Jurnal Teknologi

Full paper

Flood Risk Pattern Recognition Using Chemometric Technique: A Case Study In Kuantan River Basin

Ahmad Shakir Mohd Saudi,[a,b] Hafizan Juahir,[a] Azman Azid,[a*] Mohd Khairul Amri Kamarudin,[a] Mohd Fadhil Kasim,[a] Mohd Ekhwan Toriman,[a] Nor Azlina Abdul Aziz,[a] Che Noraini Che Hasnam,[a] Mohd Saiful Samsudin.[c]

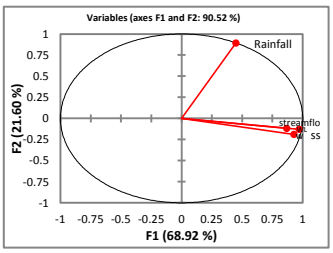
[a]*East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300, Kuala Terengganu Terengganu, Malaysia*
[b]*Faculty Science And Technology, Open University Malaysia, 40100 Shah Alam, Selangor, Malaysia*.
[c]*Environmental Forensics Research Centre (ENFORCE), Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor Malaysia*

*Corresponding author: azmanazid@unisza.edu.my

Article history

Received: 22 May 2014
Received in revised form:
28 October 2014
Accepted: 1 December 2014

Graphical abstract



Abstract

Integrated Chemometric and Artificial Neural Network were being applied in this study to identify the main contributor for flood, predicting hydrological modelling and risk of flood occurrence at the Kuantan river basin. Based on the Correlation Test analysis, the relationship for Suspended Solid and Stream Flow with Water Level were very high with Pearson correlation of coefficient value more than 0.5. Factor Analysis had been carried out and based on the result, variables such as Stream Flow, Suspended Solid and Water Level turned out to be the major factors and had a strong factor pattern with the results of factor score with >0.7 respectively. Time series analysis was being employed and the limitation had been set up where the Upper Control Limit for Stream Flow, Suspended Solid and Water Level where at this level, it was predicted by using Artificial Neural Network (ANN) to be High Risk Class. The accuracy of prediction from this method stood at 97.8%.

Keywords: Integrated chemometric, artificial neural network, factor analysis, time series analysis.

© 2014 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Kuantan is being ranked as the ninth out of tenth largest cities in Malaysia. Since 2005, petrochemical industries, which has become one of the main industries in the city, alongside with the rapid and progressive development of other various industrial sectors, tourism and transportation has led to drastic urbanisation and modernization of the city. Based on the National Physical Plan 2005, this city has been considered as one of the commercial hub for East Coast of Peninsular Malaysia in regard to its strategic and pivotal location.

This study was conducted to measure the impact caused by human development affecting hydrological modelling at Kuantan river basin. Kuantan has been undergoing rapid development and thus affecting the water quality as the area is being developed with miscellaneous kinds of development.[1]Unsustainable development generates destruction to hydrological stability in the location. As the human activities are being closely related to huge impact towards environment, they undeniably cause abundant changes to the assemblages and biodiversity of the river fauna.[2-4] The industrial and domestic wastes have become major reason in contributing heterogeneous types of diseases which are being spread from polluted water.[5] Kuantan river basin's length is about 86 km and it provides the water supply to 607,778 populace of Kuantan. Greater pollution is deemed to intimidate along the river basin as the unsustainable development is widely spread and thus will eventually contribute manifold kind of pollutants flow into the water body system. The transformation of particular types of land use such as agricultural and forest areas into industrial or municipal areas will change the types of pollutant loadings into the river system.[6] Monsoon seasons as well as the swift physical development will lead to the change of hydrological modelling at the study area.

Chemometric analysis is one of the best methods by which the method itself is easy, uncomplicated and able to produce significant results.[7] This technique is capable to classify which variables will become the main contributor for the changing of other variables and it is regarded as helpful in decision making and problem solving in the local environmental issues.[8] The techniques are widely adopted and are being chosen for multi-criteria decision making which are very efficacious in determining the main contributor for the changes of variable from other variables and classifying variables into its own class.[9] This study utilised hydrological data obtained from the Department of

Drainage and Irrigation (DID) from the year of 1982 until 2012. The statistical analysis methods such as correlation test are able to identify relationship among variables while in factor analysis, it is apt to identify which variable is the main contributor for the changing of other variables.

Rapid physical development in the study area inflicts the negative impacts towards the rate of the surface runoff flow into the water body system. This will be affected water level at certain location in the river basin and subsequently may lead to flood. The aim of this study is to clarify surface run off to the river as the main reason for flood occurrence at 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 the year of 1982 until 2012 and to identify suitable mitigation measures for flood prevention based on prediction at the study area.

2.0 EXPERIMENTAL STUDY AREA

Kuantan river basin is located in the north-eastern of Pahang, spreading across the capital city of Pahang. Geographically, the basin is located at the coordinates of 30 12' 27.66" N and 1030 07' 39.99" E, covering the water supply for various activities in the capital of Pahang, with the population of 607,778 dwellers. The total length of the basin is about 86 km and the total area is 1638 sq. km. The Kuantan town is an urbanized area, which is situated close to the South China Sea. The basin originates from Gunung Tapis and it flows in an easterly direction through Sungai Lembing to Kuantan town before discharging into the South East China Sea. In accordance with the federal authority, all stations have been monitored by the Department of Environment, Malaysia (DOE). Table 1 and Fig. 1 portray the study area of this study. This river basin receives annual rainfall of 2470 mm and the northeast monsoon wind from November until March. The surrounding of Kuantan River is dominated by agriculture (32.05%), housing areas (2.99%), road and transportation (2.01%), industry (1.97%), institutions (0.95%), recreational areas (0.75%), infrastructure (0.57%), and forest (54.71%).

Figure 1 Location of monitoring stations at Kuantan river basin

Table 1 Location of monitoring stations Kuantan river basin

Station No.	Latitude	Longitude	Name of Station	Variables
Site 3833004	3°53'40 N	103° 23"00'E	Ladang Jeram, Kuantan	Rainfall
Site 3930501	3°55'55 N	103° 03"30'E	Sungai Kuantan at Bukit Kenau, Kuantan	Suspended Solid
Site 3930401	3°55'55 N	103° 03"30'E	Sungai Kuantan at Bukit Kenau, Kuantan	Streamflow
Site 3930401	5°18'32 N	103° 03"30'E	Sungai Kuantan at Bukit Kenau, Kuantan	Water Level

Statistical Analysis/ Preprocessing Data Correlation test

In this study, the application of Correlation Test was used to show which variables that have strong relationship for further analysis. This method is categorized as a best method as it measures two variables whose relationship ranges from -1 to 1. There are two types of products which can be used in this method, which known as Pearson Coefficient and Spearman Coefficient. However, Pearson Coefficient is widely used when there is an association of two variables. In this study, the test is used to measure the relationship among important parameters in hydrological data. Correlation test is applied in this study to ensure the relationship among the entire parameters and to find out the ones which show the strongest relationship. Until then, we are able to point out which development has the biggest impacts on the hydrological modelling at Kuantan river basin.

There are two most common types of correlation test known as Spearman rank coefficient and Pearson rank coefficient. Spearman rank coefficient may need ordinal data where its calculation will be based on rank of data. Spearman's rank correlation is able to measure the strength of the coefficient between variables which carried out in the study for further analysis. Based on the Spearman 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 negative correlation shows one variable increasing while the other decreasing in a linear condition.

The Pearson rank coefficient may need actual data to be calculated and all variables which need to be tested must be in ratio scale. In this study, both tests have been carried out using XLSTAT 2014 software, and the best result is used for discussion in this study.
this study.

Chemometric Techniques

Chemometric technique such as application of Factor Analysis is able to see the reduction of variables into a set of factors for futher analysis. Based on Floyd and Widamann,[10] researcher rarely collect and analyze data with a priori idea about how the variables are related and application of these method able to make comparison which variables that effecting the most towards the

change of the hydrological modelling at the study area with a cheap cost and quicker compare to other technique.

The reduction of variables into a set of factors for further analysis can be observed using chemometric technique, such as the utilization of Factor Analysis. It is seldom that the researcher collects and analyzes data with prior knowledge regarding the relationship of the variables, but through this technique, variables with the biggest influence in the change of the hydrological modelling at the study area can be compared in a cost effective and quicker manner compared to other techniques.[2]

Factor Analysis

The utilization of this method in the study allowed the inclusion of large number of variables into smaller set of variables, otherwise known as factors. The dimension between factor analysis variables and the measured latent construct established the dimension between these two elements and construct validity evidence of self reporting scales.[1] Other than that, factor analysis also examines the structure or relationship between variables, reduces the number of variables, and can be used for the detection and assessment of unidimensionality of theoretical construct.[4] The method also considers the existence of two or more variables that are correlated (e.g. multicollinearity), which is suitable for this study. The equation implemented in this method was:

$$z_{ji} = b_{j1}F_{1i} + b_{j2}F_{2i} + \ldots + b_{jn}F_{ni} \quad [1]$$

The common-factor approach only considers the covariation between observed variables, whereas the principal-component approach considers all variations in the observed variables.

- Factor loadings (b') represents the correlation coefficient between each factor and the observed variables.
- Factor scores are the values of each observation on the factor (F_k).

Control Chat Builder of Six Sigma using Time Series Analysis

Time Series Analysis is essential for the prediction of water level at the study area, where this method enables an efficient evaluation of the process from the performance of analyzed data. The method produces three important data e.g. Upper Control Limit (UCL), Average Value (AVG) and Lower Control Limit (LCL) for the trend and prediction of future hydrological modeling, where the Sigma is within range value of a set of data. Control Chart can detect some trends and patterns with actual data deviations from historical baseline, able to capture unusual resource usage, can determine the dynamic threshold, and also can become the best base lining to examine the actual data deviation from the historical baseline. In this study, this analysis was performed using JMP10 software.

Artificial Neural Network

Artificial Intelligent mimics the concept of human brain and it has been utilized in the method for data analysis known as Artificial Neural Network. This concept was introduced by McCulloch and Pitts in 1943, where the stimulation of structure and the performance of biological neural network in the computing system have been investigated.

An activation function is utilized to transform the weighted sum of the inputs transferred to the hidden neurons. The back propagation method is also implemented in the learning process for the purpose of error distribution, where the process can reduce the errors to the minimum level. After the error function has been minimized, the iteration is terminated when the value of error function reached the predefined goal, thus completing the process.[11]

The process of cross validating the testing data set can be used to indicate the performance of the data, where the algorithm needs to be terminated during the process using back propagation. The architecture of the network and number of hidden units affects the learning ability of ANN. The size of the network is also important in capturing the connectivity of the data, as the degree of freedom works to capture the connection, and the size of the network must be compatible with the degree of freedom or the process will fail.

Imrie et al.[12] determined the effectiveness of ANN for rainfall–runoff modelling and flood forecasting, where the ability of ANN in predicting river flow and quality of water downstream has been highlighted. As a matter of fact, the aforementioned issues were also considered in this study. This analysis was performed using JMP10 Software.

3.0 RESULTS AND DISCUSSION

To identify whether monsoon or human development has become the major reason for flood occurrence, Pearson's correlation coefficient analysis was applied and the result summarized in Table 2.

Correlation analysis showed that suspended solid and water level has the highest correlation with the result of correlation coefficient is 0.909. While, the correlation coefficient for the rainfall and water level is 0.105 and it is considered as a weak correlation. According to Herman,[13] when the correlation coefficient with more than 0.5, it is considered as a strong correlation and suitable to be taken for further analysis. The other variable such as Stream Flow also shows almost strong correlation coefficient with Water Level with correlation coefficient value of 0.599. This exhibits that relationship between surface run off is strong with the changing of Water Level at Kuantan river basin.

Table 2 Correlation test result

	F1	Initial communality	Final communality	Specific Variance
Streamflow	**0.868**	1.000	0.753	0.247
Rainfall	0.448	1.000	0.201	0.799
Suspended Solid	**0.927**	1.000	0.860	0.140
Water Level	**0.971**	1.000	0.942	0.058

Factors which contribute to flood occurrence

Factor Analysis was being exerted to identify the major factors which contribute to flood occurrence at the study area. Based on Table 3, there are three major factors which contribute to flood occurrence and those factors constituting of Stream Flow, Suspended Solid and Water Level are positively loaded with the factor and the result of 0.868 for Stream Flow, 0.927 for Suspended Solid and 0.971 for Water Level. Based on Aitchison[14] he explains that for the result of elements which is more than 0.7 being considered as strong factor in contributing the changes of its condition.

Based on this analysis, it clearly explains that Stream Flow and Suspended Solid have become the major factors for the changes in Water Level which induces the flood occurrence in the work region.

Table 3 Factor Analysis result

Variables	*Water Level*	*Rainfall*	*Suspended Solid*	*Stream flow*
Water Level	1	0.105	0.909	0.599
Rainfall	0.105	1	0.065	0.079
Suspended Solid	0.909	0.065	1	0.458
Stream flow	0.599	0.079	1	1

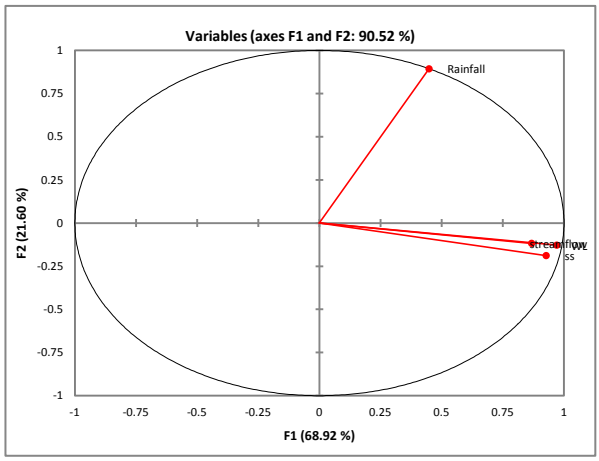


Figure 2 Result for correlation coeeficient of variables.

Flood control warning system.

Time series analysis had been carried out for further analysis in this study by applying Control Chart Builder. This method was being applied to determine the limitation for all variables which have become factors for flood occurrence at Kuantan river basin. Based on this method, the output from the analysis will be able to be used as control limit for flood control. There will be three phases for control limit and those phases are Upper Control Limit, Average Limit and Lower Control Limit.

Based on Figure 3 and Table 4, it explains that the Upper Control Limit for Water Level at Kuantan river basin comprises of 18.67m and 17.77m for Average Limit and 16.87m for Lower Control Limit. From this result, for the rate of Water Level above Upper Control Limit is considered as high risk to face flood occurrence and the most stable condition is the rate for Water Level within Average Limit and for Water Level below Lower Control Limit is being regarded to be no possibility for flood occurrence.

Table 4 Result for time series analysis based on water level at the study area.

Points Plotted	LCL	AVG	UCL	Limit	Sigma	Sample size
Individual	16.87 m	17.77m	18.67 m	Moving	Range	1

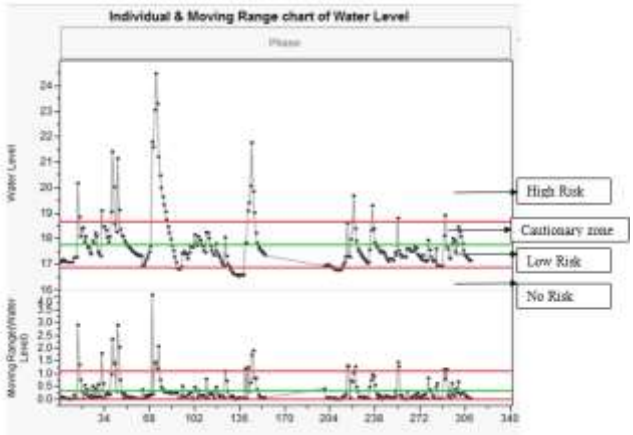


Figure 3 Result for time series based on water level at the study area.

Based on Table 5 and Figure 4, it illustrates the control limit for Suspended Solid at the study area. Result shows that the Upper Control Limit for Suspended Solid is 1455.478 sediment tones/day, 620.85 sediment tons/ day for Average Limit and 213.485 sediment tons/day for Lower Control Limit. The limit control for Suspended Solid explains the limitation for Suspended Solid discharge at Kuantan river basin. The rate which is beyond control limit for Suspended Solid will be effecting the changes of Water Level for flood occurrence and could trigger the flood occurrence at the study area.

Table 5 Result for suspended solid based on Control Chat Builder

Points Plotted	LCL	AVG	UCL	Limit	Sigma	Sample size
Individual	213.48 sed. tons/day	620.85 sed.tons/ day	1455.47sed. tones/ day	Moving	Range	1

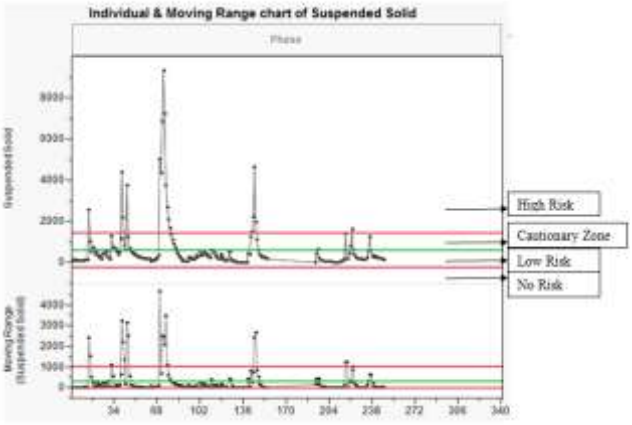


Figure 4 Result for time series analysis based on suspended solid at the study area.

Stream Flow becomes one of the variables which emerged as among the main factors for flood occurrence at the study area. Result in Figure 5 and Table 6 explains that the Upper Control Limit for Stream Flow at Kuantan river basin is 263.088m^3s, 108.44 for Average Limit and 45.399m^3/s for Lower Control Limit value. Stream Flow is one of the variables which turned out to be the factor for flood occurrence and all data which are located above Upper Control Limit is high risk for flood occurrence.

Table 6 Result for time series an based on Control Chat Builder

Points Plotted	LCL	AVG	UCL	Limit	Sigma	Sample size
Individual	45.39 9m^3/s	108.44 m^3/s	263.0 88m^3/s	Moving	Range	1

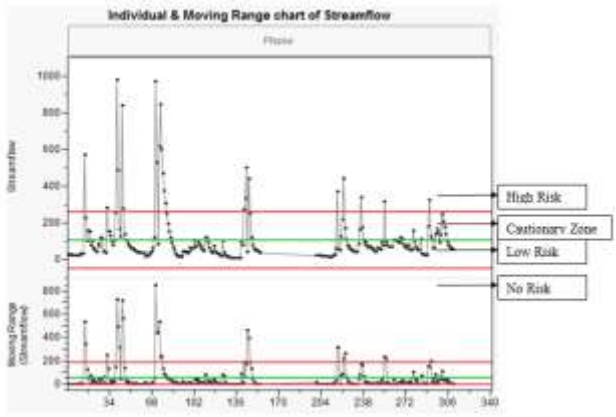


Figure 5 Result for time series analysis based on stream flow at the study area.

Risk classification is being categorized based on time series analysis in Figure 3, Figure 4 and Figure 5 and those classes are High Risk class, Cautionary Zone class, Low Risk Class and No Risk class. High Risk class is being categorized for all data which are plotted above Upper Control Limit line based on Control chart in time series analysis. Cautionary Zone class comprises of data which are being plotted between Average Limit line and Upper Control Limit line. Low Risk Zone class is being classified for all data which are being plotted between Lower Control Limit line and Average line and No Risk class represents all data which are being plotted below Lower Control Limit line.

Based on the risk class, if the condition of the river basin is within the High Risk class, the possibility for flood occurrence is high and immediate emergency response plan is severely needed which is able to reduce the cost of destruction and could save human's life if the area is flooded. The range for the river being considered as stable condition begins from Low Risk Class until No Risk class whereby within this range, the river basin is being categorized as safe from flood occurrence.

All risk classes are being predicted by using Artificial Neural Network and based on the results in Table 7, it explains that the rate of accuracy for prediction is 0.978 which is equal to 97.8% and it shows high accuracy of prediction that has been made for the risk class in this study.

Table 7 Prediction for hierarchy of flood risk

Out_1	Accuracy	Total
Train	1	108
Test	0.978	47

4.0 CONCLUSION

In conclusion, based on the analysis which has been carried out in this study, clearly even in monsoon season rainfall is not become the major factor for flood occurrence based on Factor Analysis and its correlation with changes of water level also not strong at the study area. Based on the method which being applied in this study, local authority also able to take earlier action for flood prevention and emergency response plan for citizen of Kuantan when the limitation for major contributor for flood occurrence has been set up based on Time Series Analysis.Besides, the prediction for the risk class for flood from ANN analysis able to help in constructing proper mitigating measure more efficiently for citizen of Kuantan before, during and after for flood occurrence and this will reduce cost of destruction and save life.

Local government also able to enforce more strict action towards developer especially for development along the river that are following guideline based on Environmental Quality Act where local authority able to stop the project immediately or charge compound RM 1000 for each day from beginning of project construction.

Acknowledgment

The authors acknowledge the Department of Environment Malaysia (DOE) and Department of Irrigation and Drainage Malaysia (DID) from the Malaysian Ministry of Natural Resource and Environment, Professor Dr. Mohd Talib Latif from Faculty of Science and Technology, Universiti Kebangsaan Malaysia, and Professor Dr. Sharifuddin Mohd Zain from Chemistry Department, Universiti Malaya for their permission to utilize the data, advice, guidance and support for this study.

References

[1] Rizwan, A.M., L.Y.C. Dennis, and C. Liu. 2008. *Journal of Environmental Science*. 20: 120–128. DOI: 10.1016/s1001–0742(08)60019–4.
[2] Metcalfe, J.L. 1989. History and present status in Europe. *Environ. Pollut.* 60: 101–139.
[3] Pinel-Alloul, B., G. Methot, L. Lapierre and A. Willsie. 1996. *Environ. Pollut*. 9: 65–87.
[4] Nedeau, E.J., R.W. Merritt, and M.G. Kaufman. 2003. *Environmental Pollution*. 123(1): 1–13.
[5] Dan'azumi, S., and M.H. Bichi. 2007. *International Journal of Engineering & Technology IJET-IJENS*. 10(01).
[6] Juahir H., S.M. Zain, M.K. Yusoff, T.I.T. Hanidza, A.S.M. Armi, M.E.Toriman, and M. Mokhtar. 2011. *Environ. Monitoring Assessment* 173: 625–641. DOI: 10.1007/s10661-010-1411-x.
[7] Mazlum, N., A. Ozer, and S. Mazlum. 1999. *Turkish Journal. Engineering Environmental Science*. 23: 19–26.
[8] Juahir, H., M.Z. Sharifuddin, K.Y. Mohd, H.A.S. Tengku, A. Mohd, E.T. Mohd, and M. Mazlin. 2010. *Environ Monit Assess*. 173 (1–4): 625–41. DOI: 10.1007/s10661-010-1411-x.
[9] Juahir, H., M.E. Toriman, S.M. Zain, M. Mokhtar, J. Zaihan, and M.J. Ijan Khushaida. 2008. *American-Eurasian Journal of Agricultural & Environmental Sciences*. 4(1): 258–265.
[10] Floyd, F.J., and K.F. Widaman. 1995. *Psychological Assessment*. 7 (3): 286–299.
[11] Juahir, H., M.Z. Sharifuddin, Z.A. Ahmad, K.Y. Mohd, and M. Mazlin. 2009. *Journal of Environmental Monitoring*. 12: 287–295.
[12] Imrie, C.E, Durucan, S. and Korea A. 2000. *J.Hydrol.* 233: 138–153.
[13] Herman, I. 1994. Selangor: Tekno Edar–Descriptive statistical analysis
[14] Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, London, United Kingdom