

# Flood Risk Pattern Recognition Using Chemometric Technique: A Case Study in Muda River Basin

Ahmad Shakir Mohd Saudi<sup>1,2</sup>, Hafizan Juahir<sup>1</sup>, Azman Azid<sup>1</sup>, Mohd Khairul Amri Kamarudin<sup>1</sup>, Mohd Ekhwan Toriman<sup>1</sup>, Nor Azlina Abdul Aziz<sup>1</sup>

<sup>1</sup>East Coast Environmental Research Institute, University Sultan Zainal Abidin, Kuala Terengganu, Malaysia

<sup>2</sup>Faculty Science and Technology, Open University Malaysia, Shah Alam, Malaysia

Email: [kir\\_dewa@yahoo.com](mailto:kir_dewa@yahoo.com), [hafizanj@gmail.com](mailto:hafizanj@gmail.com), [azmanazid@gmail.com](mailto:azmanazid@gmail.com), [amri.ukm@gmail.com](mailto:amri.ukm@gmail.com), [ekhwantoriman@gmail.com](mailto:ekhwantoriman@gmail.com), [naziz\\_2482@yahoo.com](mailto:naziz_2482@yahoo.com)

Received 17 April 2014; revised 31 May 2014; accepted 27 June 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This study constructs downscaling statistical model in analyzing the hydrological modeling in the study area which faces the risk of flood occurrence as the impact of climate change. The combination of chemometric method and time series analysis in this study show that even during the monsoon season, rainfall and stream flow are not the major contribution towards the changing of water level in the study area. Based on Correlation Test, it shows that suspended solid and water level show high correlation with  $p$ -value  $< 0.05$ . Factor Analysis being carried out to determine the major contribution to the changes of water Level and the result show that Suspended Solid shows a strong factor pattern with value 0.829. Based on Control Chart Builder for time series analysis, the Upper Control Limit for water level and suspended solid are 7.529 m and 1947.049 tons/day and the Lower Control Limit are 6.678 m and 178.135 tons/day. This shows that human development in the area gives high impact towards climate change and risk of flood in the study area which commonly faces flood during monsoon season.

## Keywords

Component, Chemometric Method, Time Series Analysis, Climate Change, Flood

---

## 1. Introduction

Kedah is “the rice bowl of Malaysia” and water has always been essential to the state. The rivers have been used

for irrigation, communication linkages between villages, and to transport rice grown in the vast area of paddy fields. The study area relatively is located in the district of Sik, and gentle flow from Kuala Kedah made navigation easy and Alor Setar was established as a port for the rice trade back in the 1800's.

Paddy uses a lot of water, and in spite of the reservoirs of Muda, Pedu and Ahning, water remains a scarce resource. MADA has therefore introduced water saving methods, including recycling of drainage water. Irrigation is still by far the largest consumer of water, but industrial demand is growing and the rising standard of living also means increasing domestic demand. Tourism also consumes a lot of water. Finally, Kedah supplies water to Perlis and shares Sungai Muda with Penang. The economic development thus increases the demand for a limited resource.

Water quality is the main concern. While the rivers are pristine in the upstream areas, they become increasingly polluted downstream due to discharge of urban wastewater and agricultural runoff. Solid waste often ends up in the river and this is especially a problem in urban areas. As a result, most storm water drains are very polluted and smelly, especially in the dry seasons.

Water quality status of the Sungai Muda is important to be known as to preventing potential health impacts on human beings in particular in a short period or even a long one, to protect the beneficial function of the river, and to provide relevant statistical analysis for water managers and decision makers to take effective control measures and have implementations of pollution preventing activities.

Urban wastewater and agricultural runoff are the main source of pollution in Muda River. The main functions of Sungai Muda are water supply for agriculture and also the water sources of people in Kedah and Penang. This pollution can be dangerous to human health and people rarely know about it.

Muda River basin which is located within the boundary of Kedah and Pulau Pinang with a catchment area of 4210 km<sup>2</sup> and 180 km length begins from Muda Dam and flows across the district of Baling, Sik and Kuala Muda. Water supply for agricultural, industrial and domestic sector for both Penang and Kedah is the key role of the river.

The catchment was often flooded during the rainy season from April to May and from September to November in every year. Many problems arise when flood keeps on worsening each year e.g., riverbank erosion, river pollution and reduction of water resources. The flood event which occurred on October 2003 was the worst compared to previous events in 1988, 1995 and 1998.

The main functions of Muda River basin are water supplies for agriculture and also the water sources of people in Kedah and Penang. It becomes sources of fresh water for Penang especially from Muda River where the total number of 17 such schemes is found in the 4 districts within the basin. These areas contain almost 3500 schemes that come from various sources, including Muda River tributaries and MADA canals.

During the rainy season, the catchment areas replenish the rivers and absorb large amount of rainwater, thereby minimizing risk of flooding. During the dry season, the catchment areas replenish the rivers and provide a continuous supply of water.

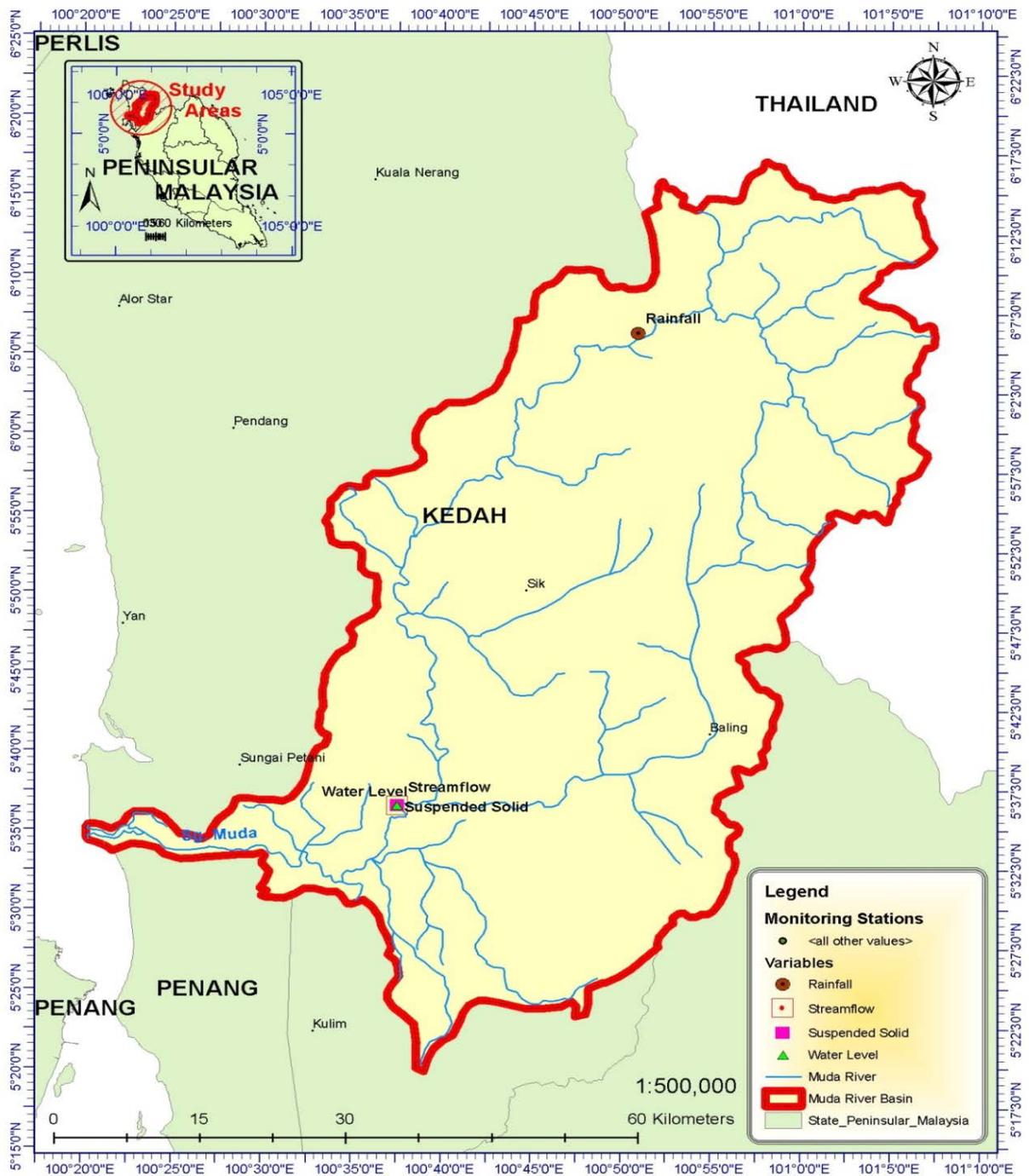
## 2. Materials and Methods

### Topography of Study Area

In a two-component gel, it is easy to modify the molecular structure of either of the two components. Situated at the coordinates of 5°06'N and 100°17'E, the natural basin covers an area of 2920 km<sup>2</sup>, with approximately 60 km wide and 80 km long, as well as ranges from 400 m high to the coastal plains, where they are the heirs of rice cultivation. Discovered over 250 years ago as the meeting point of Sungai Anak Bukit and Sungai Kedah, the main activity in the state capital was rice trade with the main area of the Muda Irrigation Scheme which consists of 966 km<sup>2</sup> coastal plains.

The topography of the Muda River Basin and the monitoring station by the Department of Drainage and Irrigation (DID) along the river basin is illustrated in **Figure 1** and **Table 1** tabulated the specific location of coordinates for monitoring stations. The secondary hydrological data were provided by the Department of Drainage and Irrigation (DID) for the year 1982-2012, which include rainfall, water level, stream flow and suspended solid.

These forests are the habitat of a diversified collection of plant and animal species. This includes the river terrapins which are threatened due to habitat destruction and excessive egg poaching. The river terrapins are protected under the Kedah Terrapin Enactment 1972. The river is also the habitat for fish species such as Channa,



**Figure 1.** Location of monitoring stations in Muda River Basin.

**Table 1.** Location of monitoring stations in Muda River Basin.

Station No.	Latitude	Longitude	Name of Station	Variables
Site 6108001	5°06'48N	100°01'55E	Kompleks Rumah Muda	Rainfall
Site 5606410	5°12'57N	100°27'16E	Sungai Muda, Jambatan Syed Omar	Suspended Solid
Site 5606510	5°17'25N	100°39'29E	Sungai Muda, Jambatan Syed Omar	Stream Flow
Site 5606410	5°18'32N	100°40'23E	Sungai Muda, Jambatan Syed Omar	Water Level

micropeltes (Toman), Labeo rohita (Rohu), Chitala chitala (Belida), Leptobarbus hoevenii (Jelawat) Pangasius nasutus (Patin), and Puntius gonionotus (Lampam).

However, rapid growth in certain area gives a negative impact towards rate of the surface runoff into the water body system and affecting water level at certain location in the river basin and leads to flooding. The study was being conducted to see the relationship between exceeded surface runoff to the water level of the river which leads to flood at high impact area, especially during monsoon season, the determine limitation of flood risk based on hydrological data from the year 1982-2012, and to identify suitable mitigation measure for flood prevention at the high impact area. Based on the **Figure 1** and **Table 1**, it explains the location of monitoring stations in the study area.

### 3. Methodology in Research

#### 3.1. Correlation Test

For this study, the Correlation test was adapted to determine the variables with strong relationship for further analysis as the test is suitable to measure two variables that have the relationship between  $-1$  to  $1$ . Pearson Coefficient and Spearman Coefficient are the types of products that can be implemented in this study, but the former was widely utilized should there be an association of two variables (Moore D.S. and McCabe G.P., 1989) [1]. It was used in this study to determine the relationship between important parameters in hydrological data, as well as to determine the parameter with the strongest relationship. Upon that, the development with the biggest influence on the hydrological modelling in Muda River Basin can be determined.

Spearman's and Pearson's rank coefficients are two of the most common types of correlation types, where the former that requires ordinal data as the calculation will be based on data ranking. In addition, it also measures the degree of strength for the coefficient between variables considered in the research (Altman, 1991) [2]. There can be either positive or negative correlations for this method, where the positive correlation indicates two variables increasing together in a linear condition. On the other hand, the negative correlation indicates one variable increasing while the other decreasing in a linear condition. Meanwhile, the Pearson rank coefficient requires actual data for the calculation and all variables considered must be in the form of ratio scale. Both tests were implemented in this study, and only the best result was used for the discussion in this study.

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

#### 3.2. Chemometric Techniques

Chemometric technique such as application of Factor Analysis is able to see the reduction of variables into a set of factors for further analysis. Based on Floyd and Widamann (1995) [3]. Based on Floyd and Widamann (1995) [3], the application of this method will make the researcher able to make a comparison of variables which give the highest impact towards the changes of water level with the lower cost compare to the other method.

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 in the study area can be compared on a cost effective and quicker manner compared to other techniques (Gorsuch, 1990) [4].

The utilization of this method in the study allowed the inclusion of a 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 (Thompson, 1996) [5]. 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 (Brett W. *et al.*, 2012) [6]. 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:

$$z_{ji} = b_j F_{1i} + b_{j2} F_{2i} \cdots + b_{jn} F_{ni} \quad (2)$$

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 represent the correlation coefficient between each factor and the observed variables.
- Factor scores are the values of each observation on the factor  $F_k$ .

### 3.3. Time Series Analysis

Time Series Analysis is essential for the prediction of water level in the study area, where this method enables an efficient evaluation of the process from the performance by analyzing 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 modelling, where the Sigma is within a range value of a set of data. Control Chart can detect some trends and patterns with actual data deviations from historical baseline, be 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 (Igor Trubin, 2008) [7]. The equation implemented in this analysis was:

$$\text{Moving Range} = \text{Plot : MRt for } t = 2, 3, \dots, m. \quad (3)$$

### 3.4. Artificial Neural Network

Artificial Intelligent mimics the concept of the human brain and it has been utilized in the method for data analysis known as an 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 the error function reached the predefined goal, thus completing the process (Juahir *et al.* 2009) [8].

The function used was given by:

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.* (2000) [9] 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.

## 4. Result and Discussion

### 4.1. Variables Which Contribute to Flood Occurrence

The correlation test being carried out for this study in order to see the relationship between variables in this study. All variables being analyzed by using Correlation Test to see whether all variables have strong correlation and based on result from **Table 2** and **Figure 2**, it shows that only water level and Suspended Solid have high correlation when the p-value for both variables is less than 0.001 (Saudi, 2014) [10].

The result for Rainfall and Stream flow shows a weak correlation with other variables when the result of the test shows that both variables show p-value close to 1. This explains that both variables have a very weak result to show no correlation with water level and Suspended Solid.

### 4.2. Factors Which Contribute to Flood Occurrence

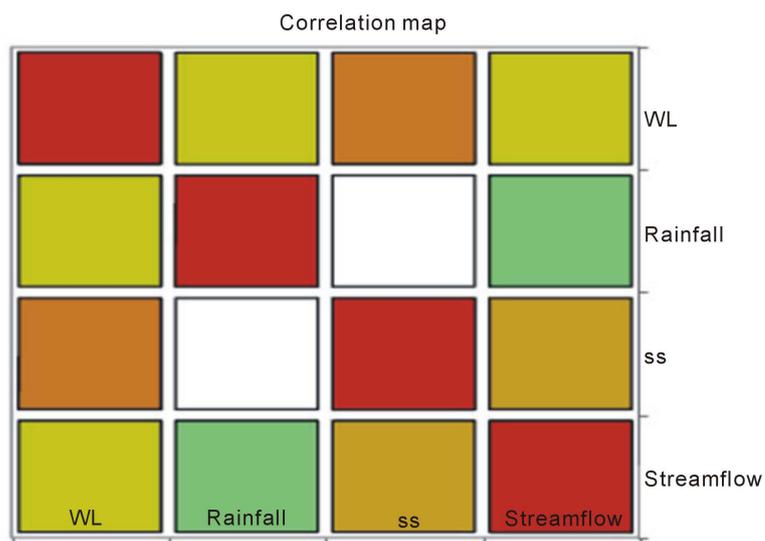
Results in **Table 3** and **Figure 3** show that there are 2 major components which affect the most of the hydrological modelling at Sungai Muda and those components are Suspended Solid and Water Level. Both variables show the strong coefficient with value more than 0.7 in Factor 1 and the result is 0.829 for Suspended

**Table 2.** Correlation test.

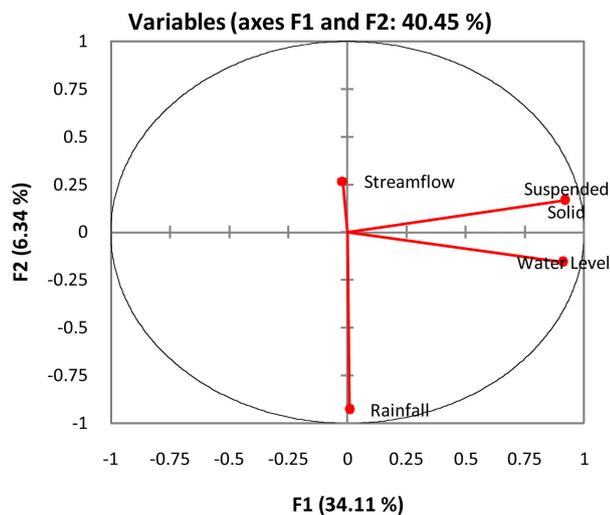
Variables	WL	Rainfall	SS	Stream Flow
WL	0	0.072	<0.001	0.061
Rainfall	0.072	0	0.230	0.026
SS	<0.0001	0.230	0	<0.001
Stream Flow	0.061	0.026	<0.001	0

**Table 3.** Factor analysis.

	F1	F2	Initial Communality	Final Communality	Specific Variance
Stream Flow	-0.018	0.136	0.005	0.019	0.981
Rainfall	0.008	-0.471	0.014	0.222	0.778
Suspended Solid	0.829	0.085	0.460	0.695	0.305
Water Level	0.822	-0.079	0.461	0.682	0.318



**Figure 2.** Correlation map.



**Figure 3.** Correlation between variables and factors.

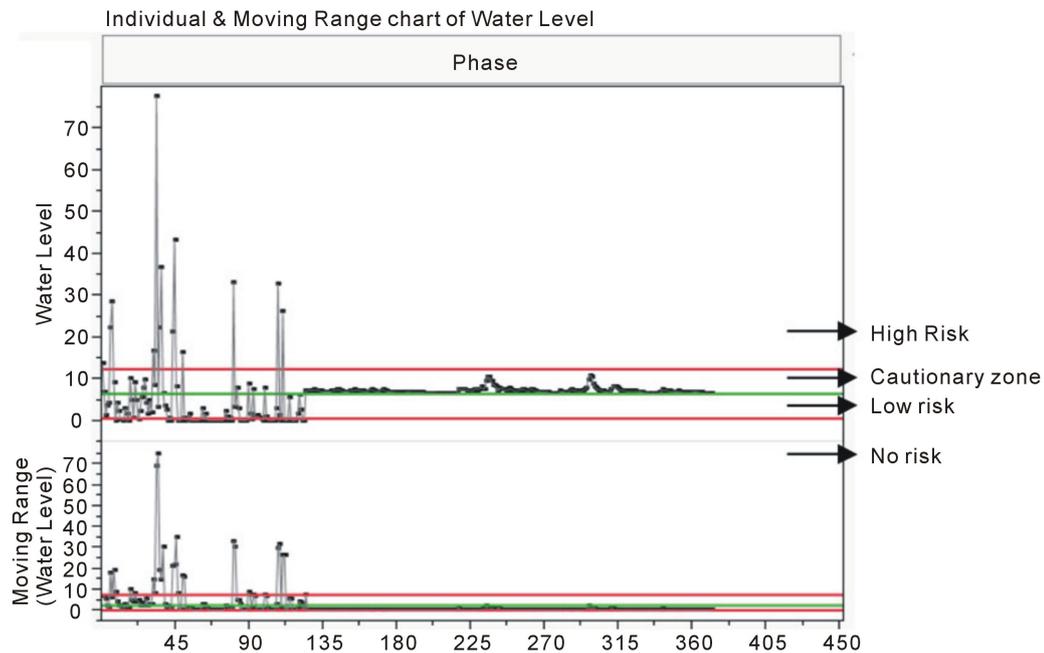


Figure 4. Result for time series based on Control Chat Builder.

Solid and 0.822 for Water Level. This concrete result shows the Rainfall is not the main factor in the changing of water level when it shows the weak coefficient with the result 0.085.

### 4.3. Flood Control Warning System

Based on **Figure 4** and **Table 4**, the average water level for Sungai Muda from year 1982-2012 is about 6.349 m and the Lower Control Limit (LCL) is 0.417 and for the Upper Control Limit is 12.282 m. This result shows that the water level above Upper Control Limit will face the risk of flood while the value of water level, which is below Lower Control Limit considered decrement on the water level at Sungai Muda where this condition will affect the role of the river as a source of water for agriculture and the source of water for citizen of Kedah and Penang.

Based on the **Figure 5** and **Table 5**, the result in Lower Control Limit for Suspended Solid from 1982 until 2012 is 178 tons ton/day, 1062 ton ton/day for average value and 1947 tons ton/day for Upper Control Limit. Result from Correlation test explains that water level and Suspended Solid show high correlation compared to Rainfall and Stream flow. This shows that when the range of Suspended Solid and water level within Upper Control Limit, the mitigating measure should be implemented in preventing flood from destroying the area even though the rate of rainfall is low within the same period. This situation can happen when the high rate of surface runoff from the water body precipitated and become the composition of the surface area of the river which cause the river turn into shallow.

The development around the study area affecting the climate in the study are not just based on the rate of rainfall anymore, but in this study area it refers to the high surface runoff will cause the high sedimentation into the river which will cause the changes of the depth of the river. This condition will cause the river easily to face flood if the heaviest rainfall occurred in a few days when the condition will cause the river become overflow and flooding.

### 4.4. Prediction of Flood Risk Classification

Based on the result of time series analysis in **Figure 4** and **Figure 5**, the risk of flood being classified into its own class based on hierarchy of risk is High Risk, Cautionary zone, Low Risk and No Risk.

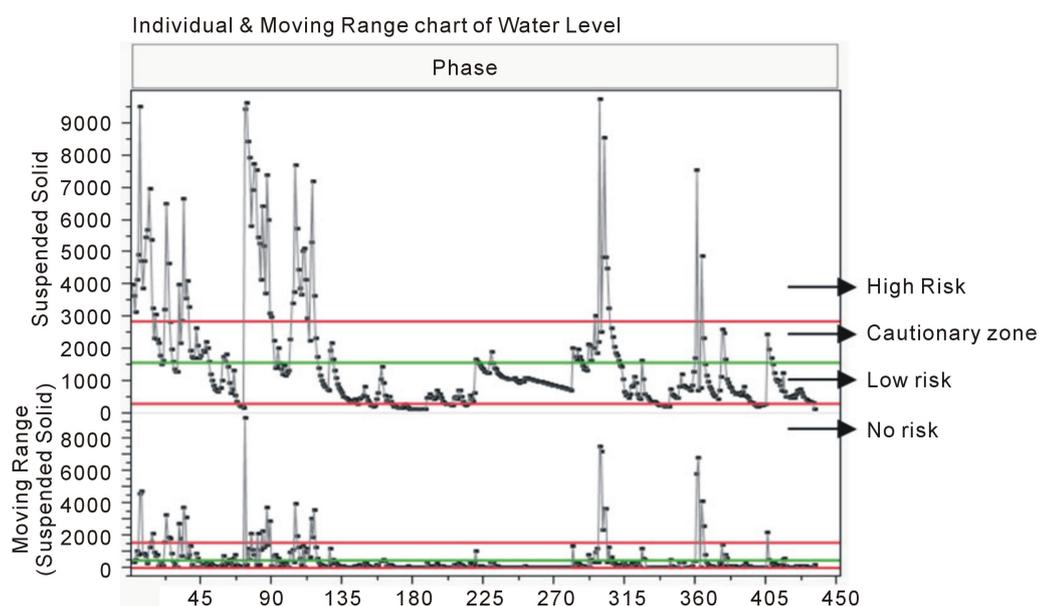
The level of High Risk are classified for all data which are pointed at the above Upper Control Limit line in Control Chart graph, followed by Cautionary Zone for data which are plotted between Average line and Upper

**Table 4.** Result for time series based on Control Chart Builder.

Points Plotted	LCL	AVG	UCL	Limit	Sigma	Sample Size
Individual	0.417	6.349	12.282	Moving	Range	1

**Table 5.** Result of suspended solid based on Control Chart Builder.

Points Plotted	LCL	AVG	UCL	Limit	Sigma	Sample Size
Individual	288.535	1563.631	2838.726	Moving	Range	1

**Figure 5.** Suspended Solid based on Control Chart Builder (time series analysis).

Control Limit, Low Risk for data which are plotted between Lower Control Limit line and Average line and No Risk for data which are plotted below the Lower Control Limit line.

Prediction of risk hierarchy being carried out by using Artificial Neural Network and the result from **Table 6** shows that the accuracy of prediction is 0.96 which is also being considered as 96% and this explains that the prediction is accurate and also can be used for future prediction in risk assessment for flood occurrence.

## 5. Conclusions

Local Authority should give a strong commitment in controlling excessive amount of surface runoff into the river. They must fit with a few conditions which are information management and performance monitoring, integrated policy and strategies, constitution legislation and standard, Erosion and Sediment Control Plan (ESCP) to control erosion and sediment, an effective enforcement by the Department of Drainage and Irrigation (DID) referring to 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. This action and legislation will be able to control the uncontrolled development along the river bank, being carried out by an irresponsible developer not following the guideline which has been set up by the government. This condition will be able to reduce the risk of flooding in the study area.

Other mitigating measure that has been implemented such as construction of the Barrage, River Bund, Pump House, Diversion, Pond, Dam and River Improvement work at study area should be well maintained and improved from time to time. The effectiveness of these mitigating measures also depends on the awareness and strong legal enforcement in controlling rate of surface runoff, which comes from uncontrolled human development in the study area and if it is not being configured well, all the structure mitigating measure means nothing in preventing of flood occurrence. Time Series Analysis is able to identify the limitation for all factors which

**Table 6.** Prediction for hierarchy of flood risk.

OUT_1	Accuracy	Total
Train	1	173
Test	0.96	75

affect the most of the changing of water level based on the results from Correlation Test and Factor Analysis, and this will not only reduce the cost of operation but also reduce the total lost from flood destruction and save lives. The application of Artificial Neural Network (ANN) is able to trigger earlier warning for citizens to take precaution for flood prevention based on level of risk from the prediction.

## Acknowledgements

I am grateful to the Ministry of Higher education for scholarship through my Ph.D. Scholarship for this research where I completely identified source and formulation in preventing flood occurrence in the study area. I would like also to thank my supervisor, Hafizan Juahir for advising me until this research completely done.

## References

- [1] Moore, D.S. and McCabe, G.P. (1989) Introduction to the Practice of Statistics. W. H. Freeman, New York.
- [2] Altman, D.G. (1991) Practical Statistics for Medical Research. Chapman & Hall, London, 285-288.
- [3] Floyd, F.J. and Widaman, K.F. (1995) Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*, **7**, 286-299. <http://dx.doi.org/10.1037/1040-3590.7.3.286>
- [4] Gorsuch, R.L. (1990) Common Factor-Analysis versus Component Analysis: Some Well and Little Known Facts. *Multivariate Behavioral Research*, **25**, 33-39. [http://dx.doi.org/10.1207/s15327906mbr2501\\_3](http://dx.doi.org/10.1207/s15327906mbr2501_3)
- [5] Thompson, B. and Daniel, L.G. (1996) Factor Analytic Evidence for the Construct Validity of Scores: A Historical Overview and Some Guidelines. *Educational and Psychological Measurement*, **56**, 197-208. <http://dx.doi.org/10.1177/0013164496056002001>
- [6] William, B., Brown, T. and Onsmann, A. (2012) Exploratory Factor Analysis: A Five-Step Guide for Novices. *Australian Journal of Paramedicine*, **8**.
- [7] Trubin, I.A. (2008) Exception Based Modelling and Forecasting. *Proceedings of the Computer Measurement Group*, Nevada, 7-12 December 2008, 353-364.
- [8] Juahir, H., Sharifuddin, M.Z., Ahmad, Z.A., Mohd, K.Y. and Mazlin, M. (2009) Spatial Assessment of Langat RIVER Water Quality Using Chemometrics. *Journal of Environmental Monitoring*, **12**, 287-295. <http://dx.doi.org/10.1039/b907306j>
- [9] Imrie, C.E., Durucan, S. and Korea A. (2000) River Flow Prediction by Using Artificial Neural Networks: Generalisation beyond Calibration Range. *Journal of Hydrology*, **233**, 138-153. [http://dx.doi.org/10.1016/S0022-1694\(00\)00228-6](http://dx.doi.org/10.1016/S0022-1694(00)00228-6)
- [10] Saudi, A.S.M., Juahir, H., Azid, A., Yusof, K.M.K.K., Zainuddin, S.F.M. and 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.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

