FOREST FIRE SUSCEPTIBILITY MAPPING BASED ON LOGISTIC REGRESSION AND FREQUENCY RATIO METHODS: A CASE STUDY OF CHIANG MAI PROVINCE, THAILAND

Laphatrada SRIRAPRATI[a] and Songkot DASANANDA[b,*]

[a] Lecturer, Udorn Thani Ratjabat University, Mueang District, Udorn Thani 40000, Thailand;
E-mail: Crystal_gookgai@yahoo.com

[b] Assistant professor, Suranaree University of Technology, Mueang District, Nakhon Ratchasima 30000, Thailand;
Tel: (66-44) 224379; Fax: (66-44) 224316; E-mail: songkot@sut.ac.th

KEY WORDS: Fire Susceptibility, Logistic Regression, Frequency Ratio, Chiang Mai

Abstract: Forest fire susceptibility mapping is a crucial activity to assist proper planning of fire prevention and warning program. However, in Thailand, report of its utilization is still rare. In this work, two established methods in the formation of landslide susceptibility maps: the logistic regression (LR) and frequency ratio (FR), are applied to produce forest fire susceptibility maps for Chiang Mai Province in northern Thailand which often experiences extensive wild fires in the protected forest areas. In the LR method, five factors most related to the occurrences of active fire spot were considered, which are, surface slope, rainfall intensity, population at sub-district level, amount of vegetation (in term of NDVI), and elevation (DEM). And in the FR method, nine factors were used, which are, vegetation, slope, aspect, distances from road/village, temperature, rain, population, NDVI. Resulted susceptibility maps from both methods indicate similar pattern of susceptibility level where high susceptibility zones concentrate mainly on the lower part of the province and the low high susceptibility zones locate mainly in the middle part of the area. These output maps were validated using the area under the curve (AUC) method where the accuracy rate of 75.88% (for FR) and 70.87% (for LR) were achieved. The obtained maps can be used to reduce forest fire hazard and assist with proper planning of land use activity in the future.

INTRODUCTION

According to the FAO terminology (Food and Agricultural Organization, 1986), forest fire risk is chance of a fire starting as determined by the presence and activity of any causative agent. This character is similar to the term “fire susceptibility”. In general, through data analysis of the satellite-based active fire maps, it is possible to assess and identify the potential prone areas to fires, which are typically illustrated in form of the fire susceptibility or fire risk maps. These maps can help decrease fire damage on human and environment through proper mitigation and preparedness. Fire risk assessment has flourished in the last two decades as evidenced by the increasing number of such studies (Zhang et al., 2010). This can be attributed to the more availability of remotely-sensed fire data and powerful capability of geographical information systems (GIS) technology in storing and processing spatial data (Burrough and McDonnel, 1998).

To construct fire susceptibility map, intricate link of causal factors, fire events and fire-prone areas, must be understood. To achieve this purpose, various fire risk assessment methods have been applied and the fire-related factors assessed (San Miguel-Ayanz and Ravail, 2005). Conceptually, probability of having fire over a particular location (PF) depends on three broad categories of influencing factors, which are:

(1) Human-related factors (HF) that indicate motive and chance of human to initiate fire at the interested locations. These factors are generally associated to the socio-economic environment and difficulty level of human to access the location, e.g. land-use pattern, human activities, agricultural practice, population density, poverty level, forest law enforcement, and proximity to road or human settlement (Dennis et al., 2005; Prasada et al., 2008).

(2) Environmental factors (EF) that can affect fire ignition and its growth or spread afterwards. These factors can be divided into 3 subgroups which are

2.1 fuel characters; e.g. vegetation type, amount, and leaf dryness
2.2 terrain characters; e.g. slope, aspect, elevation and soil quality
2.3 climate states; e.g. air humidity and temperature, wind speed and direction, amount of rainfall, and amount of incoming solar energy (insolation),

(3) Fire statistics factors (SF) that inform the likelihood of fire occurrence in the interested area based on previous fire data of that area or its neighborhood.

As there is still no consensus on which methods that are most effective in the identifying of potential fire-prone area, especially in forest ecosystem of Thailand, in this thesis, two selected methods including logistic regression and frequency ratio (FR) will be employed and their results (fire susceptibility maps) are then compared and discussed. The logistic regression is found popular among fire risk researchers but the FR method is rarely applied (e.g. Pradhan et al., 2007) but they have been widely used for the landslide susceptibility mapping (e.g. in Lee and Pradhan 2007; Oh et al., 2009). These methods take different approaches to identify fire risk area and comparison of their results may give us more insight on the complicated interaction between fire events and the environmental conditions in the study area.

For the logistic regression model, it examines relations between historical fire data and their causative factors and applies this knowledge to determine chance of having fire at a particular location. Results are typically reported as

$$PF = \frac{1}{(1+e^{-z})} \quad \text{and} \quad z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_N x_N \tag{1}$$

where PF is the probability of having fire at given location (e.g. at a specific grid cell) and z is the linear combination of the independent variables in use weighted by their regression coefficients (b); x is the value of each variable for any cell; e is the base of the natural log and N is number of the used variables (Hernandez-Leal et al., 2006; Preisler et al., 2004). PF ranges at 0-1 and coefficients of the variables can be found using logistic regression procedure (like one available in the SPSS software).

The FR method determines fire prone area by giving each considered pixel fire risk score (RS) that is computed using combination of the frequency ratio (FR) values from all used variables (N in total) at each considered grid cell, or,

$$RS = \sum_{i=1}^{N} FR_i \tag{2}$$

The FR_i value is relative fire occurrence frequency per unit area within a given range (or class) of the associated variable i (e.g. at the altitude 0-50 m or in mixed-deciduous forest) when compared to that of the total area. This can be written as

$$FR = \frac{(CFP/TFP)}{(CA/TA)} \tag{3}$$

where CFP is number of fire pixels seen in a specific class (of a certain factor), TFP is number of total observed fire pixels, CA is the associated class area and TA is the total study area. FR values can range from 0 onwards and the higher values (e.g. much greater than 1) indicate higher chance of having fire in that specified class.

RESEARCH METHODOLOGY

There are three main steps that were fulfilled in this research:

1. Generation of the fire susceptibility map based on the FR method;
2. Generation of the fire susceptibility map based on the logistic regression method;
3. Accuracy assessment of susceptibility maps from both methods.

In Step 1, the fire susceptibility map was produced based on the FR method (using Eqs. 2 and 3) where the relevant factors were divided into 2 broad groups; which are:

Static variables

1. Proximity-from road and village. 2. Population density (at sub-district level)
3. Topography- slope, aspect 4. Types of vegetation-evergreen/disturbed forest, paddy field.
5. Climate data-temperature and rainfall (30-year average during fire season, January-April)

Dynamic variable

1. Vegetation abundance-NDVI (derived from MODIS images during pre-fire period in December 2006)

The FR matrix for all these considered factor classes was constructed (Table 1) based on the 213 samples of active fire spots and the definition of FR given in Eq. 3. The fire risk scores (RS) for each image's pixel were then computed based on Eq. 2 and present in form of a classified fire susceptibility map with four levels of the severity are categorized: (1) low (VL), (2) moderate (M), (3) High (H), and (4) very high (VH).

Table 1: Assigned factor classes for the FR matrix analysis.

Factor	Class	Factor	Class	Factor	Class
Vegetation	F0 (disturbed)	Road	< 300 m	Population (sub-district)	< 6000
	F1 (evergreen)		300-500 m		6000-10000
	Paddy field		500-1000 m		10000-15000
	Others		> 1,000 m		
Slope angle	0-5 %	Village	> 5 km		15000-20000
			3-5 km		
	5 -10 %		0-3 km		>20000
	10-15 %	Temperature	20 ^{o}C	NDVI	< 0
	15-35 %		20-25 ^{o}C		0 - 0.2
	> 35 %		25-30 ^{o}C		0.2-0.4
Slope aspect	North (N)		>30 ^{o}C		0.4- 0.6
	Northeast (NE)	Rain	0 mm		0.6 -0.8
	East (E)		0-5 mm		
	Southeast (SE)		5-10 mm		
	South (S)		10-15 mm		
	Southwest (SW)		>15 mm		
	West (W)				
	Northwest (NW)				

In Step 2, the fire susceptibility map was produced based on the logistic regression method (using Eq. 1). This began with the determination of the proper variables to be used in the analysis. Then the proper function z was determined based on the relationship of the chosen variables to the used reference fire data. Finally, the probability map based on values of the p(z) function in Eq. 1 was created and modified in form of classified fire susceptibility map with 4 levels of the severity: (1) low (VL), (2) moderate (M), (3) High (H), and (4) very high (VH).

In Step 3, accuracy assessment of the gained maps from the FR and logistic regression methods was performed using the AUC method described in, e.g., Lee at al. (2004) and Intarawichian and Dasananda (2010). In this method, the computed risk score values (RS) of all pixels within the study area are sorted in descending order (from high to low). Then these ordered cell values were divided into 100 classes, with accumulated 1% interval. This results in the 100 classified fire susceptibility classes for performing the accuracy assessment. The ranking orders (1 to 100) are then given to each class beginning from the very high susceptibility towards the very low ones, respectively. To assess the predictive capability of the map quantitatively, the RS ranking orders (1-100) were plotted against accumulative amount of reference fire incidences for each specific class (given in term of percentage of the total number). This appears as a line, then, the prediction accuracy of the map can be readily evaluated from the area under the plotting curve (AUC) by assuming that perfect prediction will have maximum AUC of 1 (the ideal value of 100% accuracy).

RESULTS AND DISCUSSION

In Step 1, the FR matrix was produced first based on knowledge of 213 samples of detected fire spots and values of the associated variables (as detailed in Table 1) at each used fire spots. The FR values for each identified class factor can be calculated based on Eq. 3 and results are reported in Table 2. The FR scores indicate that the fire occurrence potential is relatively high for the paddy field (1.38), the slope > 35% (1.96) and 10-15% (1.29), E/SE aspect (1.31, 1.25), distance from road > 1km (1.36), from village > 5km (2.04). Subsequently, the risk score (RS) for each pixel on the study area can be found and classified using Eq. 3 (Figures 1 and 2).

In Step 2, 229 detected fire spots during January-March 2007 were used to find their relation with some chosen variables. In the logistic regression concept, this relationship is expressed through a linear combination of these variables, or function z in Eq. 1. From knowledge of this term (z), the probability function p(z) can be identified. In this study, five variables are found important which are slope, rain, population, NDVI, and DEM. Their relation can be written as:

$$z = 707.93 + 0.024\,[\text{slope}] - 6.41\,[\text{rain}] + 0.13\,[\text{population}] - 1086.75\,[\text{NDVI}] + 0.55\,[\text{DEM}] \qquad (4)$$

Table 2: The FR matrix for all associated class factors used in the FR analysis.

Factors	Class	Total number of pixel		Fire occurrence point		FR
		Number	%	Number	%	
Vegetation type	F0	325,478	1.33	3	1.41	1.06
	F1	18,195,713	74.24	174	81.69	1.10
	Paddy	1,337,915	5.46	16	7.51	1.38
	Others	4,651,685	18.98	20	9.39	0.49
Slope angle	0 - 5 %	5,160,617	21.05	23	10.80	0.51
	5 - 10 %	3,981,751	16.24	25	11.74	0.72
	10 - 15 %	4,377,997	17.86	49	23.00	1.29
	15 - 35 %	10,403,878	42.45	106	49.77	1.17
	> 35 %	586,548	2.39	10	4.69	1.96
Slope aspect	North (N)	2,884,857	11.85	22	10.33	0.87
	Northeast (NE)	3,075,813	12.63	24	11.27	0.89
	East (E)	3,228,154	13.26	37	17.37	1.31
	Southeast (SE)	3,096,444	12.72	34	15.96	1.25
	South (S)	3,151,381	12.95	21	9.86	0.76
	Southwest (SW)	3,220,535	13.23	33	15.49	1.17
	West (W)	3,015,176	12.39	26	12.21	0.99
	Northwest (NW)	2,671,433	10.97	16	7.51	0.68
Road	< 300 m	2,624,674	10.71	9	4.23	0.39
	300 - 500 m	3,912,183	15.96	20	9.39	0.59
	500 - 1000 m	6,426,913	26.22	48	22.53	0.86
	> 1,000 m	11,547,021	47.11	136	63.85	1.36
Village	> 5km	1,409,123	5.75	25	11.74	2.04
	3 – 5 km	7,402,253	30.20	54	25.35	0.84
	0 – 3 km	15,699,415	64.05	134	62.91	0.98
Temperature	20	24,645,599	96.88	213	100	1.03
	20-25	730,412	2.87	0	0	0
	25-30	43,114	0.17	0	0	0
	>30	20,864	0.08	0	0	0
NDVI	< 0	51313	1.03	0	0.00	0.00
	0 - 0.2	122385	2.45	0	0.00	0.00
	0.2 - 0.4	1277604	25.61	112	52.58	2.05
	0.4 - 0.6	2498507	50.08	89	41.78	0.83
	0.6 - 0.8	1039361	20.83	12	5.63	0.27
Rain	0 mm	644620	2.62	1	0.47	0.18
	0-5 mm	22330414	90.77	208	97.65	1.08
	5-10 mm	1343346	5.46	3	1.41	0.26
	10-15 mm	188281	0.77	1	0.47	0.61
	>15 mm	94302	0.38	0	0.00	0.00
Population	< 6,000	7820564	31.56	45	21.13	0.67
	6,000-10,000	9348843	37.73	87	40.85	1.08
	10,000-15,000	4070826	16.43	55	25.82	1.57
	15,000-20,000	2498507	10.08	18	8.45	0.84
	>20,000	1039361	4.19	8	3.76	0.90

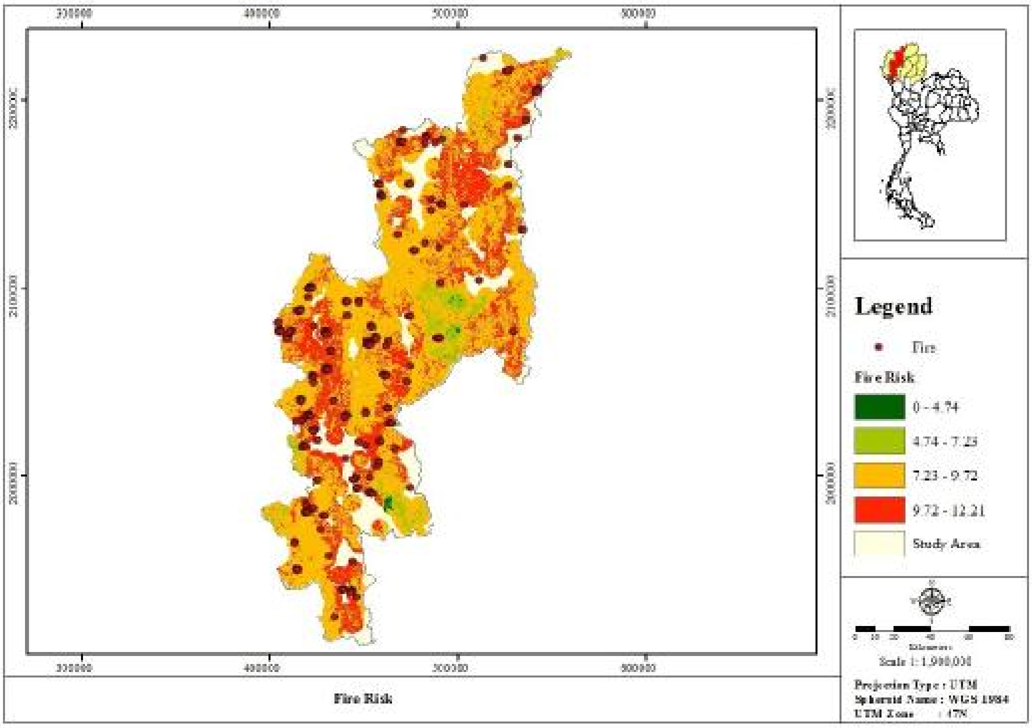

Figure 1: Map of the fire risk score (RS).

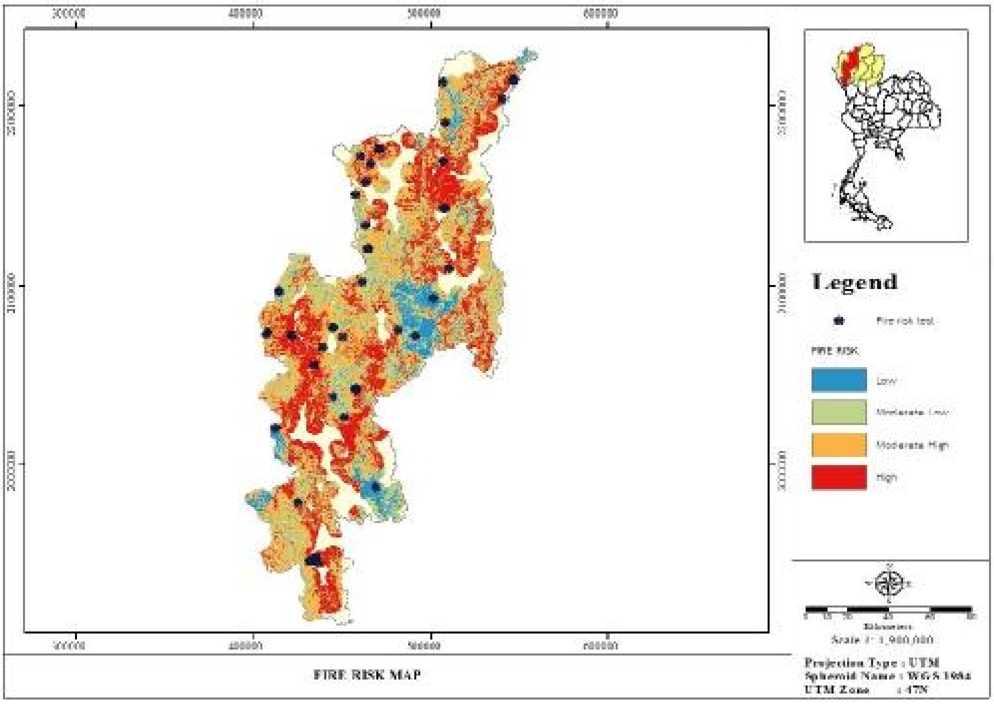

Figure 2: Classified fire susceptibility map based on the fire risk score.

which gives R^2 of 0.70. Then, the z data are normalized as follows zn = z/max(z) where max (z) is the maximum values of z calculated from Eq.4. This term (zn) is finally applied to gain the probability value p(z) based on Eq.1. These values were found to range between 0.495 and 0.731 (Figure 3).

As a consequence, 4 classes of the fire susceptibility level are identified (using the natural break concept), which are, (1) Low (L), (2) Moderately low (ML), (3) Moderately high (MH) and (4) High (H) (Figure 4 and Table 3). Primary test for the validation of this map is done by using the detected fire spots in April to calculate for their associated p(z) and values of dNBR. The accuracy assessment gets relatively high level of correlation between these two variables (R^2 = 72.35) which is satisfied the selection of dNBR to represent the existence of active fire spots. Table 3 gives the area data at class level for both maps in Figures 2 and 4.

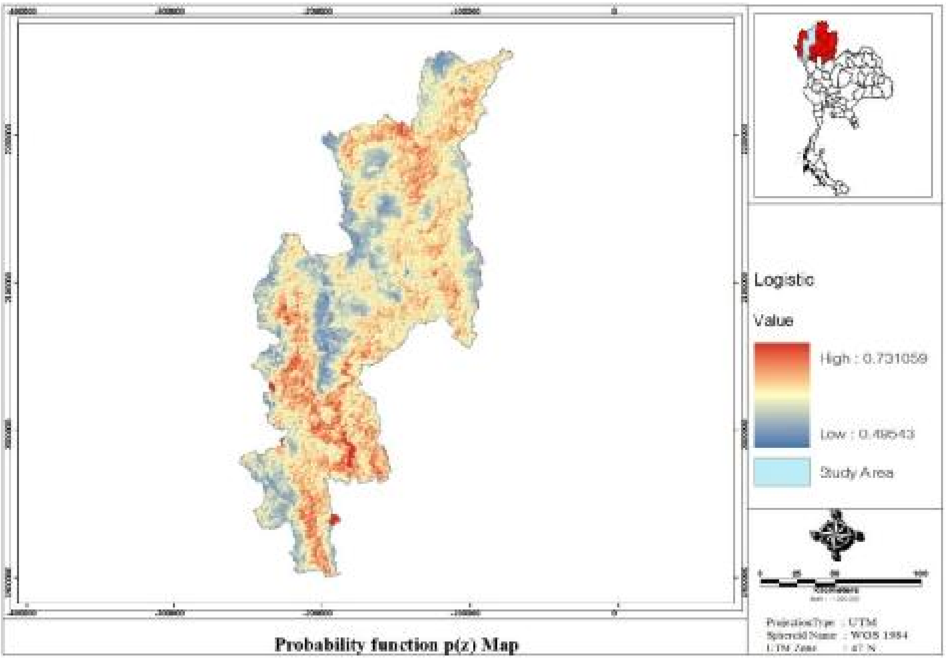

Figure 3: Map of the probability function, p(z).

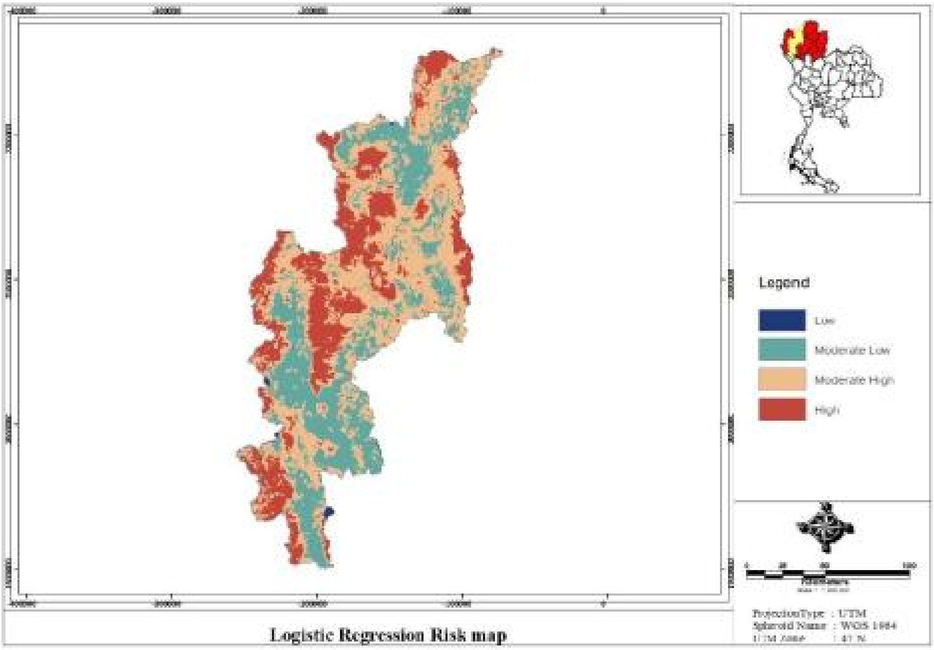

Figure 4: Classified fire susceptibility map based on the p(z) function.

Table 3: Area allocation at class level of susceptibility maps (Figures 2 and 4).

Class	Range (logistic)	Area		Range (FR)	Area	
		km^2	%		km^2	%
Low (L)	< 0.514	4872	22.33	< 7.50	1620.16	8.59
Moderately low (ML)	0.514- 0.523	8352	38.28	7.50-8.71	5087.81	26.97
Moderately high (MH)	0.523-0.563	8481.25	38.88	8.71- 9.72	6803.03	36.06
High (H)	> 0.563	110.25	0.50	> 9.72	5,353.04	28.38

In Step 3, the accuracy assessment of the two derived fire susceptibility maps (by logistic regression and FR methods) were performed based on extra 47 detected fire spots still not being used for the map production by both methods. Here, the AUC for logistic regression-based map is 70.87 % and for FR-based map is 75.88 % (Figure 5). As a result, this can be concluded that the prediction accuracies of the obtained maps in both methods are relatively high and both of them can be used to create the reliable fire susceptibility maps for the area.

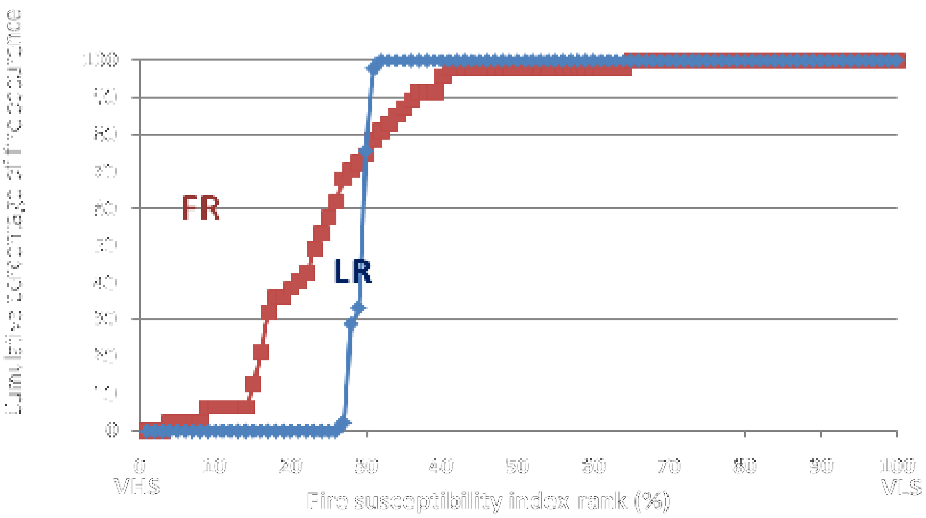

Figure 5: Prediction capabilities of the produced maps based on AUC technique.

CONCLUSIONS

In this study, the FR and logistic regression were applied to generate the fire susceptibility maps for the study area. It was found that, for the FR method, the fire occurrence potential is relatively high for the paddy field (1.38), the slope > 35% (1.96) and 10-15% (1.29), E/SE aspect (1.31, 1.25), distance from road > 1km (1.36), from village > 5km (2.04). For the FR method, most area (77.16%) was classified as being in moderate level of fire susceptibility. For the logistic regression method, the study area is classified to be in the moderately high level the most (36.06%) follows by the high level (28.38%). This is very contrast where only 0.5% of the total area is put in the high level. And for the accuracy assessment, the AUC for logistic regression-based map is 70.87 % and for FR-based map is 75.88 % which are rather high and satisfied for further use if needed.

ACKNOWLEDGEMENT

The research fund provided by Suranaree University of Technology (SUT) for this project is greatly appreciated.

REFERENCES

Burrough, P.A. and McDonnell, R., 1998. **Principles of geographical information Systems**. Oxford University Press, London.

Dennis, R.A., Mayer, J., Applegate, G., Chokkalingam, U., Pierce, C.J., Colfer, Kurniawan, I., Lachowski, H., Maus, P., Permana, R.P., Ruchiat, Y., Stolle, F., Suyanto, and Tomichet, T.P., 2005. Fire, People and Pixels: linking Social Science and Remote Sensing to understand underlying causes and impacts of fires in Indonesia. **Human Ecology**, 33(4), pp.465-504.

Food and Agricultural Organization (FAO), 1986. **Wild land fire management terminology**. Report number 70. FAO Forestry Paper, Roma.M-99.

Hernandez-Leal, P.A., Arbelo, M., and Gonzalez-Calvo, A., 2006. Fire risk assessment using satellite data. **Advances in Space Research,** 37(4), pp. 741-746.

Intarawichian, N. and Dasananda, S., 2010. Analytical Hierarchy Process for Landslide susceptibility mapping in lower Mae Chaem watershed, Northern Thailand. **Suranaree Journal of Science and Technology**, 17(3), pp. 277-292.

Lee, S. and Pradhan, B., 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. **Landslides**, 4(1), pp.33-41.

Lee, S., Choi, J., and Min, K., 2004. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea**. International Journal of Remote Sensing**, 25(11), pp.2037-2052.

Oh, H.J., Lee, S., Chotikasathien, W., Kim, C.H., and Kwon, J.H., 2009. Predictive landslide susceptibility mapping using spatial information in the Pechabun area of Thailand**. Environmental Geology**, 57(3), pp.641-651.

Pradhan, B., Suliman, M.D.H., and Awang, M.A., 2007. Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS). **Disaster Prevention and Management**, 16(3), pp. 344-352.

Prasada, V.K., Badarinathb, K.V.S., and Eaturuc A., 2008. Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India. **Journal of Environmental Management**, 86(1), 1-13.

Preisler, H.K., Brillinger, D.R., Burgan, R.E., and Benoit, J.W., 2004. Probability based models for estimation of wildfire risk. **International Journal of Wildland Fire**, 13(2), pp.133-142.

San-Miguel-Ayanz, J. and Ravail, N., 2005. Active fire detection for fire emergency management: Potential and limitations for the operational use of Remote Sensing. **Natural Hazards Journal**, 35(3), pp.361-376.

Zhanga, Z.X., Zhanga, H.Y., and Zhou, D.W., 2010. Using GIS spatial analysis and logistic regression to predict the probabilities of human-caused grassland fires. **Journal of Arid Environments**. 74(3), pp.386-393.