the rising water levels and also to improve the flood forecast. Apart from the statistical works mentioned earlier, substantial numerical works had also been done on flood monitoring and prediction. Further discussion on the numerical works can be found in the works of [11 - 14], to name a few.

METHODOLOGY

In this study, the details of the flood monitoring and response system were not shown here and it could be found in the work of [15]. Hence, only the algorithm of the prediction model was presented in this section. For the sake of simplicity, a threshold level of the river water, which acted as a warning line, was first chosen; if the predicted value was more than that level, an alert message was triggered and the community at the mentioned area was instantly notified. Two parameters, which were water level, L, and rainfall, R, had been taken into account in this study, and the work of [10] was partially adopted here. With the intention of predicting the possible river flood, the data of the parameters L and R were collected. A robust linear regression model was then computed based on the past data of the parameters L and R, in which the obtained regression model was in the form of $y = a + b_1X_1$. The present values of the rainfall were then inputted into the regression model so that the corresponding values of the water level, L_2, were predicted. The predicted water level, L_2, was then compared with the previous water level, L_1. If a decreasing pattern of the water level was noticed, $L_2 < L_1$, it indicated that no flood was anticipated and the previous procedures were repeated by inputting the present value of the water level, L_2, into the regression model. Conversely, the quadratic fit function was applied to approximate the water level trend versus time which only involved the present water level value to the first local minimum reached moving back in time and then a prediction model was generated for it. After that, the future time values were inputted into the prediction model in progressive order for predicting the future values of water level. If the predicted water level exceeds the flood line, river flood was anticipated, and vice versa. The procedures of the prediction model can be clearly illustrated with a flowchart as shown in Figure 1.

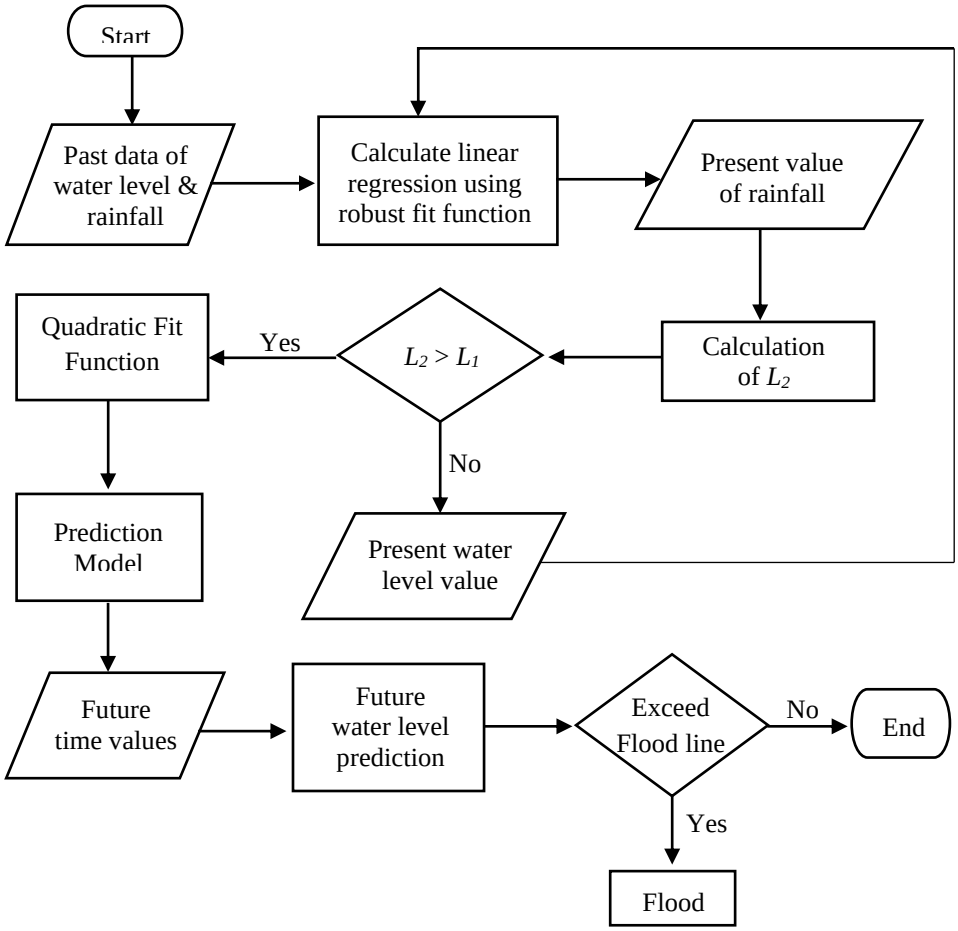

Figure 1: Flowchart of main procedures River Flood Prediction Model

Also, it is noteworthy that the robust fit function instead of ordinary least squares fit function is applied in the regression model because of the intention to dampen the influence of the outliers and to provide a

better fit of data. This is due to the large influence on outliers by the ordinary least square fit function as squaring the residuals can magnify the effects of the extreme data points. The main equations of the robust fit function are shown as following:

$$r_{adjusted} = \frac{r_i}{\sqrt{1-h_i}}, \quad (1)$$

where $r_{adjusted}$ = adjusted residuals, r_i = usual least-square residuals, h_i = weights that change the residuals by down-weighting the influence of outliers that have a large weight on the least-square fit. Besides, the equation of the standardization adjusted residuals is:

$$u = \frac{r_{adjusted}}{K_s}, \quad (2)$$

where u = standardized adjusted residuals, K = tuning constant with the value 4.685, s = robust variance which given by $\frac{MAD}{0.6745}$, in which *MAD* is the median absolute deviation of the residuals. Besides that, the quadratic fit function, in which a second order polynomial function, is applied to approximate the increased water level in order to predict the future values of water level. The equation of the second order polynomial function is shown as:

$$y = a_1x^2 + a_2x + a_3 \quad (3)$$

where x = time index that starts at the moment which water level starts to increase, y = the corresponding water level at that time, and $[a_1, a_2, a_3]^T$ is the coefficient matrix. The second order polynomial function is chosen as the simplicity of calculation over larger polynomial can be ensured. Several values (x_i, y_i) are tabulated in the form of $Y_{n\times1} = X_{n\times3}A_{3\times1}$, where X represents the Vandermonde matrix. The elements of the Vandermonde matrix are given by $X_{i,j} = x_i^{3-j}$, says, i = 1, 2, 3 and j = 1, 2, 3; the element of each row are x_i^2, x_i^1, and 1 respectively. In the form of $Y = XA$, Y and X are known; the coefficient matrix $[a_1, a_2, a_3]^T$ can then be obtained by solving for A using $X^TY = X^TXA$.

RESULTS AND DISCUSSION

The focus of this study was to develop a statistical prediction model on flood forecasting at the mentioned area based on the past data of the rainfall and water level, thus, the statistical and simulation results of the developed model were presented in this section. In this study, a data set of 24 hours rainfall and water level values at Bawang Assan area had been collected. As shown in Table 1, usually, the rainfall intensity can be divided into few categories, which were very heavy rain, heavy rain, moderate rain, light rain, and no rain [16]. To predict the possible river flood at the mentioned area, a linear regression model of 12 hours rainfall and water level data was trained by applying the robust fit and quadratic fit methods; the corresponding 12 hours rainfall and water level that acted as training data were also presented in Table 2. Figure 2 depicted a fitted robust linear regression model which represented the relationship between the rainfall and water level. The equation obtained from the regression model was $y = 1.0269x + 148.75$, where x was the rainfall values and y was the water level values. Besides that, the coefficient of determination, R-squared value, of the obtained regression model was also calculated, in which the obtained value was 0.93. Generally, the higher the R-squared values the smaller the differences between the observed data and the fitted data. In other words, the larger the R-squared, the better the regression model fits the observed data. Obviously, existence of a significant relationship between the rainfall and water level in the obtained regression model was observed as the calculated R-squared value was relatively high. Even though with a high R-squared value, sometimes, the regression model can also be a biased model. Therefore, it was important to assess the residuals plot. Figure 3 below illustrated the normal probability plot of residuals. It was clear that the normal probability plot of the residuals was approximately linear and this indicated that the error terms were normally distributed as well. Hence, this could justify the obtained regression model was indeed appropriate for the data set used in this study.

Table 1: Indicator of Rainfall Intensity [16]

Category	Intensity (mm per hour)
Very Heavy Rain	> 60 mm
Heavy Rain	31 mm – 60 mm
Moderate Rain	11 mm – 30 mm
Light Rain	0.5 mm – 10 mm
No Rain	0 mm

Table 2: Training Data

Time (hours)	Rainfall Intensity (mm/hour)	Water Level (cm)
1	27	172
2	35	185
3	31	184
4	52	197
5	70	214
6	93	238
7	82	233
8	89	238
9	72	230
10	74	232
11	69	230
12	66	214

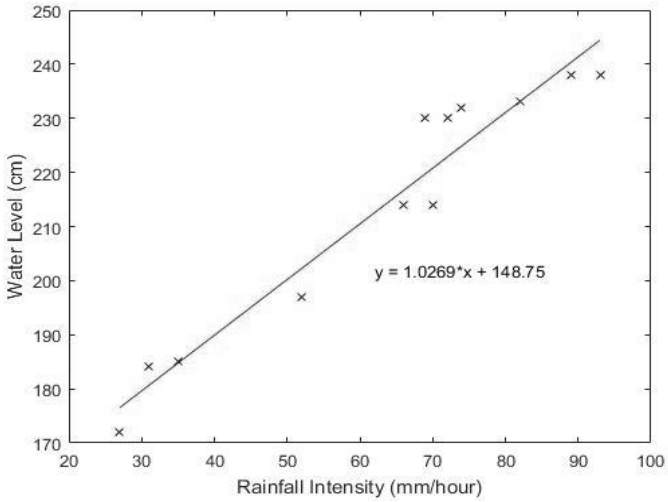

Figure 2: Fitted Linear Regression Model (—: fitted line; ×: actual data)

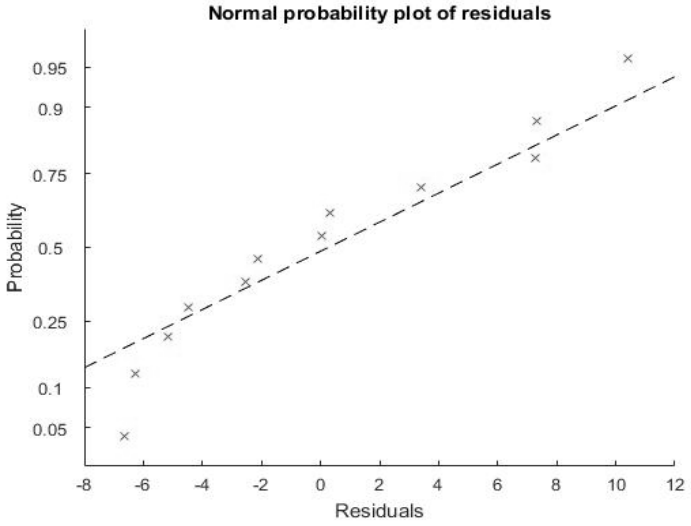

Figure 3: Normal Probability Plot of Residuals

In spite of the well-fitted regression model as mentioned above, there was still lacking information on the prediction of river floods in near future. Fortunately, this can be addressed by inputting the present rainfall values into the obtained regression model to simulate the corresponding water levels. The present rainfall values collected were shown in Table 3 and the corresponding simulation result was depicted in Figure 4. Here, an accurate threshold level for water level was set, says, the height of 250 cm; if the predicted water level is higher than the threshold value, a possible river flood is expected and a warning message is notified, and vice versa. As seen from Figure 4, a decreasing pattern of the predicted water levels after 12 hours had been observed except at the time of 16 hour. Also, a warning message of the possible river flood was not made for this case as the predicted values of water level had never cross the flood line. Overall, the predicted values of the water level agreed well with those actual values; this revealed that the accuracy of the developed prediction model in this study was truly good.

Table 3: Present Values of Rainfall

Time (hours)	Rainfall Intensity (mm/hour)
13	54
14	58
15	59
16	75
17	68
18	70
19	73
20	71
21	64
22	53
23	52
24	40

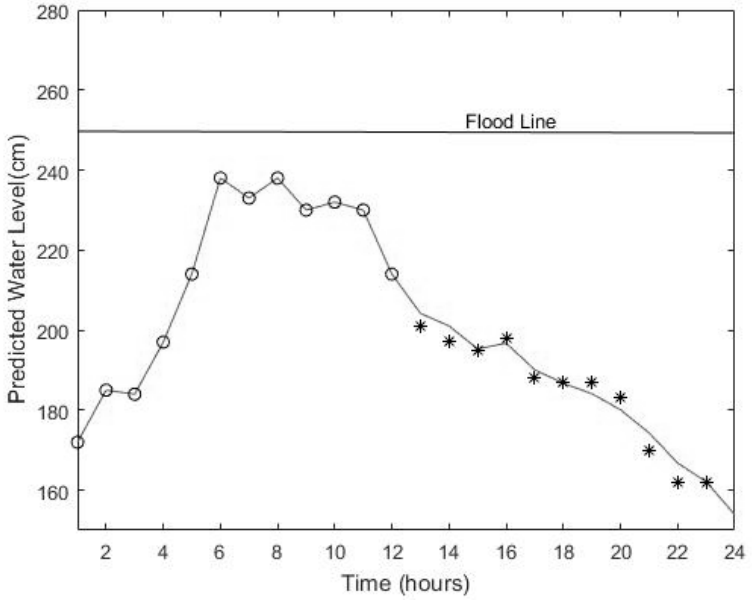

Figure 4: Comparison of the predicted and actual values of present water levels
(—: predicted results; o: training data; ∗: actual data)

CONCLUSION

In this study, a statistical flood prediction model based on the robust linear regression method had been successfully developed, in which a second-order polynomial function was also employed to approximate the rising water. The obtained results showed that the water level was strongly correlated with the rainfall intensity and the statistical model had demonstrated very precise results in predicting the possible river flood in the near future. It was anticipated that the accurate data collected was the main reason for the high precision results obtained in this study. Also, the robust fit function applied in the regression model was expected to have a high contribution to the obtained precise results as well, since negligible weights could be assigned to the outliers by robust fit function if there is any, and thus the influence of the outliers could be dampened. Apparently, the statistical prediction model developed in this study in fact is a preliminary model as it only considered two parameters, which were rainfall and water level. Thus, the future work would be to take into account of other factors such as water velocity, river discharge, water logging and etc. in order to have a more realistic prediction model on river floods.

ACKNOWLEDGEMENT

The research was funded from University of Technology Sarawak (UTS) under the Research Grant (UCTS/RESEARCH/2/2018/05). Our heartfelt gratitude extends towards the representative from the Bawang Assan longhouses community, Mr. Marcathy anak Gindau and the people of Rumah Jimbun for their full cooperation.

REFERENCES

[1] Meeyaem, K., & Polpinit, P. 2014. Mathematical Model for Flood Forecasting of the Chi River Basin. International Conference on Intelligent Agriculture, 63(2), 5-9.

[2] Smith, M. A. 2015. Output from Statistical Predictive Models as Input to eLearning Dashboards. Future Internet, 7, 170-183.

[3] Manual on Flood Forecasting and Warning, 2011. Geneva: World Meteorological Organization. Retrieved October 12, 2019, from https://library.wmo.int/doc_num.php?explnum_id=4090

[4] de Castro, J. T., Salistre Jr, G. M., Byun, Y. C., and Gerardo, B. D. 2013. Flash Flood Prediction Model based on Multiple Regression Analysis for Decision Support System. Proceedings of the World Congress on Engineering and Computer Science, 2, 23-25.

[5] Orozco, M., and Caballero, J. 2018. Smart Disaster Prediction Application using Flood Risk Analytics towards Sustainable Climate Action. MATEC Web of Conferences, 189, 10006.

[6] Noor, M. S. F. M., Sidek, L. M., Basri, H., Husni, M. M. M., Jaafar, A. S., Kamaluddin, M. H., Majid, W. H. A. W. A., Mohammad, A. H., and Osman, S. 2016. Development of Flood Forecasting using Statistical Method in Four River Basins in Terengganu, Malaysia. IOP Conference Series: Earth and Environmental Science, 32(1), 012023.

[7] Shahzad, K. M., and Plate, E. J. 2014. Flood Forecasting for River Mekong with Data-Based Models. Water Resources Research, 50(9), 7115-7133.

[8] Bolshakov, V. 2013. Regression-Based Daugava River Flood Forecasting and Monitoring. Information Technology and Management Science, 16(1), 137-142.

[9] Veiga, V. B., Hassan, Q. K., and He, J. 2014. Development of Flow Forecasting Models in the Bow River at Calgary, Alberta, Canada. Water, 7(1), 99-115.

[10] Seal, V., Raha, A., Maity, S., Mitra, S. K., Mukherjee, A., and Naskar, M. K. 2012. A Simple Flood Forecasting Scheme using Wireless Sensor Networks. International Journal of Ad hoc, Sensor and Ubiquitous Computing, 3(1), 45-60.

[11] Ivanova, O., and Ivanov, M. 2016. 1-D Mathematical Modelling of Flood Wave Propagation. Chemical Engineering Transactions, 53, 139-144.

[12] Jun, C. L., Mohamed, Z. S., Peik, A. L., Razali, S. F., and Sharil, S. 2016. Flood Forecasting Model using Empirical Method for a Small Catchment Area. Journal of Engineering Science and Technology, 11(5), 666 – 672.

[13] Azad, W. H., Sidek, L. M., Basri, H., Fai, C. M., Saidin, S., and Hassan, A. J. 2017. 2 Dimensional Hydrodynamic Flood Routing Analysis on Flood Forecasting Modelling for Kelantan River Basin. MATEC Web of Conferences, 87, 01016.

[14] Siregar, R. I. 2018. Numerical Analysis of Flood Modeling of Upper Citarum River under Extreme Flood Condition. IOP Conference Series: Materials Science and Engineering, 308, 012023.

[15] Tsai, K. F. 2019. Solar-Powered Arduino Flood Detection System. Final Year Project. University College of Technology Sarawak.

[16] Official Website of Department of Irrigation and Drainage Sarawak, n.d. Welcome to Official Website of Department of Irrigation and Drainage Sarawak. Retrieved August 17, 2019, from https://did.sarawak.gov.my