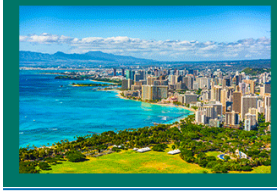
Investigation of flood pattern using ANOVA statistic and remote sensing in Malaysia



IOP Conf. Series: Earth and Environmental Science **20** (2014) 012030 doi:10.1088/1755-1315/20/1/012030

Investigation of flood pattern using ANOVA statistic and remote sensing in Malaysia

Norsuzila Ya'acob [1, 2,*], Nor Syazwani Ismail[1], Norfazira Mustafa[1] and Azita Laily Yusof[1, 2]

[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam Selangor, Malaysia
[2]Wireless Communication Technology (WiCoT), Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*Institute of Space Science, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.

norsuzilayaacob@yahoo.com,wani_nureen@yahoo.com,faziramustafa@yahoo.com, azita968@salam.uitm.edu.my

Abstract. Flood is an overflow or inundation that comes from river or other body of water and causes or threatens damages. In Malaysia, there are no formal categorization of flood but often broadly categorized as monsoonal, flash or tidal floods. This project will be focus on flood causes by monsoon. For the last few years, the number of extreme flood was occurred and brings great economic impact. The extreme weather pattern is the main sector contributes for this phenomenon. In 2010, several districts in the states of Kedah neighbour-hoods state have been hit by floods and it is caused by tremendous weather pattern. During this tragedy, the ratio of the rainfalls volume was not fixed for every region, and the flood happened when the amount of water increase rapidly and start to overflow. This is the main objective why this project has been carried out, and the analysis data has been done from August until October in 2010. The investigation was done to find the possibility correlation pattern parameters related to the flood. ANOVA statistic was used to calculate the percentage of parameters was involved and Regression and correlation calculate the strength of coefficient among parameters related to the flood while remote sensing image was used for validation between the calculation accuracy. According to the results, the prediction is successful as the coefficient of relation in flood event is 0.912 and proved by Terra-SAR image on 4th November 2010. The rates of change in weather pattern give the impact to the flood.

1. Introduction

Nowadays, at the rural area or urban area were face with the natural flood and flash flood in the certain periods. The flood happened without any sign to show that it will happen. The residents basically were not alert and beware about the flood and impact to them. Normally, they alert about flood disaster. Moreover, the increasing of development and human activities can have significant influence on the process that contributes to disaster of flood. To avoid this problem, the suitable technologies can be developed to control and prevent the flood risk and impacts for short or long terms [1]-[3]. In the flood prediction, remote sensing is been reliable tools and can be used in prediction process. It provided the researcher with a sort of multi-date satellite images for monitoring and recording the changes of process of past flood events.

Remote sensing can be defined as combination of information from the places such as distance which occur with the camera device based on the ground or sensors or cameras based on ships, aircraft, satellites, or other spacecraft. In archaeology, it generally refers to the use of aerial or satellite imagery to study cultural patterning on the ground. Aerial photography has been long used by archaeologists, but recent development of new sensing and data processing technologies has led to new possibilities for archaeological studies.

In this project, Terra-SAR image was used to compare the prediction. TerraSAR-X is a high resolution X-band radar sensor with three operational models such as Spotlight, Strip Map and Scans AR. Spotlight is the most sophisticated radar imagery available which has 1 m spatial resolution. Its steerable radar beam in flight direction (like a spotlight) illuminates the ground for the longest time possible to ensure this high resolution. For large area coverage the Scans AR mode (16 m resolution) allows areas up to 100 000 km^2 to be acquired within a week owing to its large scene size [4]-[7]. The main focus of this project are to identify the parameters involve in the prediction of flood followed by investigate the pattern and analyses or compare them with the previous flood.

2. Methodology

The main parameter for this research will focused on temperature, humidity, rainfalls and wind speed value were collect at station involved. The image was captured by Terra-SAR satellite for validation.

2.2 Data Collection

Data includes the amount off rainfalls, wind speed, temperature and humidity in Titi Gajah area as shown in Figure 1. The data of amount of rainfalls are covered from three stations such as Bukit Kayu Hitam, Kepala Batas and Alor Setar stations that nearest to the study area. For the humidity, wind speed and temperature data were taken the mean relatives for daily reading and covered for period between Augusts until November 2010 in the same station. The information the flood occurred were collected from the local newspapers and also the images from TERA-SAR at the Malaysian Remote Sensing Agency (ARSM)

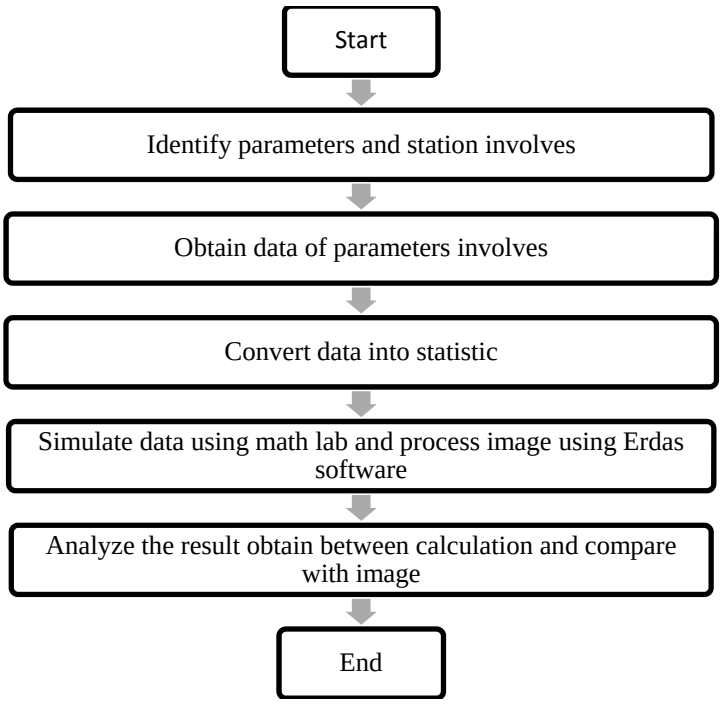


Figure 1: Flowchart of Reseach Methodology.

2.3 Analytical Techniques

In terms of the analysis part, the statistic method was used to preliminary treatment of data. The ANOVA statistic was chosen as a method to calculate the analysis for statistic. ANOVA is the techniques that used the two hypotheses to determine the results and namely as null hypothesis *(H_0)* which the assumption under investigation being tested. The H_0 is statement that "there is no effect" or "there is no change". The possible outcomes in testing a H_0 are "reject" or "fail to reject". Reject the H_0 means that there is significant evidence in the data and in the test to justify reject H_0. When H_o was rejected the data is said to be statically significant. Fail to reject H_0 means that there is NOT enough evidence in the data and the test to justify reject H_0. The alternate hypothesis *(H_1)* state that it adopt if there is strong evidence (sample data) against the null hypothesis *(H_0)*[8]. There are equations involved in ANOVA and showed in Table 1.

Table 1. ANOVA Statistic Table.

Source	**Degree of freedom**	**Means square**	**F statistic**
A	a-1	SSa/dfA	MSA/MSB
C	c-1	SSb/dfB	MSC/MSB
AC	(a-1)(c-1)	SSac/dfAC	MSAC/MSB
B	N-ac	SSb/dfB	

Where:

a= Number of levels of independent variable A.

c= Number of levels of independent variable C.

ac= Number of cell in the experiment.

N= Total number of observations in experiment

n_1= Number of observations in cell 1, etc.

$$SS_{total} = \sum X^2 - \frac{(\sum X)2}{N} \quad (1)$$

$$SS_A = \sum \frac{\sum(for\ each\ row\ )^2}{n\ for\ each\ row} - \frac{(\sum X)^2}{N} \quad (2)$$

$$SS_C = \sum \frac{\sum(for\ each\ column\ )^2}{n\ for\ each\ column} - \frac{(\sum X)^2}{N} \quad (3)$$

$$SS_{between} = \frac{(\sum X_1)^2}{n_1} + \cdots \frac{(\sum X_{ac})^2}{n_{ac}} - \frac{(\sum X)^2}{N} \quad (4)$$

$$SS_{AC} = SS_B - SS_A - SS_C \quad (5)$$

$$SS_B = SS_{total} - SS_{between} \quad (6)$$

It is very difficult to predict flood level since it fluctuate in high non-linear. The precision or complexity of traditional measures such as the hydrological statistical method and the multi-dimension mixed regression algorithm is not always satisfactory.

3. Result

This part will discuss the results and analyses from the parameters involved and their relation with the flood in November 2010 at Titi Gajah area.

3.1Parameters Analysis

There are parameters were identified related to the floods which are:

3.1.1 Temperature

Actually, the change of one weather elements can produce the changes in regional climate [6]. The increasing of temperature will be affected the amount of clouds and amount of rainfalls. Table 2 showed that the value of $F_{statistic}$ calculation 0.99 is greater than the $F_{probability}$ value 0.4887. It can be proved that the temperature also apart of parameter involved in the flood.

Table 2. ANOVA statistic of temperature.

Source	Degree of freedom	Means square	F statistic	F probability
Months	2	5592.86	16.46	0
Days	29	418.69	0.99	0.4887
Errors	87	427.88		
Total	119			

In Malaysia, the average of temperature is between 23˚C until 32˚C. If the value is more than that it will consider as extreme temperature. Figure 2 will showed that the range of temperature in August until November 2010.

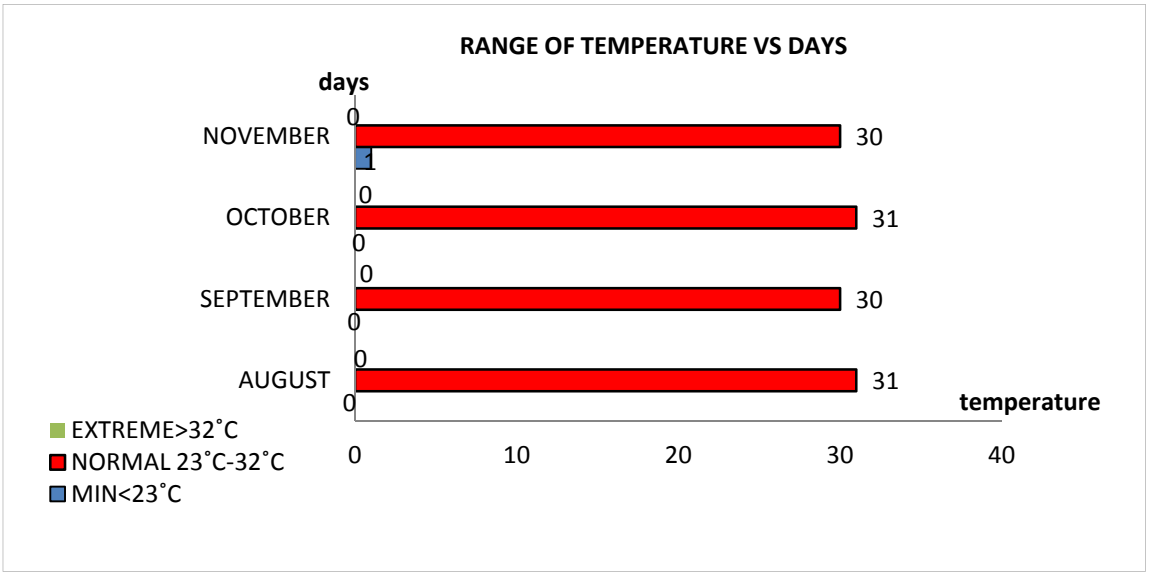


Figure 2. Range of temperature.

The extreme changes of temperature give the effect to the surrounding such as vaporization, wind speed, amount of rainfalls and also the humidity. When the temperature decrease from the normal temperature, La Nina phenomena will be happened and cause the heavy rainfall to the area involved. Based on the La-Nina Phenomena, it can be related that between August to October, Malaysia in wet and cold condition period. Addition to the monsoon season, the changing temperature gives the effect to pattern of rainfall in Titi Gajah Area. It can be concluded here that the small changes in temperature either decrease or increase were effect to the surrounding. Since ANOVA statistic showed 0.99, it can be concluded that the temperature reading has relation to the flood.

3.1.2 Humidity

Air's water vapours carrying capacity increases with increasing temperature, and decreases with decreasing temperature. Referring to Table 3, Statistic Prediction showed the probability the humidity was involved which $F_{statistic}$ value is 0.98 compared to the $F_{probability}$ 0.5082.

Table 3. ANOVA Statistic of humidity.

Source	Degree of freedom	Means square	F statistic	F probability
Months	2	559.86	13.07	0
Days	29	418.69	0.98	0.5082
Errors	87	427.88		
Total	119			

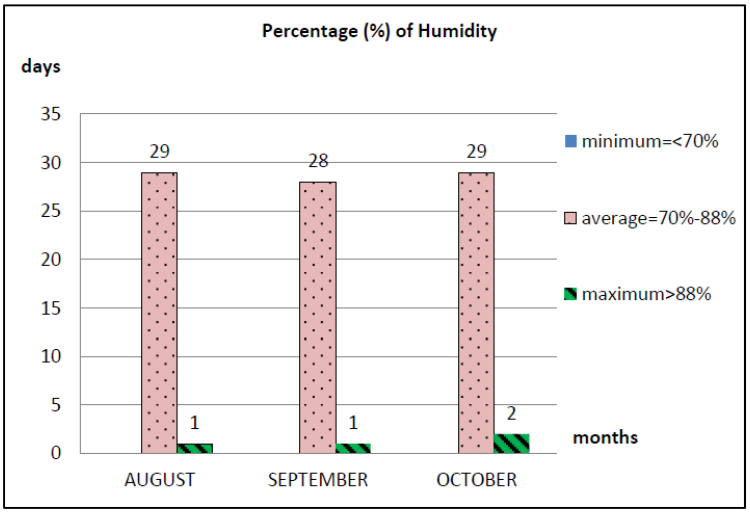


Figure 3. Percentage of humidity versus days.

According to Figure 3, the maximum readings of humidity were records every month. The maximum reading showed that heavy rains were happened on the days and surrounding becomes wet. When humidity becomes higher, process of vaporization will be increase rapidly and make the cloud becomes heavy. Then extreme rainfalls will be happened quickly. Titi Gajah area was surrounded by paddy field area. The maximum range was found in the north-western peninsular (Alor Star) where the average relative humidity varies from 72% to 87% in February.

It noted that in Peninsular Malaysia, the minimum relative humidity is usually available in January and February, but for the east coast states, namely Kelantan and Terengganu, where the relative humidity is a minimum in March. The maximum relative humidity was typically is in November. As in the case of temperature, diurnal variation of relative humidity is greater than the annual change.

The average daily minimum can reach as low as 42% during the dry months and as high as 70% during humid months. However, the mean daily maximum did not change much from one place to another, which is more than 94% and may reach as high as 100%. States in the northwest, the northern Kedah and Perlis and Sabah have biggest daily humidity changes. It can be conclude that the increase percentage of humidity proportional due to the temperature reading. Both of elements were given the impact to the flood. The increases of humidity also effect to the rainfalls pattern in Titi Gajah.

3.1.3 Rainfalls

When the humidity was increase, quantities of rainfalls also increase because the vaporization process. The probability it was involved can be shown in Table 4 analysis, system showed that the value of $F_{statistic}$ calculation is 0.92 which is greater than the $F_{probability}$ value. The null hypothesis, *Ho* state that rainfalls are not involved in flood pattern is rejected. It showed that the data is the concrete evident for the flood pattern.

Table 4. ANOVA Statistic of rainfalls.

Source	Degree of freedom	Means square	F statistic	F probability
Months	2	1332.45	3.14	0.264
Days	65	391.01	0.92	0.6425
Errors	195	424.23		
Total	263			

Heavy rains induce flooding when they occur about three or more times in a month during the period [1]-[3]. Based on Figure 4, Bukit Kayu Hitam station showed that the highest reading of rainfalls 263 mm on the October and 393 mm on November followed by Alor Setar station 344 mm on October.

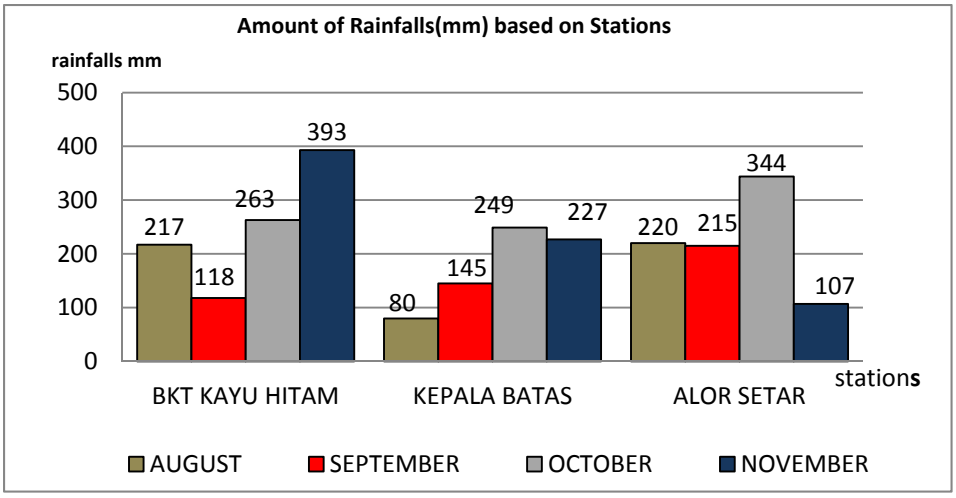


Figure 4. Amount of rainfalls based on station.

The extreme rainfalls were record in Bukit Kayu Hitam and made overflow from river nearest to Bukit Kayu Hitam and at the same time Titi Gajah also sustain heavy rain and caused the flood. The rainfall pattern in the area other than the west coast of Peninsular Malaysia showed two maximum rain period separated by two periods of minimum rainfall. Main maximum usually occurs in October and November, while the secondary maximum occurring in April to May. In the northwest, the main minimum occurs in January and February, while the secondary minimum occurs in June and July. In other places, the primary minimum occurs in June and July, while the secondary minimum occurs in February. It can be conclude the heavy and very heavy rain can be happen when the increases rapidly of the temperature readings and percentage of humidity in the air.

3.1.4 Wind Speed

Lastly, the element was identified and give the impact to the flood is wind speed. In Northeast Monsoon occurs in November until March of the year accompanied with rumbustious and waterspout. Table 5 below showed the statistic of wind speed involved. Based on Table 5, the prediction also give the result that wind speed is a part of parameter involved in the flood. Statistic results showed that the value of $F_{statistic}$ calculation 1.33 is greater than the $F_{probability}$ value of 0.1544.

Table 5. ANOVA Statistic of wind speed.

Source	Degree of freedom	Means square	F statistic	F probability
Months	2	6.19697	14.51	0
Days	29	0.56951	1.33	0.1544
Errors	87	0.42697		
Total	119			

The northeast monsoon season usually commences in early November and ends in March. During this season, steady easterly or north-easterly winds of 10 to 20 knots prevail. The winds over the east coast states of Peninsular Malaysia may reach 30 knots or more during periods of strong surges of cold air from the north (cold surges). In Malaysia, Wind movement will change the moving ocean. Low air pressure occurs due to an increase in the surface temperature of the sea water. This causes the wind patterns change from the sea shore and air moisture accumulated and subsequently causes the phenomenon of heavy rain.

As Malaysia is mainly a maritime country, the effect of land and sea breezes on the general wind flow pattern is very marked especially during days with clear skies. On bright sunny afternoons, sea breezes of 10 to 15 knots very often develop and reach up to several tens of kilometres inland. On clear nights, the reverse process takes place and land breezes of weaker strength can also develop over the coastal areas.

3.2 Total Prediction Pattern

To prove that all the parameters have relationship due to flood, all the parameters data were combined. Table 6 showed the result of combination of all parameters in flood prediction. According to Figure 5, the temperature was donating the higher percentage relation to the flood followed by Humidity then Rainfalls and lastly winds speed. The heavy cloud will be move according to wind speed from high pressure to low pressure then make the extreme rainfalls to the area involve. If the extreme rainfalls were happened in several days, chances flood might be happened also increase.

Table 6. Combination results of all parameters.

Source	Degree of freedom	Means square	F statistic	F probability
Months	2	127523	445.45	0
Days	119	443.4	1.55	0.0012
Errors	357	286.3		
Total	479			

According to Figure 5, the temperature was donating the higher percentage relation to the flood followed by humidity then Rainfalls and lastly winds speed. The climate changing gives the impact to the flood for the example when the temperature increase in 1 degree Celsius, the humidity in the surrounding also increases proportionally. It happened because the process of vaporization will happened rapidly compared to usually. The heavy cloud will be move according to wind speed from high pressure to low pressure then make the extreme rainfalls to the area involve. If the extreme rainfalls were happened in several days, chances flood might be happened also increase.

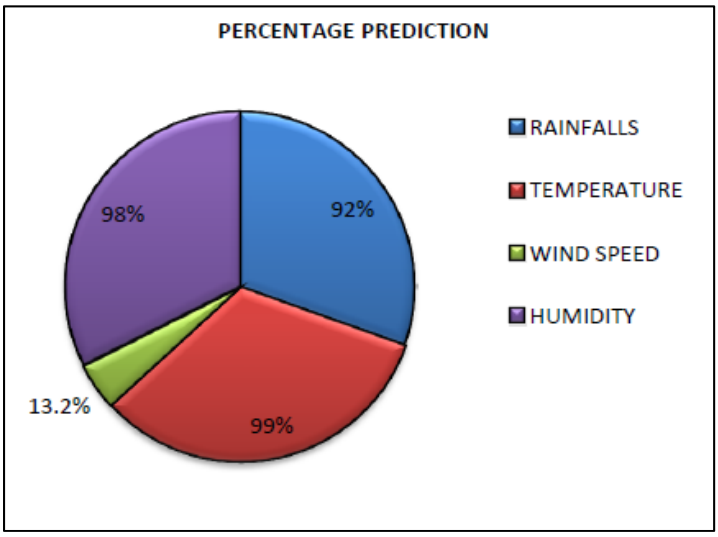


Figure 5. Percentage of parameters in flood pattern.

3.3 Regression and Correlation Test

After applying the implicitly gain in the prediction, the values of R^2 and r^2 are equal as shown in Figure 6. So the data model and the correlation have the same values of 83.21. To get the coefficient of the prediction, the values of correlation must be square root and the actual value of the coefficient is 0.912. According to coefficient table, the strength of coefficient can be classified as very high coefficient.

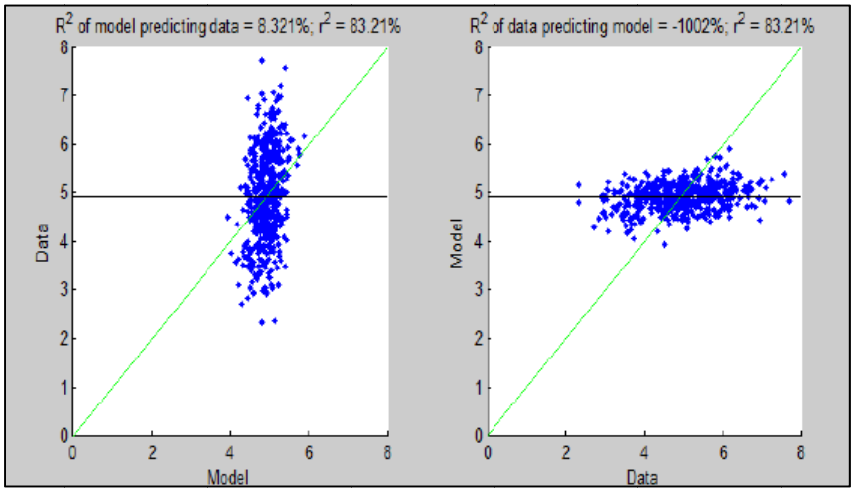


Figure 6. Values of R^2 and r^2 after implicitly gain.

3.4 Validation Image to Calculation

Image from Terra-SAR satellite showed that the Flood happen on November 2010 in Kampung Titi Gajah as shown in Figure 7. Based on the all analysis on the correlation of the parameters involved in the flood on 4th November 2010, it can be conclude that all the parameters show strong correlation. Since the value of coefficient is 0.912, all the parameters give the impact to the flood. The correlation pattern also gives impact if there is a change in one of the elements.

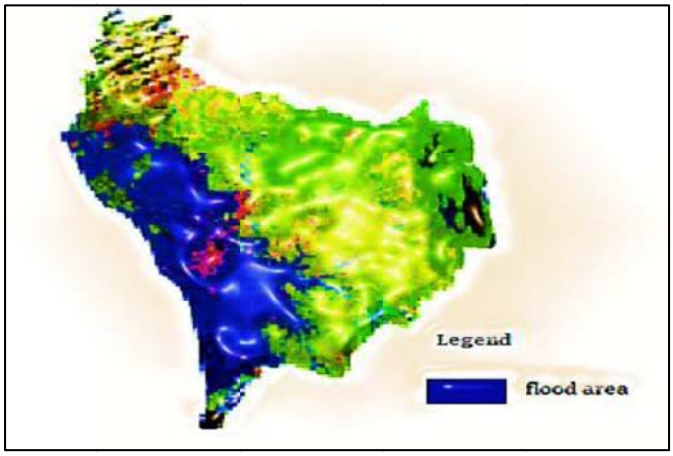


Figure 7. Terra-SAR satellite image on 4th November 2010.

4. Conclusion

It can be conclude that the rainfalls, temperature, wind speed and the humidity are the parameters involved in the process of flood. Each the element in space weather have the relation to each other. The statistic in show that each parameter satisfied to the target result less than 1. The changes in temperature will give effect to the humidity reading then followed by amount of rainfalls and lastly the wind speed at the surrounding was involved. Since the coefficient of pattern is 0.912, the correlation is very high strength and it is successful. Then the result from image of Terra-SAR satellite is a concrete evident showing that in 4th November 2010, there was flood happened at that time.

Acknowledgment
This research is sponsored by Universiti Teknologi MARA (UiTM) through the university research grant 600 RMI/DANA/5/3/RIF (194/2012). The appreciation also goes to Malaysian Remote Sensing Agency (ARSM), Department of Irrigation and Drainage Malaysia and Malaysia Meteorology Department for provided the discharge history data to complete this project.

References
[1] Akintola, F O (1986). Rainfall Distribution in Nigeria (1892 – 1983). Ibadan: Impact Publishers Nig.Ltd.
[2] Ayoade, J O (1976). A Preliminary Study of the Magnitude, Frequency and Distribution of Intense Rainfall in Nigeria. Hydrological Science Bulletin
[3] Babatola, J S (1996). Recent Changes in Rainfall Patterns and its Implication for Flood Occurrence in Ondo, Nigeria. Ondo Journal of Arts and Social Sciences, 1 (1), pp. 125 – 136.
[4] J. Mittermayer, M. Younis, B. Bräutigam, T. Fritz, R. Kahle, R. Metzig, Verification of TerraSAR-X System. IEEE Geoscience and Remote Sensing Symposium (IGARSS), Barcelona, Spain, 2007 .
[5] R. Metzig, B. Bräutigam, D. Polimeni, J. Böer, M. Bachmann, J.Mittermayer. TerraSAR-X Instrument Characterization / Verification. Accepted at: 7th European Conference on Synthetic Aperture Radar, Friedrichshafen, Germany, 2008.
[6] J. Mittermayer, U. Steinbrecher, A. Meta, N. Tous-Ramon, S. Wollstadt,M. Younis, J. Marquez, D. Schulze, C. Ortega, TerraSAR-X System Performance and Command Generation. Accepted at: 7th European Conference on Synthetic Aperture Radar, Friedrichshafen, Germany, 2008.
[7] J. Marquez-Martinez, C. Gonzalez, M. Younis, S. Wollstadt, R. Metzig,U. Steinbrecher, N. Tous-Ramon, A. Meta, J. Mittermayer, In-Orbit SAR Performance of TerraSAR-X. In: IEEE Geoscience and Remote SensingSymposium (IGARSS), Barcelona, Spain, 2007.
[8] Campolo, M., Soldati, A., and Andreussi, P. 2003. Artificial Neural Network Approach to flood Forecasting in the River Arno. Hydrological Sciences Journal/ Journal des Sciences Hydrologiques, Vol. 48(3), pp. 381-398
[9] T. Back, Evolutionary Algorithms in Theory and Practice, Oxford University Press, 1998.
[10] Campolo, M., Soldati, A., and Andreussi, P. 2003. Artificial Neural Network Approach to flood forecasting in the River Arno, Hydrological Science Journal/ Journal des Sciences Hdrologiques,Vol.48(3),pp.381-398