Jurnal Teknologi

Full paper

Flood Risk Pattern Recognition Using Integrated Chemometric Method and Artificial Neural Network: A Case Study in the Johor River Basin

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

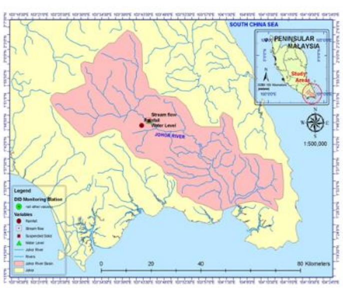
[a]*East Coast Environmental Research Institute, University Sultan Zainal Abidin, Gong Badak Campus, 21300, Kuala Terengganu, Terengganu, Malaysia*
[b]*Faculty of Science and Technology, University Science Islam Malaysia, 71900, Nilai, Negeri Sembilan, Malaysia*
[c]*Faculty Science and Technology, Open University Malaysia, 40100 Shah Alam, Selangor, Malaysia*

*Corresponding author: hafizanjuahir@unisza.edu.my

Article history

Received :9 December 2014
Received in revised form :
27 January 2015
Accepted :15 March 2015

Graphical abstract



Abstract

Flood is a major problem in Johor river basin, which normally happened during monsoon season. However in this study, it shows that rainfall did not have a strong relationship for the changes of water level compared to suspended solid and stream flow, where both variables have p-values of <0.0001 and these variables also became the main factors in contributing to the flood occurrence based on Factor Analysis result. Time Series Analysis was being carried out and based on Statistical Process Control, the limitation has been set up for mitigation in controlling flood. All data beyond the Upper Control Limit was predicted to have High Risk to face flood and Emergency Response Plan should be implemented to prevent complication and destruction because of flood. The prediction for the risk level was carried out using the application of Artificial Neural Network (ANN), where the accuracy of prediction was very high, with the result of 96% for the level of accuracy in the prediction of risk class.

Keywords: Flood; monsoon; factor analysis; time series analysis; statistical process control; artificial neural network

Abstrak

Banjir merupakan masalah utama di lembangan sungai Johor, yang biasanya berlaku pada musim tengkujuh. Namun dalam kajian ini, ia menunjukkan bahawa hujan langsung tidak mempunyai hubungan yang kuat dengan perubahan paras air berbanding pepejal terampai dan aliran sungai, di mana kedua-dua pembolehubah mempunyai nilai p <0.0001 dan pembolehubah-pembolehubah ini juga menjadi faktor utama dalam menyumbang kepada kejadian banjir berdasarkan dapatan Analisis Faktor. Analisis Siri Masa telah dijalankan dan berdasarkan kepada Proses Kawalan Statistik, had pengawalan banjir telah ditetapkan. Semua data di luar Had Kawalan Atas yang telah diramalkan akan mempunyai risiko tinggi dalam menghadapi banjir dan pelan tindakan kecemasan perlu dilaksanakan untuk mengelakkan komplikasi dan kemusnahan harta benda. Ramalan untuk tahap risiko telah dijalankan menggunakan aplikasi Rangkaian Neural Buatan (RNB), di mana ketepatan ramalan adalah sangat tinggi, dengan hasil sebanyak 96% untuk tahap ketepatan bagi ramalan kelas risiko.

Kata kunci: Banjir; monsun; analisis faktor; anaisis siri masa; proses kawalan statistik; rangkaian proses buatan

1.0 INTRODUCTION

Johor river basin is the major contributor as drinking water source and domestic use, not just for the state of Johor, but also for Singapore. This river basin actually occupies 14 % of Johor and as Johor is situated closely to Singapore, the link of sharing water source, e.g. the pipeline, has been built in fulfilling the purpose. Syarikat Bekalan Air Johor and Public Utility Board of Singapore pump out about 0.25 x 106 m^3 of water from this river basin, especially in Kota Tinggi. The supply scheme from this source has been set up since mid 1960s . Besides, there are other rivers in tributaries of Johor river basin also contribute in the supply scheme within small scale, and those rivers are Sayong, Semangor and Lebam rivers. Since 1993, Lingui dam construction has been completed and also becomes a part of water supply scheme for Johor and Singapore[1].The flood occurrence in this river basin will give high implication towards the water supply, destruction of infrastructure and communities within the radius of this river basin. In Johor, flood is most influenced by North East monsoon. The changes of land use in the particular area because of deforestation for palm plantation affect the rate of surface runoff into the river that may lead to high sedimentation. This will cause the changes of water level at the Johor river basin This condition may lead to high complication towards citizen along the river basin if the flood occurrence is not being warned to them in the early stage. Late emergency response plan may cause destruction towards buildings and facilities, where the worst condition for this complication is fatality. The aim of this study is to clarify surface run-off to the river as the main reason for flood occurrence in the study area during monsoon season. The finding in this study would assist in determining the limitation of flood risk based on hydrological data from 1982 until 2012 and to identify suitable mitigation measures for flood prevention based on prediction in the study area.

2.0 EXPERIMENTAL

Study Location

Johor river basin is situated in southern part of Peninsular Malaysia with the starting point from Mount Gemuruh and this basin totally covers an area of 2,700 km^2. The river flows into Straits of Johor and the total length of the river basin is approximately 122.7 km. By referring to Figure 1 and Table 1, the basin point begin from 1^o 27'N to 1^o 51'N latitude and from 103°42' E to 104°01' E longitude. It is being covered mostly by rubber plantation, oil palm plantation and virgin forest.

There are 4 monitoring stations being selected along the river basin. The variables being used in this study were Rainfall, Suspended Solid, Stream flow and Water Level. Data has been received from Department of Drainage and Irrigation from 1982 until 2012. All data being analyzed by using XLSTAT and JMP 10 software for integrated chemometric method, time series analysis and Artificial Neural Network.

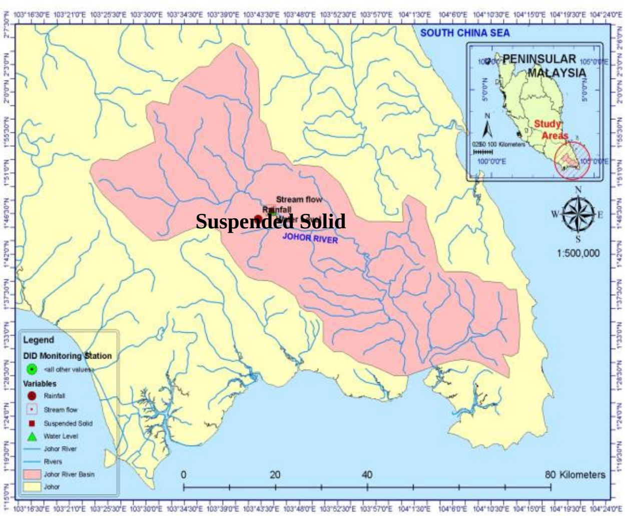


Figure 1 Location of monitoring stations in Johor river basin

Table 1 Location of monitoring stations in Johor river basin

Station No.	Latitude	Longitude	Name of Station	Variables
Site 1737001	1º45'50 N	103° 43"10'E	SMK Bukit Besar, Kota Tinggi, Johor	Rainfall
Site 1737551	1º46'50 N	103° 44"45'E	Sungai Johor at Rantau Panjang, Johor	Suspended Solid
Site 1737451	1º46'50 N	103° 44"45'E	Sungai Johor at Rantau Panjang, Johor	Streamflow
Site 1737451	1º46'50 N	103° 44"45'E	Sungai Johor at Rantau Panjang, Johor	Water Level

Correlation Test

In this study, Correlation Test was utilized to show which variables have strong relationship for further analysis. This method best suits the study as it measures two variables whose relationship ranges from -1 to 1. There are two types of coefficients that can be used in this method; Pearson and Spearman coefficients. However, Pearson Coefficient is widely used when there is an association of two variables[2]. In this study, the test was used to measure the relationship among important parameters in hydrological data. Correlation Test was applied in this study to ensure the relationship among the entire parameters and to determine the ones that shows the strongest relationship. After that, we are able to point out which development has the biggest impacts on the hydrological modelling at Johor river basin.

Spearman's rank correlation is able to measure the strength of the coefficient between variables, which was carried out in the study for further analysis[3]. Based on the Spearman's rank correlation, there are two types of correlation, namely positive and negative correlations. The positive correlation shows two variables increasing together in a linear condition, whereas the negative correlation show one variable increasing while the other decreasing in a linear condition. In this study, this tests have been carried out, and the best result was used for the discussion in this study. The equation used in this method was:

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} \quad \text{[1]}$$

r_s = the correlation coefficient

d = the difference between the rank

n = the total number of observation

Chemometric Techniques

Chemometric technique, such as the application of Factor Analysis, is able to observe the reduction of variables into a set of factors for further analysis. Based on Floyd and Widamann[4], researcher rarely collect and analyze data with a priori idea about how the variables are related and the application of these methods are able to make comparison which variables that affect the most towards the change of the hydrological modelling at the study area with a cheap cost and quicker compared to other techniques[5].

Factor Analysis

Factor Analysis is a technique where the moldings of new variables after following a few processes and these new variables have a linear combination of the original variables. FA is used to reduce the dimensionality of the data set by explaining the correlation among a large set of variables in terms of a small number of underlying factors or principal components without losing much information[6].

This method was applied in this study as it is able to define large number of variables into smaller set of variables, which are also known as factors. The measured factor analysis variables and latent construct will establish the dimension between these two elements and construct validity evidence of self-reporting scales[7]. It also reduces the number of variables, examine the structure or relationship between variables, and detection and assessment of unidimensionality of theoretical construct[8]. This method also addresses multicollinearity (two or more variables that are correlated), which is efficient to be carried out in this study. The equation used in this method was:

$$Z_{ij}=a_{i1}x_{j1}+a_{i2}x_{j2}+a_{i3}x_{j3}+a_{im}x_{mj} \quad [2]$$

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

Control Chat Builder Six Sigma (Time Series Analysis)

Time Series Analysis is very important in predicting water level at the study area. By applying this method, we are able to evaluate the process from the performance of the analyzed data efficiently. It produces three important results that are important to study the trend and also for the prediction of hydrological modelling in the future, and those results are Upper Control Limit (UCL), Average Value (AVG) and Lower Control Limit (LCL), where the Sigma is within the range value of a set of data. Control Chart has the ability to uncover some trends and patterns showing actual data deviations from historical baseline, dynamic threshold, able to capture unusual resource usage and able to become the best base lining to show how actual data deviate from historical baseline[9]. The equation used in this analysis is:

Moving Range (MR) = Plot: MR_t for t=2, 3, ..., m.

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

$$\tilde{x} = \frac{\sum_{i=1}^{m} x_i}{m} \quad [3]$$

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

Artificial Neural Network

The concept of human brain has been utilized in Artificial Intelligent and it was applied as a method in data analysis, known as Artificial Neural Network. This concept was created by McCulloch and Pitts in 1943 by simulating the structure and performance of biological neural network in the computing system[10].

In this study, the Artificial Neural Network was being computed by using JMP10 software. The Artificial Neural Network (ANN) contains the technique of back propagation where for this technique, it does requires the training algorithm for multi-layer feed forward network. The "neurons" or nodes which is also being considered as an interconnected layer in the system will then be formulated and being arranged in three layers and those layers can be categorized as the input layer, hidden layer and output layer This technique able to minimize the error function and the iteration is terminated when the error function value reached predefined goal, thus completing the process[11,12].

Performance data can be calculated by implementing cross validation to a testing data set, where the algorithm needs to be terminated during the process and this process is performed by using back propagation[13,14]. Learning ability of ANN depends on the architecture of network and number of hidden unit. The size of the network plays a major role in capturing connectivity of the data, when the function of degree of freedom is to capture the connection and the size of network must be sufficient with the degree of freedom or the process will fail.

The input layer being taken as the first section of the model as in this part it was located all of the input data and the last part of this model was an output layer war in this part the output data will also be extracted. In this study, the input layer consists of the variables which been selected based on the factors which contribute to flood occurrence in the study area. Based on this condition, the input layer will become a neuron and the output layer will be the dependent variables to be predicted (factor for flood occurrence). The layer which will be between these layers is hidden layer with the number of neurons will be based on the number of inputs (Ni) and output nodes (No)[15].

There are three groups for the data which need to be classified and those groups are trained, testing and controlling sub-set[16]. The training subset will including the whole set of data were in this part, the system will use the data for recognizing and learn the pattern of every selected variable which been used in this analysis. After that, the part of test subset will be triggered and evaluation ability of the trained data will be carried out in the test subset. The validating subset is functioning in handling the final evaluation of the training set.

3.0 RESULTS AND DISCUSSION

In this topic, the relationship between variables is clearly being shows and explain through Spearman Correlation test. The results in Table 2 explain the rate of correlation between variables studied based on p-value. In Correlation Test, if the p-value was less than 0.05, the relationship between variables was considered to be strong and significant. The condition of relationship was considered to be strong when is the value was less than 0.0517. The results in Table 2 show that the correlation of Water Level with Suspended Solid and Stream Flow was very strong and significant, based on the result of p-value <0.0001. Meanwhile, the rate of correlation between Rainfall and Water Level was less significant and showed a weak correlation with p-value of 0.211. This demonstrated that changes in Rainfall did not truly involve the rate of Water Level at Johor river basin compared to Suspend Solid and Stream Flow, which caused the changes of Water Level at the study area.

Table 2 Correlation test for selected variables based on p-values

Variables	Rainfall	Stream Flow	Suspended Solid	Water Level
Rainfall	0.0000	0.2250	0.9350	0.2110
Stream Flow	0.0250	0.0000	<0.0001	<0.0001
Suspended Solid	0.9350	<0.0001	0.0000	<0.0001
Water Level	0.2110	<0.0001	<0.0001	0.0000

Factors Which Contribute to Flood Occurrence

Factor Analysis

Flood occurrence in the study area actually occurred due to certain factors that caused changes in Water Level until beyond control limit. Factor Analysis was performed in the study to determine the main factor that caused flood occurrence at Johor river basin, and based on the results in Table 3 and Figure 2 it shows that in F1 Suspended Solid, Stream Flow and Rainfall were in the same factor with strong coefficient, with the result of 0.699 for Stream Flow, 0.980 for Suspended Solid and 0.738 for Water Level.

This shows that these elements belong in the same factor for flood occurrence compared to Rainfall, which belongs in F2 with weak coefficient of 0.147. Based on the research Saudi *et al*.[18], they explained that the result for factors beyond 0.7 was considered for interpretation in Factor Analysis. From this research, it clearly explained that the factor loading in F1 became the main factor loading for flood occurrence and clearly showed that Rainfall was not the major reason for changes of Water Level and contributed to flood occurrence.

Table 3 Factor Analysis result

	F1	F2	Initial Communality	Final Communality	Specific Variance
Rainfall	0.006	**0.147**	0.007	0.022	0.978
Stream Flow	**0.699**	-0.314	0.426	0.587	0.413
Suspended Solid	**0.980**	0.116	0.665	0.975	0.025
Water Level	**0.738**	0.142	0.548	0.565	0.435

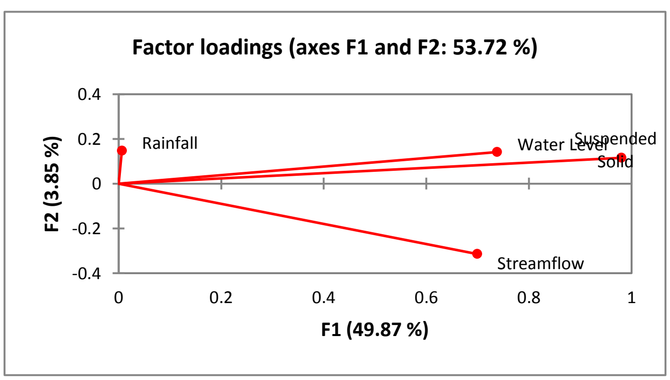


Figure 2 Result for correlation coeeficient of variables

Flood Control Warning System

Time Series Analysis

Based on the result in previous analysis, it was clear on the assumption of which variables became the main factor for flood occurrence at the study area. Correlation Test and Factor Analysis clearly stressed that Suspended Solid, Water Level and Stream Flow became the major contributors and have a strong relationship for flood occurrence. Time Series Analysis in this study practiced Control Chart Builder Six Sigma, which carried out Statistical Process Control with Individual Chart and Moving Range Type.

This analysis was set up to control the limit based on Time Series Analysis of variables that became the major contributor for flood occurrence at Johor river basin. There were three levels of control limit in this analysis, and those control limits were Upper Control Limit, Average Limit and Lower Control Limit. The analysis was conducted using JMP10 software.

Based on Figure 3 and Table 4 show the control limit for variables Stream Flow had a strong relationship and was one of the main factors for flood occurrence at Johor river basin. The result showed that the Upper Control Limit for Stream Flow that was considered as a level that may lead to flood was 68.559 m^3/s. Meanwhile, the Average Limit value considered as normal condition of Stream flow at Johor river basin was 43.439 m^3/s, and the Lower Control Limit for Stream Flow was 18.317 m^3/s.

The moving range for Stream Flow Upper Control Limit was 30.864 m^3/s, where within this range, the changes were considered beyond normal condition for the river basin. The Average Limit value for moving range of Stream Flow was 9.445 m^3/s, and the moving range within this limit was considered as normal condition at the study area. Finally, the Lower Control Limit for moving range of Stream Flow was 0 m^3/s. This condition considered that there was no possibility of flood occurrence at Johor river basin.

Table 4 Result for time series analysis based on stream flow at the study area

Points Plotted	Lower Control Limit (m^3/s)	Average (m^3/s)	Upper Control Limit (m^3/s)	Limit Sigma	Sample Size
Individual	18.317	43.439	68.559	Moving Range	1
Moving Range	0	9.445	30.864	Moving Range	1

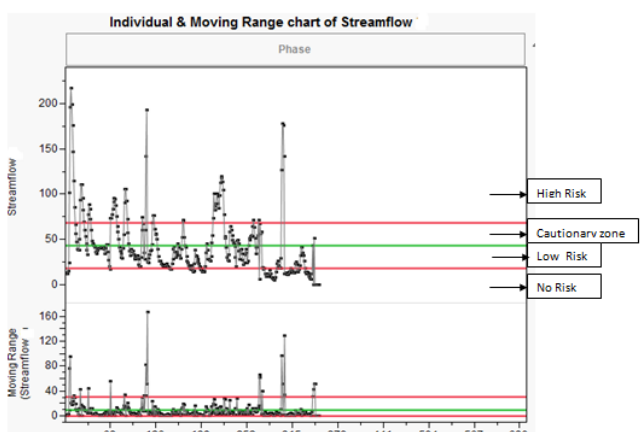


Figure 3 Result for time series based on stream flow in the study area

Time Series Analysis was also performed for Suspended Solid variable. which is also one of the main factors and had strong relationship in contributing to flood occurrence at the study area. Based on Figure 4 and Table 5 the Upper Control Limit for Suspended Solid at the study area was 285.157 sediment tons/day, where this value was considered beyond normal condition and there was high possibility in contributing flood occurrence at the study area.

The Average Limit value for Suspended Solid at Johor river basin was 43.439 sediment tons/day. This level and below was considered to be in normal condition and safe from flood occurrence. Meanwhile, the Lower Control Limit for Suspended Solid at the study area was 18.317 sediment tons/day. At this level and below, it was considered as out of range for any kind of possibilities for flood occurrence at the study area.

The Upper Control Limit for moving range of Suspended Solid was 155.858 sediment tons/day. This moving range was considered as high potential, which may lead to flood occurrence. The Average Limit for moving range of Suspended Solid under normal condition was 47.714 sediment tons/day, whereas the Lower Control Limit for moving range of Suspended Solid at the study area was 0 sediment tons/day. At this moving range, the condition of river basin showed the lowest possibility for flood occurrence.

Table 5 Result for suspended solid based on control chat builder

Points Plotted	Lower Control Limit (Sedimen tons/day)	Average (Sedimen tons/day)	Upper Control Limit (Sedimen tons/day)	Limit Sigma	Sample Size
Individual	18.317	43.439	68.559	Moving Range	1
Moving Range	0	9.445	30.864	Moving Range	1

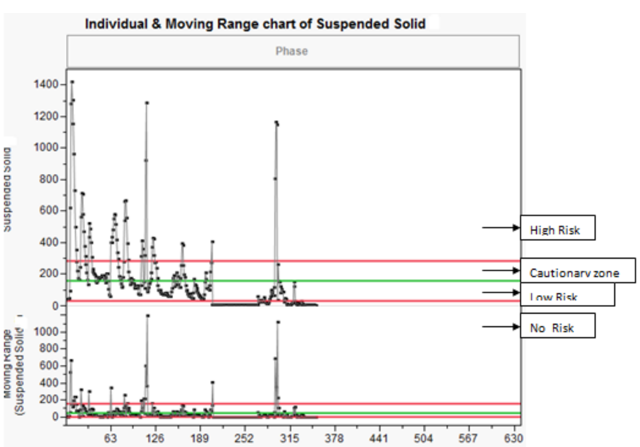


Figure 4 Result for time series analysis based on suspended solid at the study area

Water Level was also one of the variables that become one of the main factors and had strong relationship for flood occurrence based on the previous analysis result. Figure 5 and Table 6 below explain the result for Time Series Analysis for Water Level and based on the result, the Upper Control Limit for Water Level was 5.376 m. The Water Level at this limit and above can easily cause flood occurrence as it was beyond the normal condition of the Johor river basin. The normal condition for Water Level at the study area was within the Average Limit and the value for the limit was 4.502 m. The Lower Control Limit for Water Level was 3.629 m.

The Upper Control Limit of moving range for Water Level was 1.073 m and the moving range of Water Level at this point and above may easily lead to flood occurrence at the study area. The normal condition for moving range of Water Level was 0.328 m, which was within the Average Limit in Control Chart. The Lower Control Limit for moving range of Water Level at Johor river basin was 0 m and within this range, it was considered to be safe from flood occurrence in the study area.

Table 6 Result for time series an based on control chat builder

Points Plotted	Lower Control Limit (m)	Average Limit (m)	Upper Control Limit (m)	Limit Sigma	Sample Size
Individual	3.629	4.502	5.376	Moving Range	1
Moving Range	0	0.328	1.073	Moving Range	1

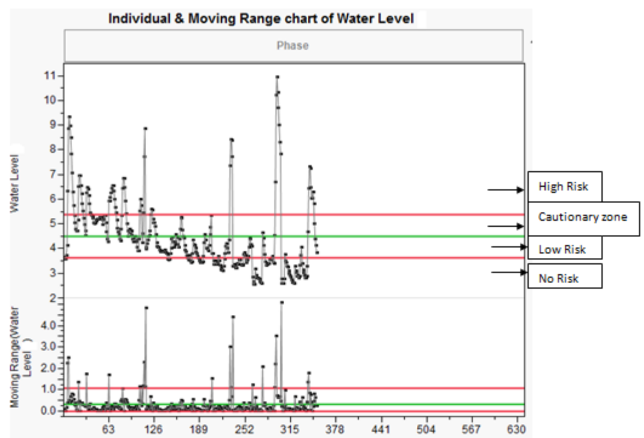


Figure 5 Result for time series analysis based on stream flow in the study area.

Prediction of Flood Risk Classification

In order to have a good mitigating measure for flood control and proper response plan before, during and after flood occurrence, we need to have a good prediction in the class of risk for flood occurrence in the study area. This study utilized Artificial Neural Network in predicting the risk class for flood occurrence, and those classes were categorized as High Risk, Cautionary Zone, Low Risk and No Risk.

The risks were classified based on the points plotted in the Control Chart graphs (Figure 3, 4, and 5 in Time Series Analysis. It is predicted that for all points plotted above the Upper Control Limit, it is classified as High Risk class with high possibilities for flood occurrence. Cautionary Zone class is classified for all points plotted between the Average Limit line and Upper Control Limit line. This class is considered for the Local Authority to prepare to take action for any possibilities of flood occurrence before the water level rises up to High Risk class.

The Low Risk class is all the data plotted between Average Limit line and Lower Control Limit line, and this is being considered as normal condition for Johor river basin, while No Risk class is classified for the data below the Lower Control Limit line and at this condition, the river basin is considered to be safe from floods**.**

Based on the result in Table 7 the accuracy for risk class prediction was 0.96, which equals to 96%.

Table 7 Accuracy of risk class prediction by using ANN

OUT_1	Accuracy	Rel. Entropy
Train	1	0.016128
Test	0.962264	0.024217

■4.0 CONCLUSION

Nowadays, flood occurrence in the study area is not much affected by rainfall even in the monsoon season, but it is more affected by the imbalance development along the river basin, which causes uncontrolled surface runoff that can trigger flood at the study area. The application of integrated Chemometric, Time Series Analysis and Artificial Neural Network makes the main factors of flood which are suspended solid and water level much clearer to be identified. Warning system can also be implemented more efficiently and more cost effective by using the limitation system and classification of risk based on the changes of variables that become the main factor for changes of water level that will able to cause flood in Johor river basin.

The Local Authority must be able to handle the complication of flood occurrence. It must be handled with a few 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 by Department of Drainage and Irrigation (DID), applying the regulation of Environmental Quality Act 1974 (Act 127) & Subsidiary Legislation and Waters Act 1920 (Act 418) in order to control human development along the river bank. Based on this condition, the developer must also take serious action in handling their construction activity along the river bank by reducing the rate of surface run-off before, during and after the construction finish by implementing suitable plan and following rules and regulations which have been set up in protecting the environment, thus able to reduce the risk of flood occurrence at the study area.

Besides that, other mitigating measures such as construction of barrage, river bund, pump house, diversion, pond, dam and river improvement work 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 surface run-off, which comes from uncontrolled human development, and if it is not being configured well, all the structure for mitigating measure means nothing in preventing flood occurrences.

Acknowledgement

I am grateful to the Ministry of Higher education for scholarship through My Phd Scholarship for this research where I am completely identified source and formulation in preventing flood occurrence at the study area. I would like also to thanks my supervisor Associate Professor Hafizan Juahir for advising me until this research completely done.

References

[1] Diane, S. 2004. Singapore's Water Trade with Malaysia and Alternatives. Harvard University, Accessed at http://www.transboundarywaters.orst.edu/publications/abst_docs/related_research/Segal-Singapore-Malaysia%2004.pdf, January 2015.

[2] Moore, D. S. and McCabe, G. P. 1989. *Introduction to the Practice of Statistics*. W. H. Freeman, New York.

[3] Altman, D. G. 1991. *Practical Statistics for Medical Research*. Chapman & Hall, London. 285–288.

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

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

[6] Meglen, R. R. 1992. Examining Large Databases: A Chemo-metric Approach Using Principal Component Analysis. *Mar. Chem.* (39): 217–237.

[7] 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.

[8] William, B., Brown, T. and Onsman, A. 2012. Exploratory Factor Analysis: A Five-step Guide for Novices. *Australasian Journal of Paramedicine.* 8(3): 1.

[9] Yegnanarayana. 1994. *Artificial Neural Networks for Pattern Recognition*. Scidhanci. 19(2): 189–238.

[10] Trubin, I.A. 2008. Exception Based Modelling and Forecasting. *Proceedings of the Computer Measurement Group, Nevada.* 7–12 December 2008. 353–364.

[11] Juahir, H., Sharifuddin, M. Z., Mohd K. Y., Tengku, H. A. S., Mohd A., Mohd E. T., Mazlin, M. 2010. Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. *Environ Monit Assess*. Doi: 10.1007/s10661-010-1411-x.

[12] 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.

[13] Azid, A., Juahir, H., Latif, Toriman, M. E., Kamarudin, M. K. A., Saudi, A. S. M. 2014. Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: A Case Study in Malaysia. *Water, Air & Soil Pollution*. 225: 2063.

[14] Goethals, P. L. M., Dedecker, A. P., Gabriels, W., Lek, S. and De Pauw, N. 2007. Applications of Artificial Neural Networks Predicting Macroinvertebrates in Freshwaters. *Aquatic Ecology*. 41(3): 491–508.

[15] Rech, G. 2002. *Forecasting with Artificial Neural Network Models*. SSE/EFI Working Paper Series in Economics and Finance 491, Stockholm School of Economics

[16] Azid, A., Juahir, H., Latif, M. T., Zain, S. M., Osman, M. R. 2013. Feed-forward Artificial Neural Network Model for Air Pollutant Index Prediction in the Southern Region of Peninsular Malaysia. *Journal of Environmental Protection*. 4(12A).

[17] Gelman, A. 2013. Two Simple Examples for Understanding Posterior P-Values Whose Distributions are Far from Unform. *Electronic Journal of Statistics.* 7: 2595–2602. ISSN: 1935-7524. DOI: 10.1214/13-EJS854.

[18] Saudi, A. S. M., Juahir, H., Azid, A., Yusof, K. M. K. K., Zainuddin, S. F. M., Osman, M. R. 2014. Spatial Assessment Of Water Quality Due To Land-Use Changes Along Kuantan River Basin. *From Sources to Solution*. 2014: 297–300.