Jurnal
Teknologi

Full Paper

Flood Risk Pattern Recognition by Using Environmetric Technique: A Case Study in Langat River Basin

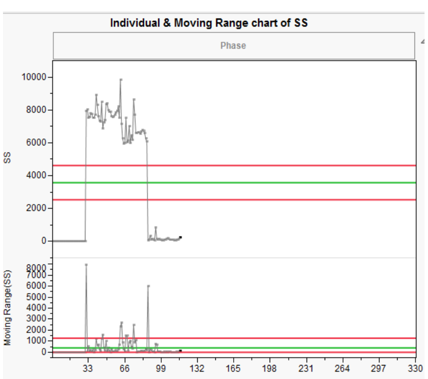
Article history
Received
25 February 2015
Received in revised form
24 March 2015
Accepted
1 October 2015

Ahmad Shakir Mohd Saudi[a], Hafizan Juahir[b*], Azman Azid[a], Mohd Ekhwan Toriman[a,c], Mohd Khairul Amri Kamarudin[a], Madihah Mohd Saudi[a], Ahmad Dasuki Mustafa[a], Mohammad Azizi Amran[a]

[a]East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300,Kuala Terengganu, Terengganu, Malaysia
[b]Faculty Science and Technology, University Science Islam Malaysia, 71800, Nilai, Negeri Sembilan, Malaysia
[c]School of Social, Development and Environmental Studies, Faculty of Social Sciences and Humanities, National University of Malaysia, 43600 Bangi, Selangor, Malaysia

*Corresponding author
hafizanjuahir@unisza.edu.my

Graphical abstract



Abstract

This study looks into the downscaling of statistical model to produce and predict hydrological modelling in the study area based on secondary data derived from the Department of Drainage and Irrigation (DID) since 1982-2012. The combination of chemometric method and time series analysis in this study showed that the monsoon season and rainfall did not affect the water level, but the suspended solid, stream flow and water level that revealed high correlation in correlation test with p-value < 0.0001, which affected the water level. The Factor analysis for the variables of the stream flow, suspended solid and water level showed strong factor pattern with coefficient more than 0.7, and 0.987, 1.000 and 1.000, respectively. Based on the Statistical Process Control (SPC), the Upper Control Limit for water level, suspended solid and stream flow were 21.110 m^3/s, *4624.553* tonnes/day, and 8.224 m/s, while the Lower Control Limit were 20.711 m, 2538.92 tonnes/day and 2.040 m/s. This shows that human development in the area gives high impact towards climate change and flood risk, and not the monsoon season. Prediction was carried out using the Artificial Neural Network (ANN) to classify risks into their own classes, and the rate of accuracy for the prediction was 97.1%. This meant that the points in the time series analysis which were located beyond Upper Control Limit were considered as High Risk class, and the probability for flood occurrence is very high. The other classes classified in this prediction are Caution Zone, Low Risk and No risk. This is important to set a trigger for warning system in the case of emergency response plan during flood.

Keywords: Hydrological, climate change, flood risk, time series analysis, Factor analysis

Abstrak

Kajian ini adalah mengamilkira penskalaan untuk menghasilkan dan meramal model hydrologi di kawasan kajian berdasarkan data sekunder yang diperoleh Jabatan Pengairan Dan Saliran daripada tahun 1982-2012. Kombinasi kaedah kemometri dan analis siiri masa dalam kajian ini membuktikan bahawa musim monsoon dan hujan tidak member kesan terhadap aras air, tetapi mendapan pepejal ,aliran sungai, dan aras air .yang member ikan hubungan yang tinggi dalam ujian korelasi dengan *p*- value <0.0001 yang memberikan kesan kepada aras air sungai. Faktor analisis bagi pembolehubah laluan air, mendapan pepejal,dan aras air menunjukkann corak faktor yang kuat dengan koefisi melebihi 0.7, dan 0.9787, 1.000, dan 1.000. Berdasarkan pengawalan proses statistikal, kawalan limit atas bagi aras air, mendapan pepejal, dan aliran sungai pada bacaan 21.110 m , *4624.553* tan/hari, dan 8.224 m/s, manakala bagi Had Kawalan Rendah adalah 20.711m, 2538.92 tan/sehari dan 2.040 m/s. Hal Ini ,menunjjukkan bahawa pembangunan manusia di kawasan kajian memberi impak yang tinggi terhadap perubahan cuaca, dan risiko kepada banjir dan bukan lagi semata-mata kepada keadaan musim monsun. Ramalan telah dijalankan dengan menggunakan Jaringan Hubungan Neural mengklasifikasikan risiko ke dalam kelas masing-masing, dan tahap akurasi bagi ramalan adalah 97.1%. Ini menunjukkan bacaan dalam analisis siri masa yang terletak di atas Had Limit Kawalan dikira sebagai kelas Risiko Tinggi, dan kemungkinan bagi berlakunya banjir di kawasan tersebut adalah sangat tinggi. Kelas-kelas lain yang terklasifikasi dalam ramalan ini adalah Zon Berjaga, Risiko Rendah dan Tiada Risiko. Ini adalah penting untuk membina system amaran banjir dalam kes pelan tindakan kecemasan ketika banjir

Kata kunci: Hydrologi, perubahan cuaca, risiko banjir, analisis siri masa, analisis faktor

1.0 INTRODUCTION

The Langat River basin is facing high urbanization activities around its surrounding area. The land-use along the river basin has been shifted from agricultural to industrial, causing changes in the hydrological structure and the river has been categorized as one of the 42 most polluted tributaries in Malaysia [1]. This shows that human development within the study area is rapidly growing and the population growth has also increased as these elements were highly correlated.

The transformation of low development area into an urbanized area at the basin causes changes in the hydrological modelling at the study area. Based on [2], the change of land-use from virgin forest and non-urban area into an industrialized urban area with high density of population will cause high impact on the discharge of pollutants that are directly flown into the water body system of the Langat River basin.

Since there is high population along the river basin, it will become a major contribution towards the rate of surface run-off and discharge into the river and the risk of flood occurrence at the study area would be high as well. The Langat river basin is a source of drinking water, domestic use and hydropower for the people in the district of Hulu Langat and it has also become one of the developed areas in the state of Selangor. The condition in the Langat river basin should focus on flood prevention. This is because high rate of surface run-off will cause the river to turn shallow and triggering flood condition that may lead to destruction of physical structure along the affected area. Furthermore, the risk of flood is classified high during the monsoon season at the study area. The South-West monsoon which begins from June until October is the major cause for flood occurrences and within this period, the rate of rainfall is considered high compared to the other months that are not affected by the monsoon period.Rapid growth of population and development gives negative impact towards climate change and water storage around the study area when people are prompting with the scarce resource in urban areas. Hence, the aim of this study are to look into the relationship of hydrological variables involved, factors for flood occurrence, control limit value for factors of flood occurrence and risk level of flood in the study area.

2.0 EXPERIMENTAL

The Langat River covers an area of approximately 1,815 km. The Langat River Basin consists of 15 sub-basins; Pangsoon, Hulu Lui, Hulu Langat, Cheras, Kajang, Putrajaya, Hulu Semenyih, Semenyih, Batang Benar, Batang Labu, Beranang, Bangi Lama, Rinching, Teluk Datok, and Teluk Panglima Garang. The basin is about 141 km in length and it has several tributaries with the principal ones being the Semenyih, Lui, and Beranang Rivers (Figure 1).

The basin also has two reservoirs; the Langat and the Semenyih Reservoirs, with catchment areas of 54 km^2 and 41 km^2, respectively. These reservoirs were built in 1982 to supply domestic and industrial water around those areas. The Langat Reservoir is also used to generate power supply at a moderate capacity for the population within the Langat Valley [5,6,7,8,9,10].

The climate in the study area is normally high in temperature, rainfall and humidity throughout the year. This climate has a dominant impact on the hydrology and geomorphology in the study area. A study has been carried out by [11] explained that the importance of land and atmosphere interaction may give great influences towards regional climate. Generally, the study area faces two types of seasons; the wet season, from April until November, and a relatively drier period from January to March. The weather is very much influenced by the South-West monsoon that blows across the Straits of Malacca. Besides, flooding is a common disaster in the study area. In this study, the location of the monitoring station for the Department of Irrigation and Drainage is shown in Table 1 and the data were collected from 1982-2012.

Figure 1 Study areas and sampling location station map

Table 1 Location of the Department of Drainage and Irrigation at Semenyih Reservoir in Langat River Basin

Variables	Station Number	Coordinates	Location
Rainfall	Site 3018101	3°5'7.31"N, 101°53'55.69"E	Semenyih Dam
Stream flow	Site 2918401	2°54'6.70"N, 101°48'28.98"E	Kg. Rinching
Suspended Solid	Site 2918501	2°55'47.96"N, 101°51'43.51"E	G. S. Rinching
Water Level	Site 2918443	2°57'19.37"N, 101°50'37.91"E	Semenyih

Statistical Analysis/ Pre-processed Data

Correlation Test

In this study, the application of Correlation Test was carried out to identify the relationship between variables for further analysis. This method best suited the study as it measures two variables and the relationship ranges from -1 to 1. The Pearson and Spearman Coefficients were used in this study. The Pearson Coefficient is normally used to test the relationship between two variables [12]. In this study, the test was used to measure the relationship between important parameters in the hydrological data. On the other hand, the correlation test was applied in this study to look into the relationship between the parameters and to identify the strongest relationship. Based on the results, we would be able to point out the development with the biggest impacts on hydrological modelling at the Langat River basin.

There are two common types of correlation tests known as Spearman and Pearson rank coefficients. The Spearman rank coefficient may need ordinal data as its calculation is based on the rank of the data. Spearman's rank correlation coefficient measures the strength of the coefficient between the variables in data analysis [13].

Based on the Spearman rank correlation, there are two types of correlations, namely positive and negative correlations. The positive correlation shows two variables increasing together in a linear condition, whereas negative correlation shows one variable increasing and the other decreasing in a linear condition. The Pearson rank coefficient needs the actual data to be calculated and all variables tested must be in ratio scale. In this study, both tests were carried out and the results are discussed.

$$r_P = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \quad [1]$$

Based on the equation above, X˘ and Y˘are the sample mean of X1, X2,..., Xn and Y1, Y2,...,Yn.

Artificial Neural Network

The concept of human brain has been utilized in Artificial Intelligent and this is also applied as a method in analysing data and it is called Artificial Neural Network (ANN)[18]. This concept was created by McCulloch and Pitts in *1943 make* as far as the process is concerned, the weighted sum of the inputs are transferred to the hidden neurons, where it is transformed using an activation function.[19] Learning process also has back propagation which applies error distribution, whereby this application can reduce errors up to a minimum level. This technique is used to minimize error functions and the iteration is terminated when the error function value reaches pre-defined goal, thus completing the process[2].

The function is:

$$f(x) = \frac{1}{1+e^{-\lambda x}}$$

$f(x)$ = Function involved in analysis

Λx = Gain parameter

The performance data can be calculated by implementing cross validation to a data set where the algorithm is terminated during the process and this process is done by using back propagation. The learning ability of ANN depends on the architecture of network and the number of hidden unit. The size of the network plays a major role in capturing the connectivity of the data when the degree of freedom is functioned to capture the connection, and the size of the network must be sufficient with the degree of freedom or the process will fail. In Addition, a study conducted by [14] discovered the effectiveness of ANN for rainfall, run-off modelling, and flood forecasting. The research also highlighted the ability of ANN in predicting river flow and the quality of water downstream, which were focused in this study as well.

Chemometric Techniques

One of the chemometric technique is Factor Analysis that being used to identify the reduction of variables into a set of factors for further analysis. According to [15], this method compares sets of data and identifies the variables that effect the most towards the change of the hydrological modelling in the study area with lower cost and less time compared to other techniques [16].

Factor Analysis

Factor analysis is applied in this study to define large number of variables into smaller sets. Factor analysis variables and latent construct that are measured will establish the dimension between these two elements and construct validity evidence of self reporting scales [17].

It also reduces the number of variables, examines the structure or relationship between variables, and detects and assesses unidimensionality of the theoretical construct [18]. This method also addresses multicollinearity (two or more variables that are correlated), which was carried out in this study as well. The equation for this method is:

$$Z_{ij} = a_{i1}x_{j1} + a_{i2}x_{j2} + a_{i3}x_{j3} + a_{im}x_{mj} \quad \ldots\ldots\ldots\ldots\ldots\text{[ii]}$$

Z = Component score
a= Component loading
x= Measured of variable
i= Component number
m= Total variable

Statistical Process Control (SPC)

Time Series Analysis is very important in predicting the water level at the study area. With this method, we were able to evaluate the process from the performance of the analysed data efficiently. It produced three important results, which were important in predicting the hydrological modelling in the future, and those results were Upper Control Limit (UCL), AverageValue (AVG) and Lower Control Limit (LCL). The Sigma in the control chart is represented within the range value of a set of data. The Control Chart has the ability to uncover some trends and patterns, showing actual data deviations from the historical baseline and dynamic threshold, being able to capture unusual resource usage, and becoming the best base lining to show how actual data are deviated from the historical baseline [20].The equation that is used in this analysis is:

Moving Range = Plot: MRt for t= 2, 3, ..., m.

MR= average moving range
t = time
m = individual values
Average Value:

$$\tilde{x} = \frac{\sum_{i=1}^{m} x_i}{m} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{[iii]}$$

$\bar{X}$= moving range
m = individual values
x_i = difference between data point

3.0 RESULTS AND DISCUSSION

The Relationship Between Variables that Contribute to Flood Occurenced

Based on Figure 2 and Table 2, the Pearson Correlation Test showed high correlation between the Stream Flow, Suspended Solid and Water level, with p-value less than 0.0001. Based from the results shown in Table 2, rainfall was not a major cause for the increase of Water Level at the study area even though the result was 0.049, as this value was still below 95% of confidence interval. It was clear that Suspended Solid showed a very strong correlation for the changing rate of Water Level where Suspended Solid described the rate of surface run-off that flowed into the water body. This explains that the risk of flood is irrelevant to the monsoon season. The result from the Correlation

Test clearly showed that the surface run-off had become the major impact for the increase of water level in the study area.

Table 2 Correlation test

Variables	Rainfall (mm)	Stream flow m/s	Suspended Solid Tonnes/day	Water Level (m)
Rainfall	0	0.037	0.095	0.049
Stream flow	0.037	0	**<0.0001**	**<0.0001**
Suspended Solid	0.095	**<0.0001**	0	**<0.0001**
Water Level	0.049	**<0.0001**	**<0.0001**	0

The Factors that Contribute to Flood Occurrence

From the result, it showed that rainfall was l.ess significant compared to the Stream flow, Suspended Solid and Water Level when the p-value for these elements were less than 0.001, compared to rainfall with p- value of 0.049. This clearly explains that the monsoon season does not pose risk of flood in the study area and the change of hydrological modelling is significant to be compared with the surface run-off at the study area rather than South-West monsoon.

Figure 2 and Table 3 show the variables with significant impact towards the hydrological modelling at the study area by using FA. From Table 3, the Stream flow, Suspended Solid and Water Level resulted in the highest coefficient; 1.000, 0.979 and 1.000, respectively. If the coefficient is more than 0.7, it is considered as strong coefficient. Hence, it is the human development that caused high rate in the surface run-off that affected the Water Level and the Stream Flow.

This analysis also reduced the variables by looking into the most significant variables for further analysis. There is high flood risk if no limitation is set for mitigating action.

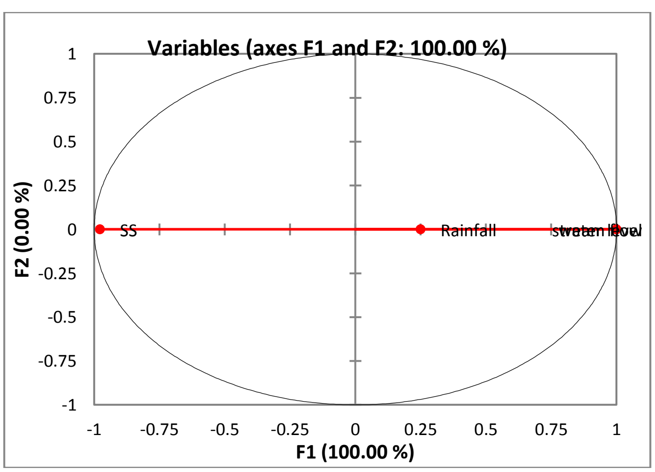


Figure 2 Correlation between variables and factors

Table 3 Correlation between variables and factors

	F1	Initial Communiality	Final Communiality	Specific Variance
Rainfall	0.249	0.109	0.062	0.938
Stream flow	1.000	0.998	1.000	0.000
Suspended Solid	0.979	0.974	0.958	0.042
Water level	1.000	0.999	1.000	0.000

Flood Control Warning System

After conducting the Factor Analysis, three variables were chosen for further analysis to set limitation for those variables related to flood risk analysis. Time series analysis was carried out through Control Building Chart to set control limit for flood prevention in the study area. Based on Figure 3 and Table 3, the result of suspended solid showed that the Lower Control Limit, mean value and Upper Control Limit were 2538 sediment/tonnes, 3581.772 sediment/tonnes and 4624.553 sediment/tonnes, respectively. From the findings, it explains that the limitation for SS discharge into the river must be below Upper Control Limit or otherwise, the risk of flood would be higher.

An area is considered safe from flood if the rate of SS is within the Average value and below the Lower Control Limit. The limitation for Suspended Solid is important for the Local Authority and legal enforcement as precaution, not only for flood control management, but also in preventing environmental pollution in the study area.

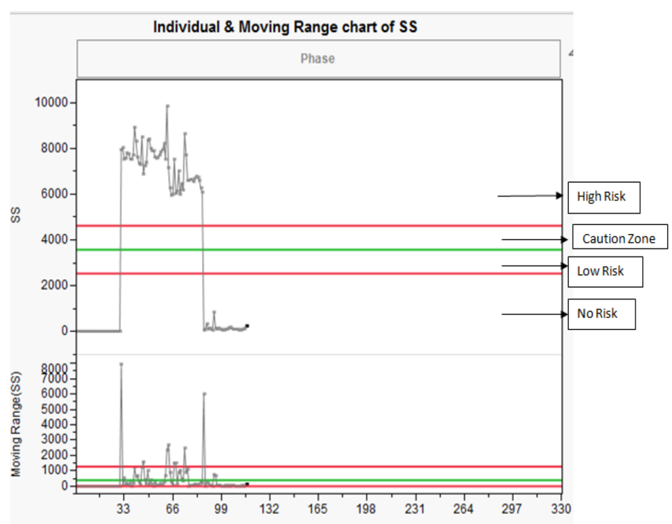


Figure 3 Control building chart for suspended solid from 1982-2012

Table 3 Results for control building chart

Points Plotted	Lower Control Limit	Aver age	Upper Control Limit	Limit Sigma	Sampl e Size
Individu al	2538.92	3581. 772	4624.55 3	Movin g Range	1

Next, Figure 4 and Table 4 portray the results of time series analysis using Control Chart for Stream Flow. The value for Lower Control Limit, mean value and Upper Control Limit for the stream flow were 2.040 m/s, 5.132m/s and 8.224 m/s, respectively. The findings explain that if the value is beyond the Upper Control Limit, it will cause destruction along the river bank. Thus, from this condition, proper mitigation should be implemented in order to control the stream flow rate and to make sure that the value is within the Average value. Hence, the Engineering and Biological Control Plan should be implemented by the Local Authority in controlling the rate of Stream Flow from exceeding the Upper Control Limit value.

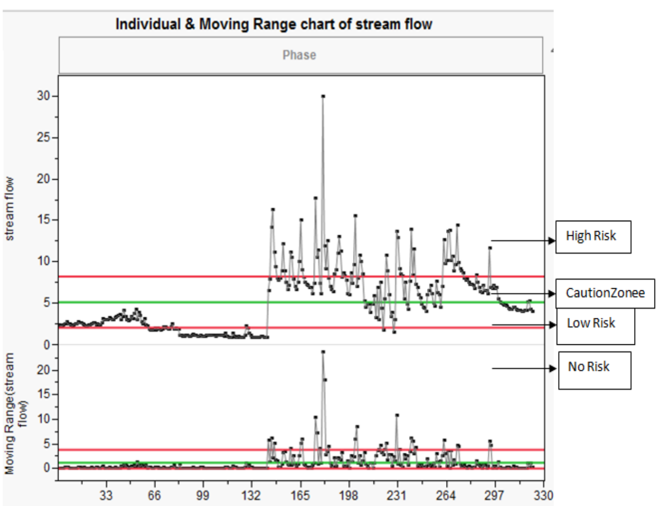


Figure 4 Control building chart for stream flow from 1982-2012

Table 4 Results for control building chart

Points Plotted	Lower Contro l Limit	Averag e	Upper Control Limit	Limit Sigma	Sample Size
Individua l	2.040	5.132	8.224	Moving Range	1

Figure 5 and Table 5 show that Water Level is an important parameter to determine flood occurrences in the study area. The limitation must be clear in order to trigger an emergency response action in preventing the destruction of human development along the study area.

Based on Table 6, the Lower Control Limit, mean value and Upper Control Limit for Water Level were 20.711 m^3/s, 20.910 m^3/s and 21.110 m^3/s, respectively. The results show that the development around the study area affects the climate change in the study area. In this study, the climate change is not based on the monsoon season anymore, but it refers to the rate of water storage that has become less when the river is shallow due to the impact of human development in the particular area. This is caused by the surface run-off, which becomes higher and precipitated on the surface ground of the river. This condition will cause the study area to face flood occurrences easily even if the rate of rainfall is low.

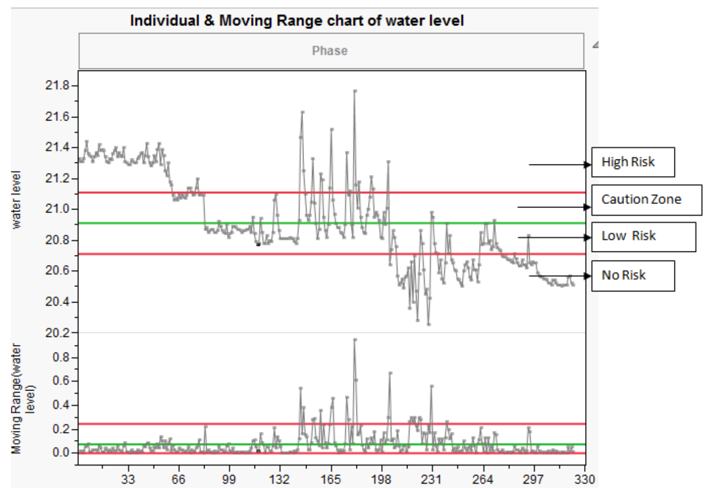


Figure 5 Control building chart result for water level from year 1982-2012

Table 5 Results for control building chart based on figure 6

Points Plotted	Lower Control Limit	Averag e	Upper Contr ol Limit	Limit Sigma	Sample Size
Individua l	20.711	20.910	21.110	Moving Range	1

Prediction for Flood Risk Classification

Based on the results from the time series analysis, the classification of risk based on its level of seriousness can be classified based on its points in the graph of control limit. The classification of risk for below the Lower Control Limit is No Risk, points between the line of Lower Control Limit and Average Limit are classified as Low Risk, points between Average and Upper Control Limit are Cautionary Zone, and points above Upper Control Limit are classified as High Risk.

All the risk points were predicted with ANN to make sure that the classification of points for Risk Class is significant to be in the group of its characteristics. Table 6 explains the results from the prediction of Risk Class:

Table 6 Prediction for classification of risk class

Flood Risk	Accuracy	Rel. Entro	Total
Train	1	0.01319	81
Test	0.971	0.0213	35

From the result, the accuracy of prediction was 0.971 and it shows that the prediction of risk in Risk Class is accurate and relevant to be taken into consideration for future prediction of flood at the study area in the future.

4.0 CONCLUSION

In conclusion, the Local Authority should show strong commitment in controlling the excessive amount of surface run-off into the river. They must be able to construct conditions such as information management and performance monitoring, integrated policy and strategies, constitution legislation and standard, Erosion and Sediment Control Plan (ESCP) to control erosion and sediment, highly enforce the Department of Drainage and Irrigation (DID), apply the regulation of Environmental Quality Act 1974 (Act 127) & Subsidiary Legislation, Waters Act 1920 (Act 418) & Water Supply (Federal Territory of Kuala Lumpur) Act 1998 (Act 581), and Water Act 1989 - Chapter 15, to control human development activities along the river bank. The developer who is responsible for the construction along the river bank must implement suitable mitigating measures to prevent excessive run-off which may cause precipitation of sediment that may become the major source for the river to become shallow and finally, to flood easily.

Other mitigating measures that should be implemented along the study area are the construction of Barrage, River Bund, Pump House, Diversion, Pond, Dam and River Improvement work that must be well maintained and improvised from time to time. The effectiveness of these mitigating measures also depends on the awareness and strong legal enforcement in controlling the rate of the surface run-off which comes from uncontrolled human development, and if this is not configured well, all the structures of the mitigating measures would mean nothing in preventing flood occurrences. The application of ANN is able to trigger warnings earlier for its citizen to take precaution for flood prevention based on the level of risk from the prediction. Time Series Analysis was able to identify the limitation for all factors that affect the most for the changing of the water level based on the results from the Correlation Test and Factor Analysis. This does not only reduces the cost of operation, but also the total loss from flood destruction and saves lives.

Acknowledgement

The authors acknowledge the Department of Drainage and Irrigation (DID) under Ministry of Natural Resources and Environment for giving us permission to utilize hydrological data, and East Coast Environmental Research Institute (ESERI), UniSZA for giving advice, guidance and support for this study.

References

[1] Aiken, R. S., Leigh, C. H., Leinbach, T. R., & Moss, M. R. 1982. *Development and Environment in Peninsular Malaysia*. Singapore. McGraw-Hill International Book Company.

[2] Juahir, H., Sharifuddin, M. Z., Ahmad, Z. A, Mohd, K. Y., Mazlin, M. 2009. Spatial Assessment of Langat River Water Quality Using Chemometrics. *Journal of Environmental Monitoring*.12: 287-295.

[3] Toriman, M. E., Md. Pauzi Abdullah, Mazlin Bin Mokhtar, Muhamad Barzani Gasim, Othman Karim. 2010. Surface Erosion and Sediment Yields Assessment from Small Ungauged Catchment of Sungai Anak Bangi, Selangor. *Malaysian Journal of Analytical Sciences*.14 (1): 12-23.

[4] Toriman, M. E., Joy Jacqueline Pereira, Muhamad Barzani Gasim, Sharifah Mastura, S. A., and Nor Azlina Abdul Aziz. 2009. Issues of Climate Change and Water Resources in Peninsular Malaysia: The Case of Northern Kedah. *The Arab World Geographer*. 12(1-2): 87-94.

[5] Toriman, M. E. 2008. Developing a Conceptual Model of Geomorphic Change for Channel Management Proposes. A case of Langat River Basin. *Journal e-BANGI*. 3(3): 1-12.

[6] Toriman, M. E., Zulkifli Yusop, Mazlin Mokhtar & Hafizan Juahir. 2006. Application of GIS for Detecting Changes of Sungai Langat channel. *Malaysian Journal of Civil Engineering*. 18(1): 58-70.

[7] Mazlin, M., Mohd Ekhwan Toriman & Md. Abdullah Abraham Hossain. 2010. Social Learning in Facing Challenges of Sustainable Development: A Case of Langat River Basin, Malaysia. *Research Journal of Applied Sciences*. 5(6): 434-443

[8] Gasim, M. B., Mazen, M. Jamil, Sahibin, A. Rahim & M. E. Toriman. 2009. Water-Quality Assessment o the Langat River at Kilometre 7, Jalan Kajang-Bangi, Selangor, Malaysia. *The Arab World Geographer*. 12(3-4): 188-196.

[9] Pielke, R., G. Dalu, J. Snook, T. Lee, and T. Kittel. 1991. Nonlinear Influence of Mesoscale Land-Use on Weather And Climate. *J.Climate*. 4: 1053-1069.

[10] Moore, D. S., & McCabe, G. P. 1989. *Introduction to the Practice of Statistics*. WH Freeman/Times Books/Henry Holt & Co. Altman, D .G. 1991. Practical Statistics for Medical Research. Chapman & Hall, London. 285-288.

[11] Imrie, C. E., Durucan, S. and Korea, A. 2000. River Flow Prediction by Using Artificial Neural Networks: Generalisation Beyond Calibration Range. *J.Hydrol*. 233: 138-153.

[12] Floyd, F. J., & Widaman, K. F. 1995. Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*. 7(3): 286-299.

[13] Gorsuch, R. L. 1990. Common Factor-Analysis Versus Component Analysis-Some Well and Little Known Facts. *Multivariate Behavioral Research*. 25(1): 33-39.

[14] Thompson, B., Daniel, L. G. 1996. Factor Analytic Evidence for the Construct Validity of Scores: A Historical Overview and Some Guidelines. *Educational and Psychological Measurement*. 56(2): 197-208.

[15] Brett William, Ted Brown, and Andry Onsman. 2012. Exploratory Factor Analysis: A Five-Step Guide for Novices. *Australasian Journal of Paramedicine*. 8(3).

[16] Jackson, J. E. 1991. *A User's Guide to Principal Components*. Wiley, New York.

[17] Sarbu, C., and Pop, H. F., 2005. Principal Component Analysis Versus Fuzzy Principal Component Analysis A Case Study: The Quality Of Danube Water (1985–1996). *Journal of Talanta*. 65: 1215-1220.

[18] Raman H., and Sunilkumar, N. 1995. Multivariate Modeling of Water Resources Time Series Using Artificial Neural Network. *Journal of Hydrological Sciences*. 40: 145-163.

[19] Yegnanarayana. 1994. Artificial Neural Networks for Pattern Recognition. *Journal of Scidhanci.* 19: 189-238.

[20] Igor Trubin. 2008. Exception Based Modeling and Forecasting. *Proceedings of the Computer Measurement Group.*