Application of Stochastic Flood Forecasting Model Using Regression Method for Kelantan Catchment

Sazali Osman[1][*], Norizan Abdul Aziz[1], Nurul Husaif[1], Lariyah Mohd Sidek[2], Aminah Shakirah[3], Faezah Hanum[3], and Hidayah Basri[2]

[1]Department of Irrigation and Drainage Malaysia (DID), Kuala Lumpur, Malaysia
[2]Universiti Tenaga Nasional (UNITEN), Kajang, Selangor, Malaysia
[3]ZHL Engineers Sdn. Bhd, Putrajaya, Malaysia

Abstract. Flood is without doubt the most devastating natural disasters, striking numerous regions in Malaysia each year. During the last decades, the trend in flood damages has been growing exponentially. This is a consequence of the increasing frequency of heavy rain, changes in upstream land-use and a continuously increasing concentration of population and assets in flood prone areas. Malaysia, periodically, have faced with huge floods since previous years. Kelantan River basin, which located in the Northeast of Peninsular Malaysia, is prone to flood events in Malaysia. Kelantan River is the principal cause of flooding because it is constricted at its lower reaches. The capacity of the river at the downstream coastal area is less than 10,000 m^3/s, therefore flood that exceeds this capacity will overspill the banks and discharge overland to the sea. Realizing the seriousness of the problems, it is vital in providing in time useful information for making crucial decisions especially to provide warning for any potential flood occurrence. In this study, stochastic flood forecasting model using stage regression method was applied to Kelantan River basin, in which the regression coefficients and equations was derived from the least square principle. The stochastic model were calibrated and validated which then shows that the equations derived are suitable to predict the hydrograph in Kelantan River basin. In conclusion, establishing a flood forecasting system would enhance the effectiveness of all other mitigation measures by providing time for appropriate actions. This has increased the importance of flood modelling for flood forecasts to issue advance warning in severe storm situations to reduce loss of lives and property damage.

1 Introduction

Flooding is a chronic natural hazard with potentially devastating consequences, giving rise to a third of all losses due to natural events. Extreme weather events over the last decade have fuelled the perception that, whether due to anthropogenic global warming or otherwise, flooding is becoming more extreme, more widespread and more frequent (WMO, 2011). The expectations for flood forecasts in terms of magnitude and timing have grown with the recognition of the importance of flood warning as a contribution to flood management. This means that past methods of simple extrapolation of forecasts from gauged sites may no longer suffice (Moore et al., 2001). While "the heart of any flow forecasting system is a hydrological model" (Serban and Askew, 1991), catchment modelling is just one of the crucial elements on which the

---

[*] Corresponding author: Aminahshakirah91@gmail.com

effectiveness and efficiency of an integrated flood forecasting and warning system (FFWS) depends.

Recently, data-driven models have been extensively used in stream flow forecasting. The models range from straight-forward empirical models, such as regression models, to soft computing models using neural and fuzzy logic techniques. Regression analysis performs and exploration of relationships between one dependent and one more independent variables under consideration. Regression methods are widely used for many hydrological applications. The regression coefficients were estimated with the help of mass balance equations. In this study, Kelantan River basin was selected as the study area for the implementation of stage regression method for flood forecasting.

In this analysis, stochastic method is adopted to predict future outcomes based on available hydrologic data. This model works as a black box interpreting inputs and predicting outputs based on historic data. Several statistical techniques have been used in stochastic models using ARMA (Autoregressive Moving Average), ARIMA (Autoregressive Integrated Moving Average), Multivariante modelling or disaggregation (Burlando P., 1993). The stochastic methods in hydrology emerged during the 1960s and 1970s and have been dominating hydrologic studies ever since and since then they have been used to evaluate the correlation between rainfall and river water level (Sivakumar B., 2010). The statistical problem in question is to prove a relationship of two variables in a time series. The correlation of these time series is a correlation in time with an existing lag, which indicates the amount of time between the first variable to change until the effect causes in the second variable is significant. There are two variables of data that the dataset contains such as rainfall information and river water level. The two pieces of information have to belong to the same river basin, with the rainfall information being taken in the same catchment, retrieving relevant rainfall information to the river water level. The stations should contain sufficient historical data as the nature of the data is seasonal from more than five years is needed. The study explains the lag between the two variables, how long it takes rain to have an effect on the river water level and the measure of this effect over time. Roland and Stuckey (2008) developed regression equations for estimating flood flows at selected recurrence intervals for ungauged catchments at Pennsylvania. These equations were developed utilizing peak flow data from 322 stream flow gauging stations in the development of the equations. The regression equations can be used to predict the magnitude of flood flows for specified recurrence intervals for most streams. However, they are not valid for streams with drainage areas generally greater than 2000 square miles or with substantial regulation, diversion, or mining activity within the basin. Study by Supriya (2015) described the relation between weighted Annual Maximum Daily Rainfall, catchment area and Annual Maximum Daily Stream Flow utilized in framing regression equation for flood forecasting. The accuracy of the results were then calculated by on performance analysis calculation. This research aims to develop a simple flood forecasting model using statistical regression technique and thus generate the regression equation for Kelantan River Basin.

2 Study Area

Kelantan cover an area of 15,099 km^2 and is located on the north eastern region of Peninsular Malaysia, bordered by parts of southern Thailand to the north, Perak to the west, Pahang to the south and Terengganu to the south-east. Hilly terrains are found on the southern parts of the State, separated by the Titiwangsa Mountain Range, with fertile coastal plains downstream defining the geography of the region. Kelantan River

is the main river to channel off water from the inundated area of Kelantan Basin to the South China Sea during wet season which is caused by the northeast monsoon. The major rivers include Sungai Nenggiri, Sungai Lebir, Sungai Galas, Sungai Pergau, Sungai Kelantan, Sungai Golok, Sungai Kemasin, Sungai Pengkalan Chepa, Sungai Pengkalan Datu and Sungai Semerak (Figure 2). The river originates in the southern rugged and steep region of the state where the elevation ranges between 1,000 m to 2,000 m LSD. Meandering through the hilly areas in the upper catchment, Sungai Nenggiri at the south-west flows in a north-easterly direction to join Sungai Galas at Bertam. From there, Sungai Galas flows north to capture Sungai Pergau, which flows south easterly from Jeli, at Dabong. From Dabong, Sungai Galas flows in northeast direction and meet Sungai Lebir, which flows in northwest direction from Gunung Gagau, at Kuala Krai. From Kuala Krai, the river is called Sungai Kelantan and flows towards north to the river mouth. The river length from Kuala Krai to the river mouth is approximately 100 km. The river mouth is situated about 15 km north Kota Bharu. The topography of Kelantan River basin is less rugged towards the main drainage lines in the central part of the basin, where most of the land is below an elevation of 75m and consists of low hills. The main high ground areas situated at western, southern and eastern part of the river catchment. These upland areas generally range from 1000mLSD to 1500mLSD in elevation with some peaks reaching more than 2000mLSD. The mountainous areas are covered with virgin jungle while rubber, oil palm and some paddy are planted in the undulating terrains and lowlands.
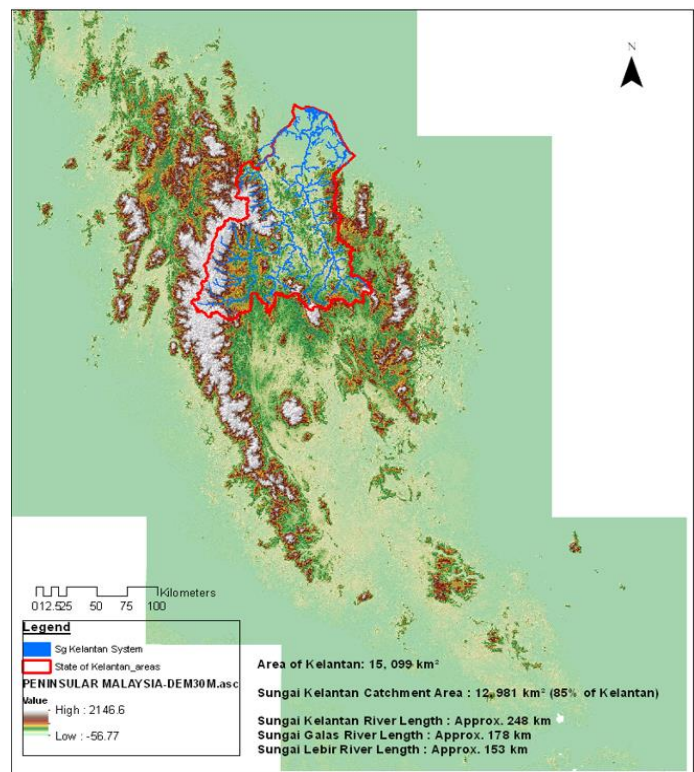

Fig. 1. Location of Sungai Kelantan in Malaysian Topographical DEM Plan.
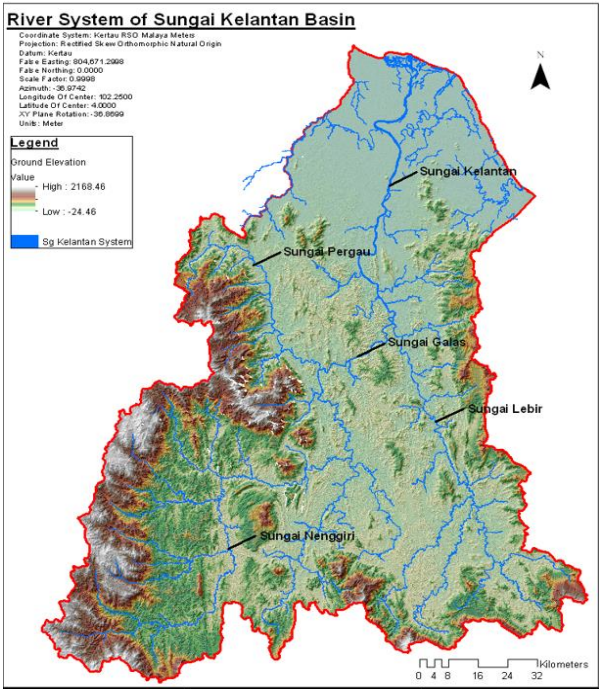
Fig. 2. Main River Systems in Kelantan River Basin.

The river valley is dominated by agricultural activities. The current landuse for state of Kelantan where the built-up area and cleared land constitute only about 18%, agricultural use at about 38.5% and forest land with more than 42%, with other use constitute the rest (Figure 3). Vegetable farming has recently been introduced at Lojing Highlands in the District of Gua Musang. The area has created much controversy due to farming activities on highlands causing large amounts of silt and sediment runoff into the receiving tributaries of Sungai Nenggiri and Sungai Kelantan.
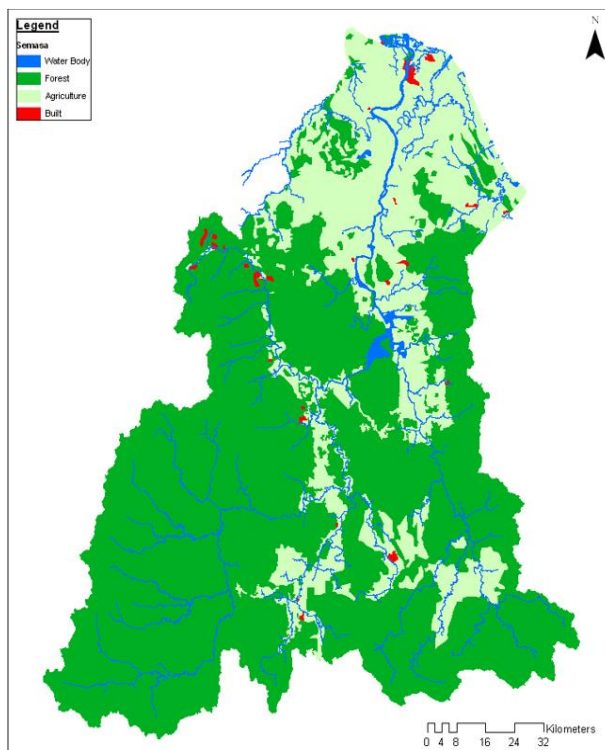
Fig. 3. Map of Kelantan Land use.

The characteristic features of the climate of Malaysia are uniform temperature, high humidity and copious rainfall. Winds are generally light. The climate of Kelantan river basin is mainly governed by the regime of Northeast and Southwest monsoon winds. Though the wind over the country is generally light and variable, there are, however, some uniform periodic changes in the wind flow patterns. Based on these changes, four seasons can be distinguished, namely, the southwest monsoon, northeast monsoon and two shorter periods of inter-monsoon seasons. The southwest monsoon season is usually established in the half of May or early June and ends in September. The prevailing wind flow is generally south-westerly and light, below 15 knots. The northeast monsoon season usually commences in early November and ends in March. The State of Kelantan comprises of about 20% flat and generally low-lying coastal plains and is subjected to frequency severe flooding. The flooding which normally occurs during the northeast monsoons is both widespread and frequent. Serious flooding problems have to certain extent dampened full development of Kelantan's total socio-economic potentials.

Historical data recorded by JPS in year 1990 stated that there was a big flooding occurred on month of November and December. The average rainfall recorded for 6 days period starting from 23rd November 28th November 1990 were 203.2mm (4 stations) and 1159mm (5 stations) for coastal area and rural area respectively. The water level at Kuala Krai reached its maximum level 26.16m which is above the danger water level on 13th December 1990. While highest water level recorded at Jambatan Guillemard was 16.72m which is above its danger level 16.00m. Villages affected in Kuala Krai were flooded for at least 2 days with flood depth ranging from 0.3m to 1.4m (DID Flood Report, 1990).

The flooding recorded in 2001, started in the middle of December 2001 until 25th December 2001. The flooding occurred during this period was severe and caused massive damage. The amount of rainfall received during this time is heavier compared to the past events. Kelantan state has recorded a total of 3,064.17mm annual rainfall which is equal to 113.49% until 31st December 2001. Water level at Kuala Krai recorded at 26.65m above danger level (25.00m) on 24th December 2001. The flooding affected almost all district in Kelantan which is Gua Musang, Kuala Krai, Machang, Tanah Merah, Kota Bharu, Pasir Mas dan Tumpat (DID Flood Report, 2001). Flooding in year 2004 occurred when Kelantan main rivers documented a high water levels. The flooding involved a flat terrain areas and areas which are close to the riverbank. Sungai Kelantan's water level at Tambatan Di Raja recorded highest water level 7.00m which is 2m above the danger level on 13th December 2004. In Tanah Merah, heavy rainfall and back water from Sungai Kelantan caused overflow at Sungai Kusial and flooding happened at nearest area (Cekok Ipoh and Kg. Kulim) with 4m flood depth (DID Flood Report, 2004).

Flood event from 14 December 2014 to 25 December 2014 is one of the most severe floods over recorded at Kelantan state. The loss of the productivity from flooding is particularly significant to Kelantan, where for an example a 50-year return period flood that occurred in 1967 as well as flood 2014 has inundated more than 50,000 hectare of land and affecting almost 330,000 people. The damages of flood 2014 event were then estimated to be more than RM 50 Million with 13 deaths (DID Flood Report, 2014). Looking at all those previous rainfall records, Kelantan River is known for the potential to have severe flood frequently. Therefore, study on flood events and attributed research are required as baseline in Kelantan river such as flood early warning system because evaluation of Kelantan river based on hydrologic and flood studies offer to provide sufficient time for the authorities to evacuate the downstream communities to safer places and take necessary measures to protect physical properties in vulnerable areas.

3 Model Description

In this section, it comprises of the overall information regarding the model description. The statistical stochastic stage regression model is applied to Kelantan River basin to simulate the simulated hydrograph based on the derived regression coefficients and equations by the least square method principle. Regression analysis is used to predict the value of a dependent variable (usually called Y) based on the value at least one or more independent variable (usually called X). It is a technique in statistics to determine the relationship between two random variables. Regression gives an equation that best describes the relationship between variables. In this study, focusing on the relationship between water level stations and to test whether such a relation is statistically significant.

For the case of two variables involved, the linear model could be an option where the relation between these two variables at single event providing may be plotted as scatter diagram. Based on the scatter diagram, we could identify the relation between the variables, where the more elongated the scatter, the better relation, and in contrast, the more circular the scatter, the worse the relation. Meanwhile, Multiple Linear Regression Analysis (MLR) is used when more than one explanatory variables as this method is described to be as a procedure of fitting an equation to a set of data that might comprise of two or more variables. This method would be solved by using the least square method, a process to obtain the best estimates of the coefficients in the linear models.

These coefficients are obtained using regression analysis. The multiple linear regression method takes the following form:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k + \varepsilon$$

where

x_i : explanatory variables

k : total of explanatory variables

ε : random error for the ith observation

β_o : intercept

β_1 : slope

The Multiple Linear Regression Method could then be solved simultaneously using the regression method by transforming the multiple power equation into a multiple linear equation model based on equation as shown below:

$$y = C x_1^a x_2^b \Rightarrow Log(y) = a Log(x_1) + b Log(x_2)$$

It is then to be decided to each individual station which equation (including or not including residual random error) that best suits the stage profile of the rivers. Suitability is defined here as having the predicted value closest to the actual readings over the longest period of time of the rivers, whilst showing similar trends. This equation is then validated with the most recent flood event recorded.
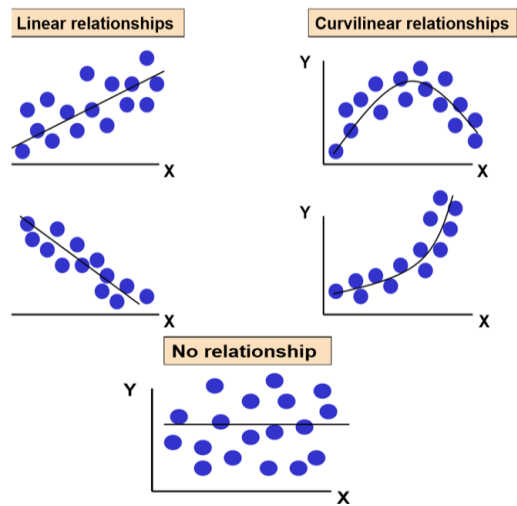

Fig. 4. Type of relationship in linear regression.

In a stage regression, the smaller the value of the standard error of the estimate, S_e, the better the predictive power of the model. A smaller Se means that all of the confidence intervals obtained from the estimated model will be narrower. S_e is an estimate of the

common standard deviation of all of the conditional distributions, namely σ which measures the scatter around the regression line. One way to improve the model's predictive power is to reduce the Se by explicitly considering additional factors as independent variables. The estimation of the regression coefficients of a multiple regression by the least square method is based on the same principle applied to a simple regression. The outputs are obtained using MS Excel's regression tool to perform the estimation.

The output is divided into three sections such as the A. Regression Statistics, which provides an overview of the model's ability to explain the variation of the dependent variable, B. the ANOVA (Analysis of Variance) gives detailed information regarding the separation of the variation of Y into the explained and unexplained components, and C. the Estimated Model presents the statistical performances of the individual independent variables.

4 Data

The implementation of statistical flood forecasting model using stage regression method is totally based on satisfactory of hydrological data for the river basin. Currently, Kelantan River basin was facilitated with approximately 15 water level stations. However, only selected water level station was assessed intensively which later calibrated and validated. Selected water level stations are Station Tualang at Lebir River, Station Dabong at Galas River, Station Kuala Krai, Station Guillemard Bridge and Station Jeti Kastam at Kelantan River.
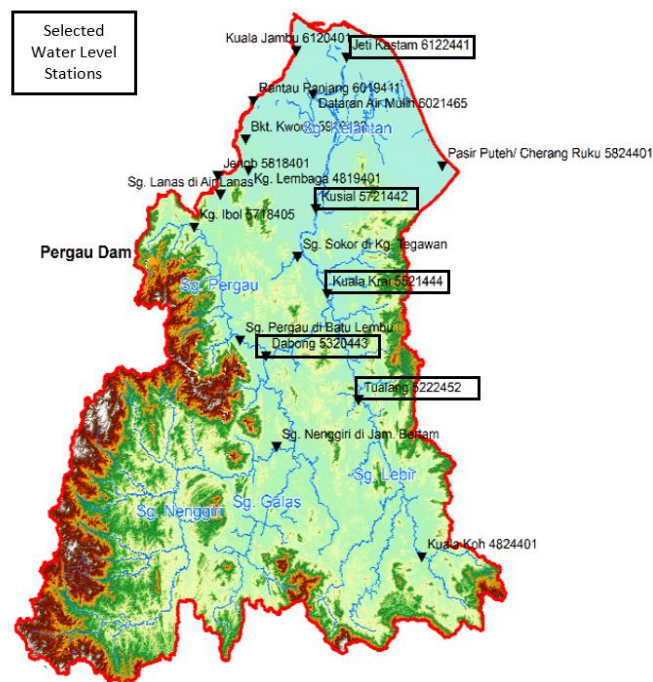

Fig. 5. Selected telemetry water level stations at Kelantan River basin for calibration and validation point.

Data employed in the stage regression model are hourly data for different flood events. The implemented flood events for calibration and validation were:

- Flood Event of December 2004, data employed –hourly data
- Flood Event of December 2007, data employed – hourly data
- Flood Event of December 2013, data employed – hourly data
- Flood Event of December 2014, data employed – hourly data

5 Model Setup

The flowchart illustrated in Figure 6 describes the methodology for the implementation of simple stage regression method.
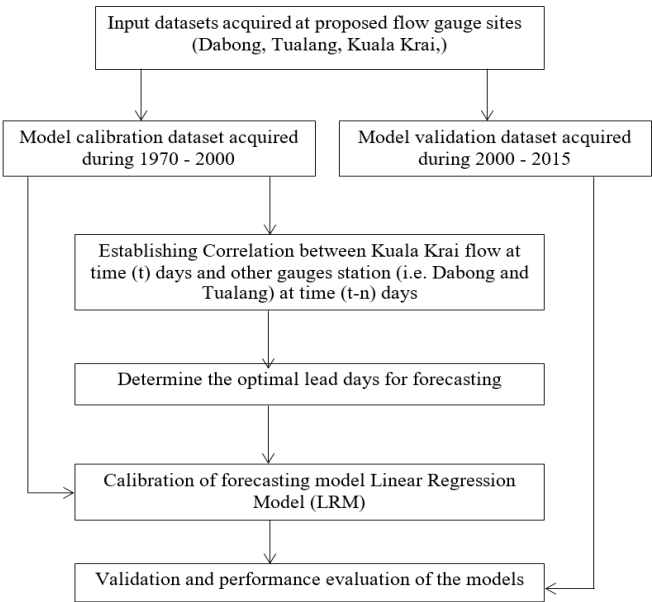

Fig. 6. Methodology for the implementation of stage regression method.

The relevant input and output stations for the selected river basins are identified to recognize the input and output values for the regression. The maximum instantaneous yearly water level is listed down for each input and output stations in order to generate the equation of the relevant stations. In the analysis, the schematic diagram of forecasted stations is depicted in Figure 7. From the diagram, it can be simplified that regression coefficients were able to be developed for both Dabong and Tualang for station at Kuala Krai (time lags are between 2 - 4 hours), for water level prediction of Sungai Kelantan at Guillemard regress by station Kuala Krai (time lags up to 6 hours), and last but not least is to estimate water level for Sungai Kelantan at Jeti Kastam correlate by water level station at Guillemard (time lags up to 6 hours).
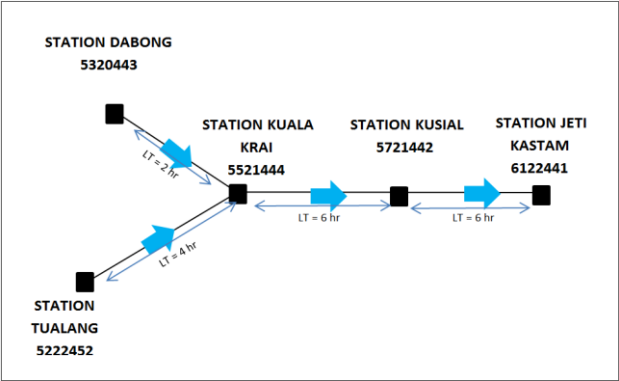
Fig. 7. Schematic diagram of stream flow stations in stage regression analysis of Sungai Kelantan.

6 Results and Discussion

The selected forecasting points for the river basin are Kuala Krai station, Guillemard station and Jeti Kastam station. The following equation was obtained to represent the relationship between the water levels of the two stations shown in Table 1. For verification purposes, 2 events were taken for calibration (2004, 2007) while 2 events for validation (2013, 2014). The instantaneous hourly data were obtained and analysed representing the most recent flood event. The resulting hydrograph for Kuala Krai, Guillemard and Jeti Kastam station are shown in the following Figure 8 until Figure 10.

Table 1. Regressed equation for stage regression.

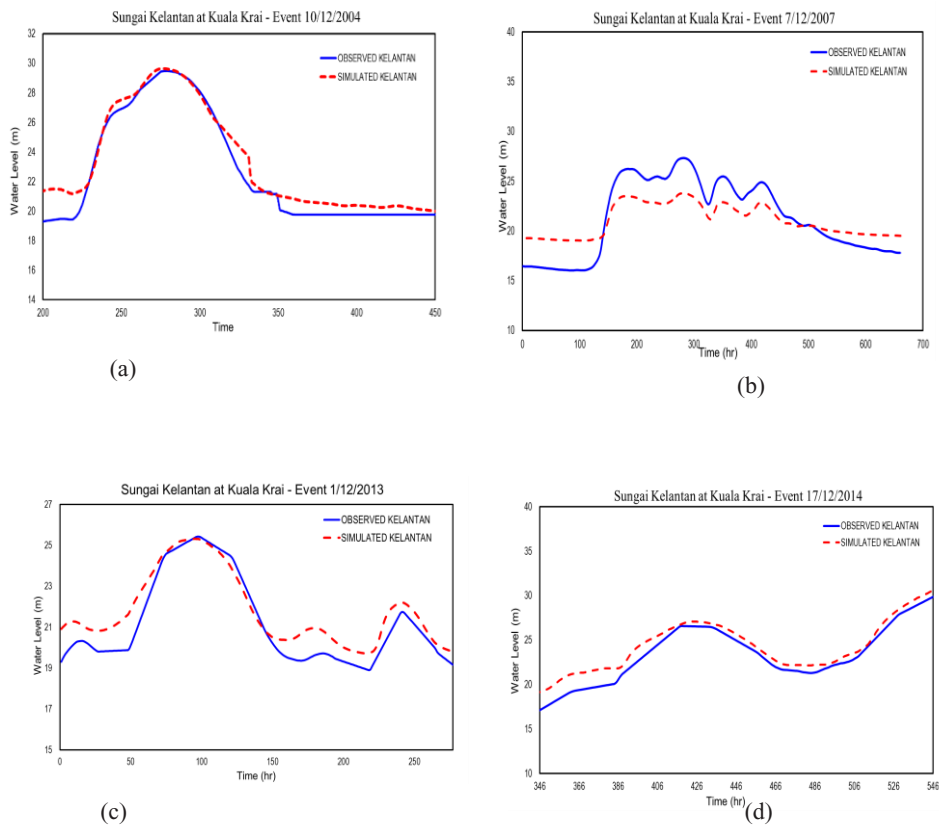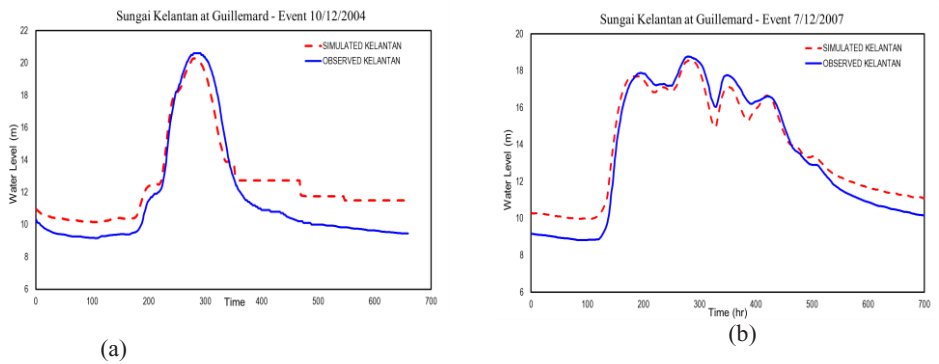| Forecasted Stations	Regression Equation
Kuala Krai	$Y_{t+T} = 0.895 \, _BX_t^{0.490348} \, _CX_t^{0.458358}$ Y_{t+T} Forecasted river stage at Kuala Krai at time t+T $_BX_t$ Observed river level at Dabong station at time t $_CX_t$ Observed river level at Tualang station at time t
Guillemard	$Y_{t+T} = 0.395043 \, _BX_t^{1.163935}$ Y_{t+T} Forecasted river stage at Guillemard at time t+T $_BX_t$ Observed river level at Kuala Krai station at time t
Jeti Kastam	$Y_{t+T} = 0.01874 \, _BX_t^{1.969169}$ Y_{t+T} Forecasted river stage at Jeti Kastam at time t+T $_BX_t$ Observed river level at Guillemard station at time t

Fig. 8. Hydrograph between observed and simulated at Kuala Krai Station for event (a) 2004, (b) 2007, (c) 2013 and 2014.
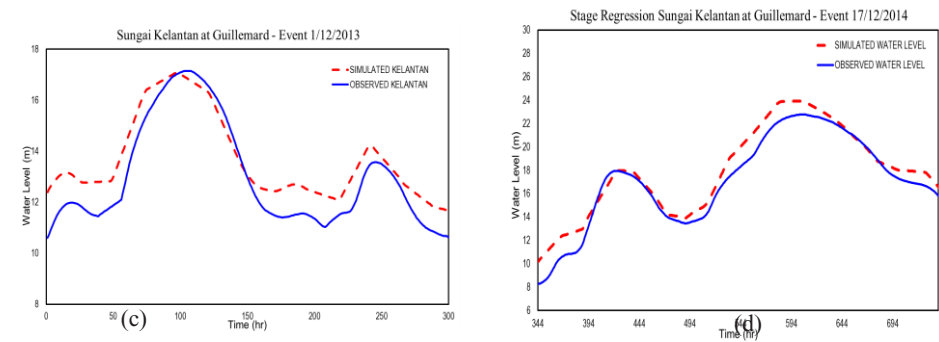
Fig. 9. Hydrograph between observed and simulated at Guillemard Station for event (a) 2004, (b) 2007, (c) 2013 and 2014.
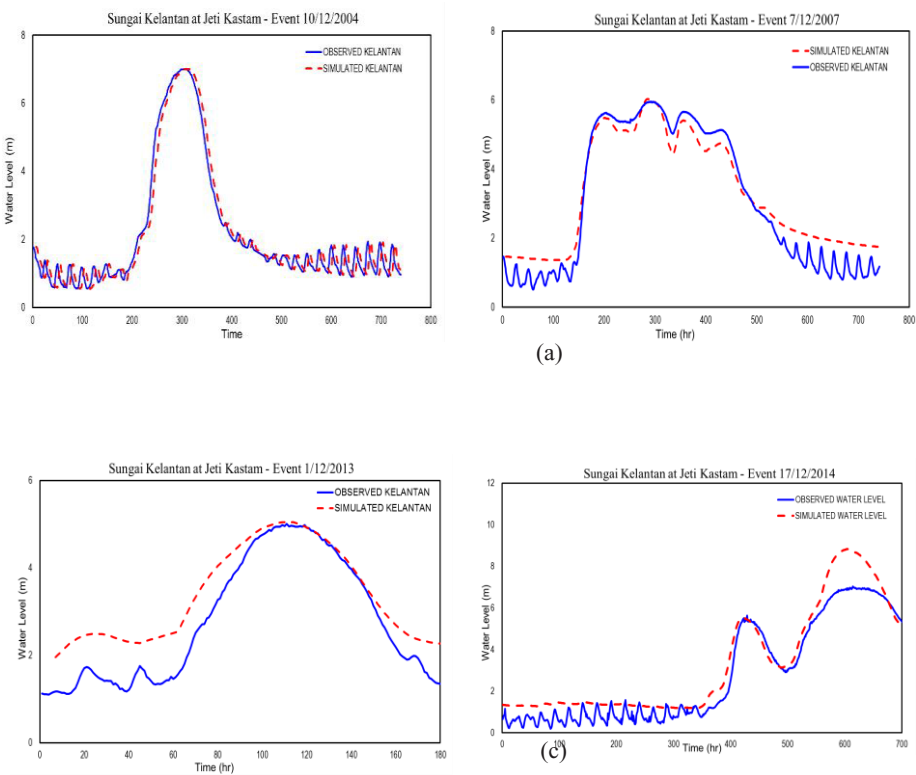

Fig. 10. Hydrograph between observed and simulated at Jeti Kastam Station for event (a) 2004, (b) 2007, (c) 2013 and 2014.

Based on the analysis done from several calibration and validation of flood event using stage regression method, it indicates that each equation generated as shown in Table 1 for each forecasted point such as Kuala Krai, Guillemard and Jeti Kastam possess a good and reliable coefficient although Jeti Kastam is exposed to tidal effects from South China Sea. The evaluation of hydrologic model behavior and performance is commonly made reported through comparisons between simulated and observed variables for multi response validation to assess overall modelling performance through few efficiency criteria such as Nash-Sutcliffe efficiency (P. Krause, 2005)

a)   The coefficient of efficiency (Nash & Sutcliffe, 1970) is define as:

$$E_c = 1 - \frac{\sum_{i=1}^{n} |Qrec - Q\,sim|}{\sum_{i=1}^{n} |Qrec - \bar{Q}rec|}$$

Where Q_{rec} (t) is the recorded discharge at time t, Q_{sim} (t) is the simulated discharge at time t, Q_{rec} is the average recorded discharge during the storm event, and n is the number of discharge records during the storm event.

b)   The error of peak discharge (Kwan Tun Lee et. al, 2008) is define as:

$$E_P = \left[\frac{Qp\,sim - Qp\,rec}{Qp\,rec} \right] x100\%$$

Where $(Q_p)_{sim}$ is the peak discharge of the simulated hydrograph, and $(Q_p)_{rec}$ is the recorded peak discharge.

Table 2. Calculated error analysis for stage regression.

Station	Event	Coefficient of Efficiency, E_C	Peak Discharge Error, E_P (%)
Kuala Krai	2004	0.99	0.0
	2007	0.97	2.6
	2013	0.95	5.5
	2014	0.97	1.9
Jambatan Guillemard	2004	0.94	2.9
	2007	0.98	1.1
	2013	0.94	9.2
	2014	0.96	5.3
Jeti Kastam	2004	0.95	1.5
	2007	0.99	0.2
	2013	0.73	1.3
	2014	0.80	5.4

7 Conclusions

By definition, regression analysis is a procedure for fitting an equation to set of data that comprises of two or more variables and the relation between these variables to represents the best fit. In this study, principle applied to solve the regression method is based on least square method. Table 2 shows promising results with coefficient of efficiency, Ec average value of 0.93 and peak discharge error, Ep average value of less

than 4% for the observed and simulated discharge. From the results, it can be seen that stage regression method provides a simplistic forecasting method to users or operators whilst maintaining a relatively reasonable amount of accuracy. The effectiveness of real time flood forecasting systems in reducing flood damage would depend upon how accurately the estimation of future water level or flow of incoming flood and its time sequence at selected points along the river. The results can be used for a timely and accurate flood forecasting and warning system of the approaching floods to save life and property as well as assisting the authorities in flood rescue operation.

8 Acknowledgement

References

1. World Meteorological Organization, WMO, (2011). Manual on Flood Forecasting and Warning, 2011 edition.
2. R. J. Moore and V. A. Bell, (2001). Comparison of Rainfall-Runoff Models for Flood Forecasting Report. Bristol.
3. P. Serban, and A.J. Askew, (1991): Hydrological forecasting and updating procedures. IAHS Publ., 201:357–369
4. P. Burlando, R. Rosso, L. G. Cadavid, and J. D. Salas, (1993), Forecasting of short-term rainfall using ARMA models, J. Hydrol. 144, 193-211.
5. B. Sivakumar, (2010). Advances in Data-based Approaches for Hydrologic Modelling and Forecasting. Edition. World Scientific Publishing Company.
6. DID (Drainage and Irrigation Department). (1990 – 2014) Annual flood report of DID for Peninsular Malaysia. Unpublished report. DID: Kuala Lumpur.
7. P. Krause, D.P Boyle, F. Base, (2005) "Comparison of different efficiency criteria for hydrological model assessment", Advances in Geosciences.
8. J. E. Nash and J. V. Sutcliffe (1970), "River flow forecasting through conceptual models part I — A discussion of principles", Journal of Hydrology, **10** (3), 282–290.
9. Kwan Tun Lee et. al, 2010. "Derivation of variable IUH corresponding to time-varying rainfall intensity during storms". Hydrological Sciences Journal, **53** (2), 323-337, ISBN 0262-6667