Article

Quantifying Forest Fire and Post-Fire Vegetation Recovery in the Daxin’anling Area of Northeastern China Using Landsat Time-Series Data and Machine Learning

Jie Qiu 1,†, Heng Wang 2,†, Wenjuan Shen 1, Yali Zhang 1, Huiyi Su 1 and Mingshi Li 1,3,*

1 College of Forestry, Nanjing Forestry University, Nanjing 210037, China; rongyu19970116@njfu.edu.cn (J.Q.); wjsen@njfu.edu.cn (W.S.); nfuzy1128@njfu.edu.cn (Y.Z.); suhuiyi@njfu.edu.cn (H.S.)
2 Land-Reserve Center of Jiangbei New Area, Nanjing 210037, China; tobyherzig92@gmail.com
3 Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: nfulms@njfu.edu.cn; Tel.: +86-25-85427327
† These authors contributed equally to this work.

Abstract: Many post-fire on-site factors, including fire severity, management strategies, topography, and local climate, are concerns for forest managers and recovery ecologists to formulate forest vegetation recovery plans in response to climate change. We used the Vegetation Change Tracker (VCT) algorithm to map forest disturbance in the Daxing’anling area, Northeastern China, from 1987 to 2016. A support vector machine (SVM) classifier and historical fire records were used to separate burned patches from disturbance patches obtained from VCT. Afterward, stepwise multiple linear regression (SMLR), SVM, and random forest (RF) were applied to assess the statistical relationships between vegetation recovery characteristics and various influential factors. The results indicated that the forest disturbance events obtained from VCT had high spatial accuracy, ranging from 70% to 86% for most years. The overall accuracy of the annual fire patches extracted from the proposed VCT-SVM algorithm was over 92%. The modeling accuracy of post-fire vegetation recovery was excellent, and the validation results confirmed that the RF algorithm provided better prediction accuracy than SVM and SMLR. In conclusion, topographic variables (e.g., elevation) and meteorological variables (e.g., the post-fire annual precipitation in the second year, the post-fire average relative humidity in the fifth year, and the post-fire extreme maximum temperature in the third year) jointly affect vegetation recovery in this cold temperate continental monsoon climate region.

Keywords: fire disturbance; post-fire recovery; VCT; SVM; SMLR; random forest

1. Introduction

Fire is a natural disturbance in ecosystems and promotes diversity and natural regeneration [1]. However, warm and dry conditions increase wildfire activity [2]. At the global scale, fire is a critical factor affecting atmospheric chemistry [3], forest ecosystems succession and health [4], carbon budgets, and land use transformation [5]. At the regional scale, fire affects the socioeconomic system, soil erosion, and the hydrological cycle [6]. The diversity of fire impacts requires a comprehensive assessment of fire vulnerability of different systems [7], including the analysis of the potential damage of fire on society and ecological systems. Additionally, knowing when and where fire occurs is of great significance for forest management, carbon cycle prediction after fire events, and local or regional scale climate effects [5]. In turn, vegetation recovery influences the climate and forest management [8]. Post-fire recovery can affect surface radiation balance, carbon budgets, and water balance by changing the surface roughness, surface albedo, soil moisture, erosion, and evapotranspiration, thus affecting the climate [9–11]. In addition, vegetation change also affects the re-establishment of ecological functions, species habitat recovery, and biota.
recolonization [12–14]. Thus, post-fire vegetation recovery and ecosystem resilience are current topics of interest for land managers and researchers [15].

Theoretically, post-fire vegetation recovery is limited by numerous on-site factors, including environmental factors [11,12], regeneration strategies [16], topography [17], local climate [15,18], forest structure [19], and resilience of the vegetation species [4]. Although existing climate change studies have improved the knowledge of the climate system [8], quantifying the relationship between climate and post-fire vegetation recovery deserves further concern for promoting the sustainable management of forests.

Combined with limited ground data, remote sensing provides a convenient, near real-time, and cost-effective method for the study of fire detection and vegetation recovery in burned areas [20] and has been extensively used in the past decades. Although high/low spatial resolution data perform well in some studies (e.g., structural damage assessments/wildfire assessments), medium resolution data are more suitable for monitoring wildfires and their impacts due to the appropriate spatial and temporal resolutions. Currently, forest change monitoring using Landsat time-series data has been performed in many forest ecosystems worldwide due to the availability of free 45+ year moderate spatial resolution data and the remarkable potential for tracking forest dynamics after fires [21].

Several highly reputed automated algorithms were widely used to detect vegetation changes in forest ecosystems and have provided good accuracy [19,22]. The Vegetation Change Tracker (VCT) algorithm was used to map forest change using Landsat time-series stacks. The VCT algorithm considers the spectral characteristics of forest disturbance recovery over time and uses a threshold-dependent criterion to determine forest disturbance events [19]. The VCT algorithm improves the efficiency for detecting forest disturbance [23] and recovery in frequently managed forest areas [24]. The mapping accuracy of forest disturbance by VCT was relatively high [25,26]. However, VCT is unable to identify specific types of disturbances, particularly fire disturbances [25]. Recently, machine learning algorithms, such as support vector machine (SVM) [24] and random forest (RF) [23], have been used in remote sensing to detect and classify disturbances [25–27]. The combination of machine learning algorithms and VCT was used to extract burned areas [25]. Although it was found that the combination of VCT products and the SVM algorithm did not perform well for extracting disturbances related to insect, snow, and wind damage [28], the proposed VCT-SVM method has proven effective for separating fire disturbance from other disturbance types [26]. However, due to different forest ecosystem types, fire regimes, and forest management approaches, the efficacy of VCT-SVM needs to be further validated in Chinese cold temperate forests to evaluate the effect of fires on human and natural systems [28].

Several algorithms for extracting burned areas have been developed to characterize and evaluate the patterns of forest ecosystem responses to fire disturbances [29]. Numerous empirical models have been used to simulate post-fire vegetation recovery [22,30]. These studies confirmed that climatic conditions played a critical role in the initial stage of vegetation recovery [30]. Extreme high temperatures were found to be highly correlated with the normalized difference vegetation index (NDVI) [31]. Additionally, post-fire climate factors, in particular those related to soil, were important to determine the degree of vegetation recovery after a fire [32].

Located in the northeast of China, the Daxing’anling area is one of the most critical areas of secondary forests and has frequently suffered from forest fires. The objectives of this study are to (a) validate the performance of the VCT-SVM algorithm for extracting burned areas in a cold temperate forest ecosystem and (b) to identify the factors contributing to vegetation recovery in a cold temperate forest ecosystem after disturbance and provide information for forest management in similar regions.
2. Materials and Methods

2.1. Study Area

The Daxing’anling area, which covers 83,000 km$^2$, is located in northeastern China at latitudes of 50°10′ to 53°33′ N and longitudes of 121°12′ to 127°00′ E (Figure 1). The area is located at an average elevation of 573 m in a cold temperate zone and has a continental monsoon climate with long, cold winters and hot, short summers. According to the meteorological administration of China, the daily mean temperature is −3 °C, the minimum temperature is −48 °C, and the maximum temperature is 36 °C. The annual rainfall is 500 mm. The minimum average monthly rainfall occurs typically in April and June, leading to a relatively dry period. The details on the Daxing’anling area were obtained from the Daxing’anling District Administration Office (http://www.dxal.gov.cn/ (accessed on 24 December 2020)).

![Figure 1. Location of the study area in Northern China. The colored tiles on the right show the Landsat footprints (p119r024, p119r025, p120r023, p120r024, p120r025, p121r023, p121r022, p121r025, p122r023, p122r024, and p123r025), and the black polygon is the exact study site.](image)

2.2. Data and Preprocessing

Annual Landsat Thematic Mapper (TM)/Enhanced Thematic Mapper (ETM+)/Operational Land Imager (OLI) observations during 1987–2016 were used in this study. The TM/ETM+/OLI images were geometrically and atmospherically corrected by the USGS EROS data center using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) in batch mode [33] or the Landsat 8 Surface Reflectance Code (LaSRC) system per customer requests when placing orders. The surface reflectance images were then freely downloaded from the USGS EROS Center Science Processing Architecture (ESPA) (https://espa.cr.usgs.gov/ (accessed on 24 December 2020)). The images had to conform to the following two rules to ensure sufficient accuracy of the forest change detection: (1) the dates of the image acquisition had to fall into the growing season (mid-June to mid-September) and (2) the images had to have low or no cloud cover. For images with a relatively high cloud cover in a given year, we collected other images from the same path/row tile in the same growing season to composite a new image if the clouds in the images were not spatially coincident. Additionally, the Shuttle Radar Topography Mission...
3 (SRTM3) digital elevation model (DEM) dataset was downloaded from the Global Land Cover Facility (GLCF) at the University of Maryland to support the analysis.

The fire data in point format used in this study was provided by the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences. The data consisted of all fire events recorded during 1967–2004 in the Daxing’anling region, including the center latitude and longitude coordinates of the fire points, the estimated burned area, and the burn severity. The fire point data after 2004 were obtained from the “13th Five-Year Plan” reports compiled by the State Forestry Administration of China and the China Ecological Remote Sensing Information Service Network (http://rsapp.nsmc.org.cn (accessed on 24 December 2020)). Additionally, from 12 August to 17 August 2017, our team conducted a field investigation of the vegetation recovery status and progress of a catastrophic forest fire that occurred on 6 May, 1987. We obtained the locations of fire points in 1987 and interviewed the local people to glean information on vegetation recovery methods. The fire point data after 2004 and the field investigation data included the geographical coordinates and the year of fire occurrence. Four large fires that occurred in 1987, 2000, 2006, and 2010 were selected to train the fire extraction model and vegetation recovery algorithms because of the large burned areas, high burn severity, and significant ecological effects.

The meteorological data were the annual climate data set of China, provided by the China meteorological data network (http://data.cma.cn/ (accessed on 24 December 2020)). A detailed description of the meteorological data is listed in Table 2. We only downloaded observations that included the annual average temperature, the extreme maximum temperature, the extreme minimum temperature, the average maximum temperature, the average minimum temperature, and the average relative moisture from the 33 meteorological stations in Heilongjiang province. The discrete station observations were interpolated to create raster layers using the co-kriging interpolation algorithm, which is based on the theory of variograms and structure analysis to obtain unbiased estimates of the regionalized variables in a limited area [34]. In the interpolation, elevation was used as a covariable of the meteorological factors, and GS+ software was used to conduct semi-variance analysis of the meteorological station data. The optimal semi-variogram function was an exponential function, with an R-squared (R²) value of above 0.9.

2.3. VCT-Based Forest Disturbance Mapping and Validation

We used the VCT algorithm to create forest disturbance time series in the study area from 1987 to 2016. The VCT algorithm considers the spectral-temporal characteristics of land cover and forest change; individual image analysis and time-series analysis are conducted. The output of the VCT algorithm includes the years of forest disturbance, forest disturbance maps, and the disturbance magnitude. The integrated forest z-score (IFZ) was used in the VCT algorithm to track forest changes. The IFZ value is defined by integrating $FZ_i$ over the spectral bands of the multi-spectral images. The calculation formulas of the IFZ and the $FZ_i$ are as follows:

$$FZ_i = \frac{b_{pi} - \bar{b}_i}{SD_i}$$

(1)

$$IFZ = \sqrt{\frac{1}{NB} \sum_{i=1}^{NB} (FZ_i)^2}$$

(2)

where $b_{pi}$ represent the value of a given pixel of band $i$, and $b_i$ and $SD_i$ represent the mean and standard deviation of the spectral value of the forest pixel of each band $i$, respectively. $NB$ represents the number of bands, which was 3 in this study (Landsat TM/ETM+ images use bands 3, 5, 7; Landsat OLI images use bands 4, 6, 7). The IFZ is an inverse measure of the likelihood of each pixel being a forest pixel. When the IFZ value of a pixel shows sustained relative stability and is lower than 0.3 (dense forests) or 0.2 (sparse forests), the pixel is identified as persisting forest. Noise observations are likely to result in an IFZ peak (a high IFZ value preceded and followed by low IFZ values), whereas disturbance observations are...
likely to result in consecutive high $IFZ$ values. Consecutive high $IFZ$ values were used to avoid noise. Forest disturbance in a year is usually accompanied by a sharp increase in the $IFZ$ value, whereas vegetation recovery was defined as the gradual reduction in $IFZ$ values. The disturbance maps could provide information on where and when disturbances had occurred; thus, we focused on the disturbance maps in this research. The background values in the disturbance maps of the VCT were removed. Persisting forest and previously disturbed forest that was spectrally similar to the forest in this year were defined as forests. The remaining categories were defined as non-forest, including disturbed in this year, persisting non-forest, persisting water, and post-disturbance non-forest.

Additionally, we adopted the accuracy validation method proposed by Li et al. [34] to assess the accuracy of the forest disturbance maps using a spatial agreement index. Specifically, we used a spatial agreement index (SAI) based on visual interpretation and Google Earth high-resolution images to verify the accuracy of the VCT forest disturbance products. We randomly selected three $3 \times 3$ km squares in the study area as the validation area (Figure 2) to derive the SAI statistics using Equation (3).

$$SAI = \frac{S_{cd}}{S_g}$$  \hspace{1cm} (3)

Figure 2. (a) Forest disturbance patterns obtained from the VCT algorithm in the study area, including persisting forest, persisting non-forest, persisting water, and forest disturbance in different years; (b) annual forest disturbance patterns of the western quadrat; (c) Landsat image of the western quadrat before the disturbance in 1986; (d) Landsat image of the western quadrat after the disturbance in 2017.

In this method, $S_d$ is the disturbed area detected by VCT, $S_g$ is the disturbed area determined by visual interpretations, and $S_{cd}$ is the overlap area of the two detected regions. We superimposed the annual maps of aggregated forest and non-forest obtained from VCT in ArcGIS to obtain a map of the annual vegetation recovery status.

2.4. Burned Area Extraction Using the VCT Products and SVM Algorithm

We combined the SVM classifier and the disturbances detected by the VCT to extract the burned areas in 1987, 2000, 2006, and 2010. We used the SVM classifier because this algorithm has the advantages of seeking the best classification solution and dealing with a small number of samples [35]. Fires that occurred between 2012 and 2016 were excluded.
from the analysis because in the subsequent vegetation recovery modeling, we used the NDVI values of the fire patches in the fifth year after the fire to characterize the vegetation recovery status. Thus, the large fire events extracted in the current work included but were not limited to the “May 6th” fire in 1987, the Huma County fire in 2000, the Kandu River fire in the Songling area in 2006, and the Huzhong fire in 2010.

Several image features were extracted to conduct the SVM-based image classification to separate fire-affected areas from non-burned areas based on the Landsat time-series observations, including the original spectral bands, NDVI [36], Enhanced Vegetation Index (EVI) [37], Normalized Difference Moisture Index (NDMI) [38], Normalized Burn Ratio (NBR) [39], Normalized Burn Ratio 2 (NBR2) [40], and the Tasseled-Cap transformation (brightness, greenness, and wetness) [41].

\[
\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}} \tag{4}
\]

\[
\text{EVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + 0.6 \times \rho_{\text{Red}} - 7.5 \times \rho_{\text{Blue}} + 1.0} \tag{5}
\]

\[
\text{NDMI} = \frac{\rho_{\text{NIR}} - \rho_{\text{SWIR1}}}{\rho_{\text{NIR}} + \rho_{\text{SWIR1}}} \tag{6}
\]

\[
\text{NBR} = \frac{\rho_{\text{NIR}} - \rho_{\text{SWIR2}}}{\rho_{\text{NIR}} + \rho_{\text{SWIR1}}} \tag{7}
\]

\[
\text{NBR2} = \frac{\rho_{\text{SWIR1}} - \rho_{\text{SWIR2}}}{\rho_{\text{SWIR1}} + \rho_{\text{SWIR1}}} \tag{8}
\]

where \(\rho_{\text{NIR}}, \rho_{\text{Red}}, \rho_{\text{Blue}}, \rho_{\text{SWIR1}}, \) and \(\rho_{\text{SWIR2}}\) are the surface reflectance in the Landsat TM/ETM+ bands 4, 3, 1, 5, and 7 and the Landsat OLI+ bands 5, 4, 2, 6, and 7.

We used the fire point data in 1987, 2000, 2006, and 2010 and the coordinates of the fire point data and grew burned pixels from the centers of 10–15 fire patches. The resulting number of burned pixels per year was around 500 (depending on the burned area in each year) (Table 1); these pixels acted as the training data for the burned area. Next, the non-burned pixels were identified from the false-color composites of the Landsat bands using visual interpretation, including forest, grassland, cropland, logging sites, built-up, and water. The number of non-burned pixels was nearly equal to that of the burned pixels. The identified pixels were then randomly divided into two groups; one group (70% of the pixels) was used for SVM classifier training, and the other group (30% of the pixels) was used for classification accuracy validation. In the SVM classification, the Gaussian radial basis function was regarded as the kernel function, the cost parameter C was 10 (representing a penalty for miss-classification errors), and \(\gamma\) was 50. Finally, a pixel was considered a fire disturbance pixel if it was labeled as a disturbance pixel by the VCT algorithm and was also classified as a burned pixel by the SVM. Thus, we were able to use fire points in each year of the Landsat time-series as training sets, and the SVM algorithm was used to extract the burned patches from the VCT-detected disturbances. As a result, the final dataset of forest fire patches in the four years was created by combining the VCT products and SVM algorithm. We used the remaining 30% of the pixels to create a confusion matrix to determine the overall accuracy, user’s accuracy, producer’s accuracy, and Kappa coefficient to quantify the reliability of the classifications.

| Year | 1987 | 2000 | 2006 | 2010 |
|------|------|------|------|------|
| Number of burned pixels | 429  | 596  | 311  | 554  |
| Number of non-burned pixels | 422  | 611  | 341  | 557  |
| Total | 851  | 1207 | 652  | 1111 |
2.5. Post-Fire Vegetation Recovery Modeling

In this study, we focused on fire disturbances that occurred in the Daxing’anling area. The NDVI is an efficient vegetation index in characterizing post-fire vegetation recovery [42]. Previous studies indicated that the NBR could be used to characterize recovery both in the short-term and long-term [43,44]; however, NDVI has the characteristics of rapid saturation and has been primarily used for shorter-term recovery assessments [45]. The increase in the post-disturbance NDVI value may be related to an increase in the cover of trees, shrubs, and herbaceous vegetation [30]. We focused on the early stages of post-fire vegetation recovery to reduce potential saturation effects. The NDVI value of the burned area in the first year was used as an independent variable to reflect the initial vegetation condition for possible vegetation recovery, whereas the NDVI values of the fifth year were used as the dependent variable to reflect the vegetative condition after recovery. Other independent variables included many topographic features (elevation, slope) [30], climatic factors (derived from meteorological station interpolation data) and fire severity (dNBR) (Table 2). Meteorological factors included the annual average temperature, precipitation, and moisture [46].

Table 2. Description of the variables used in vegetation recovery modeling.

| Type                  | Variable                                      | Unit      | Description                                                                 |
|-----------------------|-----------------------------------------------|-----------|-----------------------------------------------------------------------------|
| Dependent variable    | Post-fire NDVI of the burned area in the fifth year (NDVI_5). |           | Reflects vegetation restoration status after fire.                           |
| Disturbance effect    | Post-fire NDVI of the burned area in the first year (NDVI_1). |           | Reflects site conditions (initial growth conditions and remaining trees).   |
|                       | dNBR                                          |           | dNBR = NBR<sub>pre</sub> - NBR<sub>post</sub> Reflects the fire disturbance severity. |
| Terrain factors       | Elevation, Slope, Aspect_cos, Aspect_sin      | Meter, Degree | Elevation, Slope, Slope cosine, Slope sine                                    |
| Climate factors       | mean_tem_0, mean_tem_1, mean_tem_2, mean_tem_3, mean_tem_4, mean_tem_5, | Values*10 | Average annual temperature from the year of the fire (*_0) to the fifth year (*_5) after the fire in the burned area. |
|                       | ex_high_0, ex_high_1, ex_high_2, ex_high_3, ex_high_4, ex_high_5 |           | Extreme maximum temperature from the year of the fire (*_0) to the fifth year (*_5). |
|                       | ex_low_0, ex_low_1, ex_low_2, ex_low_3, ex_low_4, ex_low_5 |           | Extreme minimum temperature from the year of the fire (*_0) to the fifth year (*_5). |
|                       | mean_high_0, mean_high_1, mean_high_2, mean_high_3, mean_high_4, mean_high_5 |           | Average maximum temperature from the year of the fire (*_0) to the fifth year (*_5) after the fire in the burned area. |
|                       | mean_low_0, mean_low_1, mean_low_2, mean_low_3, mean_low_4, mean_low_5 |           | Average lowest temperature from the year of the fire (*_0) to the fifth year (*_5) after the fire in the burned area. |
|                       | moisture_0, moisture_1, moisture_2, moisture_3, moisture_4, moisture_5 |           | Average relative humidity from the year of the fire (*_0) to the fifth year (*_5) after the fire in the burned area. |
|                       | preci_0, preci_1, preci_2, preci_3, preci_4, preci_5 |           | Annual precipitation from the year of the fire (*_0) to the fifth year (*_5) after the fire. |

Three methods were selected to model post-fire vegetation recovery, including step-wise multiple linear regression (SMLR), SVM, and RF. Unlike the SMLR model, RF and SVM are non-parametric machine-learning algorithms that can deal with complex non-linear relationships. RF has the advantages of determining variable importance, assessing
the robustness of data reduction, high accuracy, and low sensitivity to parameter adjustments [47]. SVM is particularly well suited for a small sample size.

The SMLR was implemented by adding an important variable to or eliminating an unimportant variable from a collection of explanatory variables based on predetermined criteria. Typically, this step is performed using a series of F- or t-tests to evaluate the influence of a particular variable on the dependent variable [48]. This process was repeated twice until no new variables could be introduced into the model.

The RF algorithm has been widely used in data mining, ecology, and other disciplines [49]. An RF typically consists of a large number of regression trees, each of which has a set of constraints that are hierarchically organized and applied from the roots of the tree to the leaves. An RF starts with many randomly selected bootstrap samples that are iteratively replaced by those from the original training data set. Multiple regression trees are then generated to form an RF, and the final prediction result is obtained by averaging the predictions from all regression trees [50]. In the current analysis, we first defined two variable importance measures based on different response types before creating the RF. (1) After permutating each predictor variable randomly, the out-of-bag (OOB) error was used to compute the reduction in predictive performance. (2) The reduction in the node impurity can be quantified by the Gini index. Thus, the contribution of each predictor variable to the modeling performance was assessed and ranked. The percent increase in the mean square error (MSE) (PercentIncMSE) and the increase in the NodePurity (IncNodePurity) depend on the importance of the predictor variable. After multiple tries, we determined the following parameter values: The number of random regression trees (ntree) was 500, the number of segmentation variables (mtry) was 3, and the minimum sample size of the terminal node (node size) was 5.

The objective of the SVM is to minimize the prediction error and maximize the flatness of the function [51]. This trade-off is controlled by setting the error penalty factor C. The parameter ε is a zone defined around the regression function, and the error in the range of ±ε is ignored. If all the training tuples are in a zone with a width of 2ε, a function is output at the center of the flat channel containing all the training tuples, in which case the total error is 0. Therefore, ε controls the fitting degree between the function and the training tuples [52]. We used the same suite of independent variables derived from the stepwise regression analysis mentioned above in the SVM modeling and RF modeling to allow for the comparison of the models. A Gaussian kernel function was used; the error penalty factor C was 10, and ε-svr was the regression algorithm in the SVM regression.

We randomly selected around 100 pixels (the specific number depended on the recovery area) from the burned areas mapped by the VCT-SVM algorithm in each year and extracted the NDVI value and several impact factors for the 5 years (when the image was missing, we will use the image of next year) to fit the models. It is worth noting that collinearity usually exists among the meteorological variables. Thus, the selection of predictors used in the RF, SVM and SMLR models followed two steps: (1) we selected the predictors using the importance function in the RF model; (2) we eliminated the meteorological factors causing collinearity using SMLR [53]. An accuracy assessment was performed to assess the model’s prediction performance. In this study, we used 70% of the selected pixels for modeling and the remaining 30% for validation. We used the validation dataset to derive the mean absolute error (MAE), the coefficient of determination ($R^2$), and the root mean square error (RMSE) to evaluate the prediction performance of the models.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$
where $y_i$ is the observed NDVI value, $\hat{y}_i$ is the value predicted by the model, $\bar{y}$ is the average value of the observed samples, and $N$ is the number of samples.

### 3. Results

3.1. Forest Disturbance Mapping and Validation

Table 3 shows the validation statistics derived from the three validation quadrats. It was observed that forest disturbances frequently occurred during the study period, and no disturbance was observed in seven, six, and four years in the study period in the eastern, middle, and western quadrat, respectively. Most of the spatial agreement ranged from 70\% to 86\%; the highest spatial agreement (92.35\%) was observed in the eastern quadrat in 1987, and the lowest (60.03\%) was observed in the western quadrat in 2010. Additionally, a very large disturbance patch was detected in 1987 (Figure 2); it was the site of “the May 6th” catastrophic fire in 1987 that led to 193 deaths, one million hectares of burned land (both inside and outside the study area), and 50,000 homeless people.

Table 3. The spatial agreements between the three randomly created validation plots (9 m$^2$) and the forest disturbance map obtained from VCT.

| Disturbance Year | Spatial Agreement (%) | Disturbance Year | Spatial Agreement (%) | Disturbance Year | Spatial Agreement (%) |
|------------------|----------------------|------------------|----------------------|------------------|----------------------|
| 1987             | 72.35                | 1987             | 91.50                | 1987             | 78.15                |
| 1988             | 76.31                | 1988             | 74.34                | 1988             | No disturbance       |
| 1992             | No disturbance       | 1992             | 80.82                | 1992             | 69.31                |
| 1994             | 78.66                | 1994             | No disturbance       | 1994             | 61.01                |
| 1997             | 69.53                | 1997             | 70.00                | 1997             | 72.54                |
| 2000             | 62.31                | 2000             | No disturbance       | 2000             | 61.44                |
| 2001             | No disturbance       | 2001             | 75.93                | 2001             | 77.89                |
| 2002             | 89.53                | 2002             | 60.57                | 2002             | 89.63                |
| 2004             | 78.43                | 2004             | No disturbance       | 2004             | No-disturbance       |
| 2006             | 66.16                | 2006             | 74.24                | 2006             | 74.92                |
| 2007             | No disturbance       | 2007             | 79.42                | 2007             | 70.08                |
| 2008             | No-disturbance       | 2008             | 73.91                | 2008             | 72.25                |
| 2010             | 66.82                | 2010             | No disturbance       | 2010             | 60.03                |
| 2011             | No disturbance       | 2011             | 73.40                | 2011             | 71.30                |
| 2013             | 76.61                | 2013             | 73.93                | 2013             | 71.65                |
| 2014             | 71.52                | 2014             | 78.34                | 2014             | 70.81                |
| 2015             | No disturbance       | 2015             | No disturbance       | 2015             | No disturbance       |
| 2016             | No disturbance       | 2016             | No disturbance       | 2016             | No disturbance       |

Figure 3 shows the dynamics of the disturbance and recovery events (disturbed forest area, disturbance rate, vegetation recovery area, and recovery rate) during 1987–2016 in the Daxing’anling region. The largest forest disturbance area was observed in 1987, with a total forest disturbance area of 362,288.8 ha and a disturbance rate of 4.33\%; this result was primarily attributed to the 1987 “May 6th” catastrophic fire. The second- and third-largest disturbance areas occurred in 2001 and 2002, with similar areas of about 143,873 ha and similar disturbance rates of about 1.72\%. The smallest disturbance area was observed in 1988, with an area of 31,708.98 ha and a disturbance rate of 0.38\%. The largest recovery area was 293,112.5 ha in 1994, with a recovery rate of 3.5\%, followed by 1997, with a recovery area of 182,858.5 ha and a recovery rate of 2.19\%. The smallest recovery area was observed in 2004, with a recovery rate of 0.46\%. Overall, except for these cases, the majority of the forest disturbance rates and recovery rates were less than 2\%, and they fluctuated over time.
3.2. Fire Disturbance Mapping and Validation

Figure 4 shows the spatial distribution of the burned areas extracted from the VCT-SVM algorithms, and Table 4 shows the error matrix derived from the independent samples. The classification accuracy of the fire patches was high, with an overall accuracy of 94% or higher for the four years. The user’s and producer’s accuracies were also above 90%, except for the 2016 producer’s accuracy (88.9%) of the non-burned class. The results show that Figure 4 is accurate and reliable, and the map can be used for the subsequent vegetation recovery analysis. Only two classes (burned area and non-burned area) were considered in the SVM classification. The burned areas were flagged as disturbance events by VCT, and small burnt areas (<1 ha) were excluded from the analysis. These strategies guaranteed the high extraction accuracy of the fire patches.

Figure 4. Disturbance area obtained from VCT, persisting land cover types, and burned areas in the study region derived from the VCT-SVM algorithm.
Table 4. Error matrix for the classification of the burned areas in 1987, 2000, 2006, and 2010.

| Year | Class           | Producer’s Accuracy | User’s Accuracy | Overall Classification Accuracy | Kappa Coefficient |
|------|-----------------|---------------------|-----------------|---------------------------------|-------------------|
|      | Burned area     | 100%                | 93.9%           | 95.2%                           | 0.905             |
|      | Non-burned area | 90.3%               | 100%            |                                 |                   |
| 1987 |                 |                     |                 |                                 |                   |
|      | Burned area     | 98.7%               | 94.9%           | 96.6%                           | 0.932             |
|      | Non-burned area | 94.4%               | 98.6%           |                                 |                   |
| 2000 |                 |                     |                 |                                 |                   |
|      | Burned area     | 98.7%               | 90.3%           | 93.9%                           | 0.877             |
|      | Non-burned area | 88.9%               | 98.5%           |                                 |                   |
| 2006 |                 |                     |                 |                                 |                   |
|      | Burned area     | 97.3%               | 92.5%           | 94.6%                           | 0.891             |
|      | Non-burned area | 91.7%               | 97.1%           |                                 |                   |
| 2010 |                 |                     |                 |                                 |                   |

3.3. Model Fitting and Validation

We first used the variable importance analysis in the RF package to identify the 25 most important potential environmental variables (Figure 5). Next, we implemented a stepwise regression analysis of the 25 potential variables to address the multicollinearity between the 25 variables to obtain the final set of 11 variables (Table 5).

Figure 5. Variable importance ranking of all predictor variables. The acronyms of the predictive variable consist of two parts: abbreviation and number. Number means from the year after fire (the year of fire is 0) in the burned area. Abbreviation description: mean_tem (average annual temperature), ex_high (Extreme maximum temperature), ex_low (extreme minimum temperature), mean_high (average maximum temperature), mean_low (average lowest temperature), moisture (average relative humidity), preci (annual precipitation). A detailed description can be found in Table 2.).

Table 5. The variables selection results of the stepwise multiple linear regression (SMLR).

| Variables | Pearson Correlation Coefficient between the Selected Variables and NDVI_5 | Estimate | Std. Error | t-Value | Pr(> |t|) |
|-----------|-------------------------------------------------------------------------|----------|------------|---------|------|-----|
| Intercept |                                                                        | 2.3600141| 5.7696782 | 0.409   | 0.682871 |
| NDVI_1    |                                                                        | 0.393    | 0.2944953 | 3.723   | 0.000244 |
| mean_high_0|                                                                         | 0.300    | 0.1781534 | 1.764   | 0.076573 |
| moisture_0|                                                                        | 0.350    | 0.4454323 | 0.790   | 0.438824 |
| preci_1   |                                                                        | 0.397    | 0.0004342 | 2.164   | 0.031432 |
Table 5. Cont.

| Variables          | Pearson Correlation Coefficient between the Selected Variables and NDVI_5 | Estimate   | Std. Error   | t-Value | Pr(>|t|) |
|--------------------|---------------------------------------------------------------------------|------------|--------------|---------|---------|
| ex_low_2           | 0.208                                                                     | 0.0289014  | 0.0103279    | 2.798   | 0.005546|
| ex_high_3          | 0.341                                                                     | 0.0879369  | 0.0231587    | 3.797   | 0.000185|
| mean_low_3         | 0.198                                                                     | −0.0991507 | 0.0289456    | −3.425  | 0.00072 |
| preci_3            | 0.257                                                                     | −0.0002626 | 0.0001646    | −1.595  | 0.111959|
| mean_high_4        | 0.197                                                                     | −0.1237221 | 0.0232963    | −5.311  | 2.46 × 10⁻⁷|
| elevation          | 0.451                                                                     | −0.0002075 | 0.0001024    | −2.027  | 0.043704|
| slope              | 0.261                                                                     | −0.004464  | 0.0020547    | −2.173  | 0.030774|

The modeling results indicated that climatic factors played an important role in post-disturbance vegetation recovery of the first few years. The relative importance of the predictor variables (Figure 5) showed that the meteorological factors exhibited a good performance for vegetation recovery prediction. The precipitation factors in our study had high variable importance, indicating that precipitation was one of the limiting factors for vegetation recovery in this area. Although there was a good correlation between NDVI_5 and moisture_0, other moisture factors exhibited a poor performance for predicting the vegetation recovery status in this region. Extreme high temperatures, especially ex_high_3, performed well for vegetation recovery prediction.

Figure 6 shows the model fitting and validation results. RF outperformed SVM and SMLR regarding modeling performance. The $R^2$ of the RF model fitting was 0.926, and the MAE and RMSE were lower than those of the SVM. The $R^2$, MAE, and RMSE of the SVM fitting were 0.8288, 0.025, and 0.044, respectively. The results further showed that the validation $R^2$ of the RF was higher than that of the SVM model (0.7113), with the highest validation $R^2$ values of 0.7727.

![Figure 6](image-url)  
Figure 6. The fitting and validation results of the support vector machine (SVM), random forest (RF), and stepwise multiple linear regression (SMLR) ($P < 0.01$).

The model evaluation results showed that the modeling accuracy and validation accuracy were higher for the RF than the SVM and SMLR (Figure 6). Overestimation of low
values and underestimation of high values occurred in all three models, but these effects were least pronounced in the RF model.

4. Discussion

4.1. Fire Disturbance and Recovery

Since fire is a crucial factor resulting in forest decline in the Daxing’anling area [54], we assessed fire disturbance and recovery in this region using time-series analysis of remotely sensed spectral indices. The results were consistent with Zhang’s research describing a known burned forest area in the Daxing’anling area [55]. Our study identified a large disturbance event in 1987, which was consistent with the “May 6th” catastrophic fire in 1987. This was a catastrophic fire, leading to extensive loss of life and property and one million hectares of burned forest area (both inside and outside the study area) [56]. Post-fire recovery is affected by fire severity, biogeographic conditions, climate factors, forest management strategy, and other factors [30]. Thus, there are various recovery methods for different fires, including no land-use change, tree planting, or natural restoration; these practices have different effects on vegetation restoration [56]. Fire and post-fire recovery are crucial processes in sustainable forest management that shape the forest age, forest structure and forest type in northeastern China [30, 57].

4.2. Post-Fire Recovery Response to Climate

The climatic factors had a substantial influence on post-fire vegetation recovery in this region, especially temperature factors play an important role in vegetation recovery in this region. Minore et al. pointed out that temperature was a critical factor for forest growth when sufficient moisture is available [58]. Extremely high temperatures (e.g., ex_high_3) and extremely low temperatures (e.g., ex_low_2) may limit post-fire vegetation recovery. Extreme temperatures might be a proxy for drought stress or indicate high solar radiation and temperature [30, 59]. Precipitation factors were also important in this research, especially in the early stage (e.g., preci_1 and preci_2), which was in line with the findings by Meng et al. [30]. Liu also pointed out that higher growing-season precipitation one year after a fire had a positive influence on post-fire recovery [60]. Additionally, we found that topographic variables (especially elevation) in the modeling was important, which was supported by Bright [61]. However, this study selected the 5-year post-fire NDVI value as the post-fire state of recovery vegetation [22], and there may be a relationship between the elevation and the biomass amount [50], elevation was possibly less influential than our modeling results suggested. Lacking consideration of pre-disturbance state [22] can also explain the low importance of the dNBR, which was opposite to previous research [30]. Since the dNBR represents the difference in the NBR before and after the fire, the correlation between the dNBR and NDVI_5 (the amount of greenness at that time) was not high. However, fire creates favorable conditions for fire-adapted forest types to germinate and regenerate; therefore, some areas burned at higher severities and recovered more quickly than the areas at lower severities [31].

4.3. Vegetation Recovery Modeling

Although the overall performances of the three models are acceptable, the validation performances are inferior to the fitting performances, and overestimation for low values and underestimation for high values occurred in all three models (Figure 5). The reason may be that the fitting of the three models requires a large number of samples, and there were far more samples with intermediate values than with low values and high values, leading to inadequate or biased training of the model. In addition, the machine learning models, such as RF and SVM, performed better than SMLR. The likely reason is that most variables were non-normally distributed, causing visibly nonrandom trends in the residuals of the initial linear models [61].
4.4. Limitations and Future Improvements

Although important results were obtained in this study, the following aspects require further research: (1) Due to the limitations of VCT monitoring, ground-truth data, and the fire point data, some low-severity or small fires were not detected in the time-series data. Therefore, more sensitive spectral indices and more fixed-point field surveys are required in future research. (2) Vegetation indices are typically used for burned area extraction. However, some vegetation indices are highly correlated (such as NDVI and EVI), which may adversely affect the extraction result. (3) As mentioned above, we need to consider the pre-fire conditions to describe vegetation recovery. João used three indicators (compared with pre-fire condition) to measure different facets of the post-fire recovery process, which might be a better approach to capture recovery than NDVI_5 in our study [22]. (4) We used a small number of samples (3 quadrats) to verify the VCT disturbance results, which may have limited the ability to capture the heterogeneity and diversity of spatial, environmental, and spectral conditions in the area. Stratified random sampling will be considered to improve this problem [25]. (5) We did not consider the differences in forest type or tree species and only evaluated post-fire vegetation recovery of all forested areas. Different tree species may exhibit different recovery patterns after a fire and different responses to climate [30]. In future studies, different tree species and forest types will be considered to compare the post-fire recovery mechanism and provide comprehensive forest management suggestions after a fire.

5. Conclusions

Landsat time-series analysis can describe post-fire vegetation recovery at the landscape scale. In this study, a VCT-SVM method was used to extract fire disturbance areas and investigate vegetation recovery in burned areas. The RF, SVM, and SMLR models explained the relationship between post-fire vegetation recovery and different predictor variables, and the RF algorithm provided the best performance for predicting vegetation recovery after fire events. A good coupling relationship was observed between the climatic factors and post-fire vegetation recovery. The results of this study are applicable to other fire areas, where landscape-scale information on post-disturbance vegetation recovery is desired. The findings can inform management decisions and are suitable for carbon forecasting in forest ecosystems.

Author Contributions: Conceptualization, J.Q. and H.W.; methodology, H.W.; software, H.S., WS., and Y.Z.; formal analysis, J.Q.; investigation, H.W.; data curation, J.Q.; writing—original draft preparation, H.W.; writing—review and editing, J.Q., Y.Z, M.L., and H.S.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Natural Science Foundation of China, grant number 31971577. This research was jointly supported by the China Postdoctoral Science Foundation, grant number 2019M651842, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [https://espa.cr.usgs.gov/, http://data.cma.cn/] (accessed on 24 December 2020). Fire point data was obtained from the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences and are available with the permission of the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences.

Acknowledgments: The authors would like to acknowledge the United States Geological Survey (USGS) and National Aeronautics and Space Administration (NASA) for providing the Landsat image data. Special thanks to the Shenyang Institute of Applied Ecology at the Chinese Academy of Sciences and the China Meteorological Data Network for sharing their fire data and climate data, respectively.

Conflicts of Interest: The authors declare no conflict of interest.
29. Chu, T.; Guo, X. Remote Sensing Techniques in Monitoring Post-Fire Effects and Patterns of Forest Recovery in Boreal Forest Regions: A Review. Remote. Sens. 2013, 6, 470–520. [CrossRef]
30. Meng, R.; Dennison, P.E.; Huang, C.; Moritz, M.A.; D’Antonio, C. Effects of fire severity and post-fire climate on short-term vegetation recovery of mixed-conifer and red fir forests in the Sierra Nevada Mountains of California. Remote Sens. Environ. 2015, 171, 311–325. [CrossRef]
31. Meng, R.; Wu, J.; Zhao, F.; Cook, B.D.; Hanavan, R.P.; Serbin, S.P. Measuring short-term post-fire forest recovery across a burn severity gradient in a mixed pine-oak forest using multi-sensor remote sensing techniques. Remote. Sens. Environ. 2018, 210, 282–296. [CrossRef]
32. Certini, G. Effects of fire on properties of forest soils: A review. Oecologia 2005, 143, 1–10. [CrossRef]
33. Masek, J.; Vermote, E.; Saleous, N.; Wolfe, R.; Hall, F.; Huenemann, K.; Gao, F.; Kutler, J.; Lim, T.-K. A Landsat Surface Reflectance Dataset for North America, 1990–2000. IEEE Geosci. Remote. Sens. Lett. 2006, 3, 68–72. [CrossRef]
34. Li, M.; Huang, C.; Shen, W.; Ren, X.; Lv, Y.; Wang, J.; Zhu, Z. Characterizing long-term forest disturbance history and its drivers in the Ning-Zhen Mountains, Jiangsu Province of eastern China using yearly Landsat observations (1987–2011). J. For. Res. 2016, 27, 1329–1341. [CrossRef]
35. Huang, C.; Song, K.; Kim, S.; Townshend, J.R.; Davis, P.; Masek, J.G.; Goward, S.N. Use of a dark object concept and support vector machines to automate forest cover change analysis. Remote. Sens. Environ. 2008, 112, 970–985. [CrossRef]
36. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 1979, 8, 127–150. [CrossRef]
37. Cró, P.; Cicone, R.C. A Physically-Based Transformation of Thematic Mapper Data-The TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 1984, 22, 256–263. [CrossRef]
38. Van Leeuwen, W.J.D.; Casady, G.M.; Neary, D.G.; Bautista, S.; Alloza, J.A.; Carmel, Y.; Wittenberg, L.; Malkinson, D.; Orr, B.J. Monitoring post-wildfire vegetation response with remotely sensed time-series data in Spain, USA and Israel. Int. J. Wildland Fire 2010, 19, 75–93. [CrossRef]
39. Kennedy, R.E.; Yang, Z.; Cohen, W.B.; Pfaff, E.; Braaten, J.; Nelson, P. Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. Remote Sens. Environ. 2012, 122, 117–133. [CrossRef]
40. White, J.C.; Wulder, M.A.; Hermosilla, T.; Coops, N.C.; Hobart, G.W. A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat TM imagery. Remote Sens. Environ. 2002, 80, 385–396. [CrossRef]
41. Wulder, M. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Prog. Phys. Geogr. Earth Environ. 1998, 22, 449–476. [CrossRef]
42. Van Leeuwen, W.J.D.; Casady, G.M.; Neary, D.G.; Bautista, S.; Alloza, J.A.; Carmel, Y.; Wittenberg, L.; Malkinson, D.; Orr, B.J. Monitoring post-wildfire vegetation response with remotely sensed time-series data in Spain, USA and Israel. Int. J. Wildland Fire 2010, 19, 75–93. [CrossRef]
43. Garcia-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. Neurocomputing 2015, 167, 24–31. [CrossRef]
44. Feng, Y.; Lu, D.; Chen, Q.; Keller, M.; Moran, E.; Dos-Santos, M.N.; Bolfe, E.L.; Batistella, M. Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. Int. J. Digit. Earth 2017, 50, 1–21. [CrossRef]
45. Shen, W.; Li, M.; Huang, C.; Wei, A. Quantifying Live Aboveground Biomass and Forest Disturbance of Mountainous Natural and Plantation Forests in Northern Guangdong, China, Based on Multi-Temporal Landsat, PALSAR and Field Plot Data. Remote. Sens. 2016, 8, 595. [CrossRef]
46. Cao, X.; Chen, J.; Imura, H.; Higashi, O. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. Remote. Sens. Environ. 2009, 113, 2205–2209. [CrossRef]
47. Huang, C.; Davis, L.S.; Townshend, J.R.G. International Journal of Remote Sensing An assessment of support vector machines for land cover classification An assessment of support vector machines for land cover classi cation. Int. J. Remote Sens. 2002, 23, 725–749. [CrossRef]
48. Hidalgo-Munoz, J.M.; Gamiz-Fortis, S.R.; Castro-Diez, Y.; Argueso, D.; Esteban-Parra, M.J. Long-range seasonal streamflow forecasting over the Iberian Peninsula using large-scale atmospheric and oceanic information. Water Resour. Res. 2015, 51, 3543–3567. [CrossRef]
49. Yi, K.; Tani, H.; Zhang, J.; Guo, M.; Wang, X.; Zhong, G. Long-Term Satellite Detection of Post-Fire Vegetation Trends in Boreal Forests of China. Remote. Sens. 2013, 5, 6938–6957. [CrossRef]
55. Zhang, Y. Study on the impacts of climate change on forest fires in Daxing’anling mountains. Master’s Thesis, Northeast Forestry University, Harbin, China, 20 June 2008. (In Chinese).

56. Chen, W.; Moriya, K.; Sakai, T.; Koyama, L.; Cao, C. Post-fire forest regeneration under different restoration treatments in the Greater Hinggan Mountain area of China. *Ecol. Eng.* **2014**, *70*, 304–311. [CrossRef]

57. Cuevas-González, M.; Gerard, F.; Balzter, H.; Riaño, D. Analysing forest recovery after wildfire disturbance in boreal Siberia using remotely sensed vegetation indices. *Glob. Chang. Biol.* **2009**, *15*, 561–577. [CrossRef]

58. Minore, D.; Laacke, R.J. *Natural Regeneration*; Oregon State University Press: Corvallis, OR, USA, 1992; pp. 258–283.

59. Mildrexler, D.J.; Yang, Z.; Cohen, W.B.; Bell, D.M. A forest vulnerability index based on drought and high temperatures. *Remote Sens. Environ.* **2016**, *173*, 314–325. [CrossRef]

60. Liu, Z. Effects of climate and fire on short-term vegetation recovery in the boreal larch forests of Northeastern China. *Sci. Rep.* **2016**, *6*, 37572. [CrossRef]

61. Bright, B.C.; Hudak, A.T.; Kennedy, R.E.; Braaten, J.D.; Khalyani, A.H. Examining post-fire vegetation recovery with Landsat time series analysis in three western North American forest types. *Fire Ecol.* **2019**, *15*, 8. [CrossRef]