Journal of Spatial Hydrology
Vol.9, No.2 Fall 2009

Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing

Biswajeet Pradhan

Abstract

Recently, in the year 2006, 2007 and 2008 heavy monsoons rainfall have triggered floods along Malaysia's east coast as well as in different parts of the country. The hardest hit areas are along the east coast of peninsular Malaysia in the states of Kelantan, Terengganu and Pahang. The flood cost nearly millions of dollars of property and many lives. Foods are considered to be one of the weather-related natural disasters. Many methods exist to provide qualitative estimations of the risk level of flood susceptibility mapping within a watershed. This paper presents construction of a flood susceptible map for presumptive flood areas around at Kelantan river basin in Malaysia using a statistical model and GIS. To evaluate the factors related to flood susceptible analysis, a spatial database was constructed from a topographical map, geological map, hydrological map, Global Positioning System (GPS) data, land cover map, digital elevation model (DEM) data, and precipitation data. An attribute database was also constructed from field investigations and historical flood areas reports for the study area. Logistic regression model was applied to determine each factor's rating, and the ratings were overlaid for flood susceptibility mapping. Results indicate that flood prone areas can be performed at 1:25,000 which is comparable to some conventional flood hazard map scales. The flood prone areas delineated on these maps correspond to areas that would be inundated by significant flooding. Further, risk analysis has been performed using DEM, distance from hazard zone, land cover map and damageable objects at risk. DEM was used to delineate the catchments and served as a mask to extract the highest hazard zones of the landslide area. Qualitatively, the model seems to give reasonable results with accuracy observed was 85%.

Key words: Flood susceptibility analysis; logistic regression model; GIS; Remote Sensing

1. Introduction

Natural disasters are happened every year and their impact and frequency seem to have greatly increased in recent decades, mostly because of environmental degradation, such as deforestation, intensified land use, and the increasing population (Vincent, 1997). Floods are among the most frequent and costly natural disasters in terms of human and economic loss. As much as 90 percent of the damage related to natural disasters in Malaysia is caused by flood. Average annual flood damage is as high as US100 millions. These flooding have caused considerable damage to highways, settlement, agriculture and livelihood. In Malaysia, floods are caused by a combination of natural and human factors. Malaysians are historically river dwellers as early settlements grew on the banks of the major rivers in the peninsula. Coupled with natural factors such as heavy monsoon rainfall, intense convection rain storms, poor drainage and other local factors, floods have become a common feature in the lives of a significant number of Malaysians. Monsoon rains have a profound influence on many aspects of the lives of the people in the east coast of Peninsular Malaysia (Chan, 1995). While the rains are needed for agriculture, particularly wet rice cultivation, they are also largely responsible for bringing seasonal floods. Recently, in 2006 and 2007 heavy monsoons rainfall have triggered floods along Malaysia's east coast as well as in southern state of Johor. The hardest hit areas are along the east coast of peninsular Malaysia in the states of Kelantan, Terengganu and Pahang. The city of Johor was particularly hard hit in southern side. The flood cost nearly million dollars of property and many lives. The

Institute of Cartography, Faculty of Forestry, Hydro and Geosciences; Dresden University of Technology, 01062 Dresden, Germany, Tel: +49-351 463 33099; Fax: +49-351 463 37028
Biswajeet.Pradhan@mailbox.tu-dresden.de / biswajeet@mailcity.com

extent of damage could have been reduced or minimized if an early warning system would have been in place.

Problems related to flooding have greatly increased, and there is a need for an effective modelling to understand the problem and mitigate its disastrous effects. Human activities such as unplanned rapid settlement development, uncontrolled construction of buildings in general and major landuse changes can influence the spatial and temporal pattern of hazards. There are several factors contributing to the flooding problem ranging from topography, geomorphology, drainage, engineering structures, and climate. Most floods are caused by storms in which a lot of precipitation falls in a short period of time, of both types of rainfall, convective and frontal storms. Intensity and duration of the rain are the most influencing factors for flood hazards. In the recent years, remote sensing and Geographic information systems have been embedded in the evaluation of the geo-environmental hazards. According to Verstappen (1995), the purposes of using remote sensing include: "to investigate the susceptibility of the land and the vulnerability of the society, to construct hazard zoning maps and potential damage maps, to monitor potential hazards, and to deal with emergency situations after a disaster." Many research studies have been completed that employ remote sensing as the principal information source in the assessment of hazards/disasters. There have been many studies on flood susceptibility mapping using remote sensing data and GIS tools. Radar remote sensing data have been extensively used for flood monitoring across globe (Hess et al., 1995; Hess et al., 1990; Le Toan et al., 1997) and many of these studies have applied using probabilistic methods (Landau et al., 2000; Farajzadeh, 2001, 2002; Horritt and Bates, 2002; Pradhan and Shaffie, 2009; Pradhan et al., 2008, Pradhan and Lee, 2009). Logistic regression model, has also been applied to other natural hazard modeling such as landslide hazard and susceptibility mapping (Atkinson and Massari 1998; Dai et al. 2001; Dai and Lee 2002; Ohlmacher and Davis 2003; Lee and Pradhan, 2007). Hydrological and stochastic rainfall method for flood susceptibility mapping has been employed in other areas (Blazkova and Beven, 1997; Cunderlik and Burn, 2002; Ebisemiju, 1986; Haeng et al., 2001; Nageshwar and Bhagabat, 1997; Yakoo et al., 2001 and Villiers, 1986). Flood susceptibility mapping using GIS and neural network methods have been applied in various case studies (Honda et al., 1997; Islam and Sadu, 2001, 2002; Sanyal and Lu, 2004, 2005; Townsend and Walsh, 2005; Wadge et al., 1993; Tambunan, 2007; Profeti and Machintosh, 1997; Knebl et al., 2005; Masmoudi and Habajeb, 1993; Sinnakaudan et al., 2003; Merwade et al., 2008; Zerger, 2001).

In this paper, remote sensing data along with other tabular and meta data were used to delineate the flood susceptibility mapping for the part of the Kelantan river basin. Terrain information such as historical flooded areas extracted from RADARSAT images, DEM, slope, aspect, curvature, distance from drainage, flow direction, flow accumulation, soil, land cover, soil texture, and precipitation information have been updated to enable the quantification of flood associated attributes. Flood susceptibility mapping has been applied using logistic regression model. Further, risk analysis was carried out using the output from the flood susceptibility analysis and the socio-economic parameters in GIS environment.

2. Study area

Kelantan River Basin is selected as a pilot area because it represents typical basins and flood plains that are prone to annual monsoon floods in Malaysia. The study (Figure 1) area is part of Kelantan state which is one of the 13 states of Malaysia. Kelantan River is the major river in Kelantan state and is located in the North-East part of Peninsular Malaysia. The Kelantan River emerges at the confluence of the Galas river and Lebir river near Kuala Krai and meanders over the coastal plain until it finally debauches into the South China Sea, about 12 km north of Kota Bahru. The main reach of the Kelantan River has some further larger tributaries downstream. However, the Galas and the Lebir rivers themselves have many tributaries, which provide the majority of the flow in the main Kelantan River. These tributaries rise into the forested mountains of peninsular Malaysia. The basin covers 85 percent of the state's surface area. The river only drops 10 meters from the coast up to Guillemard Bridge with the distance of 60 km. The main river comprises of seven major Subcatchments (Kota Bahru, Gullimard, Pergau, Kuala Krai, Galas, Lebir, and Nenggri) that covers a drainage area of 13,170 km^2. Four major towns are located along the river: Kota Bahru, Pasir Mas, Tumpat, and Kuala Krai. Kota Bharu is the main city and centre of commercial trade and administration in the Kelantan state. Its ideal geographical location also makes it a gateway to the neighbouring countries of southern Thailand, and is thus a city with high tourism potential for domestic and foreign tourists who visit year-around. Due to its geographical characteristics; unplanned urbanization; and proximity to the South China sea, Kota Bharu has become extremely vulnerable to monsoon floods every year. The unprecedented flooding of November 2005, triggered by monsoon rains, has been described as one of the worst natural flood in the history of Kota Bharu.

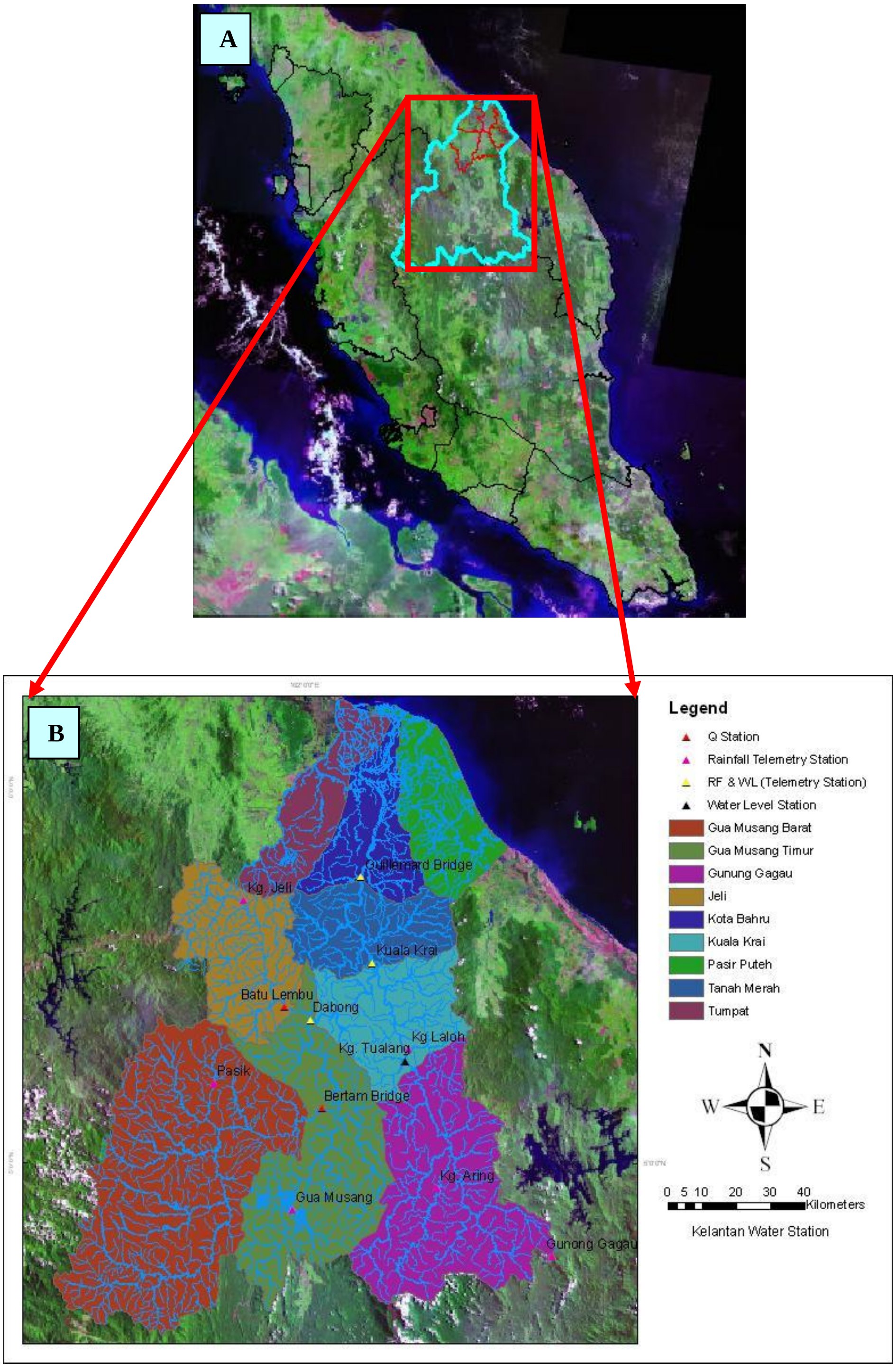


Fig.1: Study area location map; (a) satellite image of Peninsular Malaysia; (b) SPOT 5 satellite image with major catchment map of Kelantan state

Kelantan River Basin has a tropical climate receiving rainfall throughout the year. The average annual rainfall varies between 0 mm in the dry season (March-May) to 1,750 mm in the monsoon season (November – January). The average runoff for the Kelantan River basin is about five hundred (500) m^3/s. The average precipitation in the Kelantan river basin is approximately 2,500 mm/year while regional studies show that the average combined loss due to interception and evapotranspiration is about 1,200 mm/year. The resulting average runoff is therefore approximately 1300 mm/year. The average temperature is approximately 28°C at Kota Bahru. The basin is densely vegetated over the scraps and also on the valley side slopes.

In general, while localized flooding is mostly due to convectional rain storms, most of the extensive and severe floods in peninsular Malaysia are associated with the onset of the monsoon seasons. Seasonal floods caused by heavy rains during the North-East monsoon period are termed “monsoon floods”. Consequently, it is not surprising that in terms of flood frequency and magnitude, the four east coast states are the most susceptible to flooding, each experiencing various magnitudes of flood occurrence almost every year. As a result of seasonal floods occurring almost annually in one part of the peninsula or another, flood losses in terms of loss of lives and damages to properties are substantial.

The soil cover is a meter or so deep but depths of more than 18 meter may be encountered in localized areas. A fine sandy loam soil is found in the extreme east and west of the southern half of the basin. Its depth seldom exceeds a few meters. The remaining portion, comprising almost one-third of the catchment, is cloaked by a variable soil cover that varies in depth, from a few meters to more than 9 meter. The cultivation is relatively good, limited to the plains only. From a hydrological point of view, the Kelantan River Basin is made up of flat slope and moderately sloping areas. There are large level plains on the southern side and also in the south west. The steep scraps and the high slopes in the southern part of the river basin can be contributed to the major run-off zone to the Kelantan River. The drainage of the area shows a dendritic pattern in most part of the region.

Record shows that, Kelantan River regularly over spills its banks during the northeast monsoon. To monitor and issue warning for the rising water from the river, there are numerous monitoring stations installed around the Kelantan river catchment. Drainage and Irrigation Department has carried out forecast and warning activities considering hydrological parameters. However it is crucial to improve the level of preparedness and effectiveness in disaster response using real-time disaster and spatial information obtained from remote sensing and related technologies.

3. Methodology

3.1 Spatial database creation

Accurate detection of the historical flood extent is very important for statistical flood susceptibility analysis. Couple of RADARSAT images were obtained significant and cost-effective information on historical flood extent. To apply the probabilistic method, a spatial database that considers flood-related factors was designed and constructed. These data are available in Malaysia either as paper or as digital maps. The spatial database constructed is shown in Table 1. There were nine factors that were considered in calculating the probability, and the factors were extracted from the constructed spatial database. The factors were transformed into a grid spatial database using the GIS, and flood-related factors were extracted using the database. Figure 2 shows the overall methodology adopted in this analysis.

Table 1 Data layer of study area

Parameters	Sub-Classification	GIS Data Type	Scale
Historical flooded areas	Flood extent	Polygon coverage (Derived from RADARSAT images)	10 m x 10m
Basic Map	Topographic Map (DEM)	Line and Point coverage	1: 25, 000
	Slope in degrees	GRID	10 m x 10m
	Curvature	GRID	10 m x 10m
	Flow direction	GRID	10 m x 10m
	Flow accumulation	GRID	10 m x 10m
	Land Cover	GRID	10 m x 10m
	Soil	GRID	10 m x 10m
	Precipitation in cm	GRID	10 m × 10m

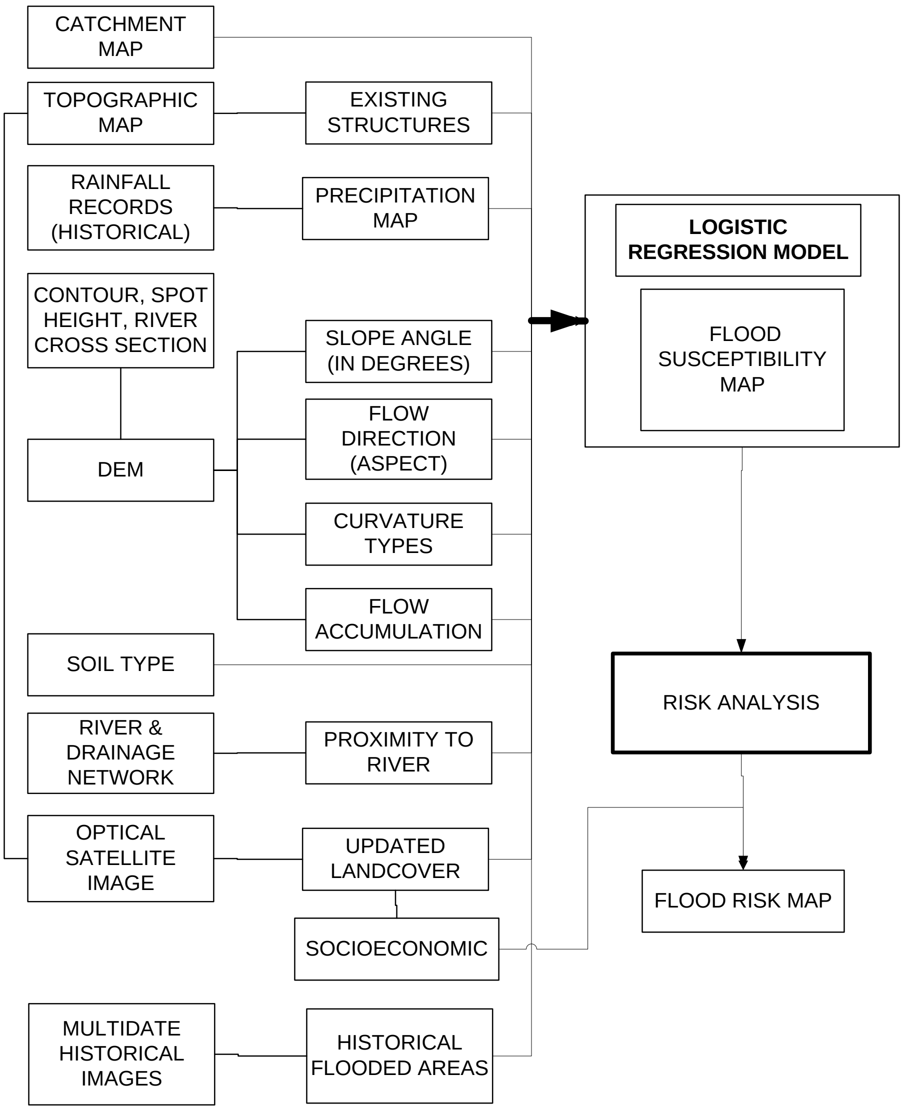


Fig 2: Flow chart of the methodology adopted in this analysis

3.2 Flood water extraction from RADARSAT images

"Before_flood" and "after_flood" RADARSAT images were acquired for wide-area flood extent mapping. A flood extent extraction model was developed in ArcGIS to extract the maximum extent of the historical floods from RADARSAT images. Water body extraction from RADARSAT images during flood includes the normal water extent, water filled paddy fields and the mountain shadow extent. To produce pure flood extent, the normal water extent, non flooded paddy field and shadow were removed; while the flooded paddy field were incorporated to the model to avoid underestimating of flood impact. Hence, the flood Extent modeling approach was involved three separate modeling processes: normal water extraction; flooded area extraction, and flooded paddy area extraction. Figure 3 shows the methodology adopted for flooded area extraction from RADARSAT images in Erdas Imagine and ArcGIS 9.2.

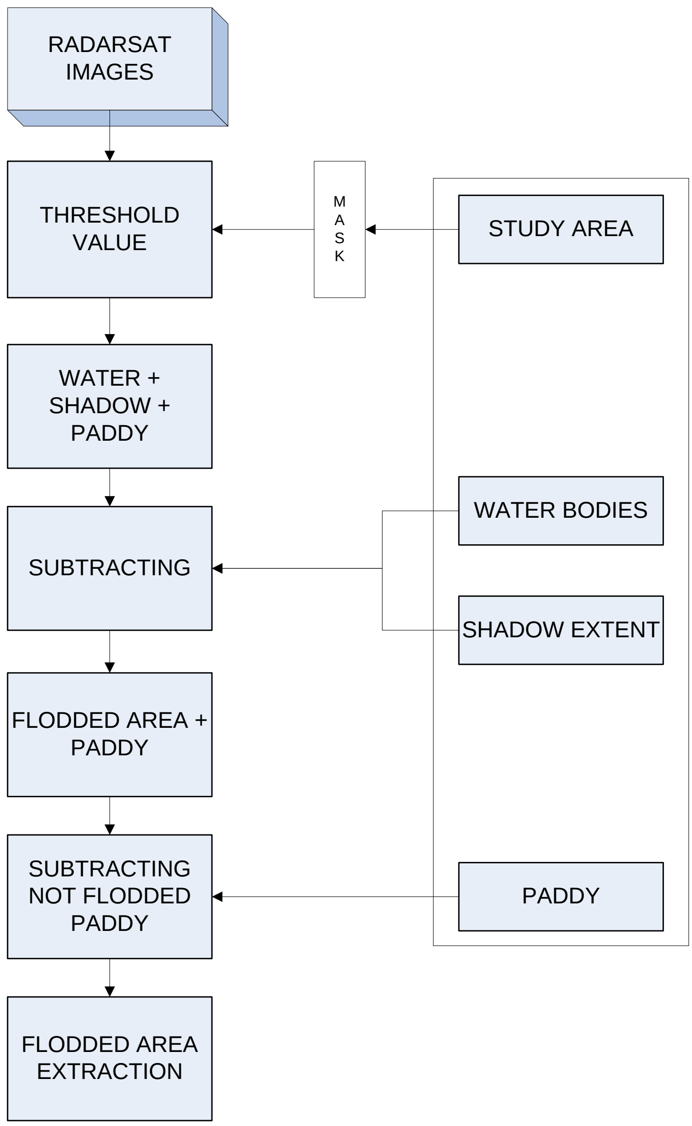
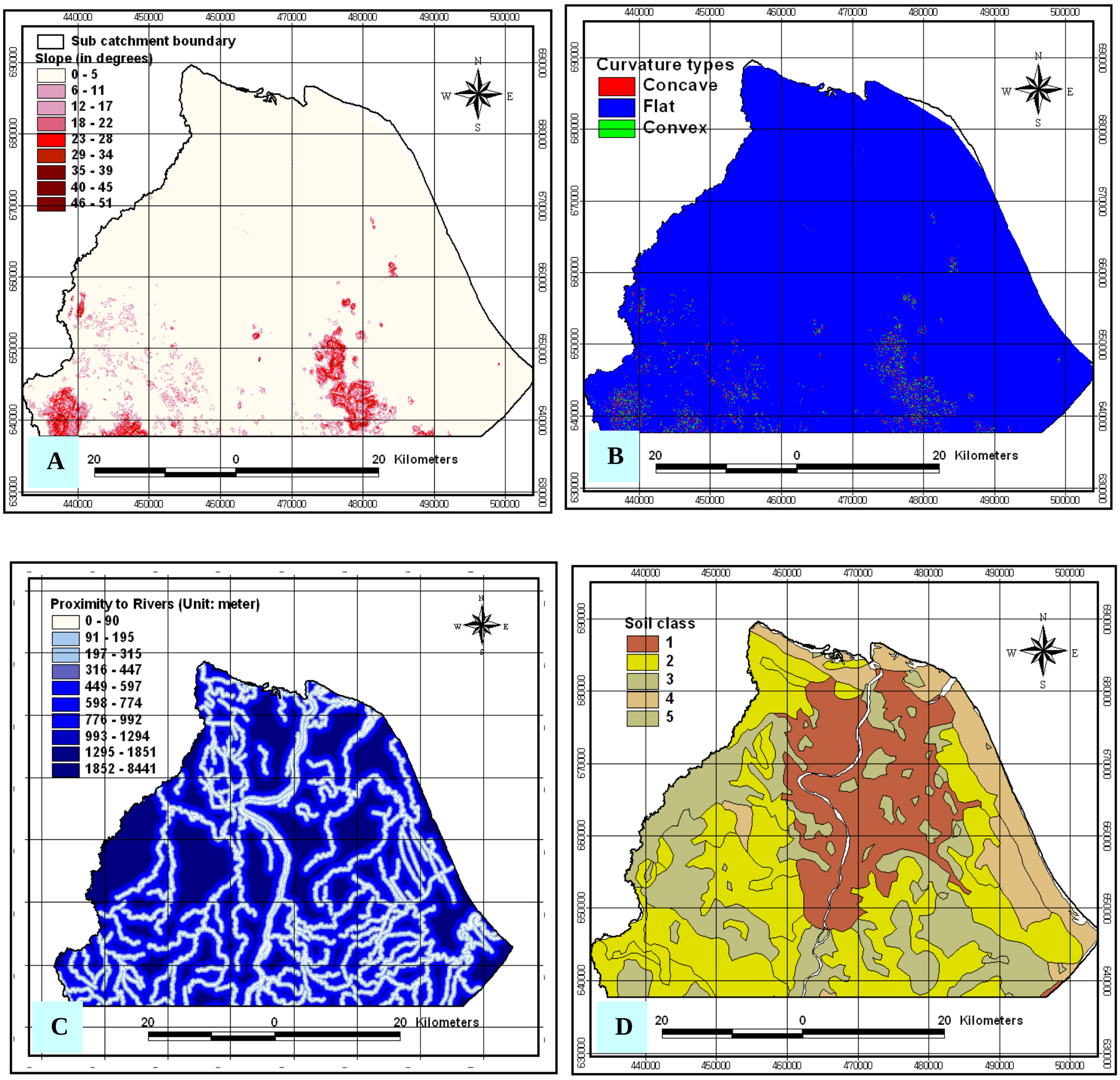


Fig.3: Flow chart for flooded area extraction from RADARSAT images

3.3 DEM and thematic layer preparation

A digital elevation model (DEM) was created first from the topographic database. Contour and survey base points that had elevation values from the 1:25,000-scale topographic maps were extracted, and a DEM was constructed with a resolution of 10 m. Using this DEM, the slope angle, slope aspect, and slope curvature were calculated. In the case of the curvature negative curvatures represent concave, zero curvature represent flat and positive curvatures represents convex respectively. The curvature map was produced using the ESRI routine in Arc View. In addition; the distance from drainage was calculated using the topographic database. The drainage buffer was calculated in 1m intervals. The soil map is obtained from a 1:250,000-scale soil map. Land cover data was classified using a SPOT 5 image employing an unsupervised classification method and topographic map. The land cover map has been classified into nine classes, such as Forest, Lake, Mangrove, Mixed Horticulture, Oil palm, Paddy, River, Rubber and urban areas were extracted for land cover mapping. Finally, precipitation data was interpolated using the meteorological station data for entire study area over last 20 years. The factors were converted to a raster grid with 10 m × 10 m cells for application of the logistic regression model. Figure 4 shows the input GIS data layers.



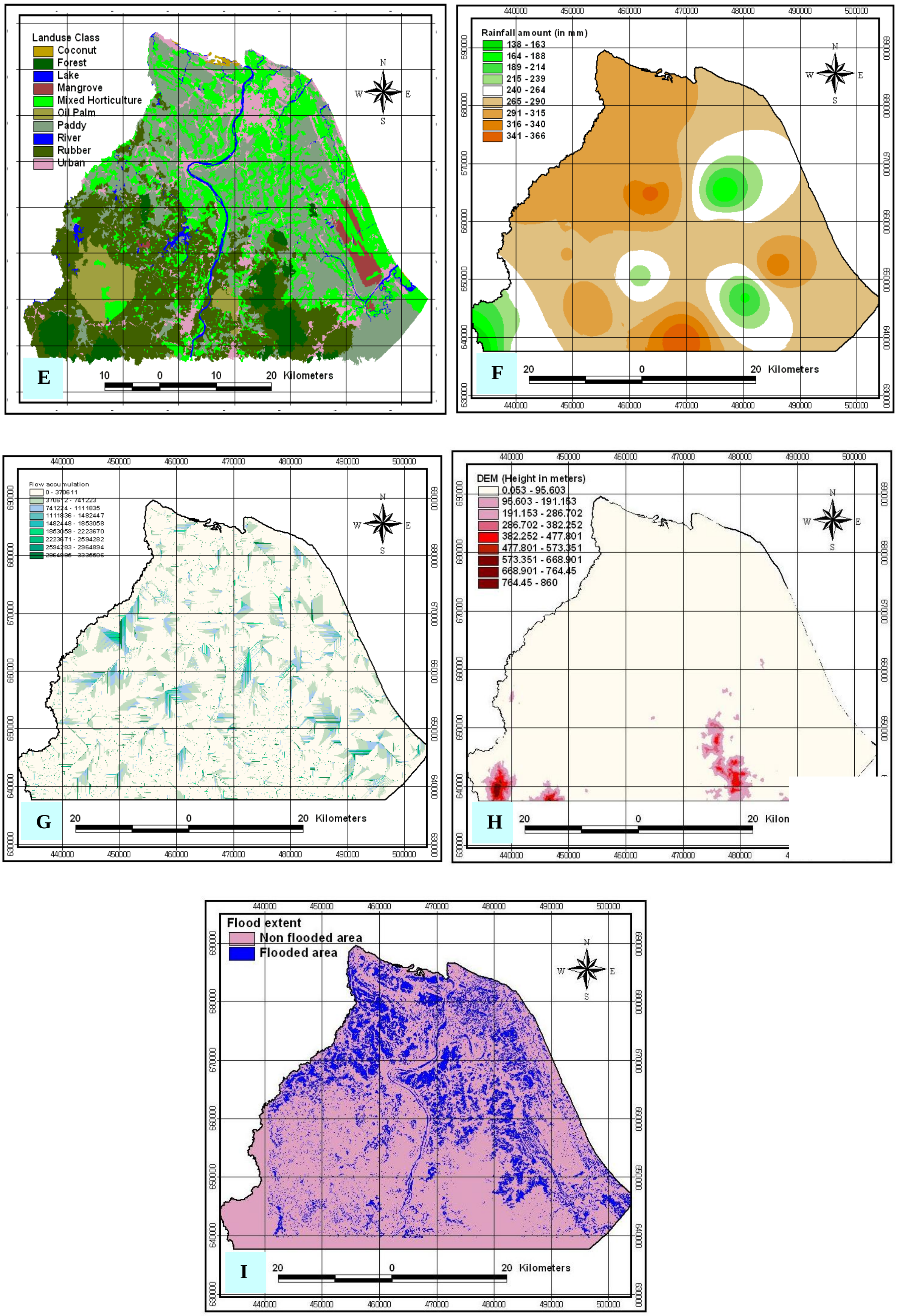


Fig.4: Input data layers (a) Slope; (b) Curvature; (c) Distance from drainage; (d) Soil; (e) Land cover; (f) Precipitation amount; (g) Flow accumulation amount; (h) DEM; and (i) Historical flood extent derived from the RADARSAT images

4. Results and Discussion

4.1 Flood susceptibility mapping using logistic regression model

A key assumption using the statistical approach is that the potential (occurrence possibility) of future flooding areas will be comparable to the actual frequency and extent of previous historical floods. Historical flooded areas were detected from RADARSAT images and field surveys. A historical flooded map was prepared from RADARSAT images, in combination with the GIS, and this were used to evaluate the frequency and extent of future floods in the area. Topography and lithology databases were constructed and lineament, land cover, vegetation index value extracted from SPOT 5 satellite image and precipitation distribution from the meteorological data for the analysis. Then, the calculated and extracted factors were converted to a 10m × 10m grid (ARC/INFO GRID type). Statistical based logistic regressions were applied using the database. Further the spatial relationships between the historic flooded areas and each flood-related factor were analyzed. Using the logistic regression model, the relationship was used as each factor's rating in the overlay analysis and a formula of flood extent possibility was extracted using the relationships. This formula was used to calculate the flood susceptibility index and the index was mapped to represent flood prone areas. The susceptibility map was made using the map overlaying techniques using equation (4). Finally, the map was verified and compared using known 2007 flood extent and success rates and ratio areas were calculated for quantitative validation. In the study, Geographic Information System (GIS) software, ArcView 3.2, and ARC/INFO 9.0 version software packages and SPSS 12.0 statistical program were used as the basic analysis tools for spatial management and data manipulation.

Logistic regression allows one to form a multivariate regression relation between a dependent variable and several independent variables. Logistic regression, which is one of the multivariate analysis models, is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. The advantage of logistic regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types and they do not necessarily have normal distributions. In the case of multi-regression analysis, the factors must be numerical, and in the case of a similar statistical model, discriminant analysis, the variables must have a normal distribution. In the present situation, the dependent variable is a binary variable representing presence or absence of flood. Where the dependent variable is binary, the logistic link function is applicable (Atkinson and Massari, 1998). For this study, the dependent variable must be input as either 0 or 1, so the model applies well to flood susceptibility analysis. Logistic regression coefficients can be used to estimate ratios for each of the independent variables in the model.

Quantitatively, the relationship between the occurrence and its dependency on several variables can be expressed as:

$$p = \frac{1}{1+e^{-z}} \tag{1}$$

Where p is the probability of an event occurring. In the present situation, the value p is the estimated probability of flooded areas. The probability varies from 0 to 1 on an S-shaped curve and z is the linear combination. It follows that logistic regression involves fitting an equation of the following form to the data:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_n x_n \tag{2}$$

Where b_0 is the intercept of the model, the $b_i \ (i = 0,1,2,\ldots,n)$ are the slope coefficients of the logistic regression model, and the $x_i \ (i = 0,1,2,\ldots,n)$ are the independent variables. The linear model formed is then a logistic regression of presence or absence of flooded areas (present conditions) on the independent variables (pre-failure conditions).

Using the logistic regression model, the spatial relationship between flood-occurrence and factors influencing flooded areas were assessed. The spatial databases of each factor were converted to ASCII format files for use in the statistical package, and the correlations between flooded areas and each parameters were calculated. There are two cases. In the first case, only one factor was used. In this case, logistic regression mathematical equations were formulated for each case. The coefficient is shown in Table 2. Finally, the probability that predicts the possibility of flooded-areas was calculated using the spatial database, data from Table 2, equations (1) and (2). In the second case, all factors were used. In this case, logistic regression mathematical equations were formulated as shown in equations (2) and (3) for each case. The coefficient is shown in Table 2.

$$\begin{aligned} Z_n &= (-0.00179 \times SLOPE \times 10000) + (-0.00562 \times CURVATURE \times 10000) + (0.00537 \times PRECIPITATION * 10000) \\ &+ (-0.00002 \times DRAINAGE \times 10000) + (-0.00080 \times DEM \times 10000) + (0.0001 \times FLOW\ ACCUMULATION \times 10000) \\ &+ FLOW\ DIRECTION_c + LANDCOVER_c + SOIL_c - 3.98050 - 34.4228 \end{aligned} \quad (3)$$

(where $SLOPE$ is slope value; $CURVATURE$ is curvature value; $PRECIPITATION$ is PRECIPITATION value; $DRAINAGE$ is distance from drainage value; DEM is elevation value; $FLOW\ ACCUMULATION$ is flow accumulation value and $FLOW\ DIRECTION_c$, $LANDCOVER_c$ $SOIL_c$, and are logistic regression coefficient value listed in Table 2 and z_n is a parameter).

Using formula (2) and (3), the possibility of flooded areas was calculated. ,

$$Susceptibility\ Index = \exp(z) / (1 + \exp(z)) \quad (4)$$

Figure 5 shows the flood susceptibility map produced by using the Equation (4).

4.2 Thematic layers and their weightage computation using data mining model

An artificial neural network model was used for the derivation of the weighatge of the various thematic layers used for flood susceptibility analysis. Before running the artificial neural network program, the training site was selected. So, the flood-susceptible (prone) area and the flood-non-susceptible area were selected as training sites computed by logistic regression model. Cells from each of the two classes were randomly selected as training cells. In the present analysis, the results from logistic regression models were selected as training sites in artitificial neural network modelling. The back-propagation algorithm was then applied to calculate the weights between the input layer and the hidden layer, and between the hidden layer and the output layer, by modifying the number of hidden node and the learning rate. Three-layered feed-forward network was implemented using the MATLAB software package. Here, “feed-forward” denotes that the interconnections between the layers propagate forward to the next layer. The number of hidden layers and the number of nodes in a hidden layer required for a particular classification problem are not easy to deduce. In this study, a 9 (input layer) x 20 (hidden layers) x 2 (output layer) structure was selected for the network, with input data normalized in the range 0.1-0.9. The nominal and interval class group data were converted to continuous values ranging between 0.1 and 0.9. Therefore, the continuous values were not ordinal data, but nominal data, and the numbers denote the classification of the input data. The learning rate was set to 0.01, and the initial weights were randomly selected to values between 0.1 and 0.3. The weights calculated were compared to determine whether the variation in the final weights was dependent on the selection of the initial weights. The back-propagation algorithm was used to minimize the error between the predicted output values and the calculated output values. The algorithm propagated the error backwards, and iteratively adjusted the weights. The number of epochs was set to 2,000, and the root mean square error (RMSE) value used for the stopping criterion was set to 0.01. Most of the training data sets met the 0.01 RMSE goal. However, if the RMSE value was not achieved, then the maximum number of iterations was terminated at 2,000 epochs. When the latter case occurred, then the maximum RMSE value was 0.051. The final weights between layers acquired during training of the neural network and the contribution or importance of each of the 9 factors are shown in Table 3.

Table 2 Coefficients of logistic regression to flooded areas

Factor	Class	Coefficients of logistic regression
Slope (in degrees percentage)	0° ~5°	-0.00179
	6°~ 11°	
	12° ~17°	
	18° ~ 22°	
	23° ~ 28°	
	29° ~34°	
	35° ~ 39°	
	40° ~ 45°	
	46° ~ 51°	
DEM (Height in meter)		-0.00080
Curvature types	Concave	-0.00562
	Flat	
	Convex	
Flow direction	North	-0.0440
	Northeast	-0.0482
	East	-0.1293
	Southeast	0.0036
	South	0.0164
	Southwest	-0.0209
	West	-0.0270
	Northwest	0.0000
Flow accumulation	0 ~370611	0.00001
	370612 ~741223	
	741224 ~ 1111835	
	1111836 ~ 1482447	
	1482448 ~ 1853058	
	1853059 ~ 2223670	
	2223670 ~ 2594282	
	2594283 ~ 2964894	
	2964895 ~ 3335506	
Distance from drainage (in meter)	0 ~ 90m	-0.00002
	91 ~195m	
	196 ~ 315m	
	316 ~ 447m	
	448 ~ 597m	
	598 ~ 774m	
	775 ~ 992m	
	993 ~ 1294m	
	1295 ~ 1851m	
	1852 ~8441m	
Soil types	Batang merbau	-0.298
	Batu hitam	-0.610
	Bungor	-0.532
	Cherang	-0.091
	Durian	-0.249
	Holyrood	-0.184
	Lubok	-0.090
	Melaka	-0.115
	Mined land	-0.194
	Peat	-1.301
	Rengam-bukit	-0.634
	Rengam-jeranga	-0.812
	Rudua-rusila	-0.948
	Serdang	-15.374
	Steepland	-0.711
	Telemong	0.000
	Tokyong	0.000
	Urban land	
Land cover types	Coconut	1.43369
	Forest	1.11290
	Lake	0.0000
	Mangrove	2.53387
	Mixed Horticulture	1.19862
	Oil palm	-17.73256
	Paddy	2.17407
	River	0.13271
	Rubber	-0.48275
	Urban	-3.20529
Precipitation amount (in cm)	138- 163cm	0.00537
	164 ~188cm	
	189 ~ 214cm	
	215 ~ 239cm	
	240 ~ 264cm	
	265 ~ 290cm	
	291 ~ 315cm	
	316 ~ 340cm	
	341 ~ 366cm	

For easy interpretation, the average values were calculated, and these values were divided by the average of the weights of the some factor that had a minimum value. The DEM value was the minimum value, 0.011, and the precipitation value was the maximum value, 0.982.

Table 3. Weights of each factor derived by data mining model

Thematic layers	Weights	Normalized weight
Slope	0.032	0,02
DEM	0.011	0,00
Curvature	0.013	0,00
Flow direction	0.063	0,05
Flow accumulation	0.431	0,43
Distance from drainage	0.274	0,27
Soil	0.316	0,31
Land cover	0.328	0,33
Precipitation	0.092	1,00

4.3 Flood risk analysis

A risk map demarcates the areas under potential consequences where consequences can be those affecting human life, having economic effects or causing environmental changes for instance. A particular surface area subject to the same hazard can face a variety of consequences, depending on landcover types. Explicitly Risk Map provides long term early warning information e.g.: risk classes and value at risk for disaster preparedness and mitigation.

Flood risk analysis takes into consideration flood susceptibilities factors, landcover information, settlement data, transportation networks, and social economic data in deriving risk categories: no risk, low risk, medium risk, and high risk. In the susceptibility map, the potential event and its probability of occurrence were combined. The susceptible categories are expressed as probability in qualitative forms (e.g. none, low, moderate, high). Therefore, the susceptible map was overlaid on damageable objects maps such as transportation network, settlement and facility centres. Figure 6 shows the landslide risk map of the northeastern part of the study area. It has been observed that, many settlements have been built up on high landslide risk areas where the probability of occurrence of landslide is very high. Those high risk areas need to be brought to the notice of the public so that people can realize the possibility of future landslides. This could save their property and life.

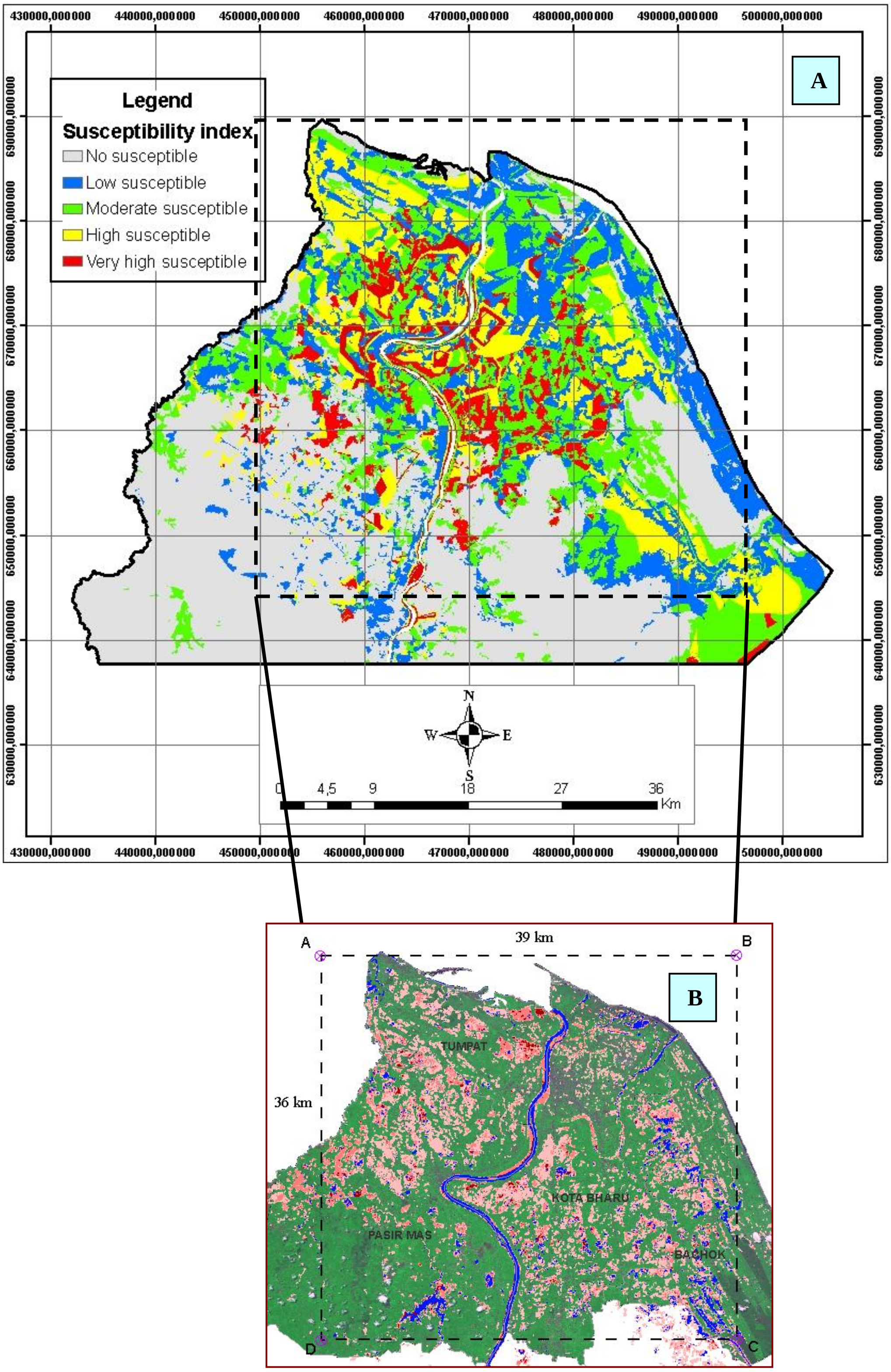


Fig. 5: Flood susceptibility map and comparison to the 2007 flood event in the study area; Upper (a) Susceptibility map using logistic regression model; (b) Flooded areas of 2007 flood in Kelantan state

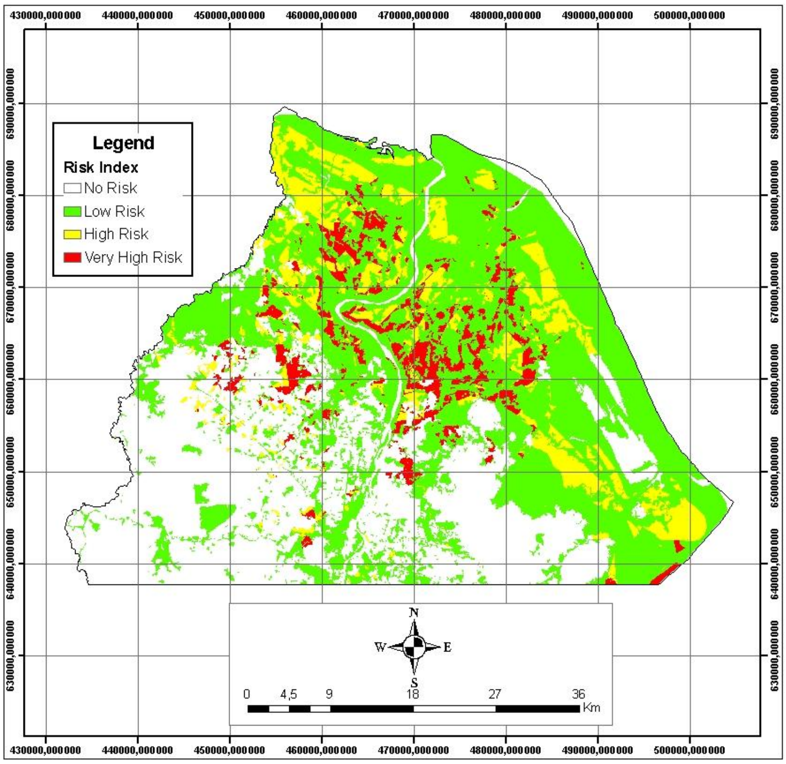


Fig. 6: Flood risk map of study area

5. Validation

For validation of flood susceptibility models, two basic assumptions are needed. One is that flooded areas are related to spatial information such as topography, soil, flow direction, flow accumulation and land cover, and the other is that future flooded areas will affected by a specific factor such as rainfall. In this study, the two assumptions are satisfied because the flooded areas were related to the spatial information and the flooded areas were triggered by heavy rainfall in the study area.

The flood susceptibility analysis result was validated using known extent of flooded areas from 2007. Figure 6b shows the validation result of the 2007 flood against the flood susceptible map. Furthermore, validation was also performed by comparing the known flood extent data with the flood susceptibility map using “Area Under Curve (AUC)” method. Each factor used and logistic regression values was compared. The rate curves were created and its areas of the under curve were calculated for all cases. The rate explains how well the model and factor predict the flooded areas. So, the area under curve in can assess the prediction accuracy qualitatively. To obtain the relative ranks for each prediction pattern, the calculated index values of all cells in the study area were sorted in descending order. Then the ordered cell values were divided into 100 classes, with accumulated 1% intervals. The rate verification results appear as a line in Figure 7. For example, in the case of logistic regression model used, 90 to 100% (10%) class of the study area where the flood susceptibility index had a higher rank could explain 30% of all the flooded areas. In addition, the 80 to 100% (20%) class of the study area where the flood susceptibility index had a higher rank could explain 56% of the flooded areas. To compare the result quantitative, the areas under the curve were re-calculated as the total area is 1 which means perfect prediction accuracy. So, the area under a curve can be used to assess the prediction accuracy qualitatively. In the case of logistic regression model used, the area ratio was 0.8476 and we could say the prediction accuracy is 84.76%.

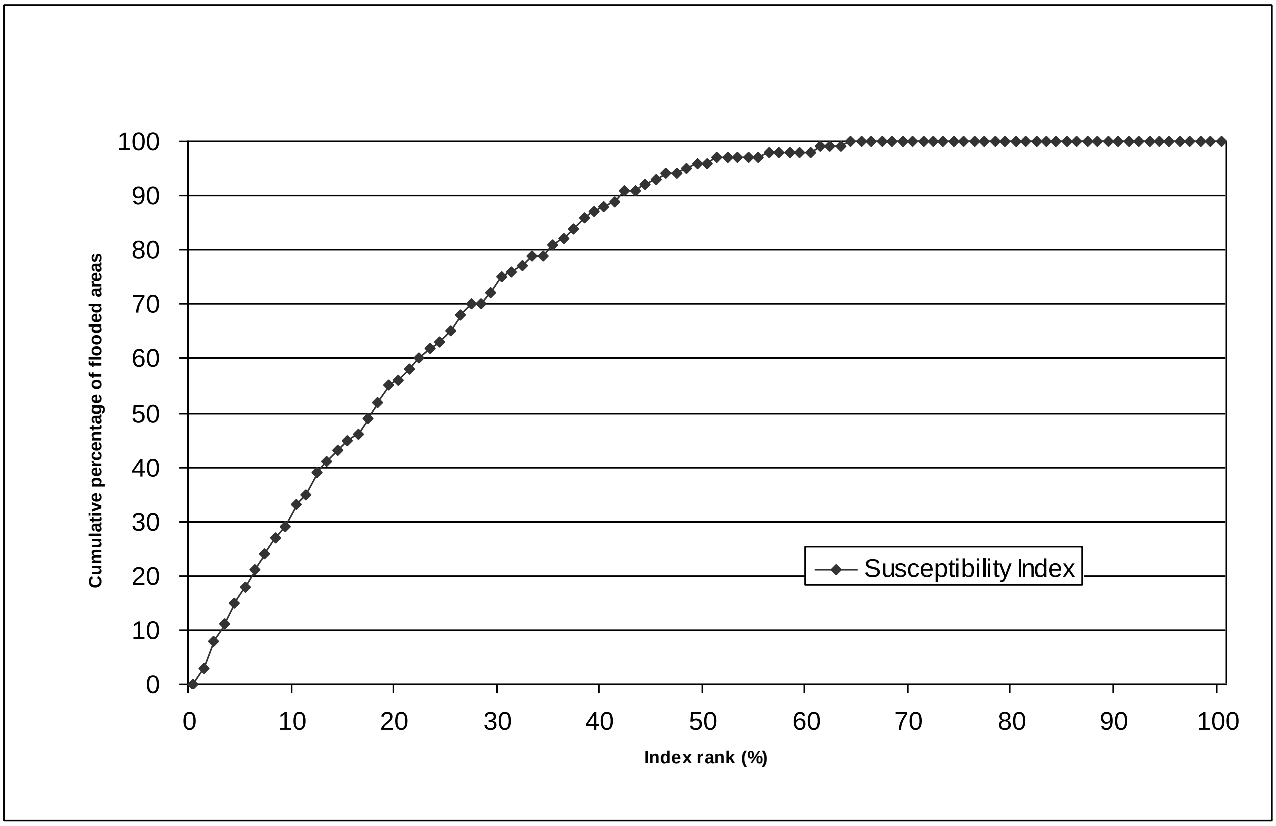


Fig. 7: Cumulative frequency diagram showing flood susceptibility index rank occurring in cumulative percent of flooded areas

6. Conclusions

In the present study, logistic regression model were applied for the flood susceptibility mapping for part of Kelantan river basin. In this research, a statistical approach to estimating the susceptible flow-flood area using remote sensing technique and the GIS was performed. For the flood susceptibility analysis, the detected historical flooded areas and the flood related database were constructed for Kelantan river basin. Using the constructed database, flood susceptibility analysis was performed using logistic regression model. It is remarked that the probability method is somewhat simplistic, and the process of input, calculation and output could be understood easily. Moreover, there is no only a simple conversion of database from GIS to ASCII is required, as the large amount of data can be processed in the GIS environment quickly and easily.

The logistic regression model is simple; the process of input, calculation and output can be readily understood. The large amount of data can be processed in the GIS environment quickly and easily. The logistic regression model requires conversion of the data to ASCII or other formats for use in the statistical package, and later re-conversion to incorporate it into the GIS database. Moreover, it is hard to process the large amount of data in the statistical package. In the case of a similar statistical model (discriminant analysis), the factors must have a normal distribution, and in the case of multi-regression analysis, the factors must be numerical. However, for logistical regression, the dependent variable must be input as 0 or 1, therefore the model applies well to flood susceptibility analysis.

Using the parameters used this research; probability method was applied to analyze the flood susceptibility analysis. The analyzed results were used to reconstruct the classified grid database, then to flood susceptibility map. The flood susceptibility map might be of great help to planners and engineers for choosing suitable locations to implement developments in Kelantan river basin. Besides, the flood susceptibility map shows five classes of susceptibility index as very high, high, medium, low, and no susceptibility index was also illustrated in Figure 5. It was noted that the city of Kota Bharu is falling under a medium- high susceptibility index. In general, the middle part of Kelantan river basin and its adjacent banks had very high to high flood susceptibility whereas the lower downstream part of the stream had very low flood susceptibility. Whereas the western and northern steep-cliff areas had a high to medium flood susceptibility whereas the main other parts else of the sub-basin have in general very low flood susceptibility.

Risk analysis was performed for the study area. The flood susceptible map was overlaid on DEM, distance from susceptibility zone, land cover map and damageable objects at risk to produce the flood risk map. These results can be used as basic data to assist flood mitigation and landuse planning. The methods used in the study are also valid for generalized planning and assessment purposes; although they may be less useful on the site-specific scale, where local landuse and geographic heterogeneities may prevail. For the model to be more generally applied, more validation data are needed.
Recently, flood susceptibility mapping has shown a great deal of importance for suitable urban developments. The results shown in this paper can help the developers, planners and engineers for slope management and land-use planning. However, one must be careful while using the models for specific site development. This is because of the scale of the analysis where other causative factors need to be considered. Therefore, the models used in the study are valid of generalized planning and assessment purposes.

Acknowledgement

Thanks to the Malaysian Center for Remote Sensing and Department of Surveying, Malaysia for providing various datasets for this research. Special thanks to the Malaysian Meteorological Service Department for providing rainfall data.

References

Atkinson, P. M., & Massari, R. (1998) Generalized linear modeling of susceptibility to land sliding in the central Apennines, Italy. *Computer & Geosciences* 24**,** 373-385

Blazkova, S., & Beven, K. (1997) Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. *Journal of Hydrology,* Vol. 195 (1-4), 256-278, IAHS Publ. UK.

Chan, N. W. (1995) Flood disaster management in Malaysia: an evaluation of the effectiveness of government resettlement schemes. *Disaster Prevention and Management*, Vol. 4 (4), 22- 29

Cunderlik J. M., & Burn, D. H. (2002) Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation, *Journal of Hydrology*, Vol. 261 (1-4), 115-131, IAHS Publ. UK.

Dai, F. C., & Lee, C. F. (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* , 42, 213– 228

Dai, F. C., Lee, C. F., Li, J., & Xu, Z. W. (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology,* 40, 381-391

Ebisemiju, F. S. (1986) Environmental constrains of the interdependent of drainage basin morphometeric properties. *Proc. of the first international conf. on geomorphology*, Part 2, 3- 19,Manchester, UK.

Farajzadeh, M. (2001) The Flood Modeling using Multiple Regression Analysis in Zohre & Khyrabad Basins. *5th International Conference of Geomorphology*, August,Tokyo, Japan.

Farajzadeh, M. (2002) Flood Susceptibility Zonation of Drainage Basins using Remote Sensing and GIS, case study area: Gaveh rod_ Iran. *Proceeding of international symposium on Geographic Information Systems*, Istanbul, Turkey, September 23-26.

Haeng, Heo J., Salas J.D. & Boes D.C. (2001) Regional flood frequency analysis based on a Weibull model, part 2 Simulations and applications. *Journal of Hydrology*, vol. 242 (3-4), 171-182 ,IAHS Publ., UK.

Hess, L. L., Melack, J. M., & Simonett, D. S. (1990) Radar detection of flooding beneath the forest canopy: A review. *International Journal of Remote Sensing*, 11, 1313−1325.

Hess, L. L., Melack, J., Filoso, S., & Wang, Y. (1995) Delineation of inundated area and vegetation along the

Amazon floodplain with the SIR-C Synthetic Aperture Radar. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 896−903.

Honda, K. C., Francis X. J., & Sah V. P. (1997) Flood monitoring in central plain of Thailand using JERS-1 SAR data. *Proc. 18th Asian Conference of Remote Sensing*, Malaysia.

Horritt, M. S., & Bates, P. D. (2002) Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology* 268, 87–99.

Islam, M. M., & Sadu, K. (2001) Flood damage and modeling using satellite remote sensing data with GIS: Case study of Bangladesh. *In: Remote Sensing and Hydrology 2000, edited by Jerry Ritchie et al.,* IAHS Publication, Oxford, 455-458.

Islam, M. M., & Sado, K. (2002) Development priority map for flood countermeasures by remote sensing data with Geographic Information System. *Journal of Hydrologic Engineering*, September-October 2002, 346-355.

Knebl, M.R., Yang, Z. L., Hutchison, K., & Maidment, D.R. (2005) Regional Scale Flood Modeling using NEXRAD Rainfall, GIS, and HEC-HMS/RAS: A case study for the San Antonio River Basin Summer 2002 storm event. *Journal of Environmental Management,* 75, 325–336

Landau, S., Mitchell, R.A.C., Barnett, V.,Colls J.J., Craigon, J., & Payne, R.W. (2000) A parsimomious, multiple-regression model of wheat yield response to environment, *Agricultural and Forest Meteorology*, Vol. 101 (2-3), 151-166., Elsevier Publi., USA.

Lee, S., Pradhan, B. (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*. 4, 33-41.

Le Toan, T., Ribbes, F., Wange, L. F., Floury, N., Ding, N., & Kong, K. H., (1997) Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 41−56

Masmoudi, M., & Habaieb, H. (1993) The Performance of Some Realtime Statistical Flood Forecasting Models Seen Through Mmulticriteria Analysis. *Water Resources Manegement* 7, 57–67

Merwade, V., Cook, A., & Coonrod, J. (2008) GIS techniques for creating river terrain models for hydrodynamic modeling and flood inundation mapping, *Environmental Modelling and Software*, 23 (10), 1300-1311

Nageshwar, R.B., & Bhagabat, P. (1997) Flood estimation for ungauged catchments using the GIUH. *Journal of Water Resources planning and management*, 228-238,Elsevier Publi., USA.

Ohlmacher, G.. C., & Davis, J. C. (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansa, USA. *Engineering Geology* 69, 331-343

Pradhan, B., & Lee, S. (2009) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environmental Earth Sciences*, on-line first, doi:10.1007/s12665-009-0245-8

Pradhan, B., Lee, S., Mansor, S., Buchroithner, M. F., Jallaluddin, N. & Khujaimah, Z. (2008) Utilization of optical remote sensing data and geographic information system tools for regional landslide hazard analysis by using binomial logistic regression model. *Journal of Applied Remote Sensing*, 2: 1-11

Pradhan, B. & Shafie M. (2009) Flood hazard assessment for cloud prone rainy areas in a typical tropical environment. *Disaster Advances*, 2 (2), 7-15

Profeti, G., & Machintosh, H. (1997) Flood management though Landsat TM and ERS SAR data: a case study. *Hydrological Processes*, 11, 1397–1408

Sanyal, J., & Lu, X. X. (2005) Remote sensing and GIS-based flood vulnerability assessment of human settlements: A case study of Gangetic West Bengal, India. *Hydrological Processes* 19, 3699-3716

Sanyal, J., & Lu, X. X. (2004) Application of remote sensing in flood management with special reference to monsoon Asia – *A Review. Natural Hazards* 33, Kluwer Academic Publishers.

Townsend, P. A., & Walsh, S. J. (1998) Modelling flood plain inundation using integrated GIS with radar and optical remote sensing. *Geomorphology* 21 (98), 295-312

Verstappen, H. T. (1995) Aerospace technology and natural disaster reduction. In Singh, R. P., and Furrer, R. (editors), *Natural Hazards: Monitoring and Assessment using Remote Sensing Technique,* Elsevier Science, Inc., New York, 3-15.

Villiers, A.B. (1986) A Multivariate evaluation of a group of drainage basin variables, African case study, *Proc. of the first international conf. on geomorphology*, Part 2,21-33, Manchester, UK.

Wadge, G., Wislocki, A. P., Pearson, J., & Whittow, J. B. (1993) Mapping natural hazards with spatial modeling system. In: *Geographic Information Handling Research and Applications, edited by Mather P. M., John Wiley and Sons Inc., New York*.

Yakoo, Y., Kazama, S., Sawamoto, M., & Nishimura, H. (2001) Regionalization of lumped water balance model parameters based on multiple regression, *Journal of Hydrology*, Vol. 246 (1-4), 209-222, IAHS Publ., UK.

Zerger, A. (2002) Examining GIS decision utility for natural hazard risk modeling*. Environmental Modelling and Software*, 17 (3), 287-294