Article

Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression

Dieu Tien Bui [1,2,*], Kim-Thoa Thi Le [3], Van Cam Nguyen [3], Hoang Duc Le [3] and Inge Revhaug [4]

[1] Geographic Information System Group, Department of Business Administration and Computer Science, University College of Southeast Norway, Hallvard Eikas Plass, N-3800 Bø i Telemark, Norway

[2] Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi 100000, Vietnam

[3] Institute of Geography, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay, Hanoi 100000, Vietnam; L_Thoa@hotmail.com (K.-T.T.L); nguyencamvandl@gmail.com (V.C.N); hoang04101989@gmail.com (H.D.L)

[4] Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003 IMT, 1432 Aas, Norway; Inge.revhaug@nmbu.no

* Correspondence: Dieu.T.Bui@hit.no or BuiTienDieu@gmail.com; Tel.: +47-3595-2724

Abstract: The Cat Ba National Park area (Vietnam) with its tropical forest is recognized as being part of the world biodiversity conservation by the United Nations Educational, Scientific and Cultural Organization (UNESCO) and is a well-known destination for tourists, with around 500,000 travelers per year. This area has been the site for many research projects; however, no project has been carried out for forest fire susceptibility assessment. Thus, protection of the forest including fire prevention is one of the main concerns of the local authorities. This work aims to produce a tropical forest fire susceptibility map for the Cat Ba National Park area, which may be helpful for the local authorities in forest fire protection management. To obtain this purpose, first, historical forest fires and related factors were collected from various sources to construct a GIS database. Then, a forest fire susceptibility model was developed using Kernel logistic regression. The quality of the model was assessed using the Receiver Operating Characteristic (ROC) curve, area under the ROC curve (AUC), and five statistical evaluation measures. The usability of the resulting model is further compared with a benchmark model, the support vector machine (SVM). The results show that the Kernel logistic regression model has a high level of performance in both the training and validation dataset, with a prediction capability of 92.2%. Since the Kernel logistic regression model outperforms the benchmark model, we conclude that the proposed model is a promising alternative tool that should also be considered for forest fire susceptibility mapping in other areas. The results of this study are useful for the local authorities in forest planning and management.

Keywords: tropical forest fire; logistic regression; remote sensing; GIS; Cat Ba; Hai Phong; Vietnam

---

1. Introduction

Forests provide resources for millions of people and make a high contribution to employment, economic development, and terrestrial biodiversity in many countries [1,2]. However, forests are sensitive to climate variations, *i.e.*, an increase of temperature and decrease of precipitation that leads to drought, and these variations make forests more susceptible to fire [3,4]. Thus, assessment and prediction of forest fire risks due to the change of climatic conditions are topics that have attracted the research community during the last decade.

For the case of Vietnam, forests occupy around 42.1% of the total land area in which forest plantations cover 3.5 million ha and 10.4 million ha are natural forests [5]. Together with tropical storms and floods, forest fires are the most common disasters that recurrently occur in the country, causing huge economic losses and devastating natural ecological systems and the environment [6,7]. According to the Department of Forest Protection of Vietnam (DoFP), there were around 704 forest fires yearly during the period of 2002 to 2010, which resulted in a loss of 5081.9 ha forest annually [8]. In addition, climate change with high temperatures and longer dry periods has a negative impact, and leads to an increasing trend in the number of forest fires [9]. Therefore, studying forest fires to understand the fire ignition distributions, in order to find prevention measures, is an urgent task.

Various approaches from simple to sophisticated models have been proposed for forest fire assessments, such as expert knowledge [10,11], statistical methods such as linear regression, multiple regression [12], logistic regression [13,14], geographically weighted regression [13], frequency ratio [14], and evidential belief function [15]. The expert knowledge method is clearly subjective and the accuracy of the results is questionable. Therefore, statistical approaches are widely used where forest fire models are developed based on the statistical assumption that the relationship between input variables and forest fire will be the same in the past and in the future [16]. However, forest fire regimes are complex and influenced by not only climatic factors (*i.e.*, temperature, humidity, wind, and rainfall) but also other factors such as fuel loads (*i.e.*, vegetation), landscape characteristics, and management policies; therefore, the accuracy of the models is not always satisfactory.

Due to the critical nature of the problem, several machine learning approaches have been proposed for forest fire assessment. Oliveira, Oehler, San-Miguel-Ayanz, Camia and Pereira [12] compared the random forests algorithm with traditional multiple linear regression for modeling spatial patterns of fire occurrence in Mediterranean Europe with the conclusion that the predictive ability of random forests was better than for multiple linear regression. However, Pourtaghi, *et al.* [17] showed that the performance of the random forests model was lower than other machine learning models. Massada, *et al.* [18] compared the generalized linear model with two machine learning models (maximum entropy and random forests) for wildfire ignition distribution modelling in the Huron–Manistee National Forest (USA), with the conclusion that the machine learning models performed better. The recent development of soft computing and geographic information systems (GIS) has introduced several new machine learning techniques, *i.e.*, kernel logistic regression and support vector machines; however, investigation of these methods for forest fire assessment has not yet been carried out.

The main objective of this study is to produce a forest fire susceptibility map for the Cat Ba National Park area, Hai Phong city (Vietnam). Cat Ba is the largest island in the Ha Long Bay and the core of the island is a tropical forest. This forest has been recognized by UNESCO as a world biodiversity conservation forest. In addition, the island is a well-known destination for tourists and receives around 500,000 travelers per year [19]. This project was carried out with support from the Vietnam Academy of Sciences and Technology and the Department of Sciences and Technology of Hai Phong city. The difference between this study and studies in aforementioned literature is that herein the kernel logistic regression is used for the forest fire assessment. In addition, a comparison of the prediction capability of the support vector machine (SVM) is provided. The data processing and visualization for this study were carried out using ArcGIS®10.2 (ESRI Inc., Redlands, CA, USA) and ENVI®4.7 (Exelis Visual Information Solutions, Boulder, CO, USA), whereas the modeling process was carried out using WEKA®3.7.10 (The University of Waikato, Hamilton City, New Zealand). In addition, an application in C++ programming that was programmed by the authors was used to transfer the modeling result to a GIS format in order to open it in ArcGIS®10.2.

2. Study Area and Data

2.1. Description of the Study Area

The Cat Ba area is located in the Ha Long Bay in the Gulf of Tonkin in northeast Vietnam, between longitudes 106°50′30″E and 107°08′49″E, and latitudes 20°42′23″N and 20°54′05″N (Figure 1). It covers an area of around 328.64 km² and is a UNESCO Biosphere Reserve since 2004 due to its high biodiversity and various ecosystems, such as tropical forest, mangroves, and wetlands [20,21]. The elevation of the area varies from 0 to 282.7 m a.s.l. with mean and standard deviation of 60.4 m and 56.6 m, respectively.
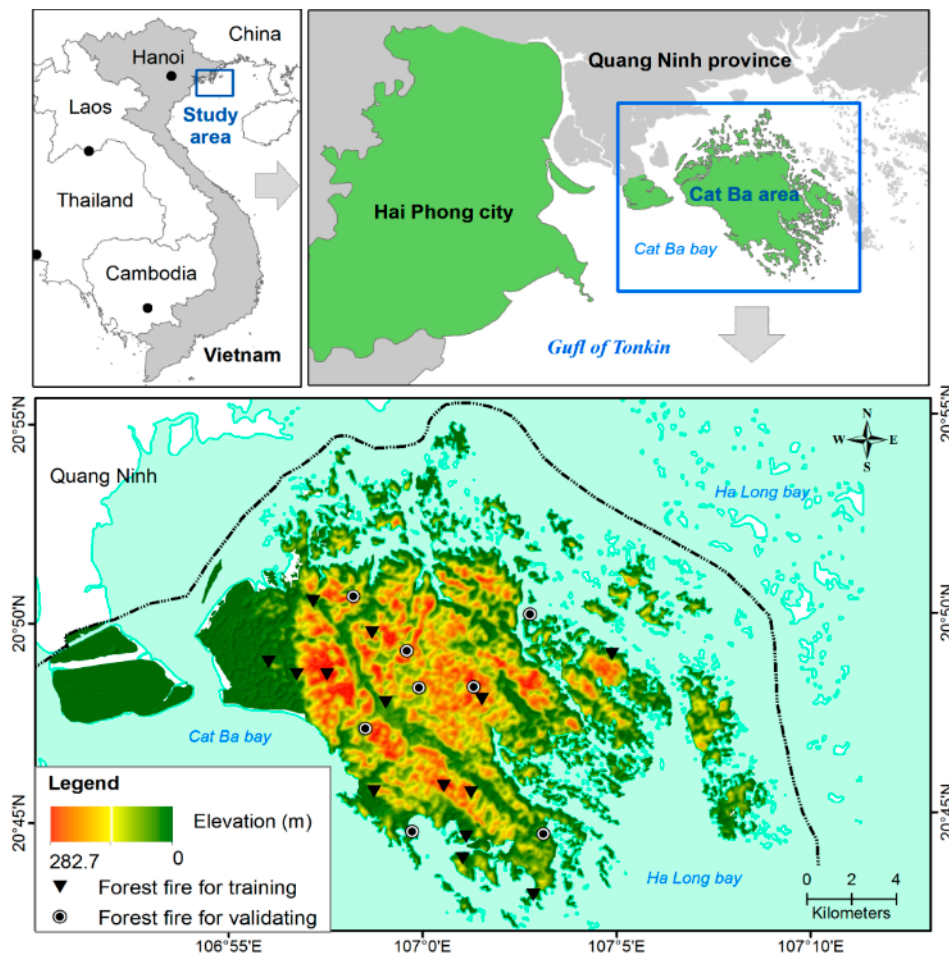
Figure 1. Location of the study area and historical forest fires.

The area is situated in the tropical monsoon region where there are two distinguished seasons, the hot wet and the dry cool seasons. The rainy season normally is from April to October with high frequency of typhoons and tropical rainstorms. The average annual temperature ranges from 24 °C to 30 °C with July as the warmest month, whereas the coolest month is January. Due to the effects of climate change, there is a higher frequency of high-temperature days (37–40 °C) and less rain, which have resulted in an increased probability of forest fires in this area [22–25].

Approximately 37.8% of the Cat Ba area is covered by dense forest, whereas areas with planted forest occupy 2.2%. Scrubland covers around 27.5%, whereas lands with mangrove forest and grass cover 4.2% and 18.2% of the total area, respectively. The populated area accounts for 2.3% of the total study area.

2.2. Data Collection and Pre-Processing

2.2.1. Forest Fire Database

The most common approaches for modeling of the forest fire susceptibility is to assume a correlation between historical fires, their locations, and driving forces, as a key for future fires [26]. Therefore, the data for historical forest fires should be carefully collected. For the Cat Ba area, data for historical forest fires were timely and effectively registered at the MODIS (*Moderate Resolution Imaging Spectroradiometer*) station established from 1 February 2007 at the Department of Forest Protection, the Ministry of Agriculture and Rural Development (Vietnam) [27]. This station has the ability to receive and process data from TERRA, AQUA, NOAA-15, NOAA-17, and NOAA-18 satellites. Forest fires were detected and processed in the TeraScan system at the station using the NASA (*National Aeronautics and Space Administration*) ATBD-MOD14 algorithm [28] for MODIS data (TERRA and AQUA) and the NOAA (*National Oceanic and Atmospheric Administration*) algorithm for the other sensors.

In this research, a total of 22 historical forest fires in the Cat Ba area was extracted, and these fires occurred in the period 2009–2013. These forest fires were checked during the fieldwork using GPS units and digital topographic maps at a scale of 1:25,000. The coordinates of these forest fires were then registered to the GIS database. A descriptive analysis of these forest fires shows that there were two fires in 2009 and eighteen fires in 2010, whereas there was one fire for each year in 2012 and 2013. Around 54.5% of the fires occurred in April and May 2010, whereas 18.2% of the fires happened in March 2010. Four months (February, September, October, and November) have no fire. Approximately 72.7% of the total forest fires occurred between 22.00 and 23.13 h, whereas 18.3% of the total forest fires occurred between 2.43 and 4.04 h. The remaining forest fires occurred between 10.42 and 1.10 h.

It could be seen that many forest fires occurred in 2010, and no forest fires occurred in 2011. It is noted that the worst drought in around 100 years occurred in Vietnam in 2010, whereas heavy rainfalls with a series of severe floods happened in 2011 [29].

Our fieldwork investigations show that most forest fires were caused by humans. There are around 16,000 inhabitants in the study area, and they mainly occupy the southern part of the Cat Ba Island [20]. The local people have poor economic conditions, and this is one of the main causes for illegal exploitations of the Cat Ba tropical forest as well as forest fires.

2.2.2. Fire Ignition Factors

Forest fire susceptibility can be expressed as the probability of a fire to occur within a specific area. The susceptibility degree for fire for each pixel is generally dependent on the contributing factors. Therefore, the determination of fire ignition factors is an important task. Since forest fires strongly depend on topography (*i.e.*, slope and aspect), fuels (*i.e.*, vegetation or NDVI), and climatic features (*i.e.*, temperature, wind, and rainfall) [26,30], these factors should be used for the analysis of fire behaviors.

Topography is considered to be an important factor that influences forest fires because topographic properties affect distributions of vegetation and local climate such as wind speeds [11,12,31,32]. Therefore, a Digital Elevation Model (DEM) was generated using national topographic maps at a scale of 1:25,000. Based on the DEM, slope, aspect, and topographic wetness index (TWI) were extracted. Slope and aspect were selected because fires may travel fast in upward-slopes but slower in areas with downward slopes, whereas aspects may influence wind speeds spreading fires [11]. Forest fires may be influenced by hydrogeological conditions [33]; therefore, TWI was included in this analysis. The slope map (Figure 2a) was constructed with five classes: 0°–5°, 5°–8°, 8°–15°, 15°–25°, and >25°, whereas the aspect map (Figure 2b) was created with nine categories as flat, north, northeast, east, southeast, south, southwest, west, and northwest. For the TWI map (Figure 2c), five classes were constructed: <7, 7–8, 8–9, 9–10, and >10.
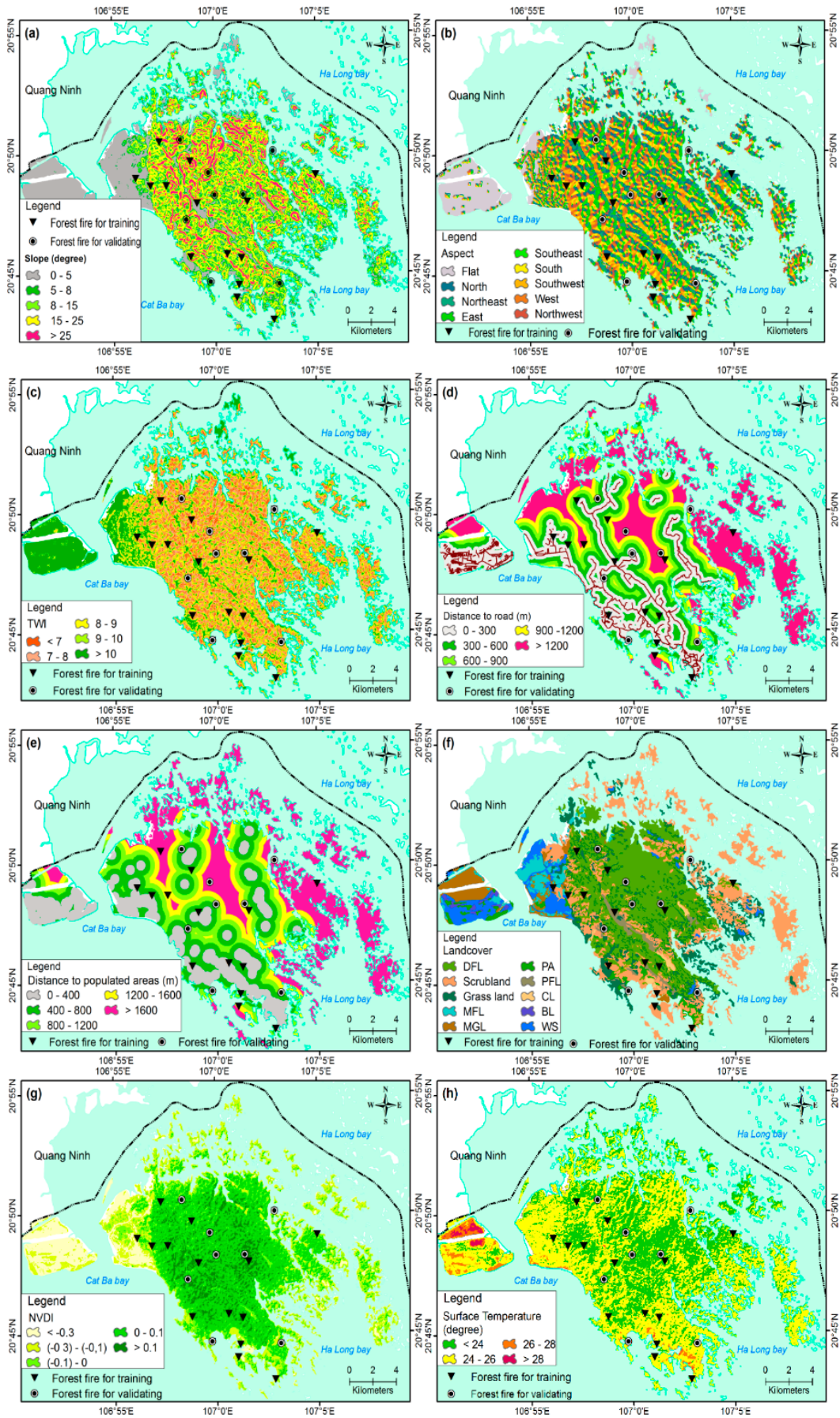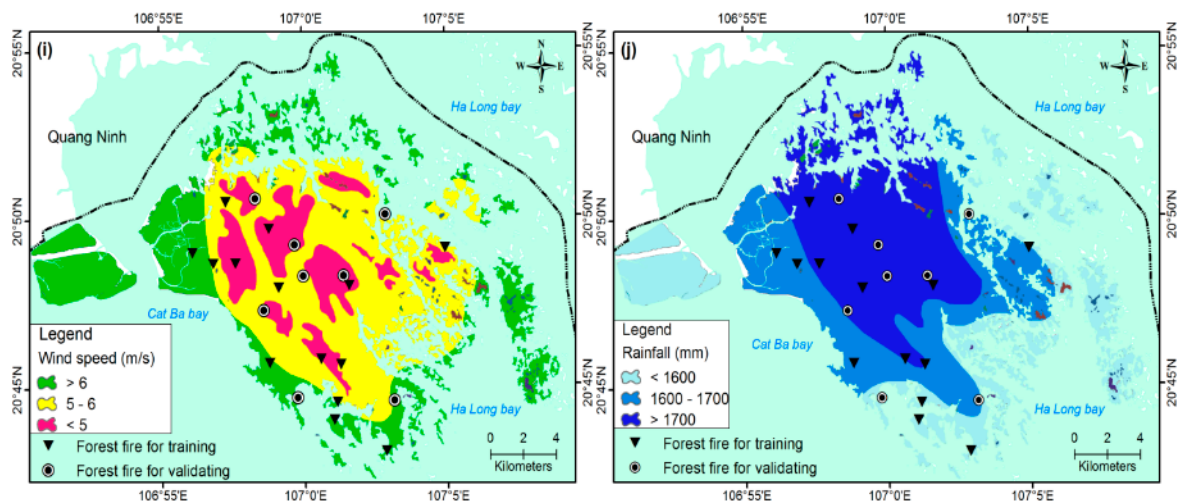
Figure 2. *Cont.*

Figure 2. Forest fire variables: (**a**) slope; (**b**) aspect; (**c**) TWI (*Topographic Wetness Index*); (**d**) distance to road; (**e**) distance to populated areas; (**f**) land cover; (**g**) NDVI; (**h**) surface temperature; (**i**) wind speed; (**j**) rainfall.

Fires may be created by vehicles traveling on roads, *i.e.*, in traffic accidents [11]; therefore, forests near roads have higher susceptible to fires. People are a factor influencing the probability of fires because they may cause accidental fires, especially near populated areas. In addition, the unemployment rate [12] and poor economic conditions may lead to exploitation of resources in forests, and the activity could cause accidental fires. Therefore, the distance to roads and the distance to populated areas were used by buffering the road network and populated areas obtained from the topographic map on the 1:25,000 scale. The distance to the road map (Figure 2d) was constructed with five classes: 0–300, 300–600, 600–900, 900–1200, >1200 m. The distance to the populated area map (Figure 2e) was created with five classes: 0–400, 400–800, 800–1200, 1200–1600, >1600 m. These classes were determined based on analysis of the historical forest fires.

Land cover and NDVI (*Normalized Difference Vegetation Index*) are main factors that have been widely used for fire occurrence analysis because land cover with different types of vegetation are considered a proxy of fuel, whereas NDVI explains for vegetation status for fires [12]. The land cover map was obtained from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) imagery with 15-m resolution (Path 126/Row 46) acquired in 27 December 2010 [34]. Control points for geo-registration of the image were collected in the field using GPS units. In addition, points from the topographic maps at scale 1:25,000 were used. Based on the field survey and available land use maps, ten typical land cover types were identified for the study area, such as dense forest land, scrubland, grass land, mangrove forest land, mangrove grass land, planted forest land, cultivated land, bare land, water surface, and populated area. The classification process was carried out using Maximum Likelihood classification method in the ENVI 4.7 software with an overall accuracy of 86%. The land cover map with the ten classes was shown in Figure 2f. NDVI for this study area was estimated from the above Landsat-7 ETM+ imagery using the following formula:

$$NDVI = (Band\,4 - Band\,3)/(Band\,4 + Band\,3) \tag{1}$$

where Band 4 is the near-infrared band (0.76–0.90 μm) and Band 3 is the red band (0.63–0.69 μm).

The NDVI map (Figure 2g) for this analysis was constructed with five classes as <−0.3, −0.3 to −0.1, −0.1 to 0, 0–0.1, and >0.1. Surface temperature is an important factor that influences forest fires [35]. In this study, the Landsat-7 ETM+ thermal infrared band (band 6, 10.4–12.5 μm) was used to derive surface temperatures using the single-channel algorithm [36]. Detailed explanation for the

calculation of surface temperatures can be found in [37,38]. The surface temperature map (Figure 2h) was constructed with four classes such as <24 °C; 24–26 °C; 26–28 °C; >28 °C.

Wind speed and rainfall are meteorological factors that heavily influence forest fires because they affect directly the evaporation and absorption of waters [39]. In this study, the wind speed map (Figure 2i) was constructed with three classes: <5, 5–6, and >6 m/s using the average wind speeds in 2010. The rainfall map (Figure 2j) was constructed with three classes <1600, 1600–1700, >1700 mm based on the total rainfall in 2010. These data were provided by the National Center for Hydro-Meteorological Forecasting, Ministry of Natural Resources and Environment of Vietnam.

3. Methodology

3.1. Kernel Logistic Regression

Kernel logistic regression (KLR) is a powerful machine learning classification method where probabilistic outcomes are estimated based on minimizing the negative log-likelihood function using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization [40]. Using kernel functions, the KLR maps the input data from the original space into a high-dimensional feature space where the data are linearly separated.

Consider a training dataset $\{x_i, y_i\}_{i=1}^{N}$ with $x_i \in \mathrm{R}^n$ as input data with n variables and N data samples. In this research context, the input variables are slope, aspect, TWI, land cover, NVDI, surface temperature, distance to roads, distance to populated areas, wind speeds, and rainfall. $y_k \in \{1, 0\}$ is the corresponding label that denotes forest fire and non-forest fire classes. KLR aims to build a non-linear decision boundary that could separate the two classes in the feature space using the following equation:

$$p(x) = \mathrm{e}^{y(x)}/(1 + \mathrm{e}^{y(x)}) = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b \tag{2}$$

where $y(x)$ is the logistic function with values in [0,1]; α_i is a vector of dual model parameters, whereas b is the intercept; $K(x_i, x_j)$ is the kernel function.

For this research, Radial Basis Function (RBF) is selected because the function is considered to be the most commonly used [41,42]:

$$K(x, x\prime) = \exp((-||x_i - x_j||^2)/2\delta^2) \tag{3}$$

where δ is the tuning parameter that control the sensitivity of the RBF kernel.

The parameters α_i and b, are obtained by minimizing the negative log-likelihood function as follows:

$$\mathrm{Min}\, \frac{1}{2}\alpha\prime\mathrm{K}\alpha + C\sum_{i=1}^{N}\log(1 + \exp(K_{1i}\alpha)) - C\sum_{i=1}^{N}y_i(K_{1i}\alpha) \tag{4}$$

where C is the regularization parameter that controls the tradeoff between the complexity of the model and degree-of-fit with the data; K_{1i} is the *i*-th row in the kernel matrix.

3.2. Preparation of the Training and the Validation Dataset

From a machine learning point of view, forest fire susceptibility mapping can be considered to be a binary classification problem with two classes: forest fire and non-forest fire. Forest fire points are coded as "1", whereas non-forest fire points are coded as "0" and they represent the dependent variable. For this analysis, the historical forest fires were split into two subsets with 65/35 ratio. The first subset includes 14 historical forest fires that occurred in the period from 15 July 2009 to 15 May 2010. These fires were used for the training of models. The second set with the remaining eight forest fires were used for the model validation and to confirm the prediction accuracy. The fires in the second set occurred from 15 May 2010 to 1 June 2013. The same amount of non-forest fire points were randomly sampled [17,43] from non-forest fire areas in the study area. Finally, values for the ten forest fire related factors were extracted to construct the training and validation datasets.

3.3. Performance Assessment

Performance of the forest fire models was assessed using five statistical evaluation measures such as overall accuracy, specificity, sensitivity, positive predictive value (*PPV*), and negative predictive value (*NPV*) [41]. Overall accuracy is the proportion of the training (or validation) samples that are correctly classified; sensitivity is the proportion the forest fires that are classified correctly; specificity is the proportion of the non-forest fires that are classified correctly. *PPV* is the probability of samples in the training (or validation) dataset that are classified to the forest fire class, whereas *NPV* is the probability of samples in the training (or validation) dataset that are correctly classified to the non-forest fire class:

$$\text{Overall accuracy} = \frac{TP + TN}{TP + TN + FP + FN}; \text{Specificity} = \frac{TN}{FP + TN} \tag{5}$$

$$\text{Sensitiviy} = \frac{TP}{TP + FN}; PPV = \frac{TP}{FP + TP}; NPV = \frac{TN}{FN + TN} \tag{6}$$

where True Positive (*TP*) and True Negative (*TN*) are the number of samples in the training dataset or the validation dataset that are correctly classified to the forest fire class and the non-forest fire class, respectively. False Positive (*FP*) and False Negative (*FN*) are the number of samples in the training dataset or the validation dataset that are erroneously classified.

The global measurement of the model performance can be assessed using the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC) [44]. The ROC curve is a descriptive graph that construct based on the sensitivity *versus* specificity. A perfect model is obtained if AUC equals 1, whereas model is non-informative if AUC is 0.

4. Results and Discussion

The prediction capability of a machine learning model may be enhanced if input variables with null or negative predictive values are removed [41,45]; therefore, the predictive ability of forest fire related factors should be quantified and assessed first. In this study, the Pearson correlation method was used to assess predictive powers of the forest fire related factors due to its efficiency.

The result (Table 1) shows that all the factors have a certain predictive power; therefore, none of these factors were excluded in this analysis. The highest predictive power is NVDI (0.702), followed by TWI (0.681), land cover (0.188), surface temperature (0.149), aspect (0.110), distance to populated area (0.099), slope (0.084), distance to roads (0.070), rainfall (0.051), and wind speed (0.001).

Table 1. Predictive power of the ten forest fire related factors using the Pearson correlation method.

No.	Forest Fire Related Factor	Predictive Power Value
1	NVDI (*Normalized Difference Vegetation Index*)	0.702
2	TWI (*Topographic Wetness Index*)	0.681
3	Land cover	0.188
4	Surface temperature	0.149
5	Aspect	0.110
6	Distance to populated area	0.099
7	Slope	0.084
8	Distance to roads	0.070
9	Rainfall	0.051
10	Wind speed	0.001

Since the performance of the KLR depends strongly on the selection of two parameters, δ and *C* (see Equations (3) and (4)); therefore, it is important to determine them properly. To obtain this purpose, the grid search technique that involves a trial and error search with a fixed grid of parameters [46] was used. With the training dataset, the best values for δ and *C* are 0.037 and 0.03, respectively. Using the best values, the KLR model was constructed and the detailed statistical evaluation measures for the

model are shown in Table 2. It could be seen that overall accuracies for the KLR model are 89.29% on the training dataset and 81.25% on the validation dataset.

Table 2. Statistical evaluation measures of the Kernel logistic regression model and the support vector machine model in this study (PPV: Positive predictive value; NPV: Negative predictive value).

Parameter	Training Data		Validation Data	
	Kernel Logistic Regression	Support Vector Machine	Kernel Logistic Regression	Support Vector Machine
Sensitivity (%)	92.86	85.71	87.50	87.50
Specificity (%)	85.71	85.71	75.00	75.00
PPV (%)	86.67	85.71	77.78	77.78
NPV (%)	92.31	85.71	85.71	85.71
Overall accuracy (%)	89.29	85.71	81.25	81.25
Kappa index	0.785	0.714	0.625	0.625

The positive predictive value is 86.67% indicating that the model correctly classifies forest fire points with a probability of 86.67%. The negative predictive value is 92.31% indicating that the model correctly classifies non-fire points with a probability of 92.31%. The sensitivity is 92.86%, indicating that 92.86% of the forest fire points are correctly classified by the KLR model, whereas 85.71% of none-fire points are correctly classified (Table 2). The Kappa index is 0.785 indicating 78.5% better than random, and a substantial high agreement between the KLR model and the training data. In general, the model performs well on both the training and validation datasets.

The global performance of the KLR model is measured using the ROC curve and the AUC. The results (Figure 3) show that the AUC is 0.959, indicating that the model has high goodness-of-fit with the training dataset, whereas the AUC is 0.922 for the validation dataset indicating that the prediction power of the model is 92.2%. Overall, the KLR model demonstrates high global performance and acceptable overall accuracy.
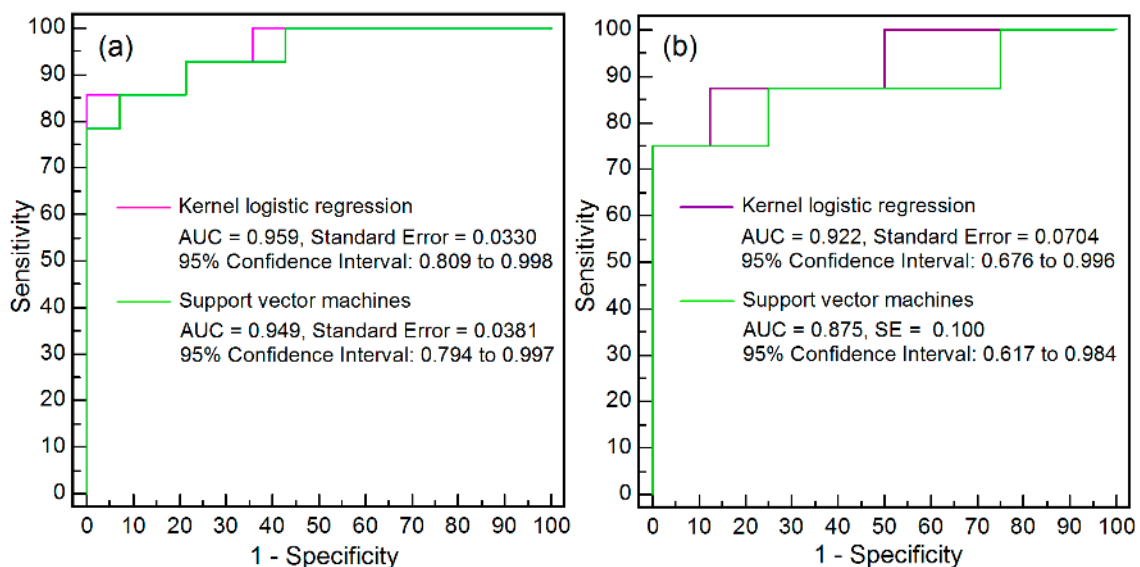

Figure 3. The Receiver Operating Characteristic (ROC) curve and area under the ROC curve (AUC) for the Kernel logistic regression model and the support vector machine model on (**a**) the training dataset and (**b**) validation dataset.

In order to evaluate the usability of the KLR model for the forest fire susceptibility mapping, the SVM has been employed as a benchmark method for the comparison purpose using the same datasets.

The SVM was selected because it is widely accepted to be an effective method for modelling of various nonlinear and complex problems [46]. For this research, the radial basic function (RBF) kernel was selected for the SVM model due to its ability to yield better results in various studies [47–51]. The best values for kernel width and regularization parameters of the SVM were obtained using the grid search method as suggested in Tien Bui, Tuan, Klempe, Pradhan and Revhaug [41], and the optimal values were found as 0.095 and 6.95 for the kernel width and regularization parameters, respectively.

The overall performance and prediction capability of the benchmark model are shown in Table 2 and Figure 3. It could be observed that the overall accuracies are 85.71% and 81.25%, indicating high performance; however, the global fit of the SVM is slightly lower than the KLR model (Figure 3a,b). Although Kappa indices of the two models are equal in the validation dataset, the Kappa index of the benchmark model is lower than around 7% compared to the KLR model in the training dataset. More importantly, the prediction power of the benchmark is 4.7% lower than the KLR model (Figure 3b). Therefore, it could be concluded that the KLR model performs better than the benchmark model.

Based on the analysis and comparison results, the KLR model is suitable for the tropical forest fire susceptibility mapping in the Cat Ba area. The model was used to calculate fire index values for all the pixels of the study area. In the next step, these pixels were converted to a GIS format using an application in C++ programming and opened in ArcGIS 10.2 software. For the purpose of comparison, a forest fire susceptible map produced by the SVM model has also been shown. These forest fire susceptible maps (Figures 4 and 5) have been visualized by six classes: extremely high (5%), very high (5%), high (15%), moderate (20%), low (25%), and very low (30%).
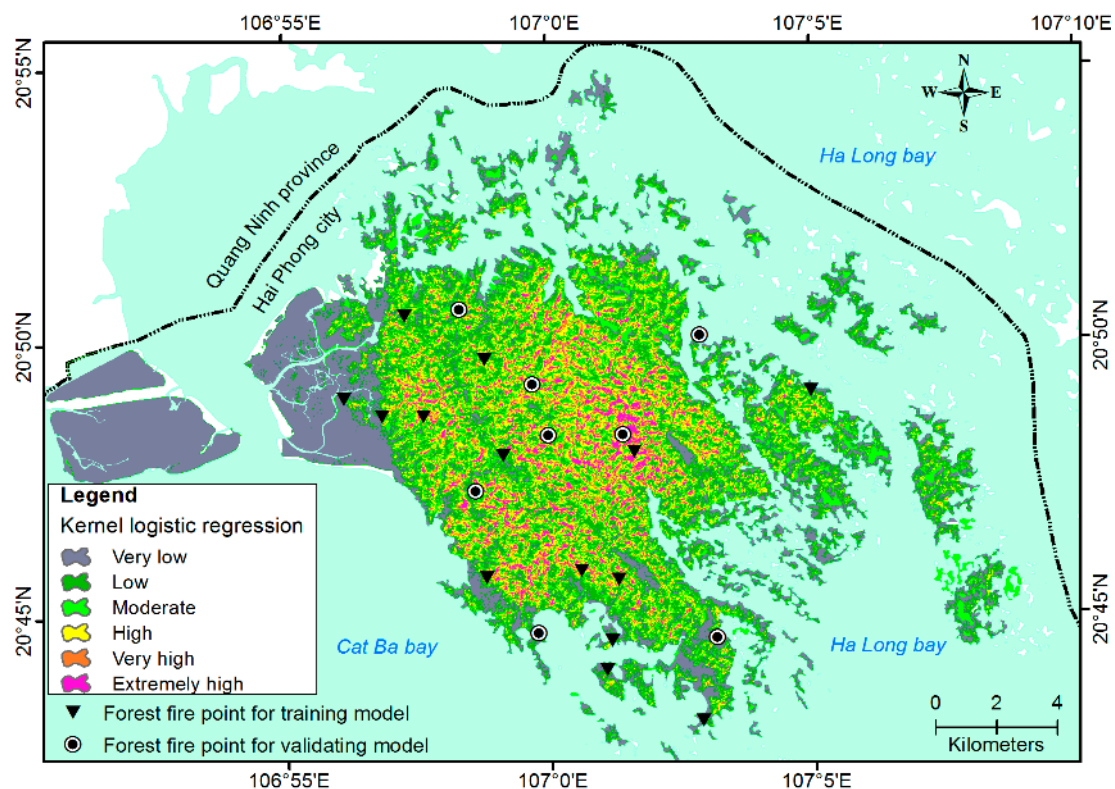
Figure 4. Forest fire susceptibility map for Cat Ba National Park area, Hai Phong City (Vietnam) using the Kernel logistic regression model.
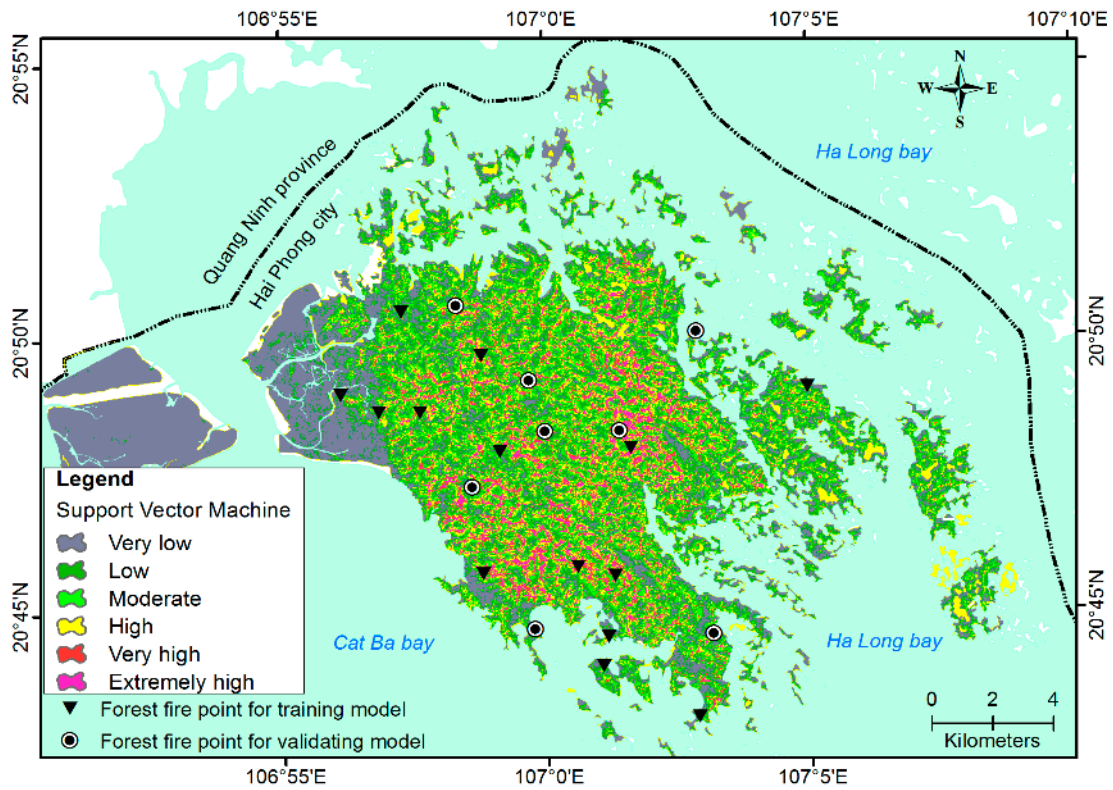
Figure 5. Forest fire susceptibility map for Cat Ba National Park area, Hai Phong City (Vietnam) using the Support Vector Machine model.

The five classes were determined by overlaying all the forest fire points on the two forest fire susceptible maps, and then a graphic curve (Figure 6) was constructed based on the percentage of the forest fire points *versus* percentage of forest fire susceptible map (sorted from high to low values). Detailed instructions on how to build the graphical curve can be seen in Chung, Fabbri and Van westen [52] and Tien Bui *et al.* [53–56]. Based on the graphic curve, susceptibility index ranges were obtained and susceptibility classes were determined (Table 3).
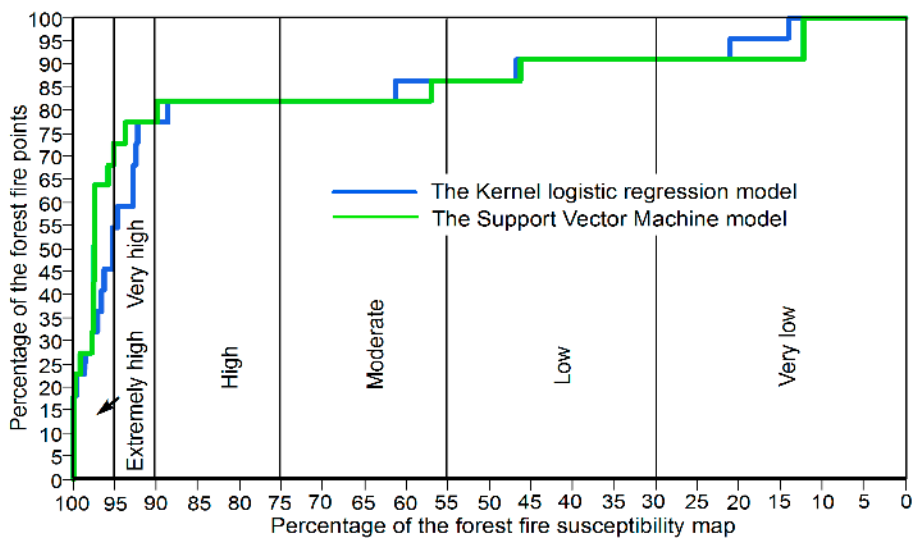

Figure 6. Cumulative percentage of the forest fire and susceptibility map.

Table 3. Forest fire susceptibility classification derived from the Kernel logistic regression model.

No.	Susceptibility Index Range	Fire Susceptibility (%)	Verbal Expression	Areas (km^2)
1	0.903–0.746	100%–95%	Extremely high	10.5
2	0.746–0.703	90%–95%	Very high	10.5
3	0.703–0.614	75%–90%	High	31.2
4	0.614–0.536	55%–75%	Medium	41.8
5	0.536–0.372	30%–55%	Low	52.2
6	0.372–0.065	0–30%	Very low	62.7

The results (Table 3 and Figure 6) show that around 81.82% of the forest fire points are located in the extremely high and very high classes for the case of the KLR model, whereas 77.27% of the forest fires are located in the extremely high and very high classes in the case of the SVM model. More specifically, 5% of the highest susceptibility map contains 74.6% and 70.5% of the total forest fires for the KLR model and the SVM model, respectively (Figure 6). These confirm that the KLR model performs well and slightly better than the SVM model.

5. Conclusions

Cat Ba National Park is the UNESCO designated biodiversity conservation area; therefore, this area has been the site for many research projects [20]; however, no attempt has been carried out for forest fire susceptibility assessment. During the last five years, the prevention of forest fires has received particular attention from the local authorities due to several forest fires that occurred. In addition, the area is a destination for tourists, with around 500,000 travelers per year. Therefore, study of forest fires is an urgent task.

We addressed the problem in this project by providing a map and a model for forest fire susceptibility. The model was developed using 22 forest fire locations and ten related factors (slope, aspect, TWI, distance to roads, distance to populated areas, land cover, NVDI, surface temperature, wind speed, and rainfall). A novel machine learning method KLR was proposed to be used for creating a forest fire model. According to current literature, this is the first time KLR has been used for forest fire modelling.

The proposed model shows high performance in both the training and validation dataset with the overall accuracy and prediction power of 89.29% and 92.2%, respectively, indicating that the proposed model is satisfactory for forest fire modeling. In addition, the ten related factors have predictive values to the forest fires, indicating that the process of collection, processing, and coding factors has been conducted successfully. NVDI, TWI, land cover, and surface temperature have the highest predictive powers for the forest fires in this study. Susceptibility index values obtained from the KLR vary from 0.065 to 0.903, which show the probabilities of fire will occur. The forest fire susceptibility map for the study was then reclassified into six classes: extremely high, very high, high, medium, low, and very low (Table 3). Interpretation of the map shows that the extremely high and very high classes occupy 20.1 km^2 (10% of the study area), but contain 81.82% of the total forest fires. The very low class (62.7 km^2) and the low class (52.2 km^2) occupy large areas but contain 9.1% and 4.6% of the total forest fires, respectively. These indicate that the KLR model produced satisfied results.

The prediction power of the KLR model has outperformed the benchmark model, the SVM. Therefore, the proposed model is a promising alternative tool that should be considered for use of forest fire susceptibility mapping in other areas. The main limitation of this study is that only ten related factors were used; therefore, the quality of the proposed model could be enhanced if other factors are considered such as humidity and drought. Despite the limitation, the forest fire susceptibility map could help the local authority in forest planning and management. In practice, the local planer could use the map to delineate areas with very high susceptibility for fires, and, based on that, a forest fire early warning system could deliver timely awareness of danger.

Author Contributions: Dieu Tien Bui, Kim-Thoa Thi Le, Van Cam Nguyen and Hoang Duc Le collected and processed the data. Dieu Tien Bui and Inge Revhaug designed the experiment, performed the analysis, and wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dieterle, G. Sustaining the world's forests: Managing competing demands for a vital resource—The role of the world bank. In *Sustainable Forest Management in a Changing World*; Springer: Medford, MA, USA, 2009; pp. 9–32.

2. Köhl, M.; Lasco, R.; Cifuentes, M.; Jonsson, Ö.; Korhonen, K.T.; Mundhenk, P.; de Jesus Navar, J.; Stinson, G. Changes in forest production, biomass and carbon: Results from the 2015 un fao global forest resource assessment. *For. Ecol. Manag.* **2015**, *352*, 21–34. [CrossRef]

3. Flannigan, M.D.; Stocks, B.J.; Wotton, B.M. Climate change and forest fires. *Sci. Total Environ.* **2000**, *262*, 221–229. [CrossRef]

4. Flannigan, M.D.; Amiro, B.D.; Logan, K.A.; Stocks, B.; Wotton, B. Forest fires and climate change in the 21st century. *Mitig. Adapt. Strateg. Glob. Chang.* **2006**, *11*, 847–859. [CrossRef]

5. Ha, C.T. Vietnam National Forest Status of 2012. *Annually Report of Ministry of Agriculture and Rural Development*; Ministry of Agriculture and Rural Development: Hanoi, Vietnam, 2013.

6. Samphantharak, K. Natural disasters and the economy: Some recent experiences from Southeast Asia. *Asian-Pac. Econ. Lit.* **2014**, *28*, 33–51. [CrossRef]

7. Le, T.H.; Thanh Nguyen, T.N.; Lasko, K.; Ilavajhala, S.; Vadrevu, K.P.; Justice, C. Vegetation fires and air pollution in vietnam. *Environ. Pollut.* **2014**, *195*, 267–275. [CrossRef] [PubMed]

8. Thuy, P.T.; Moeliono, M.; Hien, N.T.; Tho, N.H.; Hien, V.T. *The Context of REDD+ in Vietnam: Drivers, Agents and Institutions*; CIFOR Occasional Paper; CIFOR: Bogor, Indonesia, 2012.

9. Asian Development Bank. *Viet Nam: Environment and Climate Change Assessment*; Asian Development Bank: Mandaluyong, Philippines, 2013.

10. González, J.R.; Kolehmainen, O.; Pukkala, T. Using expert knowledge to model forest stand vulnerability to fire. *Comput. Electron. Agric.* **2007**, *55*, 107–114. [CrossRef]

11. Jaiswal, R.K.; Mukherjee, S.; Raju, K.D.; Saxena, R. Forest fire risk zone mapping from satellite imagery and GIS. *Int. J. Appl. Earth Observ. Geoinf.* **2002**, *4*, 1–10. [CrossRef]

12. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling spatial patterns of fire occurrence in mediterranean Europe using multiple regression and random forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [CrossRef]

13. Koutsias, N.; Martínez-Fernández, J.; Allgöwer, B. Do factors causing wildfires vary in space? Evidence from geographically weighted regression. *GISci. Remote Sens.* **2010**, *47*, 221–240. [CrossRef]

14. Guo, F.; Su, Z.; Wang, G.; Sun, L.; Lin, F.; Liu, A. Wildfire ignition in the forests of southeast china: Identifying drivers and spatial distribution to predict wildfire likelihood. *Appl. Geogr.* **2016**, *66*, 12–21. [CrossRef]

15. Pourghasemi, H.R. GIS-based forest fire susceptibility mapping in Iran: A comparison between evidential belief function and binary logistic regression models. *Scand. J. For. Res.* **2015**, *31*, 80–98. [CrossRef]

16. Eastaugh, C.S.; Hasenauer, H. Deriving forest fire ignition risk with biogeochemical process modelling. *Environ. Modell. Softw.* **2014**, *55*, 132–142. [CrossRef] [PubMed]

17. Pourtaghi, Z.; Pourghasemi, H.; Rossi, M. Forest fire susceptibility mapping in the minudasht forests, golestan province, Iran. *Environ. Earth Sci.* **2015**, *73*, 1515–1533. [CrossRef]

18. Massada, A.B.; Syphard, A.D.; Stewart, S.I.; Radeloff, V.C. Wildfire ignition-distribution modelling: A comparative study in the huron–manistee national forest, Michigan, USA. *Int. J. Wildland Fire* **2013**, *22*, 174–183. [CrossRef]

19. Le Viet, T.; Choisy, M.; Bryant, J.; Vu Trong, D.; Pham Quang, T.; Horby, P.; Nguyen Tran, H.; Tran Thi Kieu, H.; Nguyen Vu, T.; Van Nguyen, K.; *et al.* A dengue outbreak on a floating village at Cat Ba island in vietnam. *BMC Public Health* **2015**, *15*, 1–8. [CrossRef] [PubMed]

20. Nguyen, N.C.; Bosch, O.J.H. A systems thinking approach to identify leverage points for sustainability: A case study in the Cat Ba biosphere reserve, Vietnam. *Syst. Res. Behav. Sci.* **2013**, *30*, 104–115. [CrossRef]

21. Zingerli, C. Colliding understandings of biodiversity conservation in Vietnam: Global claims, national interests, and local struggles. *Soc. Nat. Resour.* **2005**, *18*, 733–747. [CrossRef]

22. Thung, D.C. Distinctive features of the property of Cat Ba archipelago, Vietnam. *J. Earth Sci. Eng.* **2014**, *4*, 271–283.

23. Van, Q.N.; Duc, T.T.; Van, H.D. Landscapes and ecosystems of tropical limestone: Case study of the Cat Ba islands, Vietnam. *J. Ecol. Environ.* **2010**, *33*, 23–36. [CrossRef]

24. Schmidt-Thome, P.; Nguyen, T.H.; Pham, T.L.; Jarva, J.; Nuottimäki, K. Climate change in Vietnam. In *Climate Change Adaptation Measures in Vietnam*; Springer: Medford, MA, USA, 2015; pp. 7–15.

25. Nguyen, Q.-T. Extreme climatic events over Vietnam from observational data and RegCM3 projections. *Clim. Res.* **2011**, *49*, 87–100.

26. Massada, A.B.; Syphard, A.D.; Hawbaker, T.J.; Stewart, S.I.; Radeloff, V.C. Effects of ignition location models on the burn patterns of simulated wildfires. *Environ. Model. Softw.* **2011**, *26*, 583–592. [CrossRef]

27. Ministry of Agriculture and Rural Development of Vietnam. The Vietnam's Firewatch System for Online Monitoring and Management of Forest Fires. Available online: http://www.Kiemlam.Org.Vn/firewatchvn (accessed on 12 April 2016).

28. Giglio, L.; Descloitres, J.; Justice, C.O.; Kaufman, Y.J. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.* **2003**, *87*, 273–282. [CrossRef]

29. Cosslett, T.L.; Cosslett, P.D. *Water Resources and Food Security in the Vietnam Mekong Delta*; Springer: Medford, MA, USA, 2014.

30. Cary, G.J.; Flannigan, M.D.; Keane, R.E.; Bradstock, R.A.; Davies, I.D.; Lenihan, J.M.; Li, C.; Logan, K.A.; Parsons, R.A. Relative importance of fuel management, ignition management and weather for area burned: Evidence from five landscape–fire–succession models. *Int. J. Wildland Fire* **2009**, *18*, 147–156. [CrossRef]

31. Oliveira, S.; Pereira, J.M.; San-Miguel-Ayanz, J.; Lourenço, L. Exploring the spatial patterns of fire density in southern Europe using geographically weighted regression. *Appl. Geogr.* **2014**, *51*, 143–157. [CrossRef]

32. Vasilakos, C.; Kalabokidis, K.; Hatzopoulos, J.; Matsinos, I. Identifying wildland fire ignition factors through sensitivity analysis of a neural network. *Nat. Hazards* **2009**, *50*, 125–143. [CrossRef]

33. Conedera, M.; Peter, L.; Marxer, P.; Forster, F.; Rickenmann, D.; Re, L. Consequences of forest fires on the hydrogeological response of mountain catchments: A case study of the riale buffaga, Ticino, Switzerland. *Earth Surf. Process. Landf.* **2003**, *28*, 117–129. [CrossRef]

34. United States Geological Survey. The United States Geological Survey Earth Resources Observation and Science Center. Available online: http://earthexplorer.Usgs.Gov (accessed on 12 April 2016).

35. Sánchez, J.M.; Bisquert, M.; Rubio, E.; Caselles, V. Impact of land cover change induced by a fire event on the surface energy fluxes derived from remote sensing. *Remote Sens.* **2015**, *7*, 14899–14915. [CrossRef]

36. Jiménez-Muñoz, J.C.; Cristóbal, J.; Sobrino, J.; Soria, G.; Ninyerola, M.; Pons, X. Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 339–349. [CrossRef]

37. Brabyn, L.; Zawar-Reza, P.; Stichbury, G.; Cary, C.; Storey, B.; Laughlin, D.C.; Katurji, M. Accuracy assessment of land surface temperature retrievals from Landsat 7 ETM+ in the Dry Valleys of Antarctica using iButton temperature loggers and weather station data. *Environ. Monit. Assess.* **2014**, *186*, 2619–2628. [CrossRef] [PubMed]

38. Walawender, J.P.; Szymanowski, M.; Hajto, M.J.; Bokwa, A. Land surface temperature patterns in the urban agglomeration of Krakow (Poland) derived from Landsat-7/ETM+ data. *Pure Appl. Geophys.* **2014**, *171*, 913–940. [CrossRef]

39. Viney, N.R. A review of fine fuel moisture modelling. *Int. J. Wildland Fire* **1991**, *1*, 215–234. [CrossRef]

40. Witten, I.H.; Frank, E.; Mark, A.H. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, VT, USA, 2011.

41. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**. [CrossRef]

42. Hong, H.; Pradhan, B.; Xu, C.; Tien Bui, D. Spatial prediction of landslide hazard at the Yihuang Area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *CATENA* **2015**, *133*, 266–281. [CrossRef]

43. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility mapping at Hoa Binh Province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput. Geosci.* **2012**, *45*, 199–211. [CrossRef]

44. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

45. Martínez-Álvarez, F.; Reyes, J.; Morales-Esteban, A.; Rubio-Escudero, C. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. *Knowl.-Based Syst.* **2013**, *50*, 198–210. [CrossRef]

46. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and Naïve Bayes models. *Math. Probl. Eng.* **2012**, *2012*, 1–26. [CrossRef]

47. Hong, H.; Pradhan, B.; Jebur, M.; Bui, D.; Xu, C.; Akgun, A. Spatial prediction of landslide hazard at the Luxi Area (China) using support vector machines. *Environ. Earth Sci.* **2015**, *75*, 1–14. [CrossRef]

48. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Application of support vector machines in landslide susceptibility assessment for the Hoa Binh Province (Vietnam) with kernel functions analysis. In Proceedings of the iEMSs Sixth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2012), Leipzig, Germany, 1–5 July 2012.

49. Tien Bui, D.; Anh Tuan, T.; Hoang, N.-D.; Quoc Thanh, N.; Nguyen, B.D.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a novel hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2016**. [CrossRef]

50. Hoang, N.-D.; Tien Bui, D. A novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. *J. Comput. Civ. Eng.* **2016**. [CrossRef]

51. Were, K.; Tien Bui, D.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an afromontane landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [CrossRef]

52. Chung, C.J.F.; Fabbri, A.G.; Van Westen, C.J. Multivariate regression analysis for landslide hazard zonation. In *Geographical Information Systems in Assessing Natural Hazards*; Carrara, A., Guzzetti, F., Eds.; Springer: New York, NY, USA, 1995; Volume 5, pp. 107–133.

53. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Spatial prediction of landslide hazards in Hoa Binh Province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **2012**, *96*, 28–40. [CrossRef]

54. Tien Bui, D.; Pham, T.B.; Nguyen, Q.-P.; Hoang, N.-D. Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of least squares support vector machines and differential evolution optimization: A case study in central Vietnam. *Int. J. Digit. Earth* **2016**. [CrossRef]

55. Tien Bui, D.; Nguyen, Q.-P.; Hoang, N.-D.; Klempe, H. A novel fuzzy k-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. *Landslides* **2016**. [CrossRef]

56. Tien Bui, D.; Pradhan, B.; Revhaug, I.; Trung Tran, C. A comparative assessment between the application of fuzzy unordered rules induction algorithm and j48 decision tree models in spatial prediction of shallow landslides at Lang son city, Vietnam. In *Remote Sensing Applications in Environmental Research*; Springer International Publishing: Cham, Switzerland, 2014; pp. 87–111.