Using GIS and Random Forests to identify fire drivers in a forest city, Yichun, China

Zhangwen Su, Haiqing Hu, Guangyu Wang, Yuanfan Ma, Xiajie Yang and Futao Guo

College of Forestry, Northeast Forestry University, Harbin, China; College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, China; Asian Pacific Forest Research Centre, Faculty of Forestry, University of British Columbia, Vancouver, Canada

ABSTRACT

Forest city (FC) usually refers to an urban area with high forest coverage. It is a green model of urban development that has been strongly advocated for by governments of many nations. Forest fire is a prominent threat in FC development, but the causes of fires in FCs are usually different and more complex than in pure forested areas since more socio-economic factors and human activity are involved in the ignition and spread of fire. The large and increasing number of lives being exposed to wildfire hazard highlights the need to understand the characteristics of these fires so that forest fire prediction and prevention can be efficient. In this study, Ripley’s K(d) function and Random Forests (RF) were applied to analyze the drivers, spatial distribution and risk patterns of fires in Yichun, a typical FC in China. The results revealed a clustered distribution of forest fire ignitions in Yichun, as well as identified the driving factors and their dynamic influence on fire occurrence. Fire risk zones were identified based on RF modelling. Improved preventive measures can be implemented in the fire prone areas to reduce the risk of fire in Yichun by considering the factors identified in this study.

ARTICLE HISTORY

Received 29 March 2018
Accepted 23 July 2018

KEYWORDS

Driving factors; Random Forests; spatial distribution; Ripley’s K function

1. Introduction

Forest fires harm forest ecosystems and pose a threat to human life, property and safety (Grogan et al. 2000; Chuvieco et al. 2008; Chas-Amil et al. 2010; Chang et al. 2013). Understanding the drivers that determine the spatial distribution and occurrence of forest fires is essential for successful forest fire prevention, including optimized fire-fighting resources and development projects, such as construction of fire lines and watchtowers (Hering et al. 2009). The drivers and spatial distribution of forest fires have been studied in many different regions of the world (Martinez et al. 2009; Grale et al. 2010; Padilla and Vega-Garcia 2011; Gralewicz et al. 2012;
Guo et al. 2016, 2017a, 2017b), though the results of these studies vary widely due to differences in meteorology, topography, vegetation, infrastructure and other socio-economic factors, making generalizations challenging (Maingi and Henry 2007; Syphard et al. 2008; Vilar et al. 2010).

A forest city (FC) is a green model of urban development that has been strongly advocated for in many nations in recent years. A FC usually refers to an urban area with high forest coverage and where forests play a leading role in urban ecosystem function. Compared to natural forests or remote forested regions, forest fires are more likely to occur in a FC due to the high frequency of human activity and the density of infrastructure (Martinez et al. 2013). Although these fires are generally small due to early detection, intense suppression efforts and better firefighter accessibility (Massada et al. 2009), every ignition source has the potential to grow into a large fire and can lead to serious damage to human health and property loss. Along with rapid advancement of FC construction nationwide, forest fire prevention and suppression will be an essential task for city security in China, making the quantification of wildfire risk and understanding of fire drivers in FCs particularly important. Identifying forest fire drivers in FCs is usually more challenging than in pure forested areas due to complex relationships between the various factors that may influence fire occurrence in FCs, including climatic, biophysical and social variables. Therefore, a fire prediction tool, which can fully account for the complexity between fire occurrence and its driving factors, is needed.

Random Forests (RF) is an ensemble learning method based on classification and regression trees (CART). RF can select important variables and calculate the relative importance of each independent variable automatically no matter how many variables are used initially (Cutler et al. 2007). Additionally, RF has been demonstrated to have a high prediction accuracy and high tolerance to outliers and ‘noise’ (Breiman 2001; Archibald et al. 2009). Due to the strengths of RF, fire occurrence studies have begun using this approach in recent years (Oliveira et al. 2012; Wu et al. 2014).
Yichun is a typical FC in Northeastern China, famous for eco-tourism. Like many of China’s FCs, Yichun faces a serious threat from forest fires in the process of urban development. To further understand the fire risks and influencing factors in Yichun, and to provide reference for other FCs, this study identifies the spatial distribution of historical fire ignitions (1980–2010) in Yichun and uses RF to examine an extensive set of factors (including four broad categories: meteorology, vegetation type, infrastructure and socio-economic factors) to assess their ability to influence the distribution and incidence of forest fire in this region. In this paper, we have carried out detailed analysis to understand the spatial distribution pattern of forest fire; identify the drivers and their importance for forest fire occurrence and map fire risk zones based on driving factors in Yichun. With its findings, this study can be used to improve forest fire management in FCs in China and serve as reference for forest fire management in other regions.

2. Materials and methods

2.1. Study area

Yichun is a key forestry region in China and prone to occurrence of large forest fires due to its climate and vegetation type. It is located in the northeastern part of China (127° 37’–130° 46’E, 46° 28’–49° 26’N) and covers an area of 39,017 km² (Figure 1). It is a hilly region, with an average altitude of 600 m. The population of Yichun is about 1.2 million, scattered across the landscape (Figure 2(a)), and rivers, railways and roads are widely distributed in the area (Figure 2(b)). Average annual air temperature is 1 °C and average annual precipitation is 750–820 mm. Forest cover if 83.8%, consisting of mixed conifer and broadleaved forests and dominated by Korean pine (*Pinus koraiensis*), along with *Larix gmelinii*, *Pinus sylvestris var. mongolica* and *Populus davidiana*. Fire seasons are April–June and September–October of each year.

2.2. Overarching study design

Ripley’s *K*(d) function was used to identify the spatial distribution of forest fires in the study area. RF algorithms were then used to identify key drivers of forest fire occurrence based on broad potential factors selected from a literature review. Potential factors (independent variables) that are likely to cause forest fire were...
classified into three categories, meteorological, non-meteorological (local factors) and combined factors, and models were then analyzed based on these three classifications accordingly. RF algorithms performed using the varSelRF package in R were used to select drivers and rank their importance, and partialPlot was then used to test the marginal effect of a variable on fire occurrence (Diaz-Uriarte and Alvarez de Andres 2005; Thompson and Spies 2009). The predictive ability of each model was evaluated based on a cut-off point that was calculated using receiver operating characteristic (ROC) curve analysis. In order to eliminate bias, we randomly selected 20% of the complete dataset as independent validation data, which was only used to assess performance of the final predictive model. The remaining 80% of the complete dataset was used for model building, for which the data (80% of the complete dataset) was randomly divided into a training dataset (60%) and an inner validation dataset (40%) (Rodrigues and Riva 2014). This ‘data division’ procedure was repeated 10 times, resulting in 10 random sub-samples of data, where each sub-sample has its own training and inner validation dataset. Ten intermediate models were obtained based on the sub-samples. The variables that were tested as important variables in at least six of the 10 intermediate models were considered as ‘key drivers’ and were used to fit the model building dataset (80% of the complete dataset) and to conduct the final validation based on the independent validation dataset (20% of the complete dataset).

The study area was divided into three zones – low, medium and high fire risk – according to the map of fire occurrence probability, and the cutoff value was
calculated using the final complete dataset of all variables from the ROC analysis. A schematic diagram of the overall model is shown in Figure 3.

2.3. Data acquisition

2.3.1. Dependent variables
The Forest Fire Prevention Office of Yichun, China provided forest fire data from 1980 to 2010. The existing database on forest fire contains information about fire location, size and cause. According to this database, 479 wildfires occurred in the study area from 1980 to 2010, and the majority of fires reported occurred due to railways (29.20%), slash burning (28.59%) and other unknown reasons (29.05%) (Figure 4).

In order to establish a discriminant model of forest fires, a certain percentage of reference points (non-fire points) were required. ArcGIS software was applied to randomly generate 1000 non-fire (i.e. control) points at a ratio of 1:2 to the fire ignition number (Catry et al. 2009). To ensure control points would not be on the same or nearby location as fire ignition points, a buffer zone of 1000 m around fire points was placed as a barrier, excluding the control points that fell into the buffer zone (Kalabokidis et al. 2007; Guo et al. 2017a, 2017b). The double random principle of time and space was adopted during the random generation process (i.e. the space coordinates were randomly generated, while the time points were selected from fire seasons during 1980–2010, then the space and time points were randomly combined together).

2.3.2. Independent variables
Independent variables included five broad categories: topography, vegetation type, infrastructure, meteorology and socio-economic factors, with a total of 27 explanatory sub-variables.

2.3.2.1. Topography. Topographic factors consisted of three explanatory variables: elevation, slope and aspect. Data for these variables are from digital topographic maps with a spatial resolution of 25 m, created in 2000 and provided by the National
Administration of Surveying, Mapping and Geoinformation of China. ArcGIS 10.0 software was used to extract the corresponding elevation, slope and aspect of each fire and non-fire point from the maps. Elevation and slope values for each point were used directly in the modelling process. The aspect was categorized as flat, North (315°–45°), East (45°–135°), South (135°–225°) and West (225°–315°). The proportion of each aspect in the study area was calculated and the corresponding value for each fire and control point was used to develop the model.

2.3.2.2. Vegetation type. A 1 km resolution digital vegetation map of China in 2000 was obtained from The Cold and Arid Regions Science Data Center, China. According to the geographical boundary of Yichun, polygons were grouped into the following four categories: (1) needle leaf deciduous and needle leaf evergreen trees (12.08%); (2) broadleaf deciduous trees and shrubs (63.52%); (3) grass and agricultural crops (21.83%) and (4) urban construction land, permanent wetland and barren or sparsely vegetated land (2.57%) (Figure 2(c)). Forest vegetation types for each fire and non-fire point were extracted from the vegetation map using ArcGIS 19.0 software. The proportion of each vegetation type located in a fire or non-fire point was used in the modelling process.

2.3.2.3. Infrastructure. Four infrastructure variables were used: distance from fire point or non-fire point to the nearest railway, nearest roads (including urban main roads and rural roads), nearest river and nearest residential area. Basic geographic data was retrieved from a 1:250,000 Digital Line Graphic (DLG) map, which was created in the year 2000 and obtained from the National Administration of Surveying, Mapping and Geoinformation of China (http://www.sbsm.gov.cn/). Distance to the nearest railway, road, river and residential area was calculated using ArcGIS.

2.3.2.4. Meteorology. Daily meteorological data from the fire season (March–November) of 1980–2010 were obtained from four national weather stations located in the four main districts of Yichun. Daily climate data were provided by the China Meteorological Data and Sharing Network (www.cma.gov.cn/2011qxfw/2011qsjgx/), and contained the following 12 climate factors: daily average, maximum and minimum ground surface temperature (°C), daily average and maximum wind speed (m s⁻¹), 24 h precipitation (mm), daily average pressure (hPa), sunshine hours (h), daily average, maximum and minimum air temperatures (°C) and average relative humidity (%). The corresponding daily climate factors for each fire and control point were retrieved under an ArcGIS19.0 environment. The daily climate variables for each fire and control point were provided by the meteorological station that was identified as being closest to each point.

Monthly weather data during the fire season, including monthly average relative humidity (%), temperature (°C) and precipitation (mm) were also obtained from the HADCM2 climate model, downloaded from the Data Sharing Infrastructure of Earth System Science (http://geodata.nju.edu.cn/Portal/index.jsp). The HADCM2 climate model was developed by the Hadley Centre for Climate Research and Prediction and the original dataset was collected from the International Panel on Climate Change.
(IPCC) in grid format. The data were converted into text format and data points for China were extracted with a spatial resolution of 0.085°. The corresponding monthly climate factors for each fire and non-fire point were retrieved under an ArcGIS19.0 environment.

2.3.2.5. Socio-economics. Socio-economic factors included three variables: population density, per capita GDP and net income of rural residents. Since the majority of rural areas in Yichun are forested, the net income of rural residents is highly dependent on forest related economic activities, which in turn influences local fire occurrence. Due to the lack of a vector dataset of socio-economic factors, we collected the recorded socio-economic data for Yichun from the Heilongjiang statistical yearbook (2006). Because this data does not contain spatial information, the same value for each socio-economic factor was assigned to the fire and non-fire points in each given year.

2.4. Methodology

2.4.1. Spatial distribution analysis
Ripley’s $K(d)$ function is a statistical method used to calculate nearest neighbour distance and has been widely used to identify the spatial distribution pattern of fire occurrence (Podur et al. 2003; Vadrevu et al. 2008; Guo et al. 2015; 2017b). The definition of Ripley’s $K(d)$ function has also been explained in detail in previous studies (Peterson and Squiers 1995; Stoyan and Penttinen 2000; Guo et al. 2015). In this study, ‘the guard area correction’, a basic edge correction method, was used in Ripley’s $K$-function and SpPack software was used to perform the $K$-function (Perry 2004). Confidence envelopes were set to 95% based on 499 replicates.

2.4.2. Random Forests model
Random Forests (RF), developed by Breiman and Cutler in 2001, is an ensemble learning method based on classification and regression trees (CART). RF has been demonstrated to have a high prediction accuracy and high tolerance to outliers and ‘noise’ (Breiman 2001). It can also evaluate the relationship between covariates and dependent variables, and calculate the relative importance of covariates (Cutler et al. 2007; Kubosova et al. 2010). RF has been applied in various fields in the past, including medicine, genetics, ecology and remote sensing. In recent years, it has been used in forest fire forecasting, for which it has demonstrated good predictive ability (Oliveira et al. 2012; Rodrigues and Riva 2014; Kane et al. 2015).

RF generates $ntree$ bootstrap samples from the original dataset. For each of the bootstrap samples, it establishes an unpruned classification or regression tree. The final outcome is the average of the results of all the $ntrees$. For each bootstrap sample, the outcome consists of about 63% of the overall samples, and retains about one-third of the samples (called out-of-bag, OOB) for validation. For each classification or regression tree, at each node RF randomly samples $mtry$ of the predictors and chooses the best split from among those variables. When using RF for data fitting and prediction, the number of trees ($ntree$) and the number of each random variable at each node ($mtry$) must be parameterized. For selecting $mtry$, Liaw and Wiener (2002)
suggests choosing $mtry = \sqrt{m}$. As the number of trees increases, the overall error rate will tend to be a stable upper bound – in other words, as long as the number of trees is large enough, the overall error rate of the forest will stabilize to ensure the convergence of the RF (Cutler et al. 2007; Genuer et al. 2010; Kubosova et al. 2010). In this study, we set the $mtry$ as $\sqrt{m}$ and $ntree$ as 2000.

RF can also be used to rank the importance of variables. The basic concept is as follows: for the $j$th variable ($X_j$), the OOB error ($errOOB^j_t$) of each tree $t$ is calculated and then the value of the $j$th variable ($X_j$) is permuted while all others are left unchanged among OOB data and the OOB error ($'OOB^j_t$) is again recalculated on this permuted dataset. RF estimates the importance of a variable by looking at how much the prediction error increases when OOB data for that variable is permuted. The importance score of $X_j$ is:

$$VI(X^j) = \frac{1}{ntree} \sum_t \left( err'OOB^j_t - errOOB_t \right)$$

where $\Sigma$ is the summation of all the trees and $ntree$ is the number of trees in the random forest (Gromping 2009; Genuer et al. 2010; Oliveira et al. 2012). For regression, the OOB error is the mean square error (MSE), while for classification, the OOB is misclassification probability.

The RF algorithm, which aims to minimize the OOB error, can be used to select featured variables in the model. In this study, the R package varSelRF was used to conduct the variable selection (Genuer et al. 2010). The partialPlot function in the varSelRF package was used to construct partial dependence plots to reveal the relationship between the dependent and independent variables (Hastie et al. 2001; Cutler et al. 2007).

2.4.3. Evaluation of model performance

The prediction accuracy of the RF model on fire occurrence was evaluated based on model-selected variables using ROC curve analysis (Fielding and Bell 2002; Rodrigues and Riva 2014). Youden criteria, calculated based on the sensitivity and specificity of ROC (i.e. Youden criterion = sensitivity + specificity − 1) (Taube 1986), were used to determine the ‘cut-off’ value, which was applied as a threshold to determine the prediction accuracy of RF. In this study, if the predicted probability of a given point in space were greater than the cut-off point, forest fires were considered to occur, otherwise there is no forest fire occurrence (Garcia et al. 1995; Catry et al. 2009; Chang et al. 2013).

In addition, the area under the ROC curve (AUC) can be considered an indicator of the goodness of model fitting, ranging from 0.5 to 1 (Zhou et al. 2002; Franklin 2010). The greater the AUC value, the better the performance of the model at fitting the data. When the AUC value equals 0.5, the model fit is considered to be equivalent to a random guess. The model is poorly fit when AUC is 0.5–0.69, moderately fit when AUC is 0.7–0.79 and well fit when AUC is 0.8–1 (McCune and Grace 2002; del Hoyo et al. 2011).
2.4.4. Mapping fire risk zones

A likelihood map of fire occurrence was created using Ordinary Kriging interpolation in ArcGIS 10.0, using the original probability scale (Isaaks and Srivastava 1989) combined with the previously determined cut-off point for forest fire occurrence. Yichun was divided into three fire risk zones: low (0–cutoff), medium (cutoff–0.5) and high (>0.5) fire risk (Chang et al. 2013). The likelihood of forest fire occurrence was analyzed and high fire risk zones within the study area were identified.

3. Results

3.1. Spatial pattern of fire occurrence

Ripley’s K function was used to analyze the spatial pattern of forest fire occurrence in Yichun, China from 1980 to 2010 (Figure 5). As shown in Figure 5, there is a clear departure from complete special randomness (CSR) towards clustering from 1980 to 2010, indicating a clustered distribution of forest fires in study area.

3.2. Identifying and ranking significant variables

3.2.1. Meteorological factors

The process of variable selection based on meteorological factors is shown in Appendix Table 2. Based on the iterative results of ten sub-training sets, eight variables were selected for the training sample, including monthly average temperature,
monthly average relative humidity, monthly average precipitation, daily maximum ground surface temperature, daily average ground surface temperature, sunshine hours, daily maximum temperature and daily average relative humidity (Appendix Figure 9). The training sample resulted in the selection of relative humidity and temperature as the main driving factors influencing forest fire during the study period. All variables in Appendix Figure 9 influenced the occurrence of forest fires, but with varying degrees of importance.

3.2.2. Local, non-meteorological factors
The process of variable selection based on local factors is shown in Appendix Table 3. The results of testing the 10 sub-training sets is that six variables were selected to fit the training sample, including distance from fire points to the nearest railway and residential area, as well as net income of rural residents, elevation, per capita GDP and population density (in descending order of importance) (Appendix Figure 10).

3.2.3. Combined factors
The process of variable selection based on combined factors is shown in Appendix Table 4. Appendix Figure 11 shows that based on the results of 10 sub-training data-sets, eight variables were used to fit the training sample. Distance from fire point to the nearest railway and daily average relative humidity were the factors with the most influence on forest fire occurrence, followed by distance from fire point to the nearest residential area and daily maximum ground surface temperature, which were the most influential of the non-meteorological and meteorological factors. Compared to other factors, population density, per capita GDP, net income of rural residents and daily maximum temperature had relatively less influence on fire occurrence, but they were still important driving factors of forest fires.

3.3. Model fitting and prediction accuracy
Figure 6 shows ROC curves and AUC values of the ten sub-training samples and the training sample based on meteorological, local and combined factors using the RF model. The AUC values for the ten sub-training samples and training sample were in the range of 0.740–0.807, indicating a moderate model fit based on meteorological factors. A good model fit was achieved for the local factors (AUC value interval of 0.82–0.851) and an excellent model fit was achieved for combined factors with an AUC of 0.886–0.906.

Table 1 shows the RF model predictive ability based on meteorological, local and combined factors. According to the sensitivity and specificity values derived from ROC curve analysis, cut-off values were obtained for the 10 sub-training samples and the training sample for each factor-based model. Combined with these cut-off values, RF model prediction accuracies were also computed. Results show that the prediction of meteorological factors is the least accurate (sub-training samples ranging from 71.2% to 76.5% and a training sample prediction of 74.2%). RF model prediction for combined factors had the highest accuracy (80.78–84.8% for the ten sub-training samples and 82.9% for the training sample) (Table 1).
Additionally, in the case of inner and independent validation, the prediction accuracy of meteorological factors range from 69.5% to 75.0% in the inner validation of ten sub-samples and that in the independent validation is 74.8%. In terms of the result of local factors test, its range is roughly equal to that of sub-training sample. Furthermore, the similar situation took place in the complete sample. Overall, the prediction for combined factors is the highest accuracy among three categories (meteorological, local factors and combined factors), while the accuracy of meteorological factors prediction is the least (Table 1).

### 3.4. Relationship between driving factors and fire occurrence

Partial dependence plots (Figure 7) show that the distance between fire points and railways does not have an obvious linear relationship with the probability of ignition; it shows a sharp downward trend within an 18km radius, a weaker trend from 18 to 38 km, and then a gradual upward trend at a radius >38 km. The probability of ignition is greatest when the average daily relative humidity is 20%, and then declines with increasing humidity, before stabilizing above 70%. The distance between fire points and the nearest residential area has an important influence on forest fire, mainly within an 18 km radius; it declines with increasing distance, then probability of ignition stabilizes at >18 km. The probability of ignition increases sharply as daily maximum ground surface temperature exceeds 20 °C and as maximum air temperature exceeds 3 °C. The probability of fire occurrence shows a sharp downward trend when the net income of rural residents is <2000 Yuan per month, and stabilizes when net income is >1200 Yuan per month. The probability of fire occurrence shows the same trend with per capita GDP as with net income, with a steep drop in

### Table 1. Prediction accuracy of RF models based on climatic, local and combined factors.

| Sample          | Cut-off (meteorological/local/combined factors) | Sub-training (60%) | Inner validation (40%) |
|-----------------|-----------------------------------------------|--------------------|------------------------|
| Sample 1        | 0.301/0.320/0.331                             | 73.2/78.2/80.8     | 75.0/79.2/84.2         |
| Sample 2        | 0.324/0.292/0.322                             | 73.2/79.2/84.8     | 74.4/79.0/81.5         |
| Sample 3        | 0.349/0.314/0.365                             | 74.2/79.9/83.3     | 71.9/78.4/80.3         |
| Sample 4        | 0.339/0.329/0.321                             | 74.3/79.3/83.4     | 69.5/78.7/79.9         |
| Sample 5        | 0.224/0.435/0.410                             | 73.6/78.9/81.7     | 72.6/77.8/82.9         |
| Sample 6        | 0.424/0.464/0.484                             | 74.3/77.8/83.4     | 70.7/80.5/81.0         |
| Sample 7        | 0.395/0.428/0.347                             | 71.2/79.6/83.2     | 70.0/78.9/82.6         |
| Sample 8        | 0.336/0.437/0.450                             | 74.6/79.8/83.0     | 71.8/79.1/81.9         |
| Sample 9        | 0.364/0.360/0.361                             | 73.1/79.3/81.1     | 74.4/80.4/83.8         |
| Sample 10       | 0.430/0.345/0.328                             | 76.5/80.6/83.8     | 70.5/77.3/80.4         |
| Final model based on selected variables based on previous 10 samples | 0.342/0.407/0.295 | 74.2/79.5/82.9         |
| Independent validation process (20% of the complete data) | Validation model based on selected variables based on previous 10 samples | 0.260/0.353/0.426 | 74.8/75.7/81.2 |

The second column represents the intermediate models (samples 1–10 including sub-training and inner validation datasets), final (training dataset) model and independent validation model; the third column represents the cut-off values that were used to determine the prediction accuracy for each intermediate and final model based on different types of factors. The final model and independent validation model were fitted using 80% and 20% of the complete dataset, respectively.

Additionally, in the case of inner and independent validation, the prediction accuracy of meteorological factors range from 69.5% to 75.0% in the inner validation of ten sub-samples and that in the independent validation is 74.8%. In terms of the result of local factors test, its range is roughly equal to that of sub-training sample. Furthermore, the similar situation took place in the complete sample. Overall, the prediction for combined factors is the highest accuracy among three categories (meteorological, local factors and combined factors), while the accuracy of meteorological factors prediction is the least (Table 1).
probability between 0 and 2500 Yuan before stabilizing at higher values. The impact of population density on forest fire is relatively significant at <30 people per km², and then declines slightly at >30 people per km².

3.5. Distribution of likelihood of fire occurrence

A map of likelihood of fire occurrence (Figure 8(a)) was produced and represents the probability of a fire occurring at a given location. Fire risk classes were also generated (Figure 8(b)), which divide the probability graph into three fire risks based on the threshold of the model fitting process and 0.5 classification points (Chang et al. 2013). These maps show a high risk of fire occurrence concentrated in the middle, southwestern and northeastern portions of the study area. They also show fire prone zones distributed throughout the southern and the northernmost regions.

4. Discussion

Distance between fire points and the nearest railway or residential area were identified as the main local factors affecting forest fire occurrence. Railways in Yichun predominantly run through forested areas, and about 29.2% of forest fires were found to
Figure 7. Partial dependence plots for variables predicting fire occurrence, selected using Random Forests. Partial plots show the dependence of the probability of fire occurrence on one predictor variable after averaging out the effects of all other predictor variables in the model.
be caused by locomotives, especially trains (Figure 4). Results show that with an increase in distance between forests and railway tracks, the influence of railways on fire ignition decreases (0–20 km), indicating that forests closer to railways have a higher chance of fire occurrence (Figure 7). Many studies attribute fire ignition near railways to accidents or negligence (Cardille et al. 2001; Martinez et al. 2009; Arndt et al. 2013). The main reasons for fire originating near railways include sparks released by the train’s steam engines, fire accidents that occur on board the train, passengers smoking on the train and a lack of control over burning activities near the tracks. The increase in fire occurrence found with increasing distance (>40 km) from tracks likely reflects the influence of other factors besides railways (Figure 7).

The relatively strong influence of residential areas on forest fire ignition (within a 10 km radius of fire points) is likely due to the high intensity of human activities, including clearing weeds from residences, agricultural activities, deliberate fires and other activities (Figure 7). As the distance from residential areas to forests increases, the influence of residential areas on fire occurrence decreases and then stabilizes (>20 km). Government statistics show that about 1.02% of total forest fire ignitions are caused by escaped fire from residents, indicating only a minor influence of local residential areas on fire occurrence and contradicting our findings (Figure 4). Interestingly, the percentage of fire ignitions stated for unknown reasons is quite a lot higher (29.20%). Based on observations and knowledge acquired from forest officials, it is suggested that these anonymous sources are most likely attributable to human-caused fire. The findings of this study, as well as local knowledge about forest fire occurrence in this region, indicate that fire prevention efforts should be more focused on local residential areas.

This study also found that socio-economic factors, including per capita GDP and population density, also affect forest fire occurrence, and these findings are supported by previous studies by Syphard et al. (2007) and Chang et al. (2013). In Yichun, most people who earn <1000 Yuan per month are highly dependent on forest resources for their income and many people with a low income engage in slash-and-burn during agriculture and wood collection in winter. Many studies have considered variables related to population density and fire risk, as there is a direct link between a growing number of people and pressure on forests, potentially increasing the risk of fire ignition (Guyette and Dey 2000; Cardille et al. 2001; Guyette and Spetich 2003). In
Yichun, a large number of people reside close to forested areas and about 28.59% of fires are caused by agricultural activities like slash-and-burn, clearly demonstrating the critical role that human populations play in forest fire occurrence.

Roads link the city to surrounding rural and forested areas, suggesting that local roads may affect forest fire occurrence (Oliveira et al. 2012), though our study did not identify any meaningful relationship between fire points and distance to the nearest road. Syphard et al. (2007) found a similar lack of relationship, in contrast to the findings of Miranda et al. (2012). The reason for this may be, as suggested by Stephens (2005), that the road network is too dense to reflect potentially important influences on forest fire, as there is a high density of roads around forested areas in Yichun (~0.21 km/km²).

Maximum ground surface temperature (GST_max), maximum air temperature (Temp_max) and average relative humidity (RH_avg) are three important climate factors influencing fire occurrence in Yichun. GST_max and Temp_max have a positive impact on fire occurrence, while RH_avg has a negative impact. High temperatures can increase evaporation from plants, as well as decrease moisture content of potential fire fuels (e.g. downed woody material), leading to increased probability of fire occurrence (Chuvieco et al. 2004). Conversely, high relative humidity induces an increase in moisture content of the fuel, thereby reducing the likelihood of fire (Guo et al. 2017a, 2017b). Previous studies have concluded that temperature and relative humidity do not have a consistent linear relationship with fire occurrence (Castro et al. 2003; Wu et al. 2015), and our results support this finding with a few threshold points at which the relationship trend changes (15 °C and 20 °C for GST_max and Temp_max, and 70% for RH_avg).

Distribution patterns of fire occurrence likelihood are critical to the management of forest fire (Saglam et al. 2008). Understanding this distribution pattern can assist in optimal allocation of resources (Wei et al. 2008) and in determination of the number and location of fire observation towers (Catry et al. 2009), ultimately contributing to more efficient use of financial and human resources. It is also an important basis for the division of forest fire risk zones (Chang et al. 2013). In recent years, forest fire management agencies have paid more attention than previously to the prevention and monitoring of forest fires (Chang et al. 2013). The results of this study show that forest fires mainly occur in the central, northeastern and southwestern regions of Yichun, while the southern region and the most northern region also have incidences of forest fire, but to a lesser extent. Fire watchtowers and checkpoints should be built and additional fire-fighting resources should be available in these high-risk fire zones in order to reduce forest fire incidence and minimize losses.

With government investment in tourism, local GDP and per capita income in Yichun expected to continue to increase in the next decades, forest fire frequency may be reduced (Figure 7). However, the population is also expected to increase from a current population density of 37 people/km² due to the opened birth policy implemented a few years ago, which will lead to increased development and promote the occurrence of fires. According to simulations with various general circulation models (GCMs), the earth’s climate will be several degrees warmer by the end of the next century due to increasing atmospheric concentrations of greenhouse gases (Flannigan et al. 1998), and it is therefore reasonable to consider an increasing temperature in Yichun over the next few decades, which may also increase fire occurrence. Although
we can predict future changes in certain factors, in general, predicting the future fire frequency in Yichun is difficult since fire occurrence results from a complex set of interactions, including fuel conditions, topography, human activity and weather including temperature, relative humidity and precipitation. Therefore, any changing factor alone does not necessarily cause a significant impact on fire frequency (Flannigan et al. 1998; Syphard et al. 2007).

More in-depth analyses of the relationship between these various factors and their joint impact on forest fire occurrence is required in the future, especially given expected changes such as climate change. This study offers a suitable approach to explore the features of forest fire occurrence in a typical forest city and provides insight on appropriate forest fire management accordingly. These findings can be widely referenced by other regions that already have forest cities or are undergoing forest city construction.

5. Conclusions

In this study, we used Ripley’s $K(d)$ function and RF models to analyze the spatial distribution of forest fire ignition and identify drivers and their effect on fire occurrence in Yichun, China. The results of our analysis are as follows: (1) forest fires are present in a clustered distribution across Yichun; (2) overall, local factors showed a stronger influence on local fire occurrence than meteorological factors, with distance from fire points to the nearest railway and residential area, elevation, net income of rural residents, per capita GDP and population density identified as the key drivers of forest fire occurrence; (3) RF performed well, forecasting fire occurrence in Yichun with a high prediction accuracy (82.9%) using all factors combined and (4) based on historical data, central, northern (north, northeast) and southern (south, southwest) regions of the study area were identified as high fire risk zones. Forest management agencies and policy makers should pay particular attention to these regions and allocate additional fire-fighting resources accordingly.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Forestry Industry Research Special Funds for Public Welfare Projects [20144402]; the Fujian Agriculture and Forestry University Funds for Distinguished Young Scholar [No. xjq201613] and the Asia-Pacific Forests Network Grant [APFnet/2010/FPF/001 Phase II].

References

Archibald S, Roy DP, van Wilgen BW, Scholes RJ. 2009. What limits fire? An examination of drivers of burnt area in Southern Africa. Glob Chang Biol. 15:613–630.
Arndt N, Vacik H, Koch V, Arpaci A, Gossow H. 2013. Modeling human-caused forest fire ignition for assessing forest fire danger in Austria. iForest. 6(6):315–325.
Breiman L. 2001. Random forests. Machine Lear. 45:15–32.
Cardille JA, Ventura SJ, Turner MG. 2001. Environmental and social factors influencing wild- fires in the Upper Midwest, United States. Ecol Appl. 11(1):111–127.
Castro FX, Tudela A, Sebastia MA. 2003. Modeling moisture content in shrubs to predict fire risk in Catalonia (Spain). Agric For Meteorol. 116:49–59.
Catry FX, Rego FC, Baçao FL, Moreira F. 2009. Modeling and mapping wildfire ignition risk in Portugal. Int J Wildland Fire. 18(8):921–931.
Chang Y, Zhu ZL, Bu R, Chen HW, Feng YT, Li Y, Hu Y, Wang ZC. 2013. Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. Landscape Ecol. 28(10):1989–2004.
Chas-Amil ML, Touza J, Prestemon JP. 2010. Spatial distribution of human-caused forest fires in Galicia (NW Spain). In: Perona G, Brebbia CA, editors. WIT transactions on ecology and the environment. Southampton: WIT Press; p. 247–258.
Chuvieco E, Cocero D, Riano D, Martin P, Martinez-Vega J, de la Riva J, Pérez F. 2004. Combining NDVI and surface temperature for the estimation of live fuel moisture content in fire danger rating. Remote Sens Environ. 92:322–331.
Chuvieco E, Giglio L, Justice C. 2008. Global characterization of fire activity: towards defining fire regimes from earth observation data. Global Change Biol. 14(7):1488–1502.
Cutler DR, Edwards TJ, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology. 88(11):2783–2792.
del Hoyo VL, Martin Isabel M, Martinez VF. 2011. Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. Eur J Forest Res. 130:983–996.
Diaz-Uriarte R, Alvarez de Andres S. 2005. Variable selection from random forests: application to gene expression data. Tech report. http://ligarto.org/rdiaz/Papers/rfVS/randomForestVarSel.htm
Fielding AH, Bell JF. 2002. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv. 24:38–49.
Flannigan MD, Bergeron Y, Engelman O, Wotton BM. 1998. Future wildfire in circumboreal forests in relation to global warming. J Veg Sci. 9:469–476.
Franklin J. 2010. Mapping species distributions. New York: Cambridge. University. Press.
Garcia CV, Woodard PM, Titus SJ, Adamowicz WL, Lee BS. 1995. A logit model for predicting the daily occurrence of human caused forest-fires. Int J Wildland Fire. 5:101–111.
Genuer R, Poggi J, Tuleau-Malot C. 2010. Variable selection using random forests. Pattern Recognit Lett. 31:2225–2236.
Grala K, Cooke WH III. 2010. Spatial and temporal characteristics of wildfires in Mississippi, USA. Int J Wildland Fire. 19:14–28.
Gralewicz NJ, Nelson TA, Wulder MA. 2012. Spatial and temporal patterns of wildfire ignitions in Canada from 1980 to 2006. Int J Wildland Fire. 21:230–242.
Grogan, P., Bruns, T.D., and Chapin, F.S. 2000. Fire effects on ecosystem nitrogen cycling in a Californian bishop pine forest. Oecologia. 122: 537–544.
Gromping U. 2009. Variable importance assessment in regression: linear regression versus random forest. Am Stat. 63(4):308–319.
Guo FT, Innes JL, Wang GY, Ma XQ, Sun L, Hu HQ, Su ZW. 2015. Historic distribution and driving factors of human-caused fires in the Chinese boreal forest between 1972 and 2005. J Plant Ecol. 8(3):1–11.
Guo FT, Selvalakshmi S, Lin FF, Wang GY, Wang WH, Su ZW, Liu AQ. 2016. Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest. Can J For Res. 46:1–13.
Guo FT, Su ZW, Tigabu M, Yang XJ, Lin FF, Liang HL, Wang, GY. 2017a. Spatial modelling of fire drivers in urban-forest ecosystems in China. Forests. 8(180):1–18
Guo FT, Su ZW, Wang GY, Sun L, Tigabu M, Yang XJ, Hu HQ. 2017b. Understanding fire drivers and relative impacts in different Chinese forest ecosystems. Sci Total Environ. 605–606:411–425.
Guyette RP, Dey DC. 2000. Humans, topography and wildland fire: the ingredients for long-term patterns in ecosystems. In: Yaussey DA, editor. Proceedings of the workshop on fire, people and the central hardwood landscape. Newton Square, PA, USA: USDA Forest Service; p. 28–35.

Guyette RP, Spetich MA. 2003. Fire history of oak-pine forests in the Lower Boston Mountains, Arkansas, USA. Forest Ecol Manag. 180:463–474.

Hastie TJ, Tibshirani RJ, Friedman JH. 2001. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. New York, USA: Springer.

Hering AS, Bell CL, Genton MG. 2009. Modeling spatio-temporal wildfire ignition point patterns. Environ Ecol Stat. 16:225–250.

Isaaks EH, Srivastava RM. 1989. Applied geostatistics. Oxford, UK: Oxford University Press.

Kalabokidis KD, Koutsias N, Konstantinidis P, Vasilakos C. 2007. Multivariate analysis of landscape wildfire dynamics in a Mediterranean ecosystem of Greece. Area. 39(3):392–402.

Kane VR, Lutz JA, Alina Cansler C, Povak NA, Churchill DJ, Smith DF, Kane JT, North MP. 2015. Water balance and topography predict fire and forest structure patterns. Forest Ecol Manag. 358:1–13.

Kubosoka K, Brabec K, Jarkovsky J, Syrovatka V. 2010. Selection of indicative taxa for river habitats: a case study on benthic macroinvertebrates using indicator species analysis and the random forest methods. Hydrobiologia. 651:101–114.

Liaw A, Wiener M. 2002. Classification and regression by random Forest. R News. 2–3:18–22.

Maingi JK, Henry MC. 2007. Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. Int J Wildland Fire. 16:23–33.

Martinez J, Chuvieco E, Koutsias N. 2013. Modeling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. Nat Hazards Earth Sys Sci. 13:311–327.

Martinez J, Vega-Garcia C, Chuvieco E. 2009. Human-caused wildfire risk rating for prevention planning in Spain. J Environ Manage. 90(2):1241–1252.

Massada AB, Radeloff VC, Stewart SI, Hawbaker TJ. 2009. Wildfire risk in the wildland–urban interface: a simulation study in northwestern Wisconsin. For Ecol Manag. 258:1990–1999.

McCune B, Grace J. 2002. Analysis of ecological communities. MJM Software Design. Glenden Beach.

Miranda BR, Sturtevant BR, Stewart SI, Hammer RB. 2012. Spatial and temporal drivers of wildfire occurrence in the context of rural development in northern Wisconsin, USA. Int J Wildland Fire. 21:141–154.

Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira J. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. Forest Ecol Manag. 275:117–129.

Padilla M, Vega-Garcia C. 2011. On the comparative importance of fire danger rating indices and their interaction with spatial and temporal fire occurrences in Spain. Int J Wildland Fire. 20:46–58.

Perry GLW. 2004. SpPack: spatial point pattern analysis in Excel using Visual Basic for Applications (VBA). Environ Modell Softw. 19:559–569.

Peterson CJ, Squiers ER. 1995. An unexpected change in spatial pattern across 10 years in an Aspen-White pine forest. J Ecol. 83:847–855.

Podur J, Martell DL, Csillag F. 2003. Spatial patterns of lightning-caused forest fires in Ontario, 1976–1988. Ecol Model. 164:1–20.

Rodrigues M, Riva JDL. 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ Model Softw. 57:192–201.

Saglam B, Bilgili E, Durmaz BD, Kadiogullari AI, Kucuk O. 2008. Spatio-temporal analysis of forest fire risk and danger using landsat imagery. Sensors. 8:3970–3987.

Statistics Bureau of Heilongjiang Province. 2006. Heilongjiang statistical yearbook. Beijing: China Statistics Press.

Stephens SL. 2005. Forest fire causes and extent on United States Forest Service lands. Int J Wildland Fire. 14:213–222.

Stoyan D, Penttinen A. 2000. Recent applications of point process methods in forestry statistics. Stat Sci. 15:61–78.
Syphard AD, Radeloff VC, Keeley JE, Hawbaker TJ, Clayton MK, Stewart SI, Hammer RB. 2007. Human influence on California fire regimes. Ecol Appl. 17(5):1388–1402.
Syphard AD, Radeloff VC, Keuler NS, Taylor RS, Hawbaker TJ, Stewart SI, Clayton MK. 2008. Predicting spatial patterns of fire on a southern California landscape. Int J Wildland Fire. 17:602–613.
Taube A. 1986. Sensitivity, specificity and predictive values: a graphical approach. Stat Med. 5:585–591.
Thompson JR, Spies TA. 2009. Vegetation and weather explain variation in crown damage within a large mixed-severity wildfire. Forest Ecol Manage. 258(7):1684–1694.
Vadrevu KP, Badarinath KVS, Anuradha E. 2008. Spatial patterns in vegetation fires in the Indian region. Environ Monit Assess. 147:1–13.
Vilar L, Woolford DG, Martell DL, Martin MP. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. Int J Wildland Fire. 19:325–337.
Wei Y, Rideout D, Kirsch A. 2008. An optimization model for locating fuel treatments across a landscape to reduce expected fire losses. Can J For Res. 38:868–877.
Wu ZW, He HS, Yang J, Liang Y. 2015. Defining fire environment zones in the boreal forests of northeastern China. Sci Total Environ. 518–519:106–116.
Wu ZW, He HS, Yang J, Liu ZH, Liang Y. 2014. Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern China. Sci Total Environ. 493:472–480.
Zhou XH, McClish DM, Obuchowski NA. 2002. Statistical methods in diagnostic medicine. New Jersey: Hoboken.

Appendix

See Figures 9–11 and Tables 2–4.

Figure 9. Variable importance measures of climate factors from Random Forests sub-samples based on Mean Decrease Accuracy (X-axis), which quantifies the importance of a variable by measuring the change in prediction accuracy when the values of the variable are randomly permuted compared to the original observations. Temp_m_avg represents monthly average temperature; Preci_m_avg: monthly average precipitation; RH_m_avg: monthly average relative humidity; GST_max: daily maximum ground surface temperature; GST_avg: daily average ground surface temperature; SSD: sunshine hours; Temp_max: daily maximum temperature; RH_avg: daily average relative humidity.
Figure 10. Variable importance measures of local factors from Random Forests sub-samples based on mean decrease accuracy (X-axis), which quantifies the importance of a variable by measuring the change in prediction accuracy when the values of the variable are randomly permuted compared to the original observations. Dis_railway represents distance to railway; Dis_residential: distance to residential; Elev: elevation; CGDP: per capita GDP; Den_Pop: density of population; Income_rural: net income of rural residents.

Figure 11. Variable importance measures of combined factors from Random Forests sub-samples based on mean decrease accuracy (X-axis), which quantifies the importance of a variable by measuring the change in prediction accuracy when the values of the variable are randomly permuted compared to the original observations. GST_max represents daily maximum ground surface temperature; RH_avg: daily average relative humidity; Temp_max: daily maximum temperature; Dis_railway: distance to railway; Dis_residential: distance to residential; Income_rural: net income of rural residents; CGDP: per capita GDP; Den_Pop: density of population.
Table 2. Variables selected based on meteorological factors using Random Forests.

| Variable       | Sample1 | Sample2 | Sample3 | Sample4 | Sample5 | Sample6 | Sample7 | Sample8 | Sample9 | Sample10 | Training dataset |
|----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------------|
| RH_m_avg       | +       |         |         | +       | −       | +       | −       | +       | +       | +       | ✓               |
| Temp_m_avg     | +       | +       | −       | +       | +       | +       | −       | +       | +       | +       | ✓               |
| Preci_m_avg    | −       | −       | −       | +       | +       | +       | +       | +       | +       | +       | ✓               |
| GST_avg        | −       | −       | −       | −       | −       | +       | +       | +       | +       | +       | ✓               |
| GST_max        | +       | +       | +       | +       | +       | +       | +       | +       | +       | +       | ✓               |
| GST_min        | −       | −       | −       | −       | −       | −       | −       | −       | −       | −       | ×               |
| Win_avg        | −       | −       | −       | −       | −       | −       | −       | −       | −       | −       | ×               |
| Win_max        | −       | −       | −       | −       | −       | −       | −       | −       | −       | −       | ×               |
| Preci          | −       | −       | −       | −       | −       | −       | −       | −       | −       | −       | ×               |
| PRS_avg        | −       | −       | −       | −       | −       | −       | −       | −       | −       | −       | ×               |
| SSD            | +       | +       | −       | +       | +       | +       | +       | −       | +       | +       | ✓               |
| Temp_avg       | +       | +       | −       | +       | −       | −       | −       | −       | −       | −       | ×               |
| Temp_max       | +       | +       | +       | +       | +       | +       | +       | +       | +       | +       | ✓               |
| Temp_min       | −       | −       | −       | +       | +       | +       | −       | −       | −       | −       | ×               |
| RH_avg         | +       | +       | +       | +       | +       | +       | +       | +       | +       | +       | ✓               |

**Note:** RH_m_avg represents monthly average relative humidity; Temp_m_avg: monthly average temperature; Preci_m_avg: monthly average precipitation; GST_avg: daily average ground surface temperature; GST_max: daily maximum ground surface temperature; GST_min: daily minimum ground surface temperature; Win_avg: daily average wind speed; Win_max: daily maximum wind speed; Preci: 20:00–20:00 precipitation; PRS_avg: daily average air pressure; SSD: sunshine hours; Temp_avg: daily average temperature; Temp_max: daily maximum temperature; Temp_min: daily minimum temperature; RH_avg: daily average relative humidity. A positive sign (+) indicates the variable was selected by the model, a negative sign (−) indicates the variable was rejected by the model. The definition of (+) and (×) in the last column is as follows: (+) means that there are at least six (+) signs (or selections) in the samples 1–10, and the sign is (−) otherwise.
### Table 3. Variables selected based on local (non-meteorological) factors using Random Forests.

| Variable          | Sample1 | Sample2 | Sample3 | Sample4 | Sample5 | Sample6 | Sample7 | Sample8 | Sample9 | Sample10 | Training dataset |
|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|-----------------|
| Dis_railway       | +       |         |         | +       | +       | +       | +       | +       | +       | +         | √               |
| Dis_subroad       | −       | −       | −       | +       | −       | −       | −       | +       | +       | +         | ×               |
| Dis_road          | −       | −       | −       | +       | −       | −       | −       | +       | +       | +         | ×               |
| Dis_residential   | +       | +       | +       | +       | +       | +       | +       | +       | +       | +         | √               |
| Dis_river         | −       | −       | −       | −       | −       | −       | −       | −       | −       | −         | ×               |
| Forest_type       | −       | −       | −       | −       | −       | −       | −       | −       | −       | −         | ×               |
| Slope             | −       | −       | −       | −       | −       | −       | −       | −       | −       | −         | ×               |
| Aspect            | −       | −       | −       | −       | −       | −       | −       | −       | −       | −         | ×               |
| Elev              | +       | +       | +       | +       | +       | +       | +       | +       | +       | +         | √               |
| CGDP              | −       | +       | +       | +       | +       | −       | −       | +       | +       | +         | √               |
| Den_Pop           | −       | +       | +       | +       | +       | −       | +       | +       | +       | +         | √               |
| Income_rural      | +       | +       | +       | +       | +       | −       | +       | +       | +       | +         | √               |

*Note: Dis_railway represents distance to railway; Dis_subroad: distance to substandard road; Dis_road: distance to standard road; Dis_residential: distance to residential area; Dis_river: distance to river; Forest_type: forest type; Elev: elevation; CGDP: per capita GDP; Den_Pop: density of population; Income_rural: net income of rural residents. Signs (+), (−), (×) and (×) represent the same meanings as in Table 2.*
Table 4. Variables selected based on meteorological and non-meteorological variables using Random Forests.

| Variable         | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5 | Sample 6 | Sample 7 | Sample 8 | Sample 9 | Sample 10 | Training dataset |
|------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------------|
| Dis_railway      | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| Dis_subroad      |          |          |          |          |          |          |          |          |          |           |                 |
| Dis_road         |          |          |          |          |          |          |          |          |          |           |                 |
| Dis_residential  | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| Dis_river        |          |          |          |          |          |          |          |          |          |           |                 |
| Forest_type      |          |          |          |          |          |          |          |          |          |           |                 |
| Slope            |          |          |          |          |          |          |          |          |          |           |                 |
| Aspect           |          |          |          |          |          |          |          |          |          |           |                 |
| Elev             |          |          |          |          |          |          |          |          |          |           |                 |
| CGDP             | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| Den_Pop          | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| Income_rural     | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| RH_m_avg         |          |          |          |          |          |          |          |          |          |           |                 |
| Temp_m_avg       |          |          |          |          |          |          |          |          |          |           |                 |
| Preci_m_avg      |          |          |          |          |          |          |          |          |          |           |                 |
| GST_avg          |          |          |          |          |          |          |          |          |          |           |                 |
| GST_max          | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| GST_min          |          |          |          |          |          |          |          |          |          |           |                 |
| Win_avg          |          |          |          |          |          |          |          |          |          |           |                 |
| Win_max          |          |          |          |          |          |          |          |          |          |           |                 |
| Win_min          |          |          |          |          |          |          |          |          |          |           |                 |
| Pred_m-avg       |          |          |          |          |          |          |          |          |          |           |                 |
| Temp_m-avg       |          |          |          |          |          |          |          |          |          |           |                 |
| Income Rural     |          |          |          |          |          |          |          |          |          |           |                 |
| Den Pop          |          |          |          |          |          |          |          |          |          |           |                 |
| CDP              |          |          |          |          |          |          |          |          |          |           |                 |
| Elec             |          |          |          |          |          |          |          |          |          |           |                 |
| Resett            |          |          |          |          |          |          |          |          |          |           |                 |
| Slope            |          |          |          |          |          |          |          |          |          |           |                 |
| Forest Type      |          |          |          |          |          |          |          |          |          |           |                 |
| Elev             |          |          |          |          |          |          |          |          |          |           |                 |
| Aspect           |          |          |          |          |          |          |          |          |          |           |                 |
| Elev             |          |          |          |          |          |          |          |          |          |           |                 |
| Temp_max         | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| Temp_min         |          |          |          |          |          |          |          |          |          |           |                 |
| RH_avg           | +        | +        | +        | +        | +        | +        | +        | +        | +        | +         |                 |
| RH_max           |          |          |          |          |          |          |          |          |          |           |                 |
| RH_min           |          |          |          |          |          |          |          |          |          |           |                 |

Note: Abbreviations and signs represent the same meanings as in Tables 2 and 3.