NEW FLOOD RISK INDEX IN TROPICAL AREA GENERATED BY USING SPC TECHNIQUE

A. S. M. Saudi[1,*], I. S. D. Ridzuan[1], A. Balakrishnan[1], A. Azid[2], D. M. A. Shukor[1] and Z. I. Rizman[3]

[1]Institute of Medical Science and Technology, University of Kuala Lumpur, 43600 Kajang, Selangor, Malaysia

[2]Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu, Malaysia

[3]Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia

ABSTRACT

This study applied four hydrology parameters. The findings from Principal Component Analysis confirmed that all selected parameters were significant to be taken as main tools for further analysis with result of $R^2 > 0.7$. SPC set up a new control limit for all selected parameters in the study area. For those data within or beyond the Upper Control Limit value, it was being considered as high risk for flood occurrence. New flood risk index within range from 0-100 was calculated using a combination of new algebraic equation and control limit values obtained from SPC analysis as variable. The accuracy of FRI was tested using ANN. The result showed the accuracy of FRI was more than 90%. It can be stipulated that the combination of chemometric techniques and SPC can produce a new standard FRI which is cost effective, accurate and flexible to be applied for the purpose of flood risk control in tropical area.

Author Correspondence, e-mail: ahmadshakir@unikl.edu.my

Keywords: flood risk index; statistical process control; chemometric technique; tropical area; control limit; prediction performance.

1. INTRODUCTION

Uncontrolled development by human is one of the major contributing factors towards the occurrence of flood. The impact of this lack of control will bring changes to the rate of sedimentation in the river and will cause changes in the climate [22]. Uncontrolled development in areas that are prone to flood and uncontrolled deforestation will trigger the increase of flood frequency as well as its magnitude [1]. The case of the flood may occur based on the climatology factor and/or weather such as the rate of rainfall, the rate of evaporation, wind drift and the natural surface of the earth [2]. More than 70% of floods in the world are within the Asia region, where this natural hazard mostly destruct vitally rural infrastructure [5]. The destruction totally creates a complicated obstacle in solving the poverty issue. The reason that makes the poverty issue worsen is the natural disaster demolishes the area of poor community and the recovery cost is very high [6]. Over the past decade, more than 10 billion USD was used to recover the annual cost of flood within Asia.

In tropical area, the reason may be different in urban areas as it specifies more on surface runoff, an occurrence that causes the river to be shallower than before and makes it easier for the river to overflow and this will cause flooding in the impact area [3]. Thus, an effective prevention method and good mitigation actions need to be carried out especially in tropical area like Malaysia in order to control the magnitude and the frequency of flood in this country.

The alteration of physical characteristics of the hydrological system caused by various type of development of heavily populated floodplains, intrusion into flood-prone areas, deforestation and hill slope expansion. The estimated area vulnerable to flood disaster is approximately 29,800 km^2 or 9% of the total Malaysia area and is affecting almost 4.82 million people, around 22% of the total population of the country [4].

The total vegetation that covers the area of tropical forest in this country is also reducing because of illegal logging. This condition will enhance soil degradation and soil erosion in which these particular problems may increase the rate of suspended solid flow into the river. Based on this fact, it will totally affect the changes of water level of affected river. This

condition will cause the river to become more polluted and may trigger flood occurrence easily.

This study is conducted to clarify the flood risk pattern recognition by applying chemometric techniques and Statistical Process Control (SPC) at Muda River Basin, Malaysia. In this study, the effects of uncontrolled human development may lead to a massive destruction caused by flood in which a high rate of surface run-off caused by the development affects the water level in the study area. This study will determine the main cause of flood occurrence, flood risk pattern, and also the prediction performance of new FRI created for flood risk control in the study area.

2. MATERIAL AND METHOD

2.1. Study Area

It is known that Muda River Basin is the source of water supply for both states of Kedah and Penang. From this river basin, the good water demand projection has been created in order to fulfill the need of water resources from the river basin [7]. This river basin has also become the main source of water supply for agriculture purposes as the area along the river basin is covered by paddy field, where it requires large amounts of water consumption from the river basin.

The river basin is located in the north-west part in Peninsular Malaysia. It covers the total area of $11,252km^2$ and the area includes three states which are Kedah, Perlis and Penang. The catchment area of the river basin is about 4,210km, where this river basin covers about 6 districts and 28 *mukims*. The study area has climate varying from the monsoon season where during the North-East monsoon season, the study area will get a small amount of rainfall due to the high mountainous area in the middle part of the Peninsular of Malaysia.

The condition is different during the southwest monsoon from May-August where within this period, there will be heavy rainfall in the study area when the monsoon from the Indian Ocean prevail towards the Peninsular Malaysia and cause high magnitude of rainfall in the particular area. The inter-monsoon season, which occurs from September until November also brings heavy rainfall to the study area when the west wind prevails causing a high rate of rainfall for two seasons in a year.

Based on previous study, more than hundreds of sand mining were allocated around the meadow river basin to extract sand from the riverbed. The amounts of extraction were far above the limit of natural sand of the river. This may result in an unstable condition of the river basin as the natural sources of the river were not being extracted in sustainable extraction. This situation leads to the destruction of the water body system and affecting the environmental condition along the river basin [8].

The flood occurs in the river basin almost every year during monsoon season, but always with different magnitude for every flood occurrence. The worse condition of the flood at the Muda River Basin within the 20 year period was in 1998 and 2003. There were 45,000 people who suffered property destruction as the impact of the flood in 2003 [9]. The flood in 2010 destroyed an estimated 45,000 hectares of rice fields with the government pledging USD8.476 million in aid to farmers, since this group is the most affected from this tragedy and also caused 50,000 people evacuated and four people died in this unexpected event [10].

The natural basin of Muda River Basin is approximately 60km wide and 80km long and covers an area of 2,920km^2 with coordinates of 5°06'N and 100°17'E. The basin ranges from 400m high to the coastal plain. The coastal plain is the center of rice cultivation. The state capital was founded over 250 years ago at the confluence of Sungai Anak Bukit and Sungai Kedah, which was the center of the rice trade. The coastal plain is dominated by the Muda Irrigation Scheme covering an area of 966km^2.

Table 1. Location of study area at Muda River Basin in Malaysia

River Basin	Station No.	Latitude	Longitude	Name of Station	Variables
	Site 6108001	05°06'48"N	100°01'55"E	Kompleks Rumah Muda	Rainfall
	Site 5606410	05°12'57"N	100°27'16"E	Sungai Muda, Jambatan Syed Omar	Suspended Solid
	Site 5606510	05°17'25"N	100°39'29"E	Sungai Muda, Jambatan Syed Omar	Streamflow
Muda River Basin	Site 5606410	05°18'32"N	100°40'23 E	Sungai Muda, Jambatan Syed Omar	Water Level

Fig.1. Location of study area at Muda River Basin in Malaysia

2.2. Principal Component Analysis

This study employed Principal Component Analysis (PCA) technique. In this technique, the moldings of new variables after following a few processes are made and these new variables have a linear combination of the original variables. PCA was 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 [11-12].

PCA is used to achieve these two aims which are; for data reduction and to interpret the data. When this technique was applied, it made it easier for us to identify a linear composite of new variables from its origin. This is important for us to calculate the variation among them in

order to cluster those variables effectively.

By using the PCA, the eigenvalues and eigenvectors are produced through the covariance matrix. PCA is considered the best tool in describing patterns of multiple measured parameters. After the production of eigenvalue and eigenvalue factors of the covariance matrix, the principal components score (PCS) is derived from the combination of eigenvectors and original normal value.

This technique helps in organizing the data by outlining the most significant data for analysis purpose. It is also able to minimize the loss of original information and this statement is supported by this equation where the technique is expressed as follows:

$Z_{ij} = a_{i1}x_{j1} + a_{i2}x_{j2} + a_{i3}x_{j3} + a_{im}x_{mj}$ (1)

where Z = Component score, a = component loading, x = measured of variable, i = component number and m = total variables.

From this analysis, the strong correlation coefficient of factor loading is being reflected for result with value greater than 0.6 [13]. Therefore, in this respective analysis, the correlation coefficient value which is greater than 0.6 will be considered as a strong factor loading thus should be taken for further analysis.

2.3. Statistical Process Control

In this study, the performance of time series involved the method of SPC. The control chart visualizes the level of quality of the selected variables constantly along with the changes of time. The control chart also positions a limit line that is considered as the measurement for the condition of quality of the product. This is done by following the upper control limit line and lower control limit line.

The control chart has the ability to reveal some trends and patterns by showing actual data deviations from the historical baseline and dynamic threshold. This in return is able to capture unusual resource usage thus able to become the best base lining to demonstrate how the actual data deviate from historical baseline [14]. The equation which was being used in this analysis is shown as follows:

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

UCL = 3.267MR

CL = MR

LCL $= 0$

where MR $=$ average moving range, t $=$ time and m $=$ individual values.

Average value,$\tilde{x} = \frac{\sum_{i=1}^{m} x_i}{m}$ (2)

where $\overline{X}$ = moving range, m = individual values and x_i = difference between data point.

In this analysis, the control limit value for the selected hydrological parameter has been formed specifically. The control limit involved in this analysis was Upper Control Limit (UCL), Center Limit (CL) and Lower Control Limit (LCL). The value of UCL outlines the maximum capacity that a river can be supported and if it is beyond the limit line, the probability for flood occurrence is likely to be very high. The method is being used as warning alert system for flood and this method never been used by other researcher for flood risk analysis purposed.

2.4. Flood Risk Index

The process of creating the risk index needs to undertake a few processes of statistical analysis. First, the selection of the best variable to be used in this model is being carried out by applying PCA. The variables with the highest correlation coefficientwill be selected to be used for the next process in the analysis.

After the selection of variable has been made, the determination for control limit value is developed by implementing the Statistical Process Control. From this method, the formation of Upper Control Limit value and Lower Control Limit will able to give a guideline in determining the ratio of the flood risk index in this research. The Upper Control Limit value is considered an intolerable value for a variable and treated as a high risk condition for flood. For the formation of flood risk index, the Upper Control Limit value was taken and risk index was calculated using the formula as follows

$\frac{UCLV}{X}$ x $100 = 70$(Value of High Risk Index)(3)

Based on the Equation (3), UCLV is Upper Control Limit Value of variable, X is the highest value of the data, 100 is the range of risk index which is from 1-100 and 70 is significant value of the index for High Risk.

Based on the above equation, the formation of formulas for the risk index is designed to obtain the best risk model for flood. The flood risk index in this study was ranging from 0-100,

where the range of 70-100 was being categorized as a High Risk Index for flood. The flood risk index from the rate of index from 0-34 considered as a Low Risk Index for flood, 35-69 was considered as Moderate Risk Index while 70-100 considered as High Risk for flood. The selection of range 70-100 was based on the Relative Strength Index (RSI), where 70 and above is considered as upper bound and an intolerable condition [15].

2.5. ANN for Flood Risk Index Prediction Model

All data are being classified into three phases which are trained, testing and validate sub-set [16]. The training subset includes the whole set of data where in this phase, the system will use the data to recognize and learn the pattern for every selected variable in this analysis. After that, the data are analyzed in the test sub-set and the ability of the trained data will be evaluated in this phase. The final phase which is the validating sub-set functions in handling the final evaluation of the training set.

In order to make precise prediction and effectiveness of each network, there are two different criteria which are taken into consideration in this respective study. These criteria are correlation of determination (R^2) and root mean square error (RMSE) [17]. The R^2 efficiency can be described as follows

$$R^2 = 1 - \frac{\sum (x-y)^2}{\sum_i y2 - \frac{\sum_i y2}{n}}. \qquad (4)$$

RMSE can be calculated by using different type of equation. The equation for RMSE is written as follows

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} \qquad (5)$$

By referring to this equation, it denotes that x_i = observed data, y_i = predicted data and n = number of observation and represent the percentage of the initial uncertainty that represent by respective model.

From both equation above, the accuracy of the prediction will depend on R^2 and RMSE value. The lower the RMSE value (RMSE = 0) and the higher the value of R^2 (R^2 = 1), the more accurate the prediction will be. From this evaluation, we are able to determine whether the flood risk index that has been predicted is accurate and significant to be taken or not for further analysis. JMP10 software was used to compute the analysis of the Artificial Neural Network (ANN). In ANN, the technique of back propagation algorithm is used whereby it

requires the training of algorithm for multi-layer feed forward network. The application of this method in this study will determine the accuracy of prediction from selected variables. The method also functions to predict the accuracy of new flood risk index that will be used to determine the level of risk for flood occurrence in the study area.

3. RESULTS AND DISCUSSION

3.1. Identification of the Significant Factor of Flood Event

The results in Table 2 show that the result of factor loading after varimax rotation with the value of eigenvalue is more than 0.7 [18]. Results of factor loading with more than 0.6 is considered to be strong factor loading and significant to be taken as strong factor in the analysis [19, 20, 23].

Based on Factor 1, the results for factor loading were 0.697 for suspended solid and 0.712 for water level. Based on the results in Factor 1, Streamflow was the strongest factor loading with the results of -0.642. While Factor 2, the strongest factor loading was Rainfall with the result of 0.705. However, the results for the other variables in Factor 2 did not exhibit any strong factor loading when the result of factor loading for streamflow were 0.164, 0.344 for the suspended solid and 0.185 for water level respectively. Factor 3 shows that streamflow was the strongest factor loading with the result of 0.710. Other variables results in this factor were 0.454 for suspended solid, followed by 0.102 for water level and -0.323 to rainfall. Factor 4 shows that the strongest factor loading was water level with the result of 0.670, followed by suspended solid with the result -0.502, stream flow with the result 0.229 and 0.012 for rainfall.

Result in Factor 1 shows that every increment of suspended solid in the river basin will cause significant increases of the water level in study areas. It also shows that for every rising of water level and suspended solid in study areas, this condition may lead to the significant reduction of the speed for stream flow in the study area. The results reflect the contribution of point source and non-point source that contributes to the rate of suspended solid cause the increasing rate of water level in most River Basins in Malaysia [21]. In Factor 2, the result demonstrates that at a high rate of rainfall, it will only give mild impact towards other variables in this study. The rate of changes of rainfall did not give significant impact towards

changes in water level, suspended solid and stream flow in study areas. When in a state of imbalance distribution of rainfall in the basin, it did not give a strong impact towards other variables, especially towards the changes of water level that will trigger the formation of floods in the study area. Refer to Factor 3, it shows that the increasing rate of streamflow was able to cause the increasing rate of suspended solid in the river. The increasing rate of these two variables did not give strong impact towards the changes of water level and rainfall in the study area as these variables did not show a strong factor loading in Factor 3. Based on Factor 4, under natural circumstances in the study area, the results statistically proved that water level was still indicated as the strong factor loading compared to other variables in this factor. The result of this analysis explains that the contribution of floods is strongly influenced by all variables involved in this study. All components showed the strong factor loading with result of factor loading more than 0.6. Rainfall component proved to have a strong impact for the formation of flood together with suspended solid and stream flow, which conveys a higher correlation coefficient in this analysis. This condition can happen because of the rain that occurs arbitrarily and localized causing readings taken at the monitoring station significant with the actual amount of rain water that flow into the river basin.

Table 2. Result of principle component analysis in Muda River Basin in Malaysia from 1982-2012

Variables	F1	F2	F3	F4
Rainfall	-0.277	0.705	-0.323	0.012
Streamflow	-0.642	0.164	0.71	0.229
Suspended solid	0.697	0.344	0.454	-0.502
Water level	0.712	0.185	0.102	0.67
Eigenvalue	1.419	0.998	0.830	0.753
Variability (%)	35.467	24.961	20.746	18.826
Cumulative %	35.467	60.428	81.174	100.000

Based on the result, for water level in Muda River Basin, the result was 7.52m of the Lower Control Limit (LCL), the Center Limit (CL) was about 8.13m and the Upper Control Limit (UCL) was 8.74m. Results in Fig. 2 and Table 3 indicate the flood patterns in the study area.

The data which is plotted above 8.744m is considered as UCL and may easily lead to flood. The water level within the range of CL which begins from 8.13m until 8.60m is considered to be adequate for the river basin to support, and the water level in this range is being categorized as within the safe zone. Water level which begins from 7.516m and above until 8.13m is being categorized as no issue for the flood and the water level is less than adequate level of the river basin support ability. The reading in the graph from 2288 upwards shows the decrement of water level in the study area was due to the increasing rate of suspended solid in Muda River Basin. This eventually caused the river turn to shallow and pose very high risk for flood occurrence if high rate of rainfall occur within this period.

By referring to Fig. 3 and Table 3, it shows the whole result for control limit value for rainfall in Muda River Basin. Based on the result, it shows that 0mm for the Lower Control Limit (LCL), the Center Limit (CL) was about 6.7mm and the Upper Control Limit (UCL) was 24.3mm. The result indicates the rainfall patterns in the study area. With reference from the data which was plotted above UCL, it was being considered as unnatural condition and beyond the ordinary limit for Muda River Basin.

Results for control limit of suspended solid in the study area are being shown in Fig. 4 and Table 3. Based on the result, it shows that 41.64 tonnes/day for the Lower Control Limit (LCL), 1221.4 tonnes/day for the Center Limit (CL) and 2401.65 tonnes/day for Upper Control Limit (UCL). From the result indicates the suspended solid patterns in the study area. The changes rate of suspended solid in Muda River Basin definitely gives impact towards the changes of water level in the study area, since this variable has a strong correlation with the variable of water level.

Based on result in Fig. 5 and Table 3, the results for streamflow in the study area are shown. From the results, it explains that the LCL for streamflow in Muda River Basin was 24.64m/s, 82.15m/s for CL and 139.65 for UCL. The data only recorded until 3100 because of DID does not provide data after that particular reading. This is because of technical problem in monitoring station after the flooding tragedy in 2010.

The results of this analysis can be used in improving the flood early warning system in the Muda River Basin. Based on the result, the control limit obtained allows a continuous monitoring system by local authorities of Kedah to execute early actions based on the level of

river basin water level degree. The establishment of early warning systems can be made more robust with high accuracy based on the current pattern formation in the area of flood. The use of the control limit system can also help the economic sector of Kedah in preparation for the possibility not only for flood, but also practical in the face of drought conditions. This is important because the Muda River Basin areas are considered as major producer of rice production in this country, and it requires optimal water supply to maintain the maximum supply of rice for local community and also for citizen of Malaysia.

Table 3. Result of SPC for Muda River Basin

| Area	Points Plotted	Lower Control Limit	Center Limit	Upper Control Limit	Limit Sigma	Sample Size
Water Level	Individual	7.516m	8.132m	8.744m	Moving Range	1
	Moving range	0m	0.2307m	0.754m	Moving Range	1
Rainfall	Individual	0mm	6.769 mm	24.315 mm	Moving Range	1
	Moving range	0 mm	6.599 mm	21.557 mm	Moving Range	1
Suspended Solid	Individual	41.641 tonnes/day	1221.403 tonnes/day	2401.165 tonnes/day	Moving Range	1
	Moving range	0 tonnes/day	443.739 tonnes/day	1449.49 tonnes/day	Moving Range	1
Streamflow	Individual	24.642m/s	82.150m/s	139.657m/s	Moving Range	1
	Moving range	0m/s	21.630m/s	70.655m/s	Moving Range	1
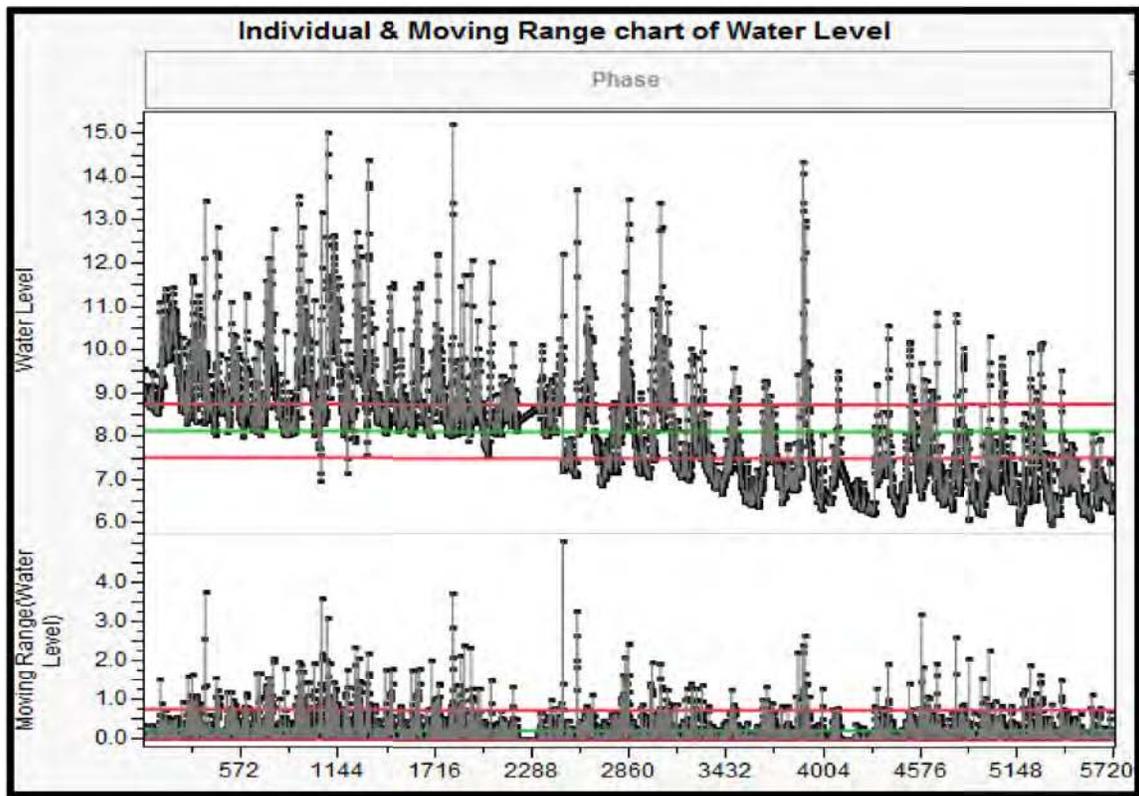
Fig.2. Result of SPC for water level in Muda River Basin
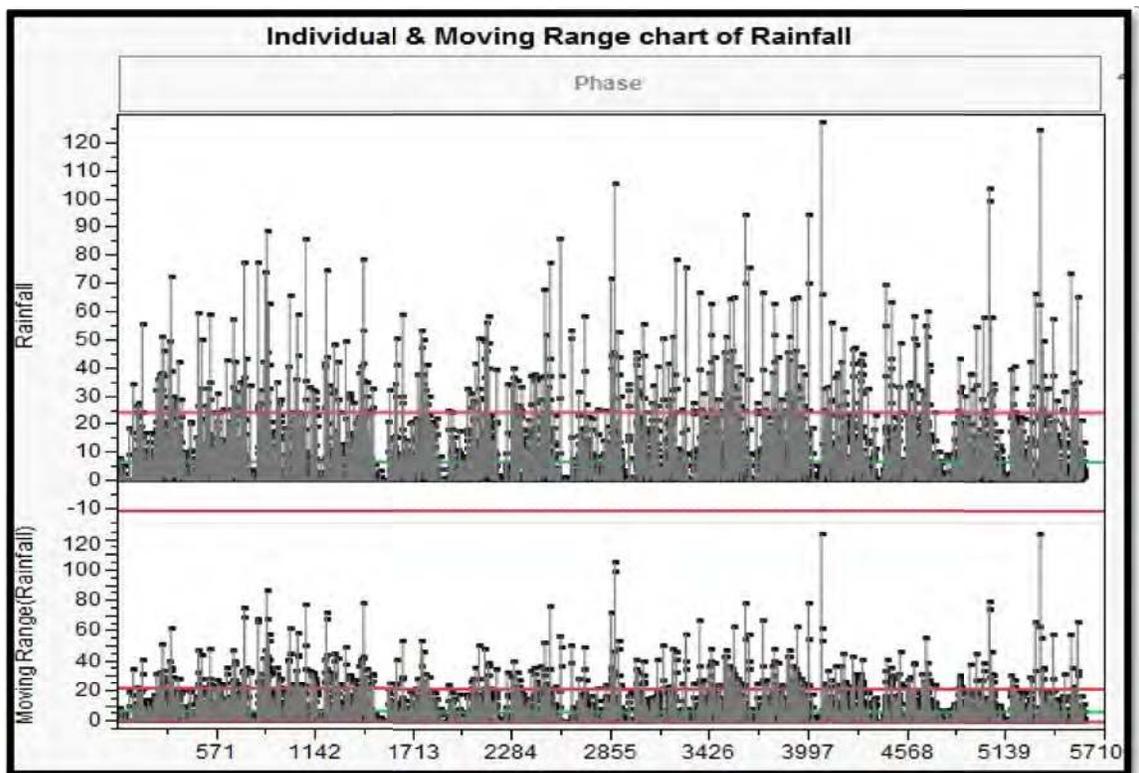

Fig.3. Result of SPC for rainfall in Muda River Basin
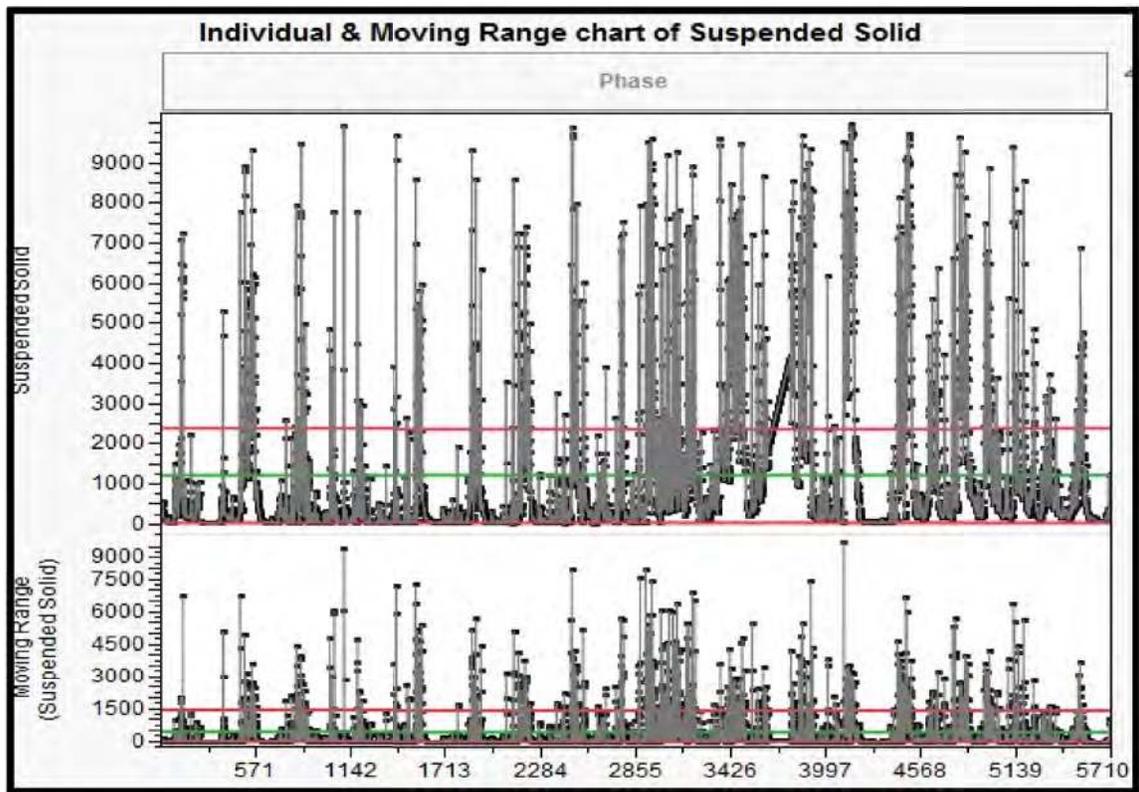
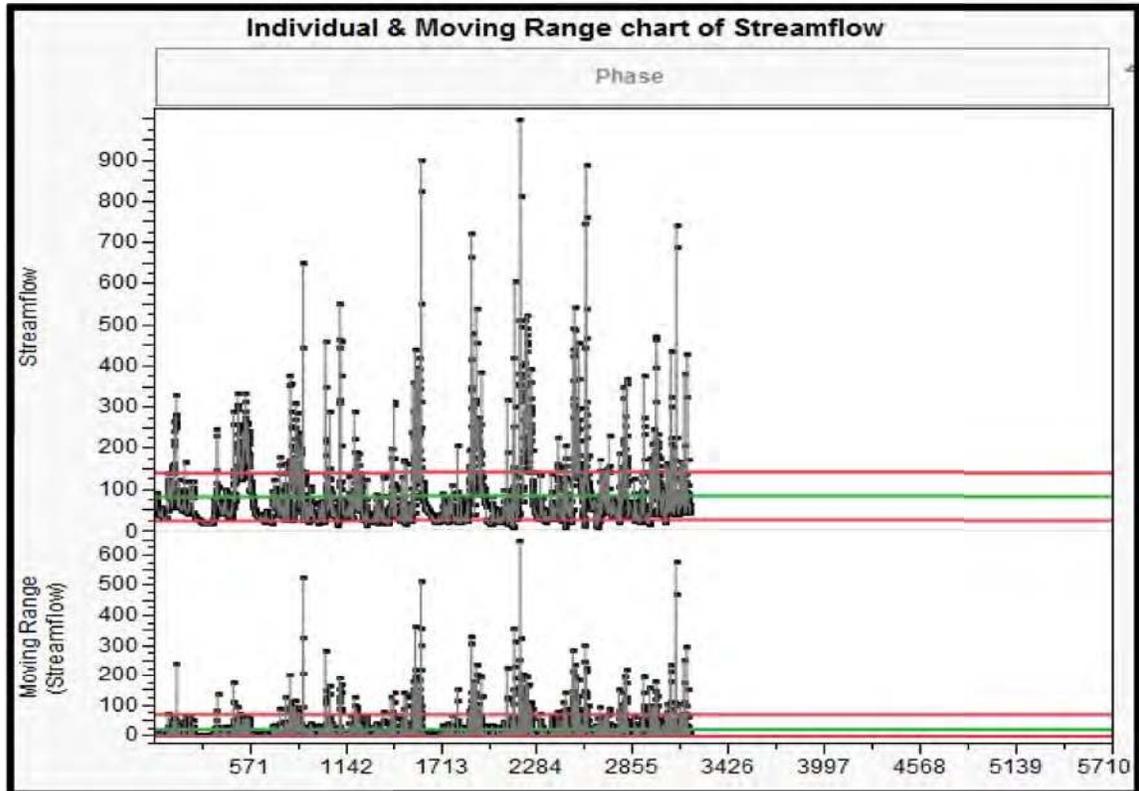**Fig.4.** Result of SPC for suspended solid in Muda River Basin

Fig.5. Result of SPC for stream flow in Muda River Basin

3.2. Flood Risk Index (FRI) in the Study Area

Index calculation was done to set the range of flood risk in the study areas. This index ranged from 0-100 rate and the rate that was set at the risk 70 and above was the high rate of water level at risk of flood. The 70 and above set for UCL and CL value were set at 35-69 and down to LCL starting with 0-34 index for the range of flood risk index [25]. High risk was classified to all data which located greater than the UCL line in the control chart graph, followed by cautionary zone for data that were plotted between CL line and UCL line, low risk for data that were plotted between LCL line and CL line, and no risk for data that were plotted below the LCL line.

Fig. 6-9 show the level of risk for the changes of water level may lead to flood occurrence in Muda River Basin and from the figure, 11.2% were categorized as a high-risk class where the risk ranged 70 and above, 26.5% were categorized as cautionary zone in which the range of risk was from 35 until 69, 40.2% were categorized as low-risk class in which the range of risk was from 0 until 34 and 22.1% were categorized as no risk class and were not included in the range of risk for flood in the Muda River Basin. The accumulative percentage of the cautionary zone class and high-risk class were 66.7%, which is considered as high percentage for risk of flood occurrence based on the water level changes in Muda River Basin for the year 1982-2012. The level of risk for the changes of rainfall was 10.2% being categorized as a high-risk class where the risk ranged 70 and above, 16.5% were categorized as cautionary zone in which the range of risk was from 35 until 69, 50.2% were categorized as low-risk class in which the range of risk was from 0 until 34 and 23.1% were categorized as no risk class and was not included in the range of risk for flood in the Muda River Basin. The result of risk level for suspended solid in the study area was 29.1% within high-risk class, 18.5%, within cautionary zone class, 30.5% within low risk class and 22.2% within no risk class. The risk level for streamflow in study area was 19.3% for high risk class, 28.4% within cautionary zone class, 34.2% for low risk class and 18.1% for no risk class.

Based on this percentage, the result shows that emergency response plan must be taken as early as in cautionary zone risk class level in which at this level the emergency response team must be ready to take proper actions before a massive destruction occurs. The destructions may be caused by flood since flood might occur within this range of risk index. The results of

risk range obtained are capable of bringing changes to the regulation of flood risk for the Muda River Basin. This enables a more detailed picture to be given to the authorities in establishing precise guidelines for the level of flood risk for Muda River Basin. Early action by authorities for issues regarding flood control can be deployed efficiently, while the effects of destruction as well as casualties can be avoided if the flood occurs in the Muda River Basin.
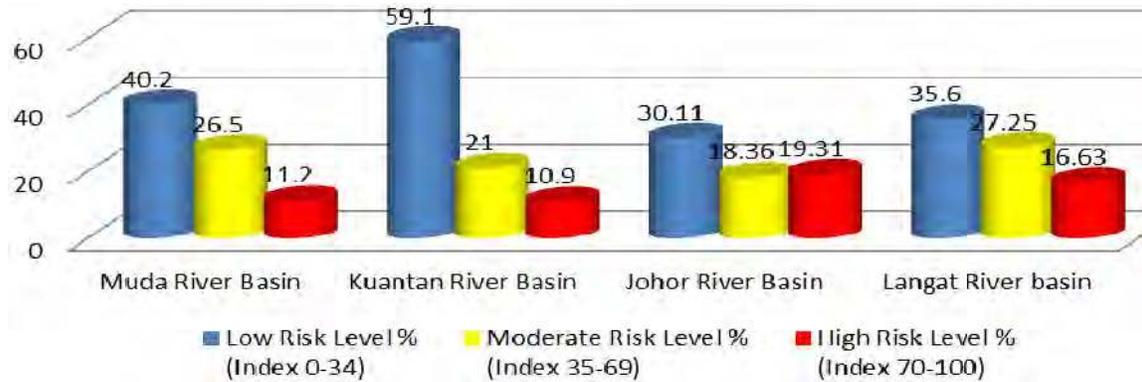

Fig.6. Percentage of flood risk index class for water level in study areas
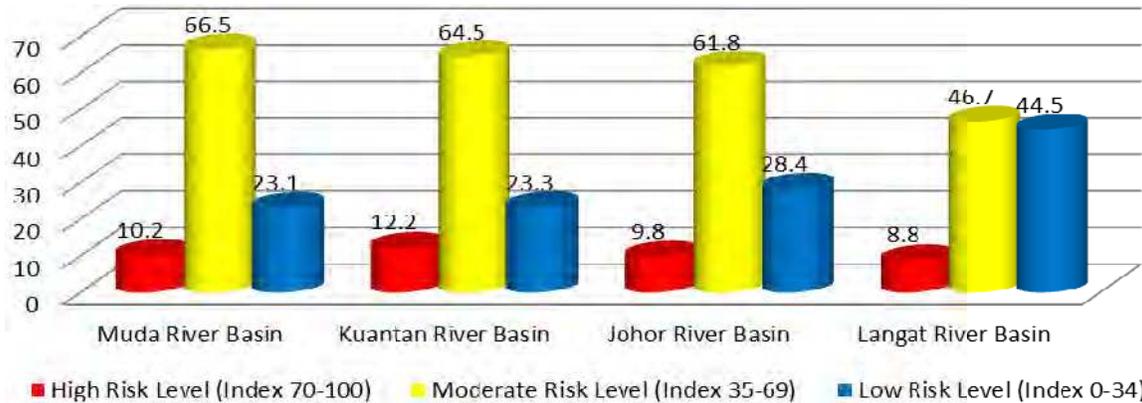
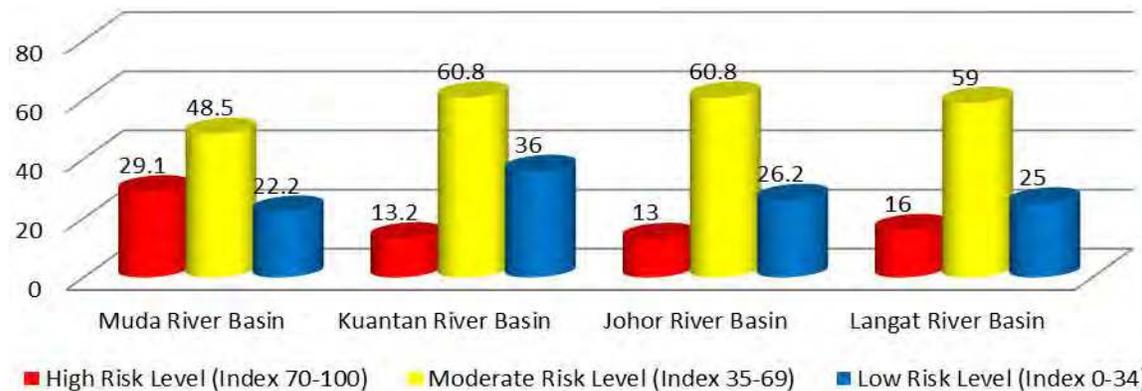**Fig.7.** Percentage of flood risk index class for rainfall in study areas

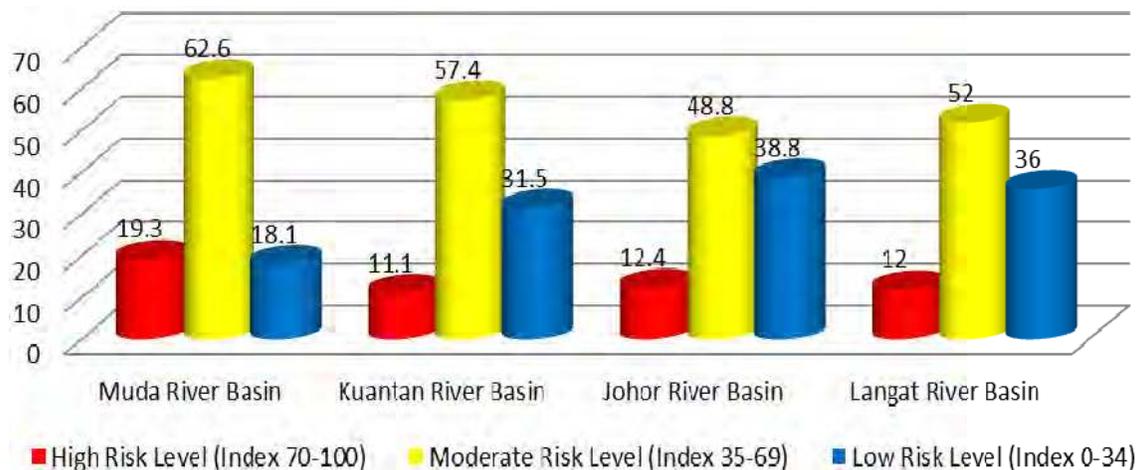**Fig.8.** Percentage of flood risk index for suspended solid in study areas

Fig.9. Percentage of flood risk index for stream flow in study areas

3.3. Prediction Performance of Flood Risk Index in the Study Area

The prediction of risk hierarchy was carried out using the ANN. The result from Table 4 and Fig. 6-9 show that the accuracy of prediction was more than 90%. This explains that the prediction is accurate and can be used for future prediction in the risk assessment for flood occurrence. The data obtained in the study were divided into three categories; the rate of 60% for validation, 20% for train and 20% for testing. The validate prediction results for water level showed the rate of R^2 in the analysis was 0.9864 and the lowest RMSE of rate was 1.4058 in three of the hidden nodes to achieve optimal. Rainfall showed the rate of R^2 in the analysis was 0.9652 with the lowest RMSE of rate with 1.783 in five hidden nodes to achieve optimal result. Suspended solid result for validate phase was 0.9117 with the lowest RMSE of rate with 2.103 in three hidden nodes to achieve optimal result. The stream flow showed the rate of R^2 in the analysis was 0.9002 with the lowest RMSE of 2.262 in three hidden nodes to achieve optimal result.

For train analysis at a rate of 20%, the results showed good results for water level variable with the results of R^2 with a value of 0.9693 and the lowest RMSE of rate with a value of 2.0501 at number six of the total number of hidden nodes in order to achieve optimal results. Rainfall shows an impressive result with R^2 0.9111, lowest RMSE of 2.1051 with four hidden nodes followed by suspended solid with R^2 0.9225, lowest RMSE of 2.1132, with four hidden nodes and stream flow with R^2 0.9018, lowest RMSE of 2.2132 with five hidden nodes to achieve optimal result.

For the analysis of testing prediction, 20% of the data were taken for future prediction. The results showed that the prediction rate for water level was 0.9987 for R^2 value with the lowest RMSE rate of 0.3314 by three numbers of hidden nodes to achieve the most optimal results. This good result followed by rainfall with R^2 0.9771, lowest RMSE of 0.3422 with four hidden nodes, suspended solid with result of R^2 0.9689, lowest RMSE of 0.3488 with four hidden nodes and stream flow with result of R^2 0.9652, lowest RMSE of 0.3331 with four hidden nodes to achieve optimal result. These results prove that the prediction made for the level of risk index for the study area is significant and accurate to be used for future risk analysis of flood. The prediction was done for the purpose of giving warnings and significant flood alert to the areas in the vicinity of Muda River Basin. The statistics proved through analysis of ANN, the prediction of the risk index and the water level of the study area has more than 90% accuracy.

Table 4. Result of prediction performance by using ANN in Muda River Basin

Prediction Area	Validation		Train		Test	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
	0.98642	1.4124	0.96623	2.0912	0.9985	0.339
	0.98646	1.4063	0.96713	2.0873	0.9986	0.3324
	0.98647	1.4058	0.96735	2.0841	0.9987	0.3314
Water Level	0.98620	1.4053	0.9689	2.062	0.9985	0.3351
	0.98510	1.3975	0.9692	2.0527	0.9981	0.3342
	0.98521	1.3881	0.9693	2.0501	0.9975	0.3241
	0.98335	1.3779	0.9678	2.055	0.9962	0.3224
	0.94213	1.9221	0.8591	0.2799	0.9653	0.3443
	0.95500	1.8999	0.8614	0.2797	0.9749	0.3433
Rainfall	0.95703	1.8225	0.8880	0.2768	0.9755	0.3429
	0.95877	1.8166	0.9111	2.1051	0.9771	0.3422
	0.965200	1.7830	0.90264	0.2431	0.9742	0.3441
	0.9600	1.7933	0.90233	0.2675	0.9737	0.3517

	0.90213	2.9221	0.8591	0.2799	0.9453	0.3643
Suspended	0.90500	2.8999	0.8614	0.2797	0.9449	0.3633
Solid	0.91170	2.103	0.8880	0.2768	0.9655	0.3529
	0.90877	2.3166	0.9111	2.1051	0.9689	0.3488
	0.88472	2.4325	0.8771	2.7614	0.94367	0.66291
	0.88489	2.4334	0.8777	2.7362	0.94467	0.56290
Stream flow	0.90020	2.2620	0.8778	2.7314	0.94536	0.56259
	0.88481	2.4336	0.8779	2.7283	0.96520	0.3331
	0.88487	2.4331	0.9018	2.2132	0.95168	0.46287
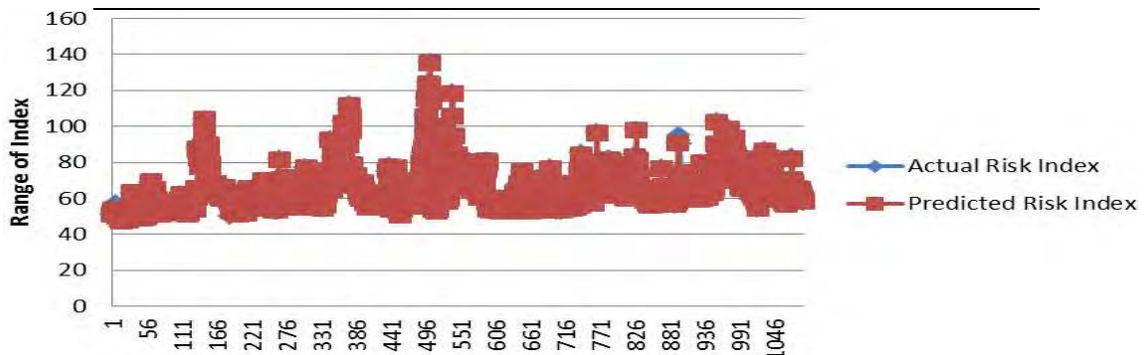
Fig.10. Overlay plot of actual risk index and predicted risk index for water level in Muda
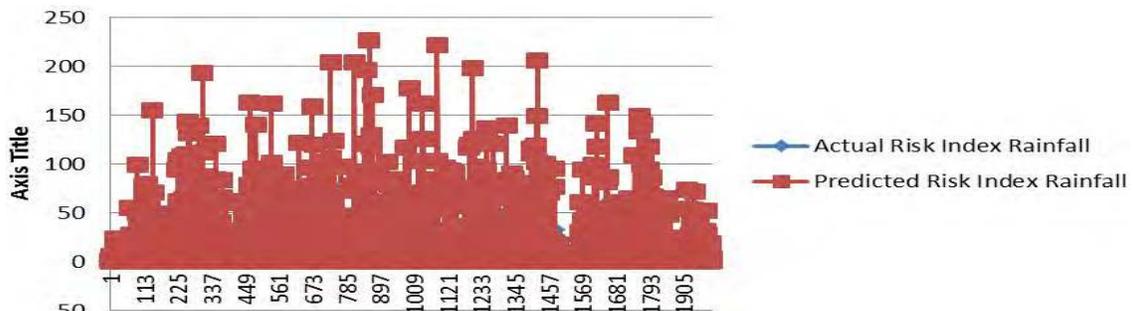
River Basin

Fig.11. Overlay plot of actual risk index and predicted risk index for rainfall in Muda River
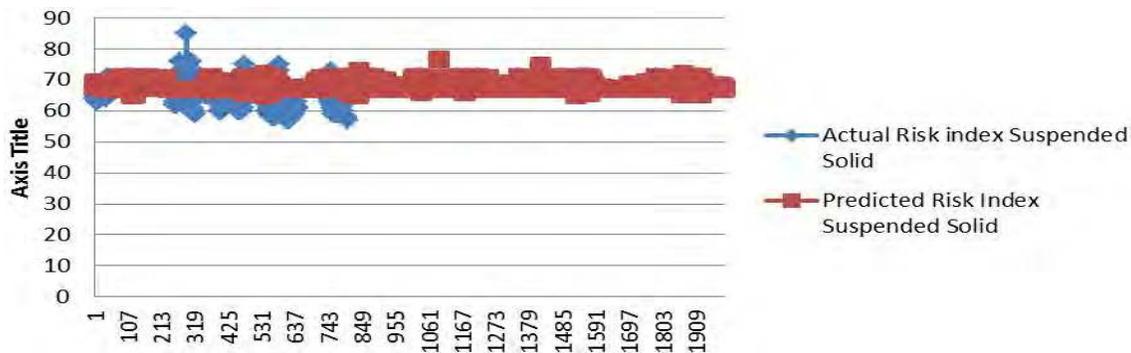
Basin

Fig.12. Overlay plot of actual risk index and predicted risk index for suspended solid in Muda River Basin
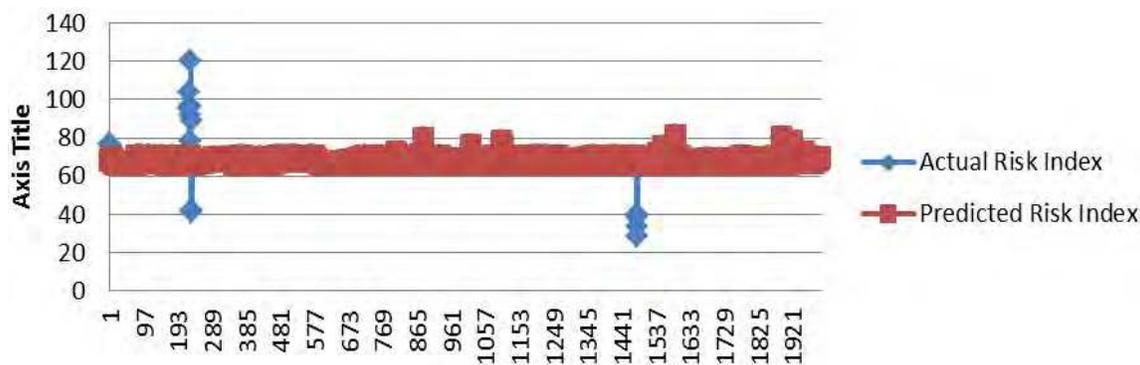

Fig.13. Overlay plot of actual risk index and predicted risk index for stream flow in Muda River Basin

4. CONCLUSION

The most significant factor for the variables in this study was obtained and applied in the development of risk models for alert warning system for floods in which variables that have been proven statistically can give emphasis to the flood pattern visualization and optimal rates for the maximum limit for flood control basin river in Malaysia, particularly in the study area. This obviously can help in planning DID flood control program in preparation for any eventuality of floods in Malaysia. The formation of an efficient control limit that is sensitive to changes in water level can enable flood warning alerts to improve the existing system used by DID in managing flood control for Malaysia .The results of the control limit obtained in the study has sparked a formation of a uniform risk index on the severity of the risk of flooding in Malaysia. This is a new finding that clearly could help the DID in flood risk analysis studies which previously has no uniformity in the form of practical risk index that turned out to be practical to be applied not only in Malaysia but also globally. The statistics clearly show that the

risk index used has very good accuracy of over 90% and this has proved that the reliability of the risk index that was developed is at a maximum. Predictive risk index for the future has also been produced and can be used by DID for planning methods and appropriate action plans to control flooding in the country. The application of integrated Chemometric, Time Series Analysis and Artificial Neural Network makes the main factor of flood 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. The classification of risk should be based on the changes of variables that become the main factor for changes [24] of water level that will able to cause flooding at the Malaysia River Basin.

5. REFERENCES

[1] Bryant E. A. Natural hazards. England: Cambridge University Press, 2005

[2] Balek J. Hydrology and water resources in tropical regions. Amsterdam: Elsevier, 1983

[3] Kesel R H. Human modifications to the sediment regime of the Lower Mississippi River flood plain. Geomorphology, 2003, 56(3):325-334

[4] Department of Irrigation and Drainage (DID) Malaysia. Flood management manual.Kuala Lumpur: DID, 2009

[5] Gilissen H K, Alexander M, Beyers J C, Chmielewski P, Matczak P, Schellenberger T, Suykens C. Bridges over troubled waters: an interdisciplinary framework for evaluating the interconnectedness within fragmented flood risk management systems. Journal of Water Law, 2016, 25(1):12-26

[6] Chan N.Chapter 29-Reducing flood hazard exposure and vulnerability in Peninsular Malaysia.In D. J. Parker (Ed.), Floods. London: Routledge, 2014, pp. 19-30

[7] Saudi A S, Juahir H, Azid A, Amri M K, Kamarudin M E, Aziz N A. Flood risk pattern recognition using chemometric technique: A case study in Muda River Basin. Computational Water, Energy, and Environmental Engineering, 2014, 3(3):102-110

[8] Japan Internation Cooperation Agency (JICA). Comprehensive management plan of Muda River Basin.Final report, Tokyo: JICA, 1995

[9] Julien P Y, Ghani A A, Zakaria N A, Abdullah R, Kiat C C, Ramli R, Dinor J, Manap A A A, Yusof M F. Design option of the flood mitigation plan of Sg. Muda, Sungai Muda, Kedah.

Final report, Pulau Pinang: Universiti Sains Malaysia, 2006

[10]Chan N W. Impacts of disasters and disaster risk management in Malaysia: The case of floods. In D. Aldrich, S. Oum, & Y. Sawada (Eds.), Resilience and recovery in Asian disasters. Tokyo: Springer, 2015, pp. 239-265

[11]Jackson J E. A user's guide to principal components. New York: Wiley, 1991

[12]Meglen R R. Examining large databases: a chemometric approach using principal component analysis. Marine Chemistry, 1992, 39(1-3):217-237

[13]Azid A, Juahir H, Aris AZ, Toriman ME, Latif MT, Zain SM, Yusof KM, Saudi AS. Spatial analysis of the air pollutant index in the southern region of Peninsular Malaysia using environmetric techniques. In A. Aris, T. T. Ismail, R. Harun, A. Abdullah, & M. Ishak (Eds.), From sources to solution. Singapore: Springer, 2014, pp. 307-312

[14]Sultana F, Razive N I, Azeem A. Implementation of statistical process control (SPC) for manufacturing performance improvement. Journal of Mechanical Engineering, 2009, 40(1):15-21

[15]MIng K L, Jais M, Karim B A. Is technical analysis profitable and capable for stock price prediction? Evidence from Malaysian stock market. UNIMAS Review of Finance and Accounting, 2016, 1(1):79-98

[16]Chaudhuri T D, Ghosh I. Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework. Journal of Insurance and Financial Management, 2016, 1:92-123

[17]Nasir M F, Juahir H, Roslan N, Mohd I, Shafie N A, Ramli N. Artificial neural networks combined with sensitivity analysis as a prediction model for water quality index in Juru River, Malaysia. International Journal of Environmental Protection. 2011, 1:1-8

[18]Yong A G, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology, 2013, 9(2):79-94

[19]Yıldırım G, Tokalıoğlu Ş. Heavy metal speciation in various grain sizes of industrially contaminated street dust using multivariate statistical analysis. Ecotoxicology and Environmental Safety, 2016, 124:369-376

[20]Tziritis E, Skordas K, Kelepertsis A. The use of hydrogeochemical analyses and multivariate statistics for the characterization of groundwater resources in a complex aquifer

system. A case study in Amyros River basin, Thessaly, central Greece. Environmental Earth Sciences, 2016, 75(4):1-11

[21] Gharibreza M, Raj J K, Yusoff I, Ashraf M A, Othman Z, Tahir W Z. Effects of agricultural projects on nutrient levels in Lake Bera (Tasek Bera), Peninsular Malaysia. Agriculture, Ecosystems and Environment, 2013, 165:19-27

[22] Wahab N A, Kamarudin M K, Gasim M B, Umar R, Ata F M, Sulaiman N H. Assessment of Total Suspended Sediment and Bed Sediment Grains in Upstream Areas of Lata Berangin, Terengganu. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):757-763

[23] Abdullahi M G, Kamarudin M K, Toriman M E, Gasim M B, Endut A, Garba I. Assessment of natural groundwater recharge in Terengganu, Malaysia. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):781-786

[24] Kamarudin MK, Nalado AM, Kasmuri A, Toriman ME, Juahir H, Umar R, Jamil NR, Saudi AS, Rizman ZI, Gasim MB, Hassan AR. Assessment of river plan changes in Terengganu River using RS and GIS method. Journal of Fundamental and Applied Sciences, 2017, 9(2S):28-45

[25] Saudi A S, Kamarudin M K, Ridzuan I S, Ishak R, Azid A, Rizman Z I. Flood risk index pattern assessment: Case study in Langat River Basin. Journal of Fundamental and Applied Sciences, 2017, 9(2S):12-27