

Flood Forecasting for Klang River at Kuala Lumpur using Artificial Neural Networks

Jer Lang Hong and KeeAn Hong

Taylor's University, Hong and Associates
jerlanghong@hotmail.com, keeahong@yahoo.com

Abstract

This study evaluates the use of Multi-Layer Perceptron (MLP) neural network models to forecast water levels of a gauging station located at the Kuala Lumpur city centre in Malaysia using records of upstream multiple stations. Cross correlation analysis of water level data was performed to determine the input vectors which include the current and antecedent water level data of the upstream stations to ensure that of the data available, the most influential values corresponding to different lags are selected for analysis. Twelve well recorded storm events were used to train, test and validate the MLP models. The best performance based on MSE, MAE and R^2 was achieved with a model of 15 input vectors of upstream current and antecedent water levels, 7 hidden nodes and an output vector for the station at Kuala Lumpur centre. The R^2 values for training, testing and validation datasets are 0.81, 0.85 and 0.85 respectively.

Keywords: Drought, Neural Network

1. Introduction

As Kuala Lumpur was built in the flood plain of the Klangriver, it was subjected toflooding from the earliest days. The earliest flood recorded is the 1926 flood. The largest recorded recently is the 1971 flood. The particular 1971 flood was widespread and lasted for 5 days and caused extensive damages toproperty, infrastructures and agriculture. In the past two decades, flooding has become more frequent and Kuala Lumpur has been affected by flash floods with hardly any warning.

As a result of the 1971 flood, structural and non-structural measures were taken by relevant authorities (Abdullah K. ,2004) to mitigate the adverse effects of floods, these include:

- (a) Improvement of river channel
- (b) Construction of levees
- (c) Construction of flood by-passes
- (d) Construction of sediment traps
- (e) Improve hydrological data as basis for flood forecasting and warning systems

In view of the fact that reliable flood level forecasts are typically important for warning against the occurrence of dangerous real time floods, which can occur in a short time for urban catchments, techniquesproviding forecast of the arrival of peak water level within minutes or a few hours, are necessary. As large quantities of additional hydrological data have been collected in the past three decades and with the development of accurate and robust methods, it is considered appropriate to develop a flood forecasting model for the Klang river using the available hydrological data.

Flood forecasting has been based on different mathematical models. The models applied for real time flood forecasting can be broadly classified as physical or black-box models. Black –box models are empirical and they are able to map the data from the input to output by using transfer functions. Artificial neural networks (ANN) are examples of the black-box model which have recently been widely used in flood forecasting. In a

physical model, the internal description of the various sub-processes is modeled so as to represent, in a simplified form, the known physical processes. The black box model does not attempt to simulate processes occurring in a catchment but the physical model attempts to describe the catchment known processes.

ANN are artificial intelligence computational algorithms that mimics the biological processes of a human brain. ANN models have the advantage of their capability to reproduce the unknown relationship between a set of input variables and a set of output variable. Without a priori knowledge of the actual physical process and the exact relationship between the input and output data, the ANN network can be trained to learn such a relationship.

ANN has been successfully used in realtime flood forecasting, examples are: Sudheer, 2000, Thirumalaiah and Deo, 2000, Phien, *et al.*, 2003, Kim, *et al.*, 2009, Campolo, *et al.*, 1999. It has also been used for forecasting floods at downstream stations using records of upstream stations. (Mutulu, *et al.*, 2008, Wu, *et al.*, 2005).

In this study, the simple Multi-Layer Perceptron (MLP) neural networks were used for flood level forecast for the Klang river. Most of the earlier works carried out elsewhere in flood forecasting used this ANN model.

For flood forecasting, storm rainfall and the state (catchment conditions) are important as these affect the flow and stage at a gauging station. However, as the Klang river is gauged at three tributaries not far from the key gauging station (with approximately the same distance to the downstream key station) and not all rainfall data are available for the events selected for this study, the 3 upstream gauging records are used to forecast the flood level of the target gauging station. As recorded river flow is the net quantity of water measured at a particular location after abstraction/retention at the basin have been deducted under catchment and hydrological conditions at the time, and the catchment size of the sub-catchments of the 3 upstream stations is 90% of the catchment area at the key station, it is considered using the streamflow /water level of the 3 upstream stations alone will not compromise in the accuracy in flood forecasting for the downstream station.

2. Materials and Methods

2.1. The Study Area

Records of rainfall stations of the Selangor river basin were used for this drought analysis. Figure 1 shows a location map of Selangor basin up to the Public Works Department (PWD) water intake point. Selangor river basin up to the intake point has an area of 1450 km² and the maximum length and width of the basin are 48 km and 39 km respectively. About 30% of the basin is steep mountainous country above 600 m, 38% is in hilly country and the remainder undulating low terrain. A large portion (two-thirds) of the basin is under jungle and the remainder under rubber, oil palm, paddy, maize and vegetable cultivation. In the eastern half fine to coarse granite and other allied rocks are found and sandstone is found. The city of Kuala Lumpur is located at the confluence of the Klang and the Gombak rivers in the Klang river basin (see Figure 1). The Klang basin has an area of approximately 1288 km² and 244 km² of it makes up the federal capital of Kuala Lumpur and the rest of the basin are in the state of Selangor.

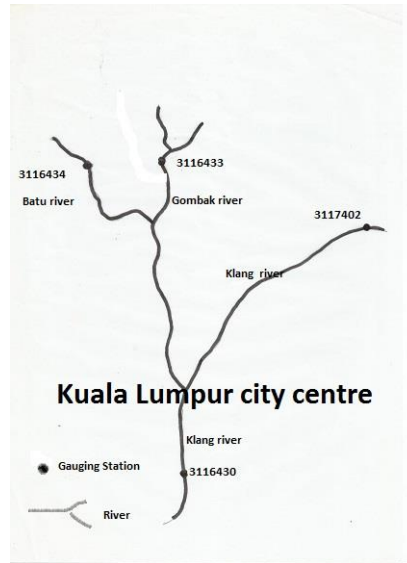


Figure 1. Klang River Basin

Kuala Lumpur has grown from a small mining town to the biggest city in Malaysia, and it is the country’s economic centre. It is estimated that about 50% of the Klang basin has been developed for residential, commercial and industrial use. About 1.5 million people of the 3.7 million population live in the city. Kuala Lumpur is therefore a densely populated city. Due to the rapid development, occurrences of flash floods because of clogged drains and overtopping of river banks are not uncommon.

The Kuala Lumpur city is surrounded by hills which rise to 1430 m in some location. TheKlang river is 120 km long from the hill to its outlet at sea. The two main tributaries are Gombak river which is 27 km and governs an area of 122 km².The Batu river is 24 km with a catchment area of 145 km².The mean annual rainfall of the basin ranges from 1900 mm to 2600 mm.

2.2. Water Level Data

Water level records are available for the gauging stations shown in Figure 1 and presented in Table 1.

Table 1

<i>Station name</i>	<i>Station no</i>	<i>Catchment area km²</i>	<i>Periods of record</i>
<i>Sg Klang at Sulaiman bridge</i>	<i>3116430</i>	<i>468</i>	<i>1974-2000</i>
<i>Sg Gombak</i>	<i>3116433</i>	<i>122</i>	<i>1960-2000</i>
<i>Sg Batu</i>	<i>3116434</i>	<i>145</i>	<i>1960-2000</i>
<i>Sg Klang at Lorong Yap Kwan Seng</i>	<i>3117402</i>	<i>160</i>	<i>1977-1997</i>

2.3. Input Dataset Preparation

To use the water level data of the 3 upstream stations as input to forecast the levels of the gauging station at Sulaiman bridge, it is necessary to estimate:

- (a) the travel time from the upstream stations to Sulaiman bridge
- (b) suitable time interval for accurate estimate of flood level at Sulaiman bridge
- (c) correlation between water levels at Sulaiman bridge and water levels and antecedent water levels at upstream stations

Water level data were carefully examined and 12 stage hydrographs for the period 1987 to 1997 for each station were chosen for this study. This period was chosen to represent the more urbanized catchment conditions considering future forecasting requirements. The stage hydrographs of the selected historical events of Klang river at Sulaiman bridge were plotted in Figure 2 to Figure 13. The time to peak of the hydrographs are calculated and presented in Table 2.

Unit hydrograph theory allows a calculation of hydrograph ordinates using time increment of not more than $0.2 T_p$ (USSCS 1972) where T_p is the time to peak. For this study, in order to obtain more accurate flood forecasting, a time increment of 15 minutes is adopted. The time of travel from the upstream station to Klang at Sulaiman bridge was estimated using the distance from the respective location to the downstream gauging station and a flow velocity of $1 \text{ m}^3/\text{s}$. Results are shown in Table 3. As the time of travel are not very much different for the 3 stations, adopting an average of 150 minutes will lead to the estimation of water level at Sulaiman bridge 10 intervals ahead of the upstream water level.

To check the influence of upstream water level on the forecast of Sulaiman bridge water level, a cross correlation analysis was carried out using water levels of Sulaiman bridge and water levels and antecedent levels with different lags of the 3 stations upstream to decide water level values to be included in the model. The cross correlation shows the degree of linear relationship between the data values. The cross correlation results are shown in Table 4. The average lag for the 3 stations are 4,4,3. For convenience, in the ANN flood forecasting for Sulaiman bridge, flows (water levels) at the 3 upstream stations up to lag 4 (at 15 minutes' interval) were used for modeling. Results of cross correlation analysis for the upstream/ downstream stations are also presented in Figure 14 to Figure 25.

Table 2. Time to Peak of the Flood hydrograph Klang River at Sulaiman Bridge

Flood hydrograph	Time to peak, hours
21 8 May 7	7.67
30 Oct 94	2.62
29 Nov 94	3.55
16 Apr 95	5.93
29 Mar 95	3.24
31 May 95	1.77
9 Dec 96	3.18
10 Oct 96	3.34
12 Feb 97	1.69
18 may 95	6.43
20 Mar 97	2.35
11 Oct 96	3.85

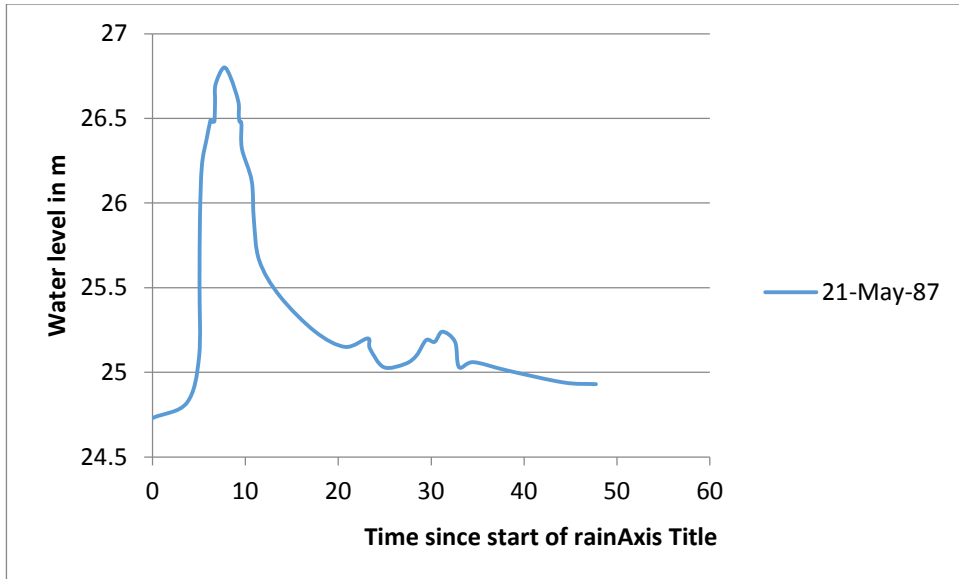


Figure 2. Stage Hydrograph of 21 may 87 Klang at Sulaiman Bridge

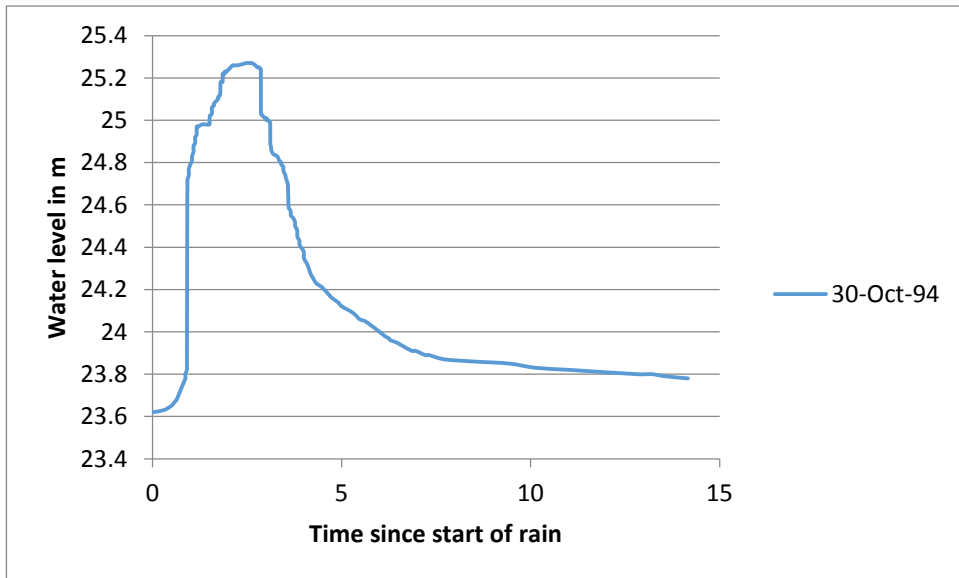


Figure 3. Stage Hydrograph of 21 may 87 Klang at Sulaiman Bridge

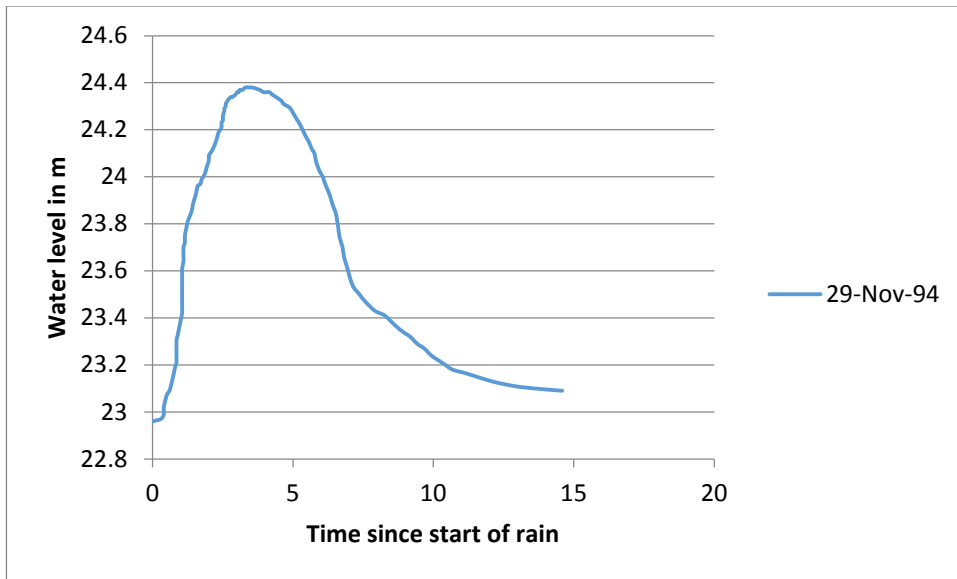


Figure 4. Stage Hydrograph of 29 Nov94 Klang at Sulaiman Bridge

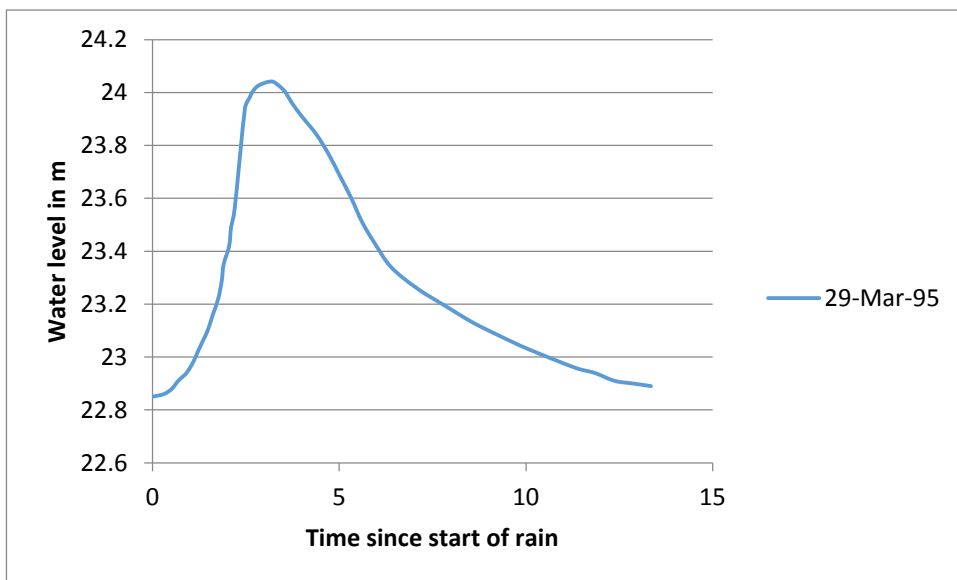


Figure 5. Stage Hydrograph of 29 March 95 Klang at Sulaiman Bridge

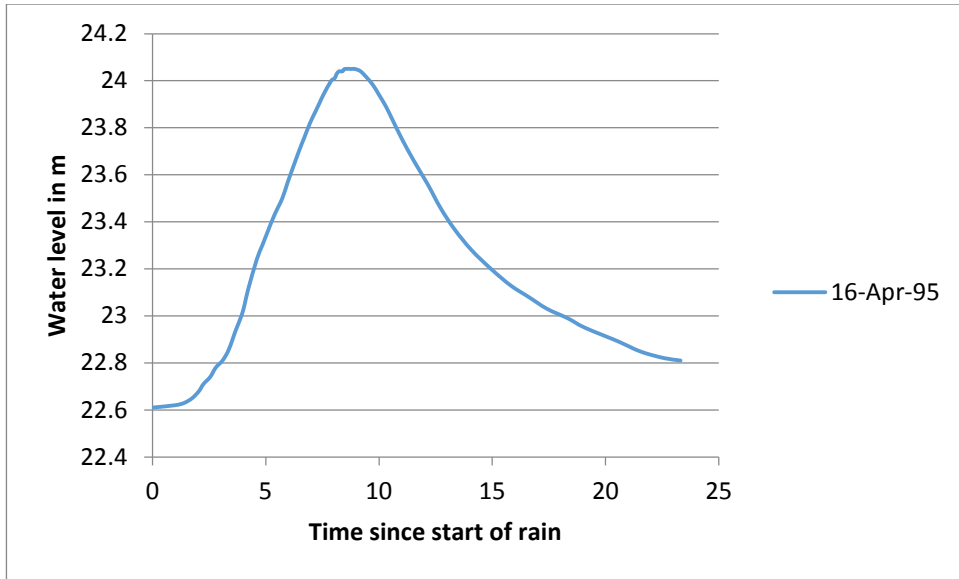


Figure 6. Stage hydrograph of 16 Apr 95 Klang at Sulaiman bridge

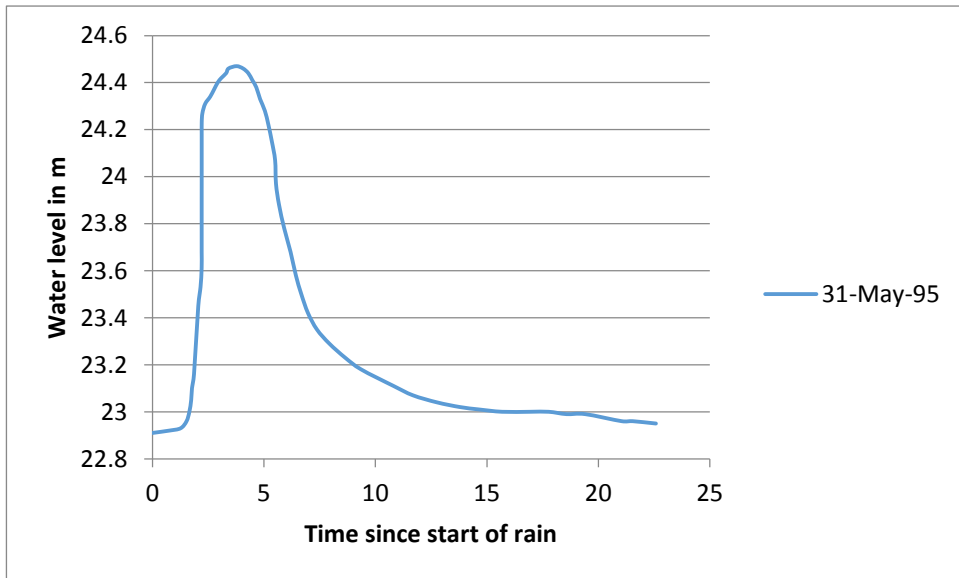


Figure 7. Stage Hydrograph of 31 May 95 Klang at Sulaiman Bridge

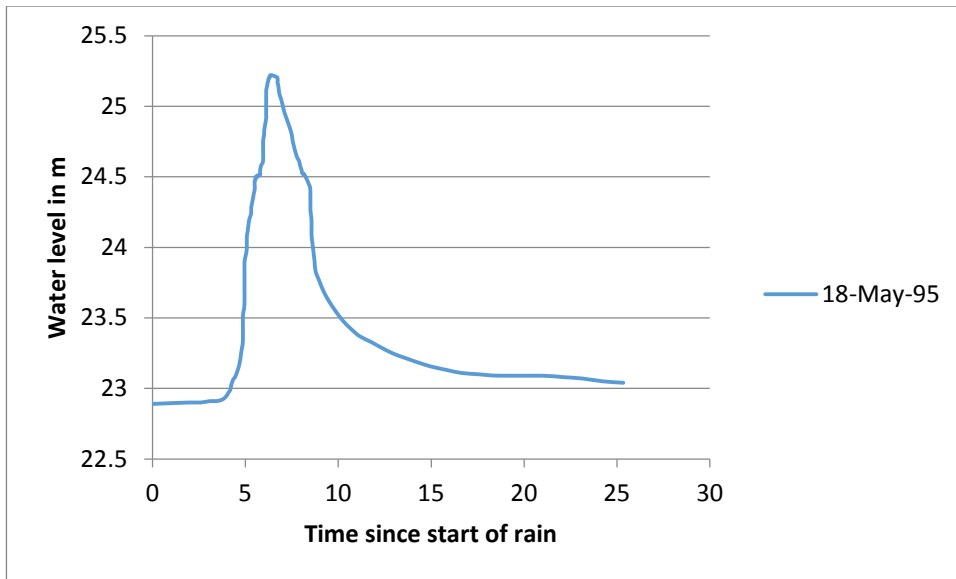


Figure 8. Stage Hydrograph of 18 May 95 Klang at Sulaiman Bridge

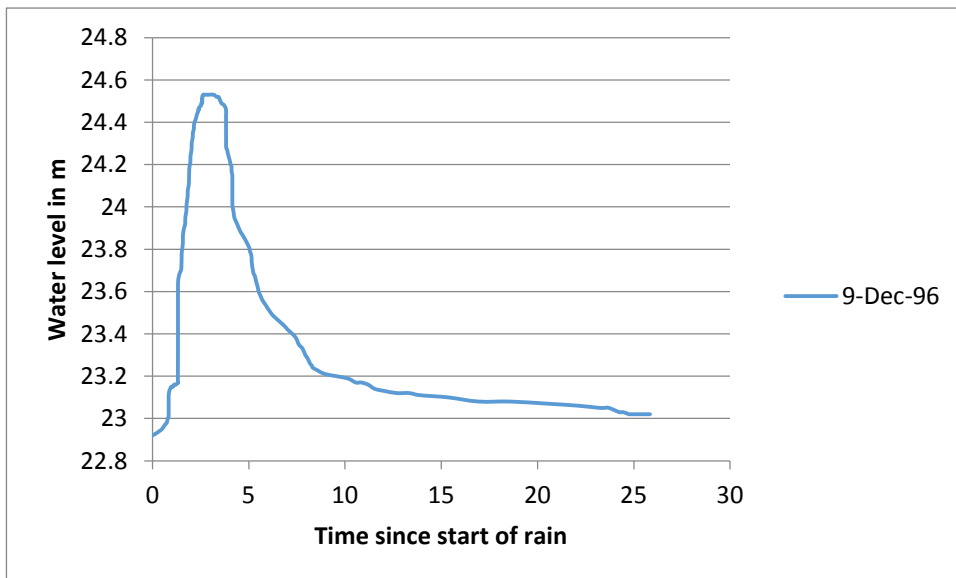


Figure 9. Stage Hydrograph of 9 Dec 96 Klang at Sulaiman Bridge

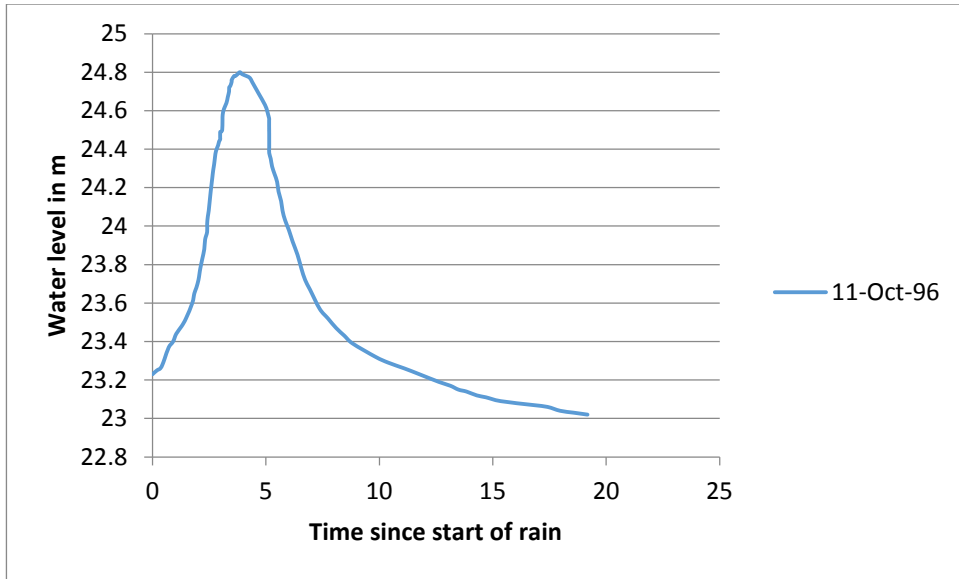


Figure 10. Stage hydrograph of 11 Oct 96 Klang at Sulaiman Bridge

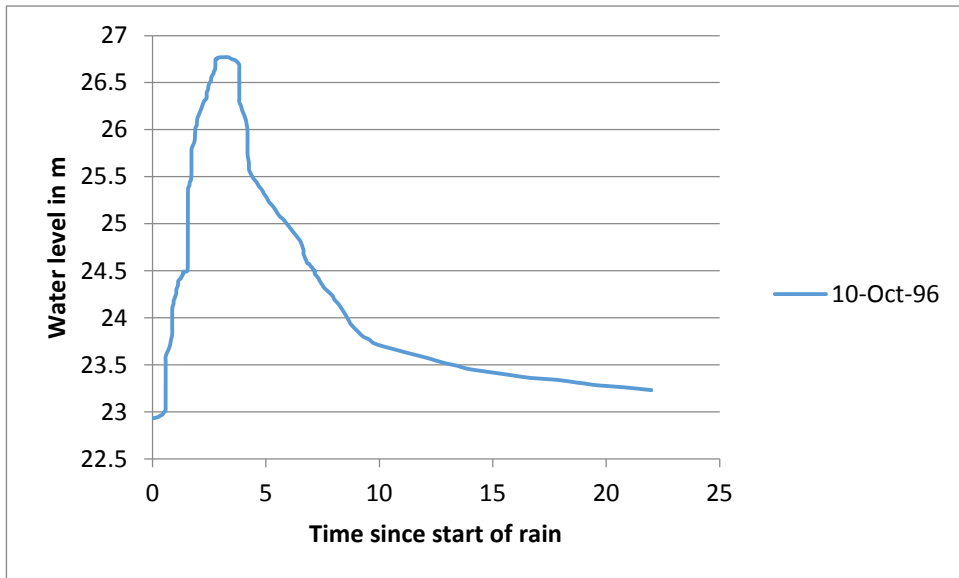


Figure 112. Stage Hydrograph of 10 Oct 96 Klang at Sulaiman Bridge

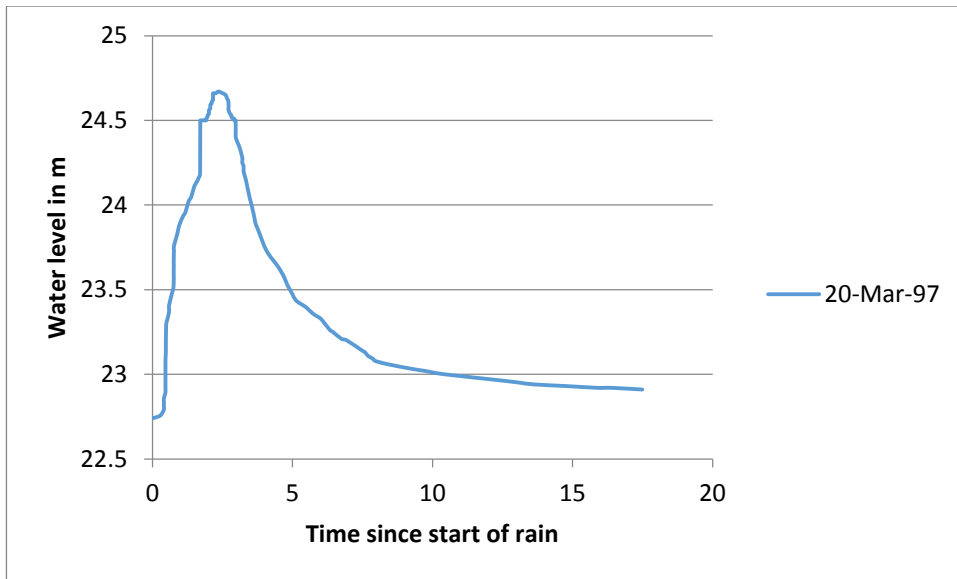


Figure 12. Stage Hydrograph of 20 Mar 97 Klang at Sulaiman Bridge

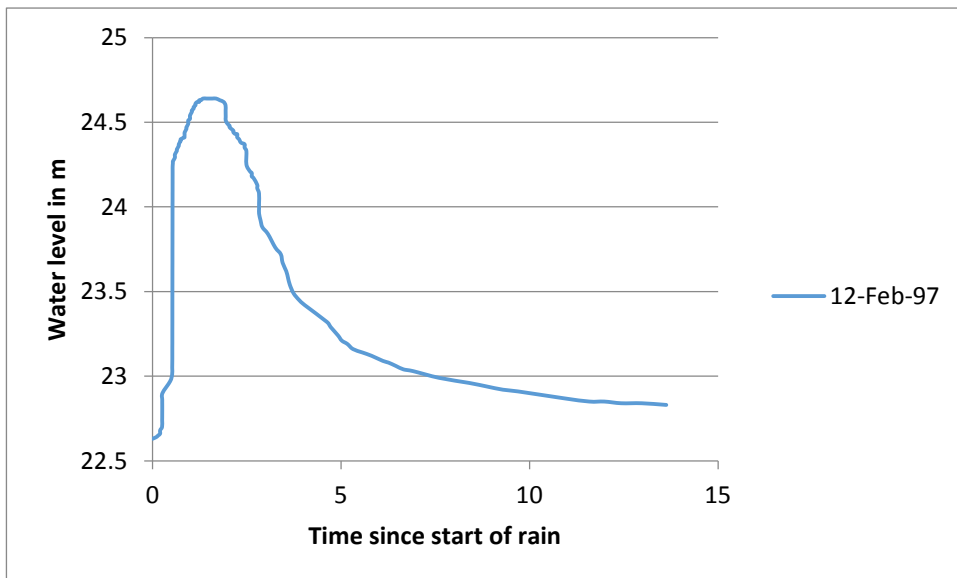


Figure 13. Stage Hydrograph of 12 Feb 97 Klang at Sulaiman Bridge

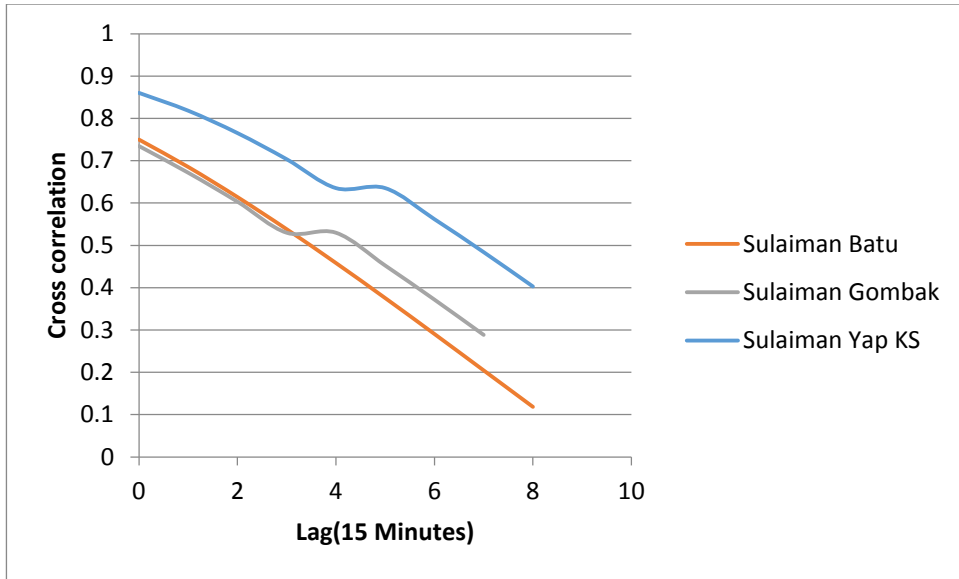


Figure 14. Cross Correlation Plots for Station Water Levels 87 May 21 Event

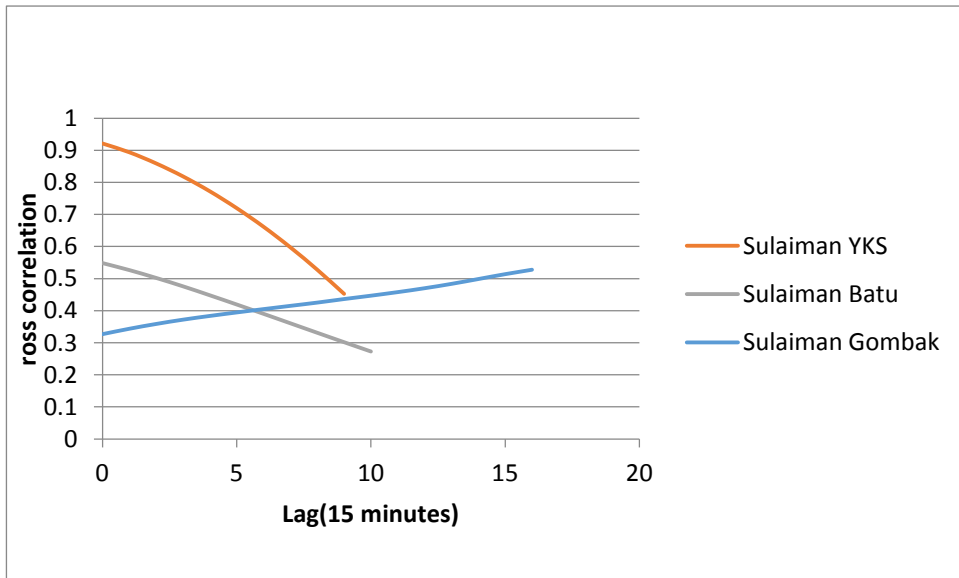


Figure 15. Cross Correlation Plots for Station Water Levels 95 Apr 16 Event

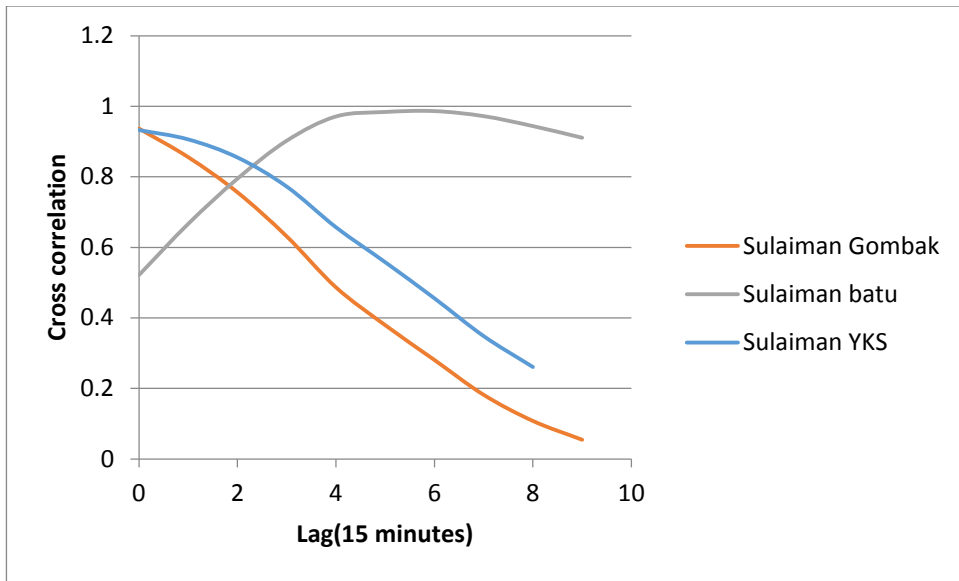


Figure 16. Cross Correlation Plots for Station Water Levels 94 Oct 30 Event

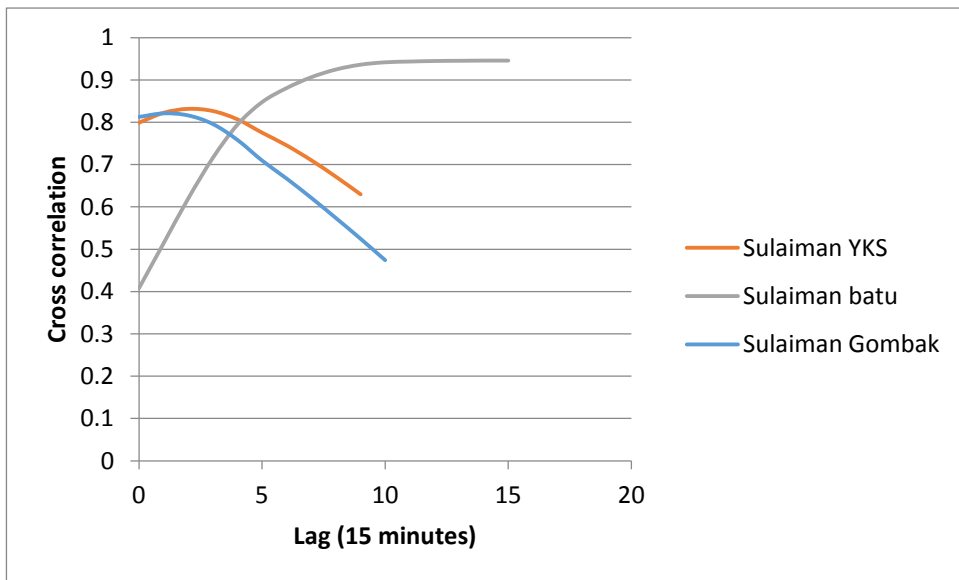


Figure 17. Cross Correlation Plots for Station Water Levels 94 Nov 29 Event

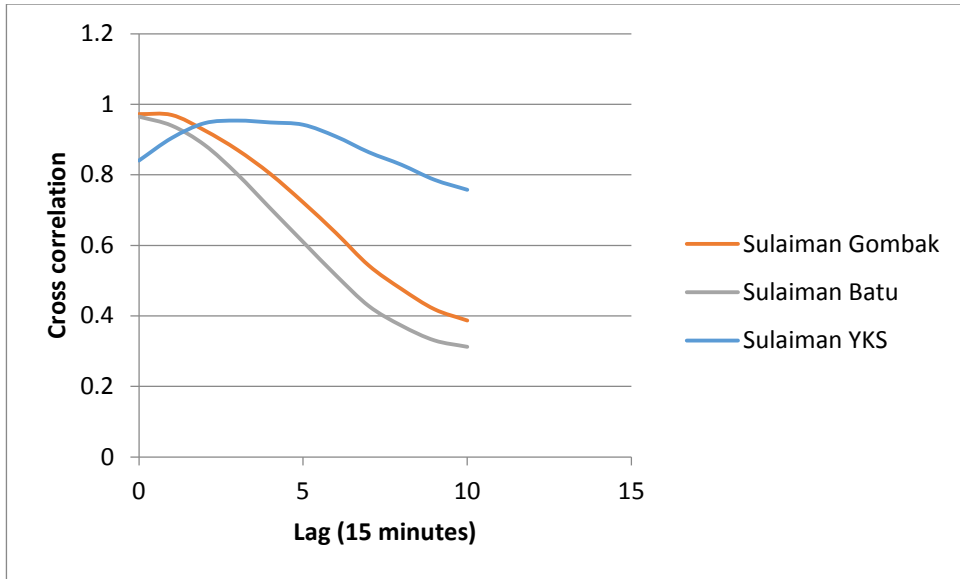


Figure 18. Cross Correlation Plots for Station Water Levels 97 Mar 20 Event

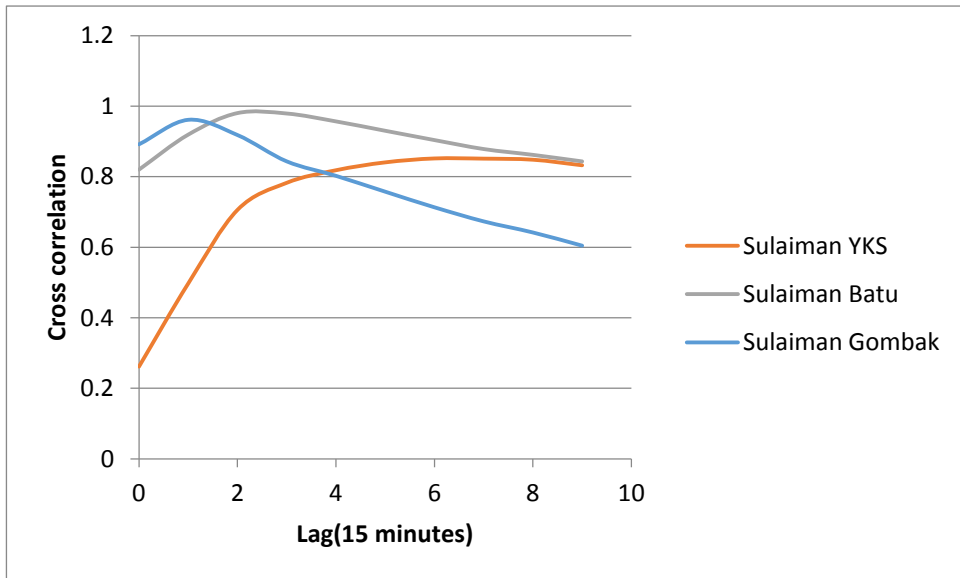


Figure 19. Cross Correlation Plots for Station Water Levels 97 Feb 12 Event

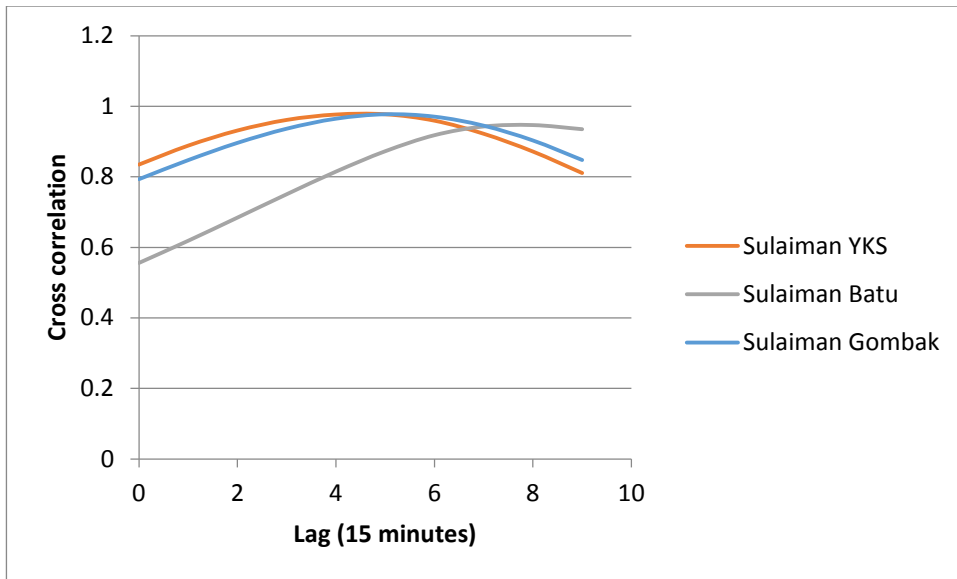


Figure 20. Cross Correlation Plots for Station Water Levels 96 Oct 11 Event

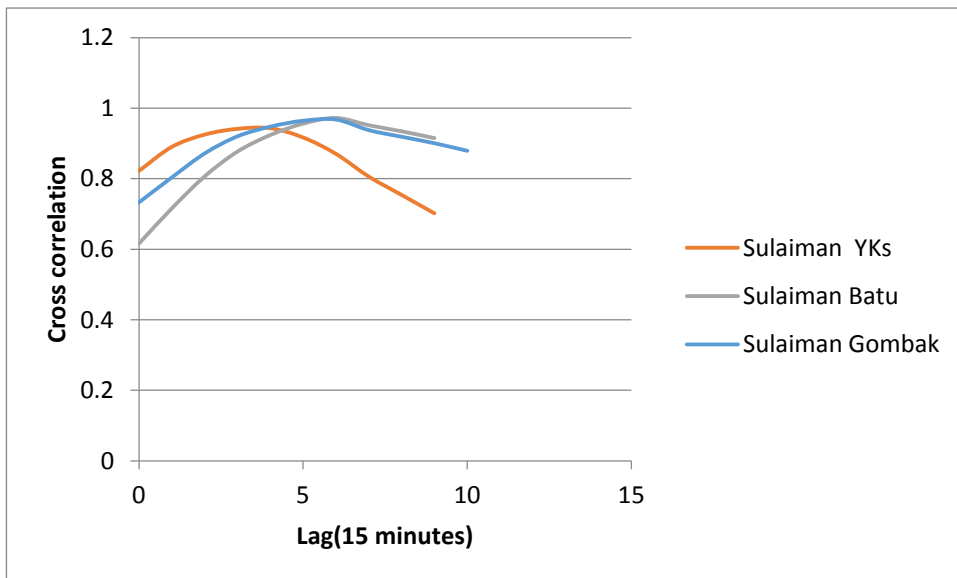


Figure 21. Cross Correlation Plots for Station Water Levels 96 Oct 10 Event

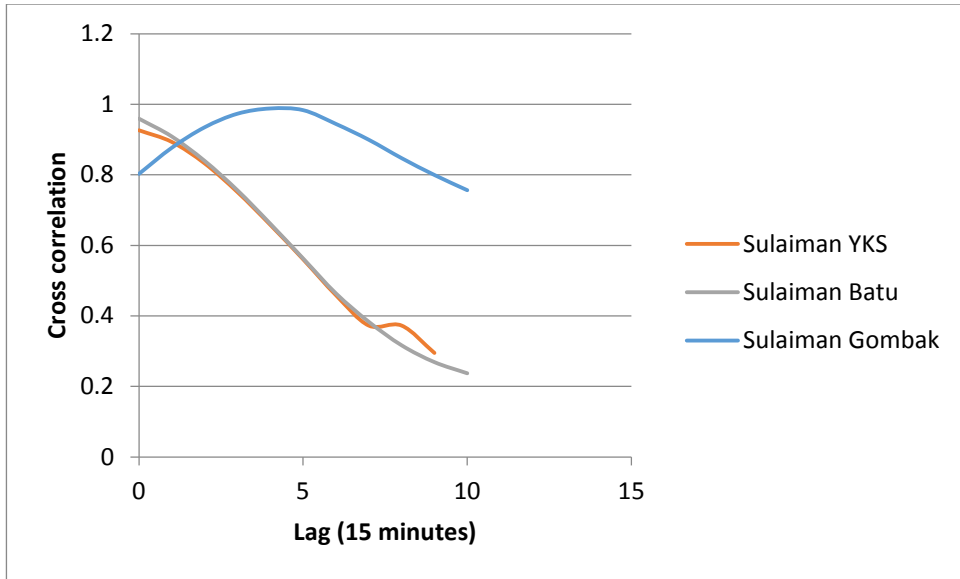


Figure 22. Cross Correlation Plots for Station Water Levels 96 Oct 9 Event

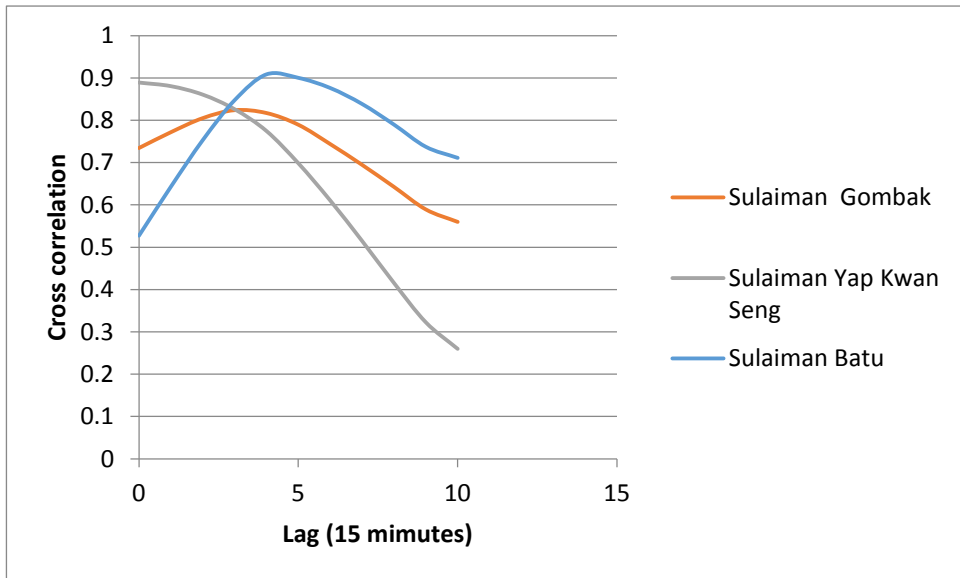


Figure 23. Cross Correlation Plots for Station Water Levels 95 May 31 Event

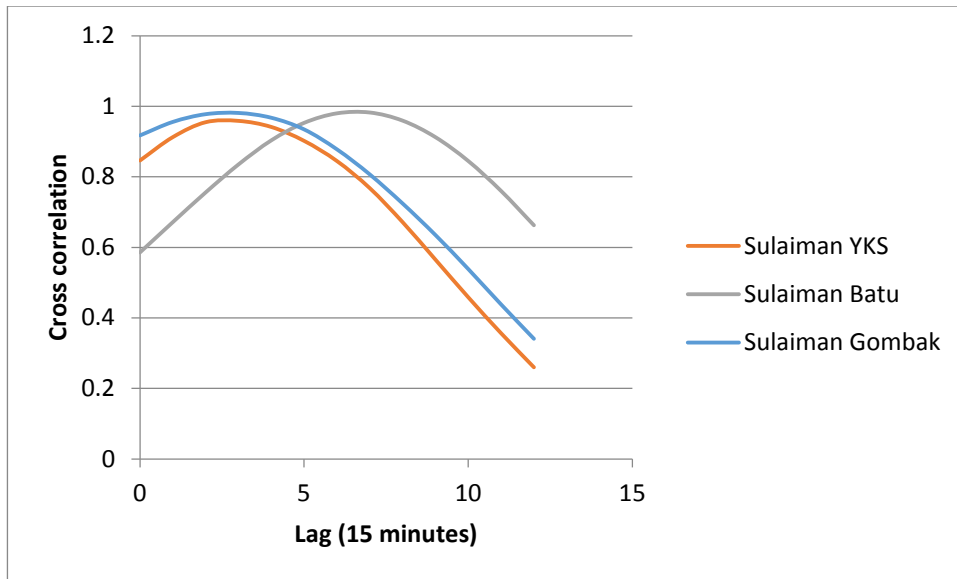


Figure 24. Cross Correlation Plots for Station Water Levels 95 May 18 Event

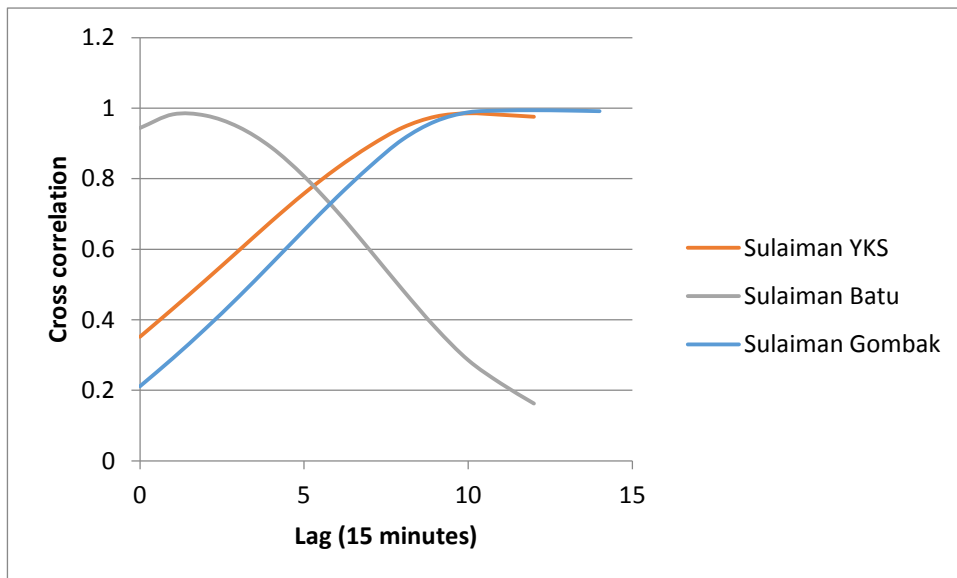


Figure 25. Cross Correlation Plots for Station Water Levels 95 Mar 29 Event

Table 3. Time of Travel from Upstream Gauging Stations to Klang at Sulaiman Bridge

Station	Distance from Sulaiman bridge km	Time of travel minutes
Gombak	7.8	130
Batu	9.8	163
Klang at Yap Kwan Seng	7.5	125

Table 4. Cross Correlation between Upstream and Downstream Stations

Event	Gombak/Sulaiman		Batu/Sulaiman		Yap K seng/Sulaiman	
	Correlation	Lag(15mins)	Correlation	Lag(15mins)	Correlation	Lag(15mins)
21May87	0.79	0	0.32	0	0.86	0
29Nov94	0.82	1	0.95	14	0.83	2
30Oct94	0.94	0	0.98	6	0.93	0
16Apr95	0.52	16	0.54	0	0.92	0
29Mar95	0.99	12	0.98	1	0.98	10
18May95	0.98	3	0.98	7	0.96	3
31May95	0.82	3	0.91	4	0.89	0
9Dec96	0.99	4	0.96	0	0.93	0
10Oct96	0.97	6	0.97	6	0.94	4
11Oct96	0.98	5	0.95	8	0.98	5
12Feb97	0.96	1	0.98	2	0.85	6
20Mar97	0.97	0	0.97	0	0.95	3

2.4 Methodology

The aim of the current study is to use Multi-Layer Perceptron (MLP) artificial neural network model to forecast flood levels of Klang river at Kuala Lumpur using the Sulaiman bridge and the records of 3 upstream gauging stations. The model constructed can be adopted to provide real time flood warning and for formulating strategies by planners and water managers of the relevant authorities to mitigate and reduce the adverse effect of the occurrence of flash floods which are not uncommon in the city centre.

2.4.1 The MLP Neural Network Model

Artificial Neural Networks (ANN) are nonlinear and flexible massively parallel distributed information processing system. For a number of nonlinear processing units, it is possible to train the neural networks to learn from experience and compute the complex functional relationships with accuracy. A number of neural networks has been proposed in the literature but the most commonly method used in hydrology for flood forecasting is the feed forward Multi-Layer Perceptron (MLP) model. As an example, a typical three-layer feed forward MLP model is shown in Figure 26. For most flood forecasting studies carried out in the past two decades, the three layer MLP model was used with the input nodes consist the lagged rainfall and streamflow or water level values and the output is the forecasted future value. In this particular MLP model, hidden nodes are used to process the information transmitted from the input nodes with a particular nonlinear transfer function. For this study, the model considered is a single output MLP model. The network is processed through training, testing and validation stages in order to forecast the flood level using the input upstream data. Back propagation (BP) algorithm (Ruminate 1986) is usually used to correct the weights of the interconnecting neurons. Back propagation (BP) uses the steepest gradient descent method to correct the weight of the interconnecting neurons. This method (BP) solves the interconnection of the processing of processing elements by adding hidden layers. For the learning process in the back propagation method, the interconnection weights are adjusted using the error convergence

method to obtain a desired output from an input. The BP algorithm propagates the error at the output to the input layer through the hidden nodes to obtain the final output. The gradient technique is used to calculate the weight of the network and adjust the weight of the interconnections to minimize the output error.

BP uses the following equation (ASCE Task Committee, 2000) to correct the weighting factor:

$$\Delta w_{ij}(n) = \alpha \Delta w_{ij}(n - 1) - \varepsilon \left(\frac{\partial E}{\partial w_{ij}} \right) \quad (12)$$

Where $\Delta w_{ij}(n)$ and $\Delta w_{ij}(n - 1)$

are weights interconnecting nodes i and j during the n th and $(n - 1)$ th steps
 α is the momentum factor used to speed up training in flat regions of the error surface and helps to prevent oscillations in the weights

A learning rate ε is used to increase the chance of avoiding the training process being trapped in a local minimum instead of a global minimum

The number of neurons in the input and output layers are problem dependent and decided by the number of input and output variables in the MLP model. The size of hidden neurons is an important factor in solving the problems using MLP. There are no fixed rules in determining the number of hidden neurons required for the model and trial and error experiments are normally adopted to determine the hidden node that gives the model the best performance. However, empirical relationships between optimum hidden neurons and number of input and output elements were given by some authors e.g. Mishra et. al (2006) used $2n+1$ for estimating the number of hidden neurons

Where n is the number of input neurons

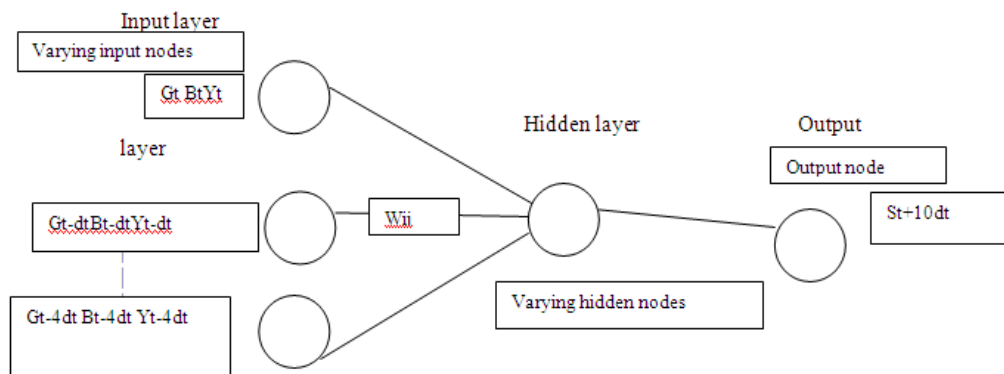


Figure 26. A three layer MLP neural network method

The neural network add-in version 1.5 software developed by the University of Adelaide (2014) was used for the forecasting of flood level with a 10 (15 minutes' interval) lead time.

The activation function used is the logical sigmoid function.

For training and validation purposes, data are normalized using the scaling method.

Input variables are selected using the partial mutual information selection option and data are split randomly with 60% ,20% ,20% for training, testing and validation purposes using the program.

The data is trained using learning rate of 0.01 and momentum coefficient of 0.9.

The performances of the MLP network model in predicting the flood level are assessed using:

$$R^2 = \left[\frac{\sum(Q_o - \bar{Q}_o)(Q_p - \bar{Q}_p)}{\sqrt{\sum(Q_o - \bar{Q}_o)^2 \sum(Q_p - \bar{Q}_p)^2}} \right]^2 \quad (1)$$

$$MSE = \frac{\sum_1^n (Q_p - Q_o)^2}{n} \tag{2}$$

$$MAE = 1/n \sum_1^n |(Q_o) - Q_p| \tag{3}$$

Where Q_o and Q_p represent the observed and predicted water level n = total number of observations

MAE is the mean absolute error

MSE is mean square error

R^2 is the square of correlation coefficient

2.4.2 MLP Model Architecture

The goal of the MLP is to generalize a relationship in the form

$$Y^m = f(X^n) \tag{4}$$

Where X^n is the n dimensional input vector consisting of variables $X_1, \dots, X_i, \dots, X_n$; Y^m is an m -dimensional output vector consisting of the resulting variables $Y_1, \dots, Y_i, \dots, Y_m$. In this study, values of X_i are the upstream station water levels with different lags and Y_i is the water level at Sulaimanbridge ten time steps ahead of the upstream station water levels at 15 minutes' increment.

The MLP model consists of water level input data from upstream stations up to a lag of four 15 minutes intervals, for 12 selected flood events. Details are shown in Table 5.

Table 5. The MLP Model for Klang River

Input	Variable
Gombak water level	Gt ,Gt-dt,Gt-2dt,Gt-3dt ,Gt-4dt
Batu water level	Bt, Bt-dt,Bt-2dt, Bt-3dt Bt-4dt
Klang at Yap Kwan Seng	Yt , Yt-dt,Yt-2dt,Yt-3dt, Yt-4dt
Output	
Klang at Sulaiman bridge	St=Q'(t+10dt)=f(Gt,Gt-dt,...Gt-4dt,Bt,Bt-dt,...,Bt-4dt,Yt,Yt-dt ,...,Yt-4dt)
Time interval dt=15 minutes	

Note: Gt=Gombak water level at time t

Gt-dt = Gombak water level at time t-dt

Bt=Batu water level at time t

Yt=Klang at Yap Kwan Seng water level at time t

St=Klang at Sulaiman water level at time t

Q'(t+10dt) = Water level 10 time steps ahead of upstream stations at Sulaiman bridge

Determining the number of input variables (water levels upstream stations) involved finding the lags of upstream water levels that have a significant influence on the predicted water level at the Sulaiman bridge. These influencing values corresponding to different lags are discussed in Section 2.3.

After inputs were determined, the MLP models were optimized to obtain the best prediction model. The number of nodes was changed in the hidden layer to determine the optimum number for the MLP model.

3.0 Results and Discussion

Data of current and antecedent water levels of the upstream gauging stations as presented in section 2.4.2 were used as input for the MLP model. Output of the model is the water level of Klang river at Sulaiman bridge. The MLP model was simulated by varying the nodes of the hidden layer from 1 to 33 and the MSE, MAE and R^2 values were noted. The models which produce the minimum MSE and maximum R^2 for the training, testing and validation datasets are listed in Table 6. It is noted that the R^2 value

varied from 0.81 to 0.85 and the MSE varied from 0.20 to 0.26. Figure 27 shows the observed and predicted water levels of Klang river at Sulaiman bridge for the validation simulation run with a 15-7-1 structure. The model is generally able to predict the water levels close to the observed values. The model is considered satisfactory when used for the Klang river water level forecasting.

Table 6. Performance of MLP Model for the Klang River at Sulaiman Bridge

Simulation mode	MSE	MAE	R ²	Model structure
Training	0.26	0.35	0.81	15-7-1*
Testing	0.20	0.32	0.85	15-7-1
Validation	0.23	0.35	0.85	15-7-1

*Denotes a network with 15 input nodes ,7 hidden nodes and 1 output node

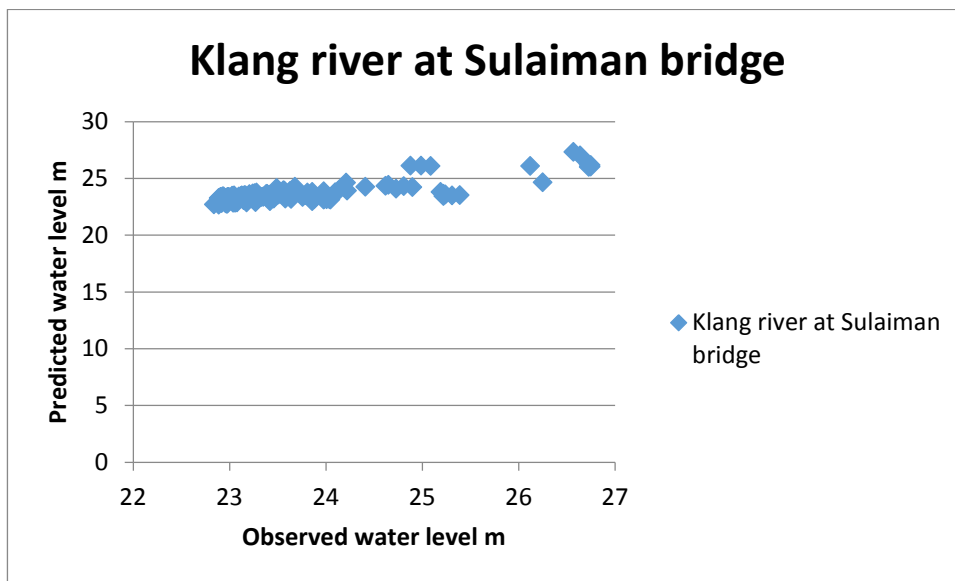


Figure 27. Observed and Predicted Water Level Klang River at Sulaiman Bridge

4.0 Summary and Conclusions

The purpose of this study is to use the artificial neural networks model to forecast flood levels of Klang river at the centre of Kuala Lumpur city for the warning, planning, management and formulating strategies to reduce and mitigate the adverse effects of flood impacts in the Klang basin. Forecasting was made for Klang river at Kuala Lumpur with a 150 minutes (travel time from upstream gauging stations) ahead lead time. Cross correlation analysis of water level data was performed to determine the input vectors which include the current and antecedent water level data of the upstream stations to ensure that of the data available, the most influential values corresponding to different lags are selected for analysis. Twelve well recorded storm events were used to train, test and validate the MLP models. The best performance based on MSE, MAE and R² were achieved with 15 input vectors from upstream current and antecedent water levels, 7 hidden nodes and an output vector for the station at Kuala Lumpur centre. The R² values for training, testing and validation datasets are 0.81, 0.85 and 0.85 respectively.

References

- [1] K. Abdullah, "Kuala Lumpur re-engineering a flood conference", Professor Chin Fung Kee Memorial Lecture, (2004).
- [2] "ASCE Task Committee Artificial neural networks in hydrology 1", Preliminary concepts Journal Of hydrologic engineering , vol. 5, no. 2, (2000), pp. 115-123.
- [3] M. Campolo, P. Andreussi and A. Soldati, River flood forecasting with a neural network water resources research, vol. 35, no. 4, (1999).
- [4] S. Kim, J. H. Kim and K. B. Park, "Neural networks models for the flood forecasting and disaster prevention system in the small catchment", Disaster advances, vol. 2, no. 3, (2009) July.
- [5] A. K. Mishra and V. R. Desain, "Drought forecasting using feed forward recursive neural network", Ecological model ing, vol. 198, no. 2, (2006), pp. 117-138.
- [6] E. Mutulu, I. Chaubey, H. Hexmoor and S. G. Bajwa, "Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed", Hydrological process, (2008).
- [7] H. N. Phien and N. D. A. Kha, "Flood forecasting for the upper reach of the Red River Basin North Vietnam", Water SA, vol. 29, no. 3, (2003).
- [8] D. E. Rumelhart, G. E. Hilton and R. J. William, "Learning representations by back propagation errors", Nature 313, 533-536 School of Engineering ,Adelaide University Neural networks add in, version 1.5, (2014).
- [9] K. P. Sudheer, "Modeling hydrological processes using neural computing technique", Ph D thesis Indian Institute of Technology, (2000).
- [10] K. Thirumailah and M. C. Deo, Hydrological forecasting using neural network Journal of hydrologic engineering U.S. Soil conservation Service (1972), SCS national engineering handbook Section 4 Hydrology, (2000).
- [11] J. S. Wu, J. Han, S. Annambhotla and S. Bryant, "Artificial neural networks for forecasting watershed runoff and stream flow", Journal of Hydrologic Engineering ASCE, (2005) May/June.

