Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest

Journal:	*Canadian Journal of Forest Research*
Manuscript ID	cjfr-2015-0373.R2
Manuscript Type:	Article
Date Submitted by the Author:	20-Jan-2016
Complete List of Authors:	Guo, Futao; Fujian Agriculture and Forestry University, College of Forestry Selvaraj, Selvalakshmi; Fujian Agriculture and Forestry University Lin, Fangfang; Fujian Agriculture and Forestry University, Computer Science Wang, Guangyu ; University of British Columbia, Faculty of Forestry Wang, Wenhui; Fujian Agriculture and Forestry University Su, Zhangwen; Fujian Agriculture and Forestry University Liu, Aiqin; Fujian Agriculture and Forestry University
Keyword:	Daxing'an Mountains, wildfire, geographically weighted logistic model, driving factors, Human-caused fire

Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest

Futao Guo[a], Selvaraj Selvalakshmi[a], Fangfang Lin[b], Guangyu Wang[a,c], Wenhui Wang[a], Zhangwen Su[a], Aiqin Liu[a*]

[a] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

[b] College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

[c] Faculty of Forestry, University of British Columbia, Canada, V6T 1Z4

*Corresponding author. Tel: +86-591-83780261; Fax:+86-591-83780261; E-mail: fjlaq@126.com (A. Liu)

Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest

Abstract:

We applied a classic logistic regression (LR) together with a geographically weighted logistic regression (GWLR) to determine the relationship between anthropogenic fire occurrence and potential driving factors in the Chinese boreal forest, and to test whether the explanatory power of the LR model could be increased by considering geospatial information of geographical and human factors using a GWLR model. Three tests, "all variables", "significant variables" and "cross-validation", were applied to compare model performance between the LR and GWLR models. Our results confirmed the importance of distance to railway, elevation, length of fire line and vegetation cover on fire occurrence in the Chinese boreal forest. In addition, GWLR model performs better than LR in terms of model prediction accuracy, model residual reduction and spatial parameter estimation by considering geospatial information of explanatory variables. This indicates that the global LR model is incapable of identifying underlying causal factors for wildfire modeling sufficiently. The GWLR model helped identify spatial variation between driving factors and fire occurrence, which can contribute better understanding of forest fire occurrence over large geographic areas and the forest fire management practices may be improved based on it.

Keywords: human-caused fire, wildfire, geographically weighted logistic model, driving factors, geospatial.

Introduction

Forest fires are both a critical process in natural forests and a major cause of timber loss and human suffering (Fernandez-Pello 1994). Frequency of forest fires has increased significantly globally over the last decades. Forest fires have been the dominant disturbance regime in boreal forests since the last Ice Age, influencing energy flows and biogeochemical cycles, including local and global carbon cycling (Weber and Flannigan 1997). As in other boreal forests of the world, the Chinese boreal forest is threatened by fire (Stocks 1993; Shvidenko and Goldammer 2001). Anthropogenic fire is linked to human-mediated disturbance (encroachment, smoking, agricultural practices, hunting, fireworks, fire escape from locomotives and resident homes, recreation, industrial activities, tourism, intentional fires, etc.), which have played a critical role in fire occurrence in boreal forests. In Canada, two-thirds of all forest fires are caused by humans (Wotton et al. 2003). In the Siberian boreal forest, more than 85% of fires are linked to anthropogenic sources (Mollicone et al. 2006). In the Chinese boreal forest, human-caused forest fires account for approximately 60-80% of the total number of fires annually (Guo et al. 2015). Because of the significant influence of forest fire on the forest ecosystem and socio-economy, governments and researchers have devoted considerable effort to fire prediction.

Understanding the driving factors that influence human-caused fire occurrence is the key to fire prediction. Korovin (1996) indicated the influence of road distribution on anthropogenic fire. Niklasson and Granstrom (2000) and Wallenius et al. (2004) found that expansion of human settlements and increased population density drives fire occurrence in the boreal forest of northern Europe. Factors such as topography (*i.e.* elevation and slope) and forest type have also been found to be meaningful drivers (Mollicone et al. 2006; Syphard et al. 2007; Romero-Calcerrada et al. 2008; Martinez et al. 2009; Romero-Calcerrada et al. 2010).

Logistic regression (LR) is the most commonly used model to predict the probability of fire occurrence and has been used successfully both at local (Vega-Garcia et al. 1995; Vasconcelos et al. 2001; Liu et al. 2012) and continental scales (Prasad et al. 2008; Martinez et al. 2009; Padilla and Vega-Garcia 2011). However, such a statistical model may be inappropriate for identifying the relationships between fire occurrence and influence factors because the assumptions of the non-spatial (stationarity) models, such as individual errors being independent of each other, may be violated. This leads to biased estimation of standard errors of model parameters and significance levels of statistical tests, and over-estimation of model R^2 (Anselin and Griffith 1988). Researchers have indicated that spatial heterogeneity in the model parameters are more consistent with real-world situations and can be the result of random sampling variations and spatially varying relationships (Fotheringham et al. 1996; Foody 2003; Wang et al. 2005). Koutsias et al. (2005) found that the explanatory power of a global linear regression model increased considerably when it assumed varying relationships instead of constant ones.

Geographically weighted regression (GWR) is a useful exploratory analytical tool that can provide information on spatial non-stationarity in relationships between variables (Huang and Leung 2002; Matthews and Yang 2012). Zhang, Gove and Health (2005) found that GWR produced more accurate predictions for the response variable, and the residuals of the GWR model had more desirable spatial distributions including lower spatial autocorrelation compared to non-spatial models. In addition, GWR is an effective technique that enables regression model parameters to vary in space (Huang and Leung 2002; Foody2004; Wang et al. 2005). In the past decades, GWR has been applied successfully in many research areas, including forestry, ecology and social science (Zhang and Shi 2004; Wang, Ni, and Tenhunen 2005; Nakaya et al. 2005). In recent years, a GWR logistic model has also been used for forest fire prediction and its superiority has been identified by researchers (Koutsias et al. 2010; Martinezet al. 2013; Rodrigues et al. 2014).

Taking the above perspectives into consideration, this study applies classical LR together with a GWR logistic model to explore long-term forest fire occurrence patterns in the Chinese boreal forest over the last decades. Specifically, the objectives are to: (1) identify potential driving factors of fire occurrence, from topography, vegetation and human factors; (2) explore whether the relationship between fire occurrence and influence factors are globally constant or spatially variable; (3) further explore spatial variability by identifying significant factors that can eventually explain spatially varying parameters; (4) compare the performance of GWLR and LR models for fire occurrence. Our hypothesis is that geospatial information of factors will increase the explanatory power of a forest fire prediction model.

Materials and Methods

Study area

China's boreal forests, located in the Daxing'an Mountains of northeastern China (50°10 - 53°33′N and 121°12′ - 127°00′E), are the southernmost part of the global boreal forest biome (Jiang et al. 2002). The total area of the forest covers 8.46×10^6 ha (Fig. 1). The dominant species is Dahurian larch (*Larix gmelinii* Rupr.) and is normally accompanied by white birch (*Betula platyphylla* Suk.), Mongolian pine (*Pinus sylvestris* L. var. *mongolica* Litv.), and Mongolian oak (*Quercus mongolica* Fisch. ex Ledeb.). The Daxing'an Mountains are located in a cold-temperate zone, with mean annual temperatures between -2 °C and 4 °C, and a range extending from -52.3 °C to 39.0 °C. The mean total annual precipitation is between 350-500 mm.

This region has the largest average annually burned area in China and is generally exposed to extremely high fire risk. Between 1980 and 2005, there were more than 1,000 forest fires, including more than 600 human-caused fires, and a total area of burned forest amounting to 1,300,000 ha (Guo et al. 2015). In recent years, fires have become smaller (burned area), but occur more frequently and more intensely than before (Chang et al.2007).

Data collection

Dependent variable: fire record

In this study, anthropogenic causes of forest fire included smoking, hunting, fireworks, and escaped fire from locomotives and resident homes. Anthropogenic fire data for the Daxing'an Mountains from 1980-2004 were provided by the Forest Fire Prevention office of the Heilongjiang Forestry bureau, China, including fire location, size, cause and date of occurrence. The fire data were acquired in a geo-database format and contained geographically referenced point locations of forest fires in the Daxing'an Mountains. Before 1990, the records of fire location were determined by the fire chief, who identified each fire location through a combined approach of fixed observation points in the forest and the Terrain and Forest Instruction Map (1:100 000)(Guo et al. 2015). Locations of fires after 1990 were recorded by GPS.

We randomly generated non-fire (*i.e.* control) points at a ratio of 1:1.5 as the fire ignition number (Catry et al. 2009; Chang et al. 2013) to meet the binary-variable requirement of LR and GWLR models. In addition, in order to improve the robustness of the method, random generation was applied one hundred times and then the non-fire points were extracted from all the randomly generated points as the 1:1.5 ratio. Consequently, the values for the dependent variables were assigned as '0' and '1' for control points (n=905) and fire points (n=620) respectively. We excluded non-fire (*i.e.* control) points located in water bodies or urban areas.

Independent variables

The independent variables consist of four categories, including topographic, vegetation, infrastructure and socio-economic factors. Details are provided in Table 1. The criteria for independent variable selection were based on previous studies of fire occurrence.

Vegetation type

A digital vegetation map of China (1km resolution) was downloaded from the Cold and Arid Regions Science Data Center, China (http://westdc.westgis.ac.cn/). The data were gathered in 2000.We grouped polygons into the following five categories: needleleaf deciduous tree cover (30.6%), broadleaf deciduous tree (12.8%), needleleaf evergreen tree (11.5%), broadleaf deciduous shrub (7.45%), grass and agricultural crop (37.7%). Forest vegetation types for each fire and control point (*i.e.* non-fire) were extracted from the vegetation map layer using ArcGIS 10.0. We used the proportion of each vegetation type in which either a fire or control point was located in the study area to develop the model.

Vegetation cover

We used fractional vegetation cover (FVC) to represent the corresponding fuel amount of each fire or control point. FVC is the vertical projection of the crown or shoot region of vegetation to the ground surface within a unit area, expressed as a fraction or percentage (Purevdor 1998). The FVC was calculated based on the normalized difference vegetation index (NDVI), which is a simple graphical indicator that can be used to assess whether the target being observed contains live green vegetation or not. Gutman and Ignatov (1997) proposed the relationship model between NDVI and FVC as:

$$FVC = (NDVI - NDVI_{soil}) / (NDVI_{veg} - NDVI_{soil}) \qquad (1)$$

where, $NDVI_{veg}$ and $NDVI_{soil}$ are the NDVI of dense vegetation canopy and bare soil, respectively. The NDVI dataset has a spatial resolution of 1km and was provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn).

Topography

Elevation, slope and aspect were retrieved based on high-resolution (25 m) digital elevation model (DEM) data that was built in 2000 and collected from the National Administration of Surveying, Mapping and Geoinformation of China. Aspect was extracted as flat, north (315-45^o), east (45-135^o), south (135-225^o) and west (225-315^o). The proportion of each aspect in the study area was calculated and used in developing the model, along with the other two topographic variables, elevation and slope.

Infrastructure

Human infrastructure has been identified as a main driver of wildfire occurrence by many studies (Martinez et al. 2009; Oliveira et al. 2012; Mundo et al. 2013; Oliveira et al. 2014). In this study, a number of tested variables (anthropogenic and environmental factors) were used based on previous studies, but other unique untested variables like number of inspection stations and length of fire line were also included. Other variables such as distance to the nearest railway, distance to the nearest road and others that can reflect the impact of distance between infrastructure and forest fire occurrence were also retrieved from a 1:250,000 Digital Line Graphic (DLG) map which was built in 2000 and collected from the National Administration of Surveying, Mapping and Geoinformation of China. All data extraction was done in ArcGIS 10.0.

Socio-economic factors

Socio-economic factors included annual funding for forest fire prevention and per capita GDP. These variables have been used in other similar studies to represent the trend of potential changes in human activity, which may influence fire occurrence (Maingi and Henry 2007; Oliveira et al. 2014). The data was collected from the Heilongjiang Statistical Yearbook (2006) and the Local Chronicles of Forest Fire Prevention of Daxing'an Mountains (2005).

Model description

Logistic Regression (LR)

LR has been used for fire occurrence prediction and to examine the driving factors of fire occurrence in different regions of the world at various scales (Martell et al.1987; Vega-Garcia et al.1995; Martinez et al. 2009). It considers p_j as the probability of fire occurrence:

$$p_j = \frac{\exp(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}{1 + \exp(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)} \quad (2)$$

The corresponding probability p_j can be transformed to a linear function as below:

$$\mathrm{logit}(p_j) = \log\left(\frac{p_j}{1-p_j}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \quad (3)$$

where, X_i represents explanatory variables, β_0 is the intercept, and β_i are coefficients of variables.

Geographically Weighted Logistic Regression (GWLR)

GWLR models can be considered an expansion of the standard LR model that incorporates geographical location into the models. The formula can be written as follows:

$$\mathrm{logit}(p_j) = \log\left(\frac{p_j}{1-p_j}\right) = \beta_0(\mu_j, \gamma_j) + \sum_{i=1}^{n} \beta_i(\mu_j, \gamma_j) x_i \quad (4)$$

where, β_0 and β_i are GWLR model parameters specific to a location at (μ_j, γ_j) coordinates.

An important aspect of GWR modeling is determining a weighting function for estimating local model parameters. In GWR, the weighting function uses a distance function, resulting in observations closer in space, which are generally assumed to have a greater effect on local parameter estimates. In software GWR4, there are four kinds of kernel functions which can be used to determine the weighting matrix, including:① Fixed Gaussian,②Adaptive Gaussian, ③ Fixed bi-square, and ④Adaptive bi-square. Zhang et al. (2014) found that the Adaptive Gaussian has a better performance than other functions. The adaptive Gaussian can be expressed as follows:

$$W_{ij} = \begin{cases} (1 - d_{ij}^2 / \theta_{i(k)})^2 & d_{ij} < \theta_{i(k)} \\ 0 & d_{ij} > \theta_{i(k)} \end{cases} \qquad (5)$$

where, W_{ij} is the weight value of an observation at location j for estimating the coefficient at location i; d_{ij} is the Euclidean distance between locations i and j; θ is the bandwidth size; $\theta_{i(k)}$ is the kernel bandwidth size defined as the kth nearest neighbor distance. In this paper, we selected an Adaptive Gaussian function to fit the model (Wang et al. 2005; Wu and Zhang 2013; Zhang et al.2014).

Model fitting and evaluation

Dataset division

In this study, three types of tests (i.e. all variables test, significant variables test and cross-validation test) were set up to compare the fitting effect of LR and GWLR models. To avoid the influence of sample distribution on the test results, the original dataset was randomly divided into training (60%) and validation (40%) samples (Rodrigues and de la Riva 2014). This procedure was repeated five times, applying a sampling with replacement method, resulting in five random sub-samples of data, each one with its own training and validation dataset.

Model comparison procedure

In the all variables test, all 13 variables are selected to test the model fit of the five sub-samples and the complete samples. In the significant variables test, to select the significant variables, the forward Wald method was used in the LR model. To test the model fit of five sub-samples, the principle of the quartile range of the estimated coefficients of GWLR is greater than±1 standard deviation range of the estimated coefficients of LR (Zhang et al. 2014) was applied in the GWLR model. Variables that appeared to be significant in at least three of five intermediate models were included in the final model. To better compare the fitting results of the two models, a variable cross-validation test was conducted (i.e.significant variables from the GWLR model were used to fit the LR model, while significant variables from the LR were used to fit the GWLR model). We applied local polynomial interpolation in ArcGIS 10.0 to estimate the coefficients of each variable for the entire study area (Rodrigues et al. 2014).

Model evaluation approaches

SPSS19.0 was used for fitting the LR model and software GWR4.0 was used for fitting the GWLR model. The predictive performance of the two models was assessed by employing AIC, AICc and SSE. The smaller the value of AIC, AICc and SSE, the better the performance of the model fitting. Additionally, Receiver Operating Characteristic (ROC) curve analysis (Fielding Bell 1997) was used to evaluate prediction accuracy of the two models. The ROC curve was obtained by plotting sensitivity versus specificity for various probability thresholds. The area under the curve (AUC) is also often used to evaluate performance (Jimenez-Valverde 2012). An AUC of 0.5 indicates no discrimination; 0.5-0.69 poor discrimination; 0.7-0.79 reasonable discrimination; and 0.8-0.9 excellent discrimination (delHoyo et al. 2011). In other words, higher AUC indicates better performance of model fitting. Moreover, the Youden criterion can be derived from ROC curve analysis (Youden criterion=sensitivity + specificity-1). This can then be used to determine the cut-off point. If the predicted probability of fire occurrence is greater than the cut-off point, forest fires are considered to occur; otherwise, there is no occurrence of forest fires (Garcia et al. 1995; Catryet al. 2009). The prediction accuracy of five sub-samples and complete samples of two models were also calculated and results were compared.

We expressed the significant coefficient of each variable as a separate spatial layer by dividing the entire study area into 1x1 km grids. In addition, we performed interpolation using Kriging method on the estimated coefficients of each variable with complete dataset in ArcGIS 10.0 software. These different significant-coefficient layers were overlaid to identify the level of spatial variability of variables in each grid. Finally, the Global Moran's I was used to calculate spatial autocorrelation coefficients for the residuals of each model. The smaller the value of the Global Moran's I, the smaller the residual spatial dependence, and the better the performance of the model fitting, including more spatial relations (Zhang et al. 2014). The Global Moran's I was calculated using the software package Rookcase added in Excel (Wu and Zhang 2013).

In order to eliminate the bias of model fitting, a multicollinearity analysis regarding independent variables was performed before fitting the two models. We used the variance inflation factor (VIF) to perform the multicollinearity diagnosis. A value of VIF>5 indicates a significant collinearity between the independent variables, and the variables should be eliminated (Wu and Zhang 2013).The total random points generated for fire points (620) and non-fire points (905) were extrapolated for the entire study area.

Results and Analysis

Model fitting and prediction accuracy comparison

Comparison based on all variables

Absence of multicollinearity between our 13 variables meant that all variables could be used in model fitting. All 13 variables were used to fit the LR and GWLR models. Estimated coefficients in both models showed that elevation, number of inspection stations, and fire line were negatively correlated; whereas variables like per capita GDP, distance to settlement, slope and funding correlated positively with fire occurrence (Appendix Table A1). The relationship between fire occurrence and the remaining variables was different between the two models.

Coefficients of the other six variables (*i.e.* distance to railway, distance to river, distance to road, forest type, aspect and vegetation cover) had both positive and negative effects in both the models, but varied in space (Appendix Fig. A1).

Model fitting evaluation (Appendix Table A2) shows that, compared to the LR model, the GWLR model has smaller AIC, AICc, SSE and higher prediction accuracy in each sample and complete dataset, indicating a better model fit. The ROC curve (Appendix Fig. A2) also clearly showed the superiority of GWLR model fit relative to LR.

Comparison based on significant variables

Table 2 shows the variables from the LR and GWLR models that were tested, their inclusion in the model being based on prior determination of significance. The LR model selected three significant variables: elevation, per capita GDP and fire line. The GWLR model selected four significant variables: elevation, per capita GDP, distance to railway and vegetation coverage. The selected variables were used to fit the complete sample dataset in the LR and GWLR models (Table 2).

Estimated coefficients of the two models, using the complete sample dataset, shows that elevation and fire line were negatively correlated with forest fire occurrence. There was a positive relationship between fire occurrence and per capita GDP using the LR model. The spatial heterogeneity coefficients of two variables, such as distance to railway and vegetation coverage, showed both positive and negative relationships across the study area in the GWLR model (Fig. 3).

Model evaluation using significant variables (Table 3) showed that, compared to the LR model, the GWLR model had smaller AIC, AICc, SSE, and higher prediction accuracy in each samples and complete dataset. This indicates that the GWLR model had better goodness of fit. ROC curve analysis (Fig. 2) showed that the GWLR model had better predictive ability than the LR model. The prediction accuracy of the LR model ranged from 58.6% to 70.5%, while the prediction accuracy of the GWLR model ranged from 65.2% to 79.9% (Table 3). As with the comparison based on all variables, the GWLR model showed higher prediction accuracy in each sample and the complete dataset compared to the LR model.

Comparison based on cross-validation

In order to better compare the fitting performance of the LR and GWLR models, a cross-validation test was conducted. In the cross-validation test, each model was tested using the variables that were significant in the other model to identify the relative accuracy of each model (*i.e.* significant variables obtained from the GWLR were used to fit the LR model, and *vice versa*).The estimated coefficients of the two models using the samples and complete sample dataset are shown in Appendix Table A3.

The fitting results of the cross-validation test (Appendix Table A4) shows that, compared to the LR model, the GWLR model has smaller AIC, AICc, SSE, and higher prediction accuracy, indicating the advantage of its model fitting. ROC curve analysis (Appendix Fig. A3) shows that the GWLR model has better predictive accuracy than the LR model.

Exploring the spatial variability of significant variables

We focused this analysis on our GWLR model, since LR is incapable of showing the spatial variability of variable coefficients. In order to better show the spatial variation of coefficients of significant variables using GWLR for the entire study area, we performed spatial interpolation on estimated coefficients of each variable using the complete sample dataset in ArcGIS 10.0.

Fig. 3 shows that distance to railway, elevation and fire line have a significant negative influence in most parts of the study area. However, vegetation cover has a strong negative correlation with fire occurrence at the south corner of the study area, but a positive correlation in the northeast of Daxing'an Mountains. It is worth noting that only four variables showed a significant coefficient in space over the entire study area.

We identified the regions exhibiting different spatial variability based on GWLR model. All coefficients were combined into one layer to identify regions of high and low spatial variability (Fig. 4). Fig. 4 shows that the entire region had spatial variability, and the highest variation in space occurred in the north and south portions of the study area.

Spatial autocorrelation of residuals

Global Moran's I of the GWLR model was smaller than for the LR model (Fig. 5). This indicates that, compared to the LR model, the GWLR model considers spatial autocorrelation more, which could be better for modeling and forecasting forest fire occurrence.

Discussion

We identified the strength of the GWLR model for predicting forest fire occurrence relative to the global LR using a series of tests (*i.e.* "all variables test", "significant variables test" and "cross-validation test"). We determined that geospatial information of explanatory factors should be considered when modeling anthropogenic fire occurrence in the Chinese boreal forest. Distance to railway, vegetation cover, elevation, and fire line were identified as the probable reasons underlying fire occurrence. Distance to railway was negatively correlated with anthropogenic fire occurrence in the Daxing'an Mountains, indicating that higher fire probabilities might be expected in areas close to railways. This is likely attributable to a variety of different sources, including errant sparks released from the steam engine, fire accidents that occur in the trains, and smokers travelling by train who throw lit cigarettes out the windows, as well as a lack of controlled burning and burning activities near the tracks. The map representing the distance to railway and locations of fire ignition is shown in the Appendix Fig. 4. It clearly indicates that fire incidence happens closer to railway tracks. Railways reflect the transportation corridor and according to the official record of fire causes, as well as other studies, some fires are caused directly by human activities around railways (Stephens 2005; Romero-Calcerrada et al. 2008; Chang et al. 2013). The coefficient of distance to railway is significant in most of study area (Fig. 3).

According to our study, anthropogenic fire is more likely to occur at low elevations. It is well acknowledged that intensive human activities tend to be focused at low altitudes, which may increase the likelihood of human-caused fire ignitions (Syphard et al. 2008; Oliveira et al. 2012; Chang et al. 2013). In addition, the effects of altitude in weather conditions, vegetation cover and soil moisture are less favorable to fire occurrence as altitude increases (González et al. 2006; Vilar et al. 2010). The significant coefficient of the variable shows that the influence of elevation is spread over most of the study area, with the strongest impact focused at the western region of Daxing'an Mountains, which has high elevated terrains. It can be concluded that elevation may result in greater influence on fire occurrence when compared to the other regions of the study area.

Fire line represents local fire prevention activity. The local forest fire agencies burn the fire line regularly, especially during the fire season, to remove understory vegetation cover and to slow or stop the spread of forest fire, as well as identify the cause and origin of fire occurrence. According to our findings, the length of the fire line is negatively correlated with fire occurrence. This is because the length of burning the fire line increases every year if the intensity and area of fire occurrence was higher in the previous year. This indicates that using fire lines works efficiently for fire spread control.

Vegetation cover represents the amount of available fuel during forest fires, which has been found to be an important indicator positively related to the forest fire ignition in previous study (Chuvieco et al. 2004). In this study, vegetation cover showed a significant, inverse relationship with fire occurrence in the southern region of the study area (Fig. 3). This seemed contrary to our previous understanding, and suggests that the spatial variation of forest fuel may not be the primary cause of local forest fire occurrence in the southern Daxing'an Mountains. Previous study by Lampin-Maillet et al. (2010) highlighted that, a low level of fire density occurs even if the vegetation cover is dense accompanied with more human settlements, which may explain our inconsistent findings.

Analysis on spatial variation indicates that, the variables specific to the regions play a major role in wildfire occurrence. Fig. 4 shows the areas influenced by a number of factors that vary on the spatial scale and have different spatial variability. Spatially varying relationships are associated with geographic characteristics of the dependent and independent variables. A very high spatial variability is found in the north and south parts of Daxing'an Mountains, with some regions impacted by two to three variables. However, Fig. 4 shows that only one variable influences fire occurrence in eastern Daxing’an Mountains. Due to the complexity of spatial influence of variables on fire, one global model seems to be insufficient to describe the relationship between fire occurrence and the underlying explanatory variables. The lower the number of predictors is easier to choose and apply fire prevention measures. However, it does not mean that higher number of significant variables cause more fire risk.

In addition, compared to the LR model, we found that the GWLR model reduced spatial autocorrelation errors, although it does not directly address spatial autocorrelation issues (Propastin and Kappas 2008). These findings are consistent with other researchers (Koutsias et al.

2005). Wu and Zhang (2013) pointed out that a GWR model is not designed to model spatial autocorrelation. However, it estimates local rather than global parameters explicitly at each data point, which can account for spatial heterogeneity, especially for data with high spatial variability. In this study, the GWLR model shows better performance and accuracy than the LR model in relation to fire occurrence. However, GWLR may not suitable for making the general inference about the relationship between variables and fire occurrence. It is likely that the relationships are in fact global, but it is possible that the effects of these variables appear to vary locally due to interactions terms (e.g. the effect of vegetation cover on fire occurrence may switch from positive to negative depending on different human influence). As Jetz et al. (2005) concluded that GWR method may not be a complete alternative, but rather a good complement to global spatial regression modelling. Its power in illustrating local performance of predictor variables and their interaction with scale makes it a useful tool for forest fire analyses at the broad scale.

The study includes several caveats. There was no past or current information for variables such as distance to roads, distance to settlements, which likely change over time. The findings also limited by the data range, which only includes up to the year 2004. Future studies would benefit in testing model applicability by also including climatic variables that promote fire occurrence. The characteristics of the data collected from different data sources need to incorporate up to date information on variables for improving the paradigm of forest fire modeling. Addressing these caveats will help towards providing fundamental information upon which more sound forest fire management practices may be developed in a changing socio-economic and environmental landscape.

Conclusions

In this study, classical LR modeling was used along with GWLR modeling to explain wildfire occurrence patterns in the Chinese boreal forest. Results indicate the importance of distance to railway, vegetation cover, elevation and fire lines as underlying factors of fire occurrence, with the latter two variables being negatively correlated with fire incidence. Compared to the LR model, the GWLR model has a better performance in model prediction accuracy, model residual reduction and spatial parameter estimation, indicating spatially varying relationships enhance the explanatory power of global methods that do not appear sufficient to fully describe the relationship between wildfire occurrence and the underlying explanatory factors. Indeed, the GWLR model can complement the global LR model in helping to overcome the problem of non-stationary variables, which means the geospatial information of explanatory variables should be considered to improve anthropogenic fire occurrence modeling in the Chinese boreal forest.

Geographically weighted regression methods can be used to identify different variables, both spatially and locally. This will help to isolate and interpret the specific factors responsible for fire occurrence, which in turn provides information to design improved fire deployment activities for local forest managers. In addition, the identification of regions with spatially varying relationships can provide insight into fire management and policy and help to further our understanding of the fire problem over large areas, while at the same time recognizing its local character.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (Grant No. 31400552); The Asia-Pacific Forests Net (APFnet/2010/FPF/001) Phase II.

References

Anselin, L., and Griffith, D.A. 1988. Do spatial effects really matter in regression analysis? Papers in Regional Science. **65(1)**: 11-34. doi: 10.1111/j.1435-5597.1988.tb01155.x.

Catry, F.X., Rego, F.C., Bacao, F.L., and Moreira, F. 2009. Modeling and mapping wildfire ignition risk in Portugal. Int. J. Wildland Fire. **18(8)**: 921-931. doi.org/10.1071/WF07123.

Chang, Y., He, H.S., Bishop, I., Hu, Y., Bu, R., Xu, C., and Li X. 2007. Long-term forest landscape responses to fire exclusion in the Great Xing'an Mountains, China. Int. J. Wildland Fire. **16(1)**: 34-44. doi.org/10.1071/WF05093.

Chang, Y., Zhu, Z.L., Bu, R.C., Chen, H.W., Feng, Y.T., Li, Y.H., Hu, Y.M., and Wang, Z.C. 2013. Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. Landscape Ecol. **28(10)**: 1989-2004. doi.10.1007/s10980-013-9935-4.

Chuvieco, E., Cocero, D., Riano, D., Martin, P., Martinez-Vega, J., de la Riva, J., and Perez, F. 2004. Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. Remote Sens. Environ. **92(3)**: 322-331. doi:10.1016/j.rse.2004.01.019.

delHoyo, V.L., Martin Isabel, M., and Martinez Vega, F. 2011. Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. Eur. J. For. Res. **130(6)**: 983-996. doi.10.1007/s10342-011-0488-2.

Fernandez-Pello, A.C. 1994. The Challenge of fire prediction. Combust. Sci. Technol. (special silver anniversary issue). **98(4-6)**: 281-290. doi:10.1080/00102209408935413.

Fielding, A.H., and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence absence models. Environ. Conserv. **24(1)**: 38-49. Available from: http://journals.cambridge.org/ [accessed on 10 August 2015].

Foody, G.M. 2003. Geographical weighting as a further refinement to regression modeling: An example focused on the NDVI-rainfall relationship. Remote Sens. Environ. **88(3)**: 283-293. doi:10.1016/j.rse.2003.08.004.

Foody, G.M. 2004. Spatial non-stationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecol. Biogeogr. **13(4)**: 315-320. doi: 10.1111/j.1466-822X.2004.00097.x.

Fotheringham, A.S., Charlton, M., and Brunsdon, C. 1996. The geography of parameter space: an investigation of spatial non-stationarity. Int. J. Geogr. Inf. Syst. **10(5)**: 605-627. doi: 10.1080/02693799608902100.

Garcia, C.V., Woodard, P.M., Titus, S.J., Adamowicz, W.L., and Lee, B.S. 1995. A logit model for predicting the daily occurrence of human caused forest fires. Int. J. Wildland Fire. **5(2)**: 101-111. http://www.publish.csiro.au/ [accessed on 24 May 2014].

González, J.R., Palahi, M., Trasobares, A., and Pukkala, T. 2006. A fire probability model for forest stands in Catalonia (north-east Spain). Ann. For. Sci. **64(5)**: 169-176. doi.org/10.1051/forest:2007047.

Guo, F.T., Innes, J.L., Wang, G.Y., Ma, X.Q., Sun, L., Hu, H.Q., and Su, Z.W. 2015. Historic distribution and driving factors of human-caused fires in the Chinese boreal forest between 1972 and 2005. J. Plant Ecol. **8(5)**: 480-490. doi: 10.1093/jpe/rtu041.

Gutman, G., and Ignatov, A. 1997. The derivation of the green vegetation fraction from NOAA/AVHRR Data for use in numerical weather prediction models. Int. J. Remote Sens. **19(8)**: 1533-1543. doi:10.1080/014311698215333.

Huang, Y., and Leung, Y. 2002. Analysing regional industrialisation in Jiangsu province using Geographically Weighted Regression. J. Geogr. Syst. **4(2)**: 233-249. doi:10.1007/s101090200081.

Jetz, W., Rahbek, C., and Lichstein, J.W. 2005. Local and global approaches to spatial data analysis in ecology. Global Ecol. Biogeogr. **14(1)**: 97-98. doi:10.1111/j.1466-822X.2004.00129.x.

Jimenez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modeling. Global Ecol. Biogeogr. **21(4)**: 498-507. doi: 10.1111/j.1466-8238.2011.00683.x.

Johson, E.A., and Larsen, C.P.S. 1991. Climatically induced change in fire frequency in the southern Canadian Rockies. Ecology. **72(1)**: 194-201. doi: 10.2307/1938914.

Korovin, G.N. 1996. Analysis of distribution of forest fires in Russia. *In* Fire in ecosystems of Boreal Eurasia. *Edited by* J.G. Goldammer and V.V. Furyaev. Kluwer Academic Publishers. Boston. pp. 112-128.

Koutsias, N., Martinez-Fernandez, J., and Allgower Britta. 2010. Do factors causing wildfires vary in space? Evidence from Geographically Weighted Regression. GIsci. Remote Sens. **47(2)**: 221-240. doi: 10.2747/1548-1603.47.2.221.

Koutsias, N., Martinez, J., Chuvieco, E., Alligower, B. 2005. Modeling wildland fire occurrence in Southern Europe by a Geographically Weighted Regression approach. *In* Proceedings of the 5th International workshop on remote sensing and GIS applications to forest fire management: Fire Effects Assessment, Universidad de Zaragoza, Spain: 57-60. ISBN: 84-96214-52-4.

Lampin-Maillet, C., Jappiot, M., Long, M., Bouillon, C., Morge, D., and Ferrier, J. 2010. Mapping wildland-urban interfaces at large scales integrating housing density and vegetation aggregation for fire prevention in the South of France. J. Environ. Manage. **91(3)**: 732-741. doi:10.1016/j.jenvman.2009.10.001.

Liu, Z., Yang, J., Chang, Y., Weisberg, P. J., and He, H.S. 2012. Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of northeast China. Glob. Chang. Biol. **18(6)**: 2041-2056. doi: 10.1111/j.1365-2486.2012.02649.x.

Maingi, J.K., and Henry, M.C. 2007. Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. Int. J. Wildland Fire. **16(1)**: 23-33. doi.org/10.1071/WF06007.

Martell, D.L., Otukol, S., and Stocks, B.J. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. Can. J. For. Res. **17(5)**: 394-401. doi:10.1139/x87-068.

Martinez, J., Chuvieco, E., and Koutsias, N. 2013. Modelling long-term fire occurrence factors in Spain by accounting for local variations with Geographically Weighted Regression. Nat. Hazard. Earth Syst.Sci. **13**: 311-327. doi: 10.5194/nhess-13-311-2013.

Martinez, J., Vega-Garcia, C., and Chuvieco, E. 2009. Human-caused wildfire risk rating for prevention planning in Spain. J. Environ. Manage. **90(2)**: 1241-1252. doi: 10.1016/j.jenvman.2008.07.005.

Matthews, S.A., and Yang, T. 2012. Mapping the results of local statistics: using geographically weighted regression. Demogr. Res. **26(6)**: 151-166. doi: 10.4054/DemRes.2012.26.6.

Mollicone, D., Eva, H.D., and Achard, F. 2006. Human role in Russian wild fires. Nature. **440(7083)**: 436-437. doi:10.1038/440436a.

Mundo, I.A., Thorsten, W., Rajapandian, K., and Thomas, K. 2013. Environmental drivers and spatial dependency in wildfire ignition patterns of northwestern Patagonia. J. Environ. Manage. **123**: 77-87. doi:10.1016/j.jenvman.2013.03.011.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., and Charlton, M. 2005. Geographically Weighted Poisson Regression for disease association mapping. Stat. Med. **24(17)**: 2695-2717. doi: 10.1002/sim.2129.

Niklasson, M., and Granstrom, A. 2000. Numbers and sizes of fires: long-term spatially explicit fire history in a Swedish boreal landscape. Ecology. **81(6)**: 1484-1499. doi: 10.2307/177301.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., and Pereira, J.M.C. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. For. Ecol. Manage. **275**: 117-129. doi:10.1016/j.foreco.2012.03.003.

Oliveira, S., Pereira, J.M.C., San-Miguel-Ayanz, J., and Lourenco, L. 2014. Exploring the spatial patterns of fire density in southern Europe using Geographically Weighted Regression. Appl. Geogr. **51**: 143-157. doi:10.1016/j.apgeog.2014.04.002.

Padilla, M., and Vega-Garcia, C. 2011. On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. Int. J. Wildland Fire. **20(1)**: 46-58. doi.org/10.1071/WF09139.

Prasad, V.K., Badarinath, K.V.S., and Eaturu, A. 2008. Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India. J. Environ. Manage. **86(1)**: 1-13. doi:10.1016/j.jenvman.2006.11.017.

Propastin, P. 2012. Modifying geographically weighted regression for estimating above biomass in tropical rainforest by multispectral remote sensing data. Int. J. Appl. Earth Obs. **18(1)**: 82-90. doi:10.1016/j.jag.2011.12.013.

Purevdorj, T.S., Tateishi, R., Ishiyama, T., and Honda, Y. 1998. Relationships between percent vegetation cover and vegetation indices. Int. J. Remote Sens. **19(18)**: 3519-3535.

doi:10.1080/014311698213795.

Rodrigues, M., and de la Riva, J. 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ. Model. Softw. **57**: 192-201. doi:10.1016/j.envsoft.2014.03.003.

Rodrigues, M., de la Riva, J., and Fotheringham, S. 2014. Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using Geographically Weighted Logistic Regression. Appl. Geogr. **48**: 52-63. doi: 10.1016/j.apgeog.2014.01.011.

Romero-Calcerrada, R., Barrio-Parra, F., Millington, J.D.A., and Novillo, C.J. 2010. Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). Ecol. Model. **221(1)**: 34-45. doi:10.1016/j.ecolmodel.2009.08.008.

Romero-Calcerrada, R., Novillo, C.J., Millington, J.D.A., and Gomez-Jimenez, I. 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). Landscape Ecol. **23**: 341-354. doi: 10.1007/s10980-008-9190-2.

Shvidenko, A.Z., and Goldammer, J.G. 2001. Fire situation in Russia. IFFN. **24**: 41-59. http://www.fire.uni-freiburg.de/iffn/country/rus/rus_26.htm.

Stephens, S.L. 2005. Forest fire causes and extent on United States forest service lands. Int. J. Wildland Fire. **14(3)**: 213-222. doi: 10.1071/WF04006.

Stocks, B. J. 1993. Global warming and forest-fires in Canada. For. Chron. **69(3)**: 290-293.doi: 10.5558/tfc69290-3.

Syphard, A.D., Radeloff, V.C., Keely, J.E., Hawbaker, T.J., Clayton, M.K., Stewart, S.I., and Hammer, R.B. 2007. Human influence on California fire regimes. Ecol. Appl. **17(5)**: 1388-1402. doi:10.1890/06-1128.1.

Syphard, A.D., Radeloff, V.C., Keuler, N.S., Taylor, R.S., Hawbaker, T.J., Stewart, S.I., and Clayton, M.K. 2008. Predicting spatial patterns of fire on a southern California landscape. Int. J. Wildland Fire. **17(5)**: 602-613. doi.org/10.1071/WF07087.

Vasconcelos, M.J.P., Silva, S., Tome, M., Alvim, M., and Pereira, J.M.C. 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogramm. Eng. Rem. S. **67(1)**: 73-81.

Vilar, L., Woolford, D.G., Martell, D.L., Pilar, and Martin, M. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. Int. J. Wildland Fire. **19(3)**: 325-337. doi.org/10.1071/WF09030.

Wallenius, T.H., Kuuluvainen, T., and Vanha-Majamaa, I. 2004. Fire history in relation to site type and vegetation in Vienansalo wilderness in eastern Fennoscandia, Russia. Can. J. Forest Res. **34(7)**: 1400-1409. doi: 10.1139/x04-023.

Wang, Q., Ni, J., and Tenhunen, J. 2005. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Global Ecol. Biogeogr. **14(4)**: 379-393. doi: 10.1111/j.1466-822X.2005.00153.x.

Weber, M.G., and Flannigan, M.D. 1997. Canadian boreal forest ecosystem structure and function in a changing climate: Impacts on fire regimes. Environ. Rev. **5**: 145-166. doi: 10.1139/a97-008.

Wotton, B.M., Martell, D.L., and Logan, K.A. 2003. Climate change and people-caused forest fire occurrence in Ontario. Climate Change**.** **60(3)**: 275-295. doi: 10.1023/A:1026075919710.

Wu, W., and Zhang, L.J. 2013. Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in northeastern Puerto Rico. Appl. Geogr. **37**: 52-62. doi:10.1016/j.apgeog.2012.10.012.

Wu, Z.E., He, H.S., Yang, J., Liu, Z.H., and 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. doi:10.1016/j.scitotenv.2014.06.011.

Xu, H.C., Li, Z.D., and Qiu, Y. 1997. Fire disturbance history in virgin forest in northern region of Daxinganling Mountains. Acta Ecologica sinica **17(4)**: 337-343. (in Chinese). http://en.cnki.com.cn/Article_en/CJFDTOTAL-STXB704.000.htm.

Zhang, H.J., Qi, P.C., and Guo, G.M. 2014. Improvement of fire danger modelling with geographically weighted logistic model. Int. J. Wildland Fire. **23(8)**: 1130-1146. doi.org/10.1071/WF13195.

Zhang, L., Gove, J.H., and Heath, L.S. 2005. Spatial residual analysis of six modeling techniques. Ecol. Model. **186(2)**: 154-177. doi:10.1016/j.ecolmodel.2005.01.007.

Zhang, L., and Shi, H. 2004. Local modeling of tree growth by geographically weighted regression. Forest Sci. **50(2)**: 225-244.

Figure Captions

Figure 1. Location map of study area with fire points

Figure 2. ROC Curves of the LR and GWLR models for five samples and complete dataset (CS) with significant variables. The upper curve (GWLR) has greater area under the curve (AUC) than that of LR, indicating GWLR has a relatively higher model fitting ability. AUC values of the respective sample are given in parenthesis.

Figure 3. Significant areas of the estimated coefficient maps resulting from GWLR. If the *t*-value of the estimated coefficient for a particular variable is < -1.96 or > 1.96, then the variable has a significant effect on fire presence; otherwise, the variable is considered as having no significant effect on fire presence. Negative coefficients are mapped with cold colors (blue) and positive coefficients with warm colors (orange to red). Full variable names, represented by abbreviations here, are given in Table 1.

Figure 4. Regions where the coefficient of variables of the GWLR model were significant. The degree of spatial variability is expressed by the number of variables exhibiting spatial variability.

Figure 5. Residual spatial autocorrelation coefficients plots of the LR and GWLR models, where a: all variables test; b: significant variable test; c: cross-validation test.

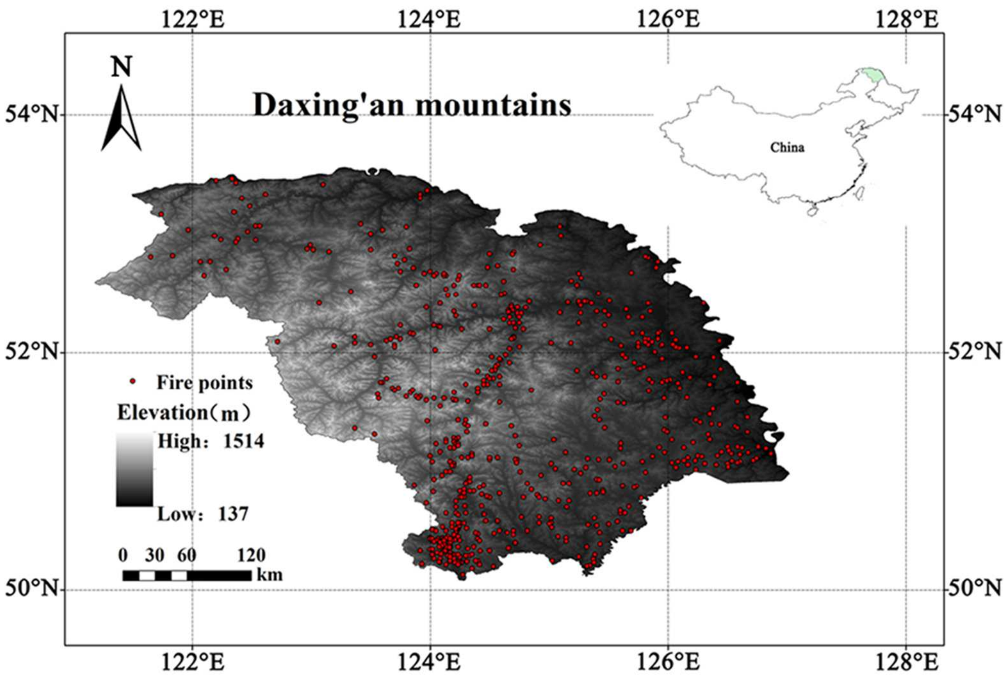

Figure 1
99x67mm (300 x 300 DPI)

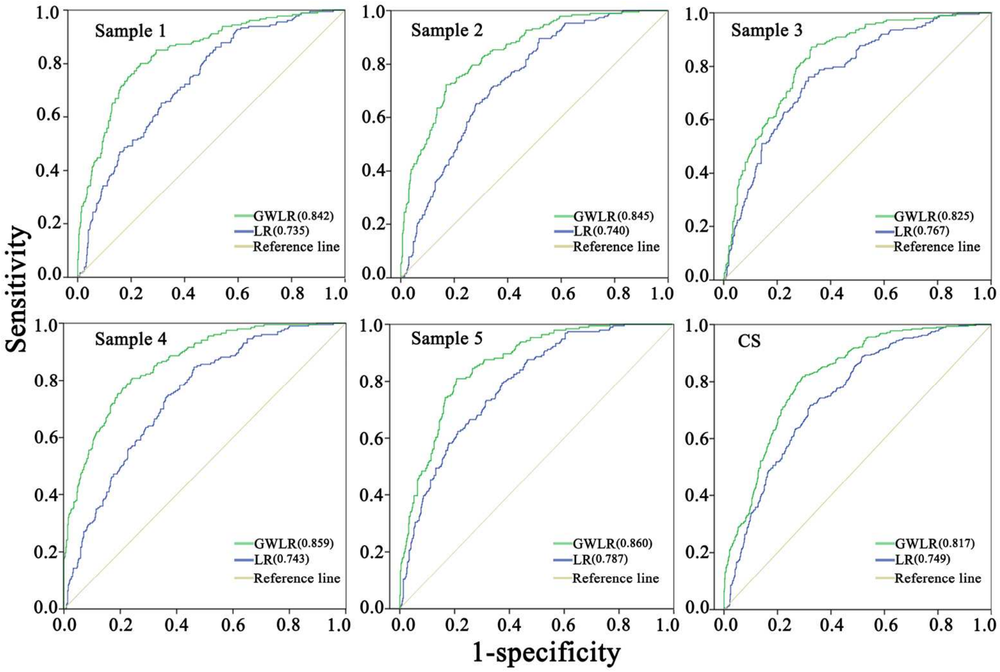

Figure 2
180x120mm (300 x 300 DPI)

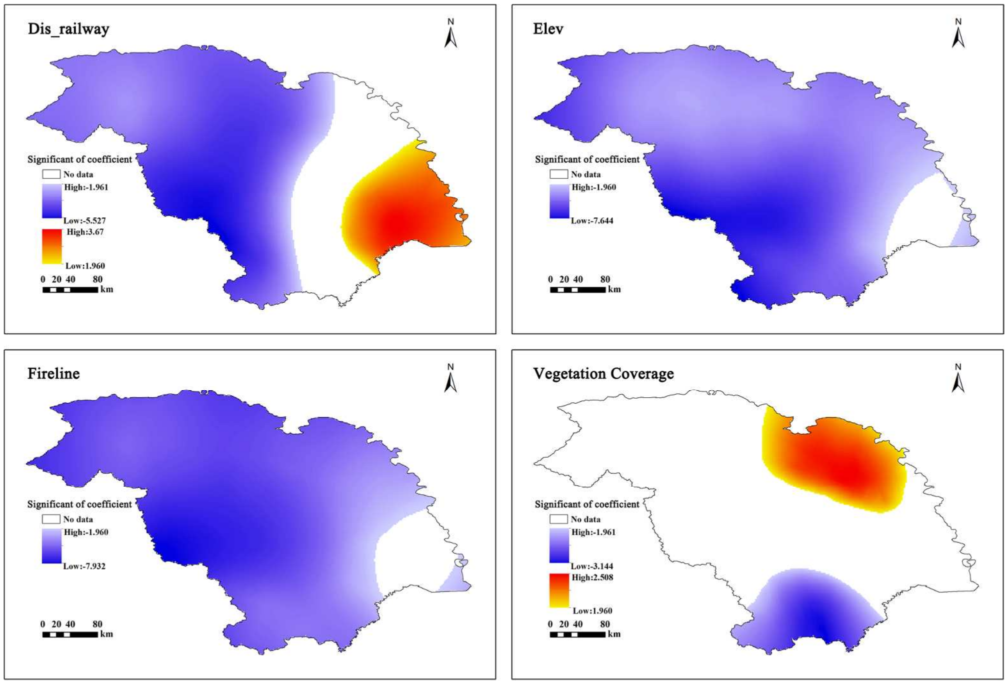

Figure 3
180x121mm (300 x 300 DPI)

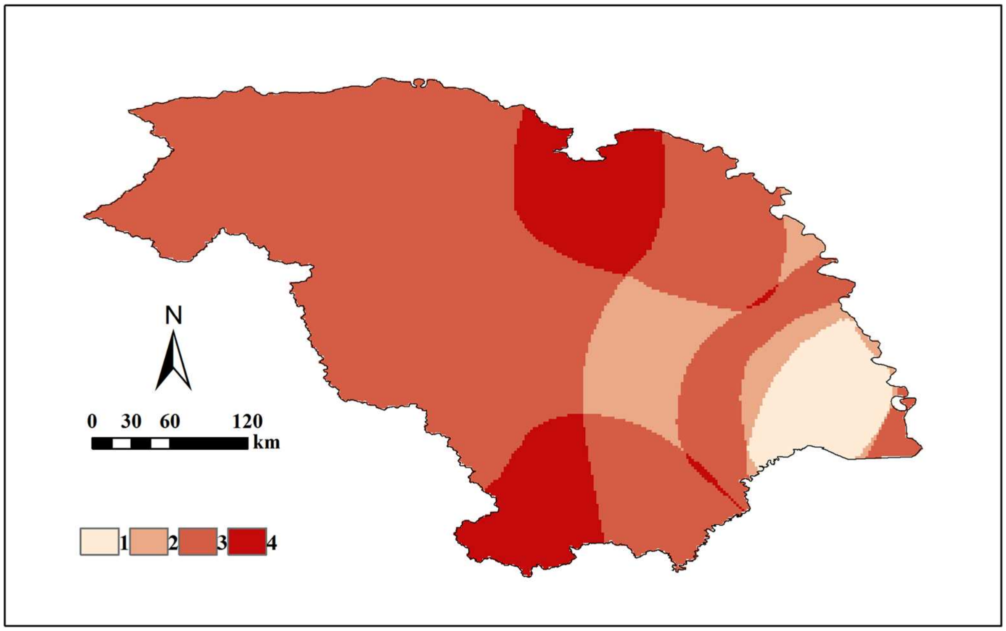

Figure 4
160x99mm (300 x 300 DPI)

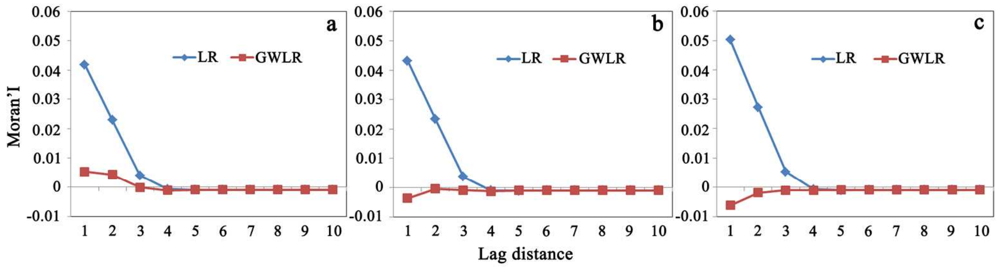

Figure 5
199x52mm (300 x 300 DPI)

Table 1. Independent or predictor variables included in forest fire model development for Daxing'an Mountains

Variable Type	Variable Name	Code	Description
Topographic	Elevation	Elev	The elevation of each fire point and control extracted from a raster map of study area
	Slope	Slope	The slope of each fire point and control extracted from a raster map of study area
	Aspect	Aspect	Proportion of each aspect class (flat, N, E, S, W) in the study area
Vegetation	Forest type	Forest_type	Proportion of each forest type in the study area
	Vegetation cover	Veg_cover	The fractional vegetation cover over the entire study area at 1 km resolution
Infrastructure	Distance to railway	Dis_railway	The distance between railway and fire point
	Distance to river	Dis_river	The distance between river and fire point
	Distance to road	Dis_road	The distance between road and fire point
	Distance to settlement	Dis_sett	The distance between settlement and fire point
	Number of inspection stations	LNS	The number of inspection stations that were used to inspect the potential fire source with people who will enter the mountains during the fire season
	Length of fire line	Fireline	The length of fire line for fire prevention
Socio-economic	Per capita GDP	CGDP	Per capita GDP of the study area
	Funding	Funding	Annual funding for forest fire prevention

Table 2. Coefficient estimates of significant variables from LR and GWLR models

Model	Statistics	$\beta_{intercept}$	$\beta_{dis\text{-}railway}$	β_2- β_7	$\beta_{elevation}$	β_{CGDP}	$\beta_{10\text{-}11}$	$\beta_{fire\ line}$	$\beta_{vegetation\ cover}$
LR	Estimate	2.8736	/	/	-0.0034	0.00014	/	-0.0003	/
	Standard deviation (s.d.)	0.3458	/	/	0.0005	0.00003	/	0.0001	/
	Estimate -1 s.d.	2.5278	/	/	-0.0038	0.00011	/	-0.0003	/
	Estimate +1 s.d.	3.2195	/	/	-0.0029	0.00016	/	-0.0002	/
GWLR	Minimum	-0.8350	-0.0336	/	-0.0058	/	/	-0.00068	-2.0262
	25% quartile	0.5161	-0.0210	/	-0.0041	/	/	-0.00052	-0.8489
	Mean	1.9052	-0.0106	/	-0.0032	/	/	-0.00037	0.2487
	Median	1.5866	-0.0124	/	-0.0030	/	/	-0.00037	0.6086
	75% quartile	3.2335	0.0013	/	-0.0025	/	/	-0.00023	1.2608
	Maximum	5.3785	0.0099	/	-0.0010	/	/	-0.00005	1.9520

Note:β_2: Distance to river; β_3: Distance to road; β_4: Distance to settlement; β_5: Forest type; β_6: Slope; β_7: Aspect; β_{10}: Funding; β_{11}: Number of inspection stations.

Table 3. Comparison of the LR and GWLR models based on significant variables

Data	Model	AIC	AICc	SSE	Cut-off	Prediction accuracy (%)	
						Training data (60%)	Validation data (40%)
Sample 1	LR	740.04	740.10	114.06	0.216	60.3	61.6
	GWLR	610.27	612.11	75.54	0.329	77.4	65.2
Sample 2	LR	730.48	730.61	118.84	0.205	60.6	58.6
	GWLR	619.71	621.66	92.88	0.394	79.9	69.3
Sample 3	LR	748.97	749.06	111.50	0.279	70.5	63.2
	GWLR	640.65	641.45	100.44	0.296	72.9	67.3
Sample 4	LR	725.03	725.26	119.86	0.242	63.4	66.8
	GWLR	639.80	646.26	89.82	0.338	77.36	71.9
Sample 5	LR	712.71	712.80	108.84	0.254	67.6	62.0
	GWLR	608.60	611.42	90.42	0.378	79.6	67.0
Complete sample	LR	1194.80	1194.86	193.05	0.299	68.9	
	GWLR	1043.80	1044.34	167.66	0.337	73.9	

Appendix Tables

Table A1. Coefficient estimates of the all variables test from LR and GWLR models

Model	Statistics	$\beta_{\text{Intercept}}$	$\beta_{\text{Dis\_railway}}$	$\beta_{\text{Dis\_river}}$	$\beta_{\text{Dis\_road}}$	$\beta_{\text{Dis\_sett}}$	$\beta_{\text{Forest\_type}}$	β_{Slope}	β_{Aspect}	β_{Elev}	β_{CGDP}	β_{Funding}	β_{LNS}	β_{Fireline}	$\beta_{\text{Veg\_cover}}$
LR	Estimate	3.4763	-0.0064	-0.0121	-0.0179	0.0071	-0.1638	0.0400	-2.2540	-0.0038	0.00012	0.00030	-0.0199	-0.00023	-0.3423
	Standard deviation (s.d.)	0.6753	0.0041	0.0913	0.0151	0.0051	0.7578	0.0220	4.0006	0.0005	0.00003	0.00017	0.0034	0.00006	0.4258
	Estimate -1 s.d.	2.8010	-0.0105	-0.1034	-0.0329	0.0020	-0.9216	0.0181	-6.2545	-0.0043	0.00009	0.00013	-0.0233	-0.00030	-0.7681
	Estimate +1 s.d.	4.1516	-0.0023	0.0792	-0.0028	0.0122	0.5941	0.0620	1.7466	-0.0032	0.00015	0.00047	-0.0165	-0.00017	0.0834
GWLR	Minimum	0.6195	-0.0251	-0.1081	-0.0524	0.0033	-0.0482	0.0174	-10.3828	-0.0053	0.00007	0.00002	-0.0225	-0.00044	-1.4495
	25% quartile	2.2786	-0.0185	-0.0653	-0.0386	0.0068	0.1901	0.0349	-7.0371	-0.0045	0.00010	0.00016	-0.0204	-0.00035	-0.7495
	Mean	3.3931	-0.0117	-0.0072	-0.0250	0.0091	0.3779	0.0399	-3.5474	-0.0037	0.00012	0.00029	-0.0184	-0.00027	-0.1432
	Median	3.7563	-0.0151	-0.0303	-0.0274	0.0090	0.3277	0.0400	-2.9585	-0.0038	0.00012	0.00026	-0.0180	-0.00026	-0.0202
	75% quartile	4.5201	-0.0042	0.0494	-0.0177	0.0114	0.5285	0.0483	-0.1119	-0.0030	0.00014	0.00042	-0.0169	-0.00019	0.5212
	Maximum	5.4462	0.0039	0.1289	0.0167	0.0160	1.2734	0.0560	2.1111	-0.0020	0.00015	0.00055	-0.0134	-0.00014	0.8459

Table A2. Comparison of LR and GWLR models with all variables

Data	Model	AIC	AICc	SSE	Cut-off	Prediction accuracy (%) Training data (60%)	Prediction accuracy (%) Validation data (40%)
Sample 1	LR	705.67	706.32	113.83	0.279	67.4	68.2
	GWLR	623.56	627.96	87.93	0.327	78.4	70.0
Sample 2	LR	720.70	721.36	116.98	0.288	69.1	69.3
	GWLR	642.43	646.15	93.71	0.342	76.6	72.5
Sample 3	LR	686.29	686.95	109.27	0.281	71.1	66.7
	GWLR	626.05	629.22	91.15	0.343	78.6	71.9
Sample 4	LR	728.93	729.58	118.11	0.270	65.9	72.2
	GWLR	653.91	659.93	92.39	0.242	70.7	58.6
Sample 5	LR	666.25	666.91	106.04	0.300	72.0	65.8
	GWLR	619.51	622.10	92.12	0.342	77.4	70.7
Complete sample	LR	1160.38	1160.77	189.84	0.281	67.8	
	GWLR	1055.96	1057.80	162.90	0.348	75.4	

Table A3. Coefficient estimates of variables from cross-validation of LR and GWLR models

Model	Statistics	$\beta_{Intercept}$	$\beta_{Dis\_railway}$	$\beta_{2\text{-}7}$	β_{Elev}	β_{CGDP}	$\beta_{10\text{-}11}$	$\beta_{Fireline}$	β_{Veg_cover}
LR	Estimate	3.2210	-0.00189	/	-0.0035	/	/	-0.0002	-0.2584
	Standard deviation (s.d.)	1.0280	0.00194	/	0.0005	/	/	0.0001	0.1979
	Estimate -1 s.d.	2.1931	-0.00383	/	-0.0040	/	/	-0.0003	-0.4563
	Estimate +1 s.d.	4.2490	0.00005	/	-0.0031	/	/	-0.0002	-0.0605
GWLR	Minimum	-0.7366	/	/	-0.0069	-0.00018	/	-0.00067	/
	25% quartile	1.1867	/	/	-0.0043	0.00001	/	-0.00045	/
	Mean	2.6304	/	/	-0.0028	0.00008	/	-0.00033	/
	Median	2.5722	/	/	-0.0031	0.00008	/	-0.00031	/
	75% quartile	3.9588	/	/	-0.0012	0.00018	/	-0.00024	/
	Maximum	6.6053	/	/	0.0034	0.00024	/	0.00005	/

Note: The corresponding coefficients for variables are β_2: Distance to river; β_3: Distance to road; β_4: Distance to settlement; β_5: Forest type; β_6: Slope; β_7: Aspect; β_{10}: Funding; β_{11}: Number of inspection stations.

Table A4. Comparison of LR and GWLR models based on cross-validation

Data	Model	AIC	AICc	SSE	Cut-off	Prediction accuracy (%) Training data (60%)	Prediction accuracy (%) Validation data (40%)
Sample 1	LR	695.31	695.40	120.73	0.245	62.5	59.5
	GWLR	618.86	620.88	92.60	0.362	79.3	66.4
Sample 2	LR	708.73	708.82	124.24	0.250	62.0	62.9
	GWLR	644.88	645.81	100.94	0.346	75.5	73.2
Sample 3	LR	681.06	681.15	118.62	0.262	64.1	60.3
	GWLR	637.39	638.69	98.43	0.299	74.6	70.0
Sample 4	LR	722.49	722.58	126.29	0.270	62.6	65.4
	GWLR	668.82	670.28	104.08	0.321	72.5	68.9
Sample 5	LR	660.26	660.36	120.33	0.271	63.8	61.0
	GWLR	607.35	609.57	91.53	0.242	72.5	67.7
Complete dataset	LR	1153.77	1153.83	201.79	0.254	62.5	
	GWLR	1040.49	1044.17	156.39	0.339	77.0	

Appendix Figures

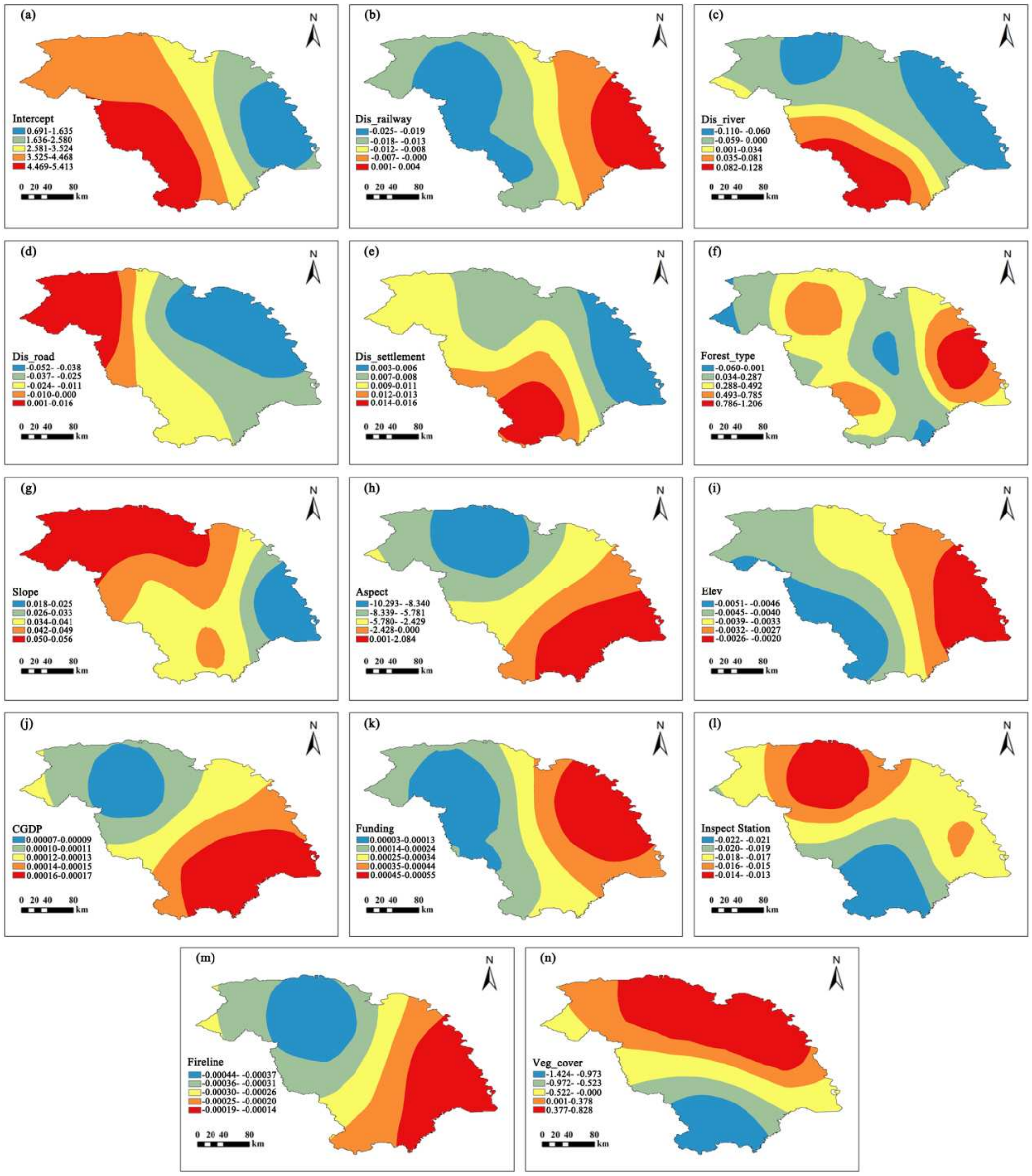

Fig. A1. Regression coefficients for all explanatory variables in the GWLR model. Negative coefficients are mapped with cold colors (blue) and positive coefficients with warm colors (red).

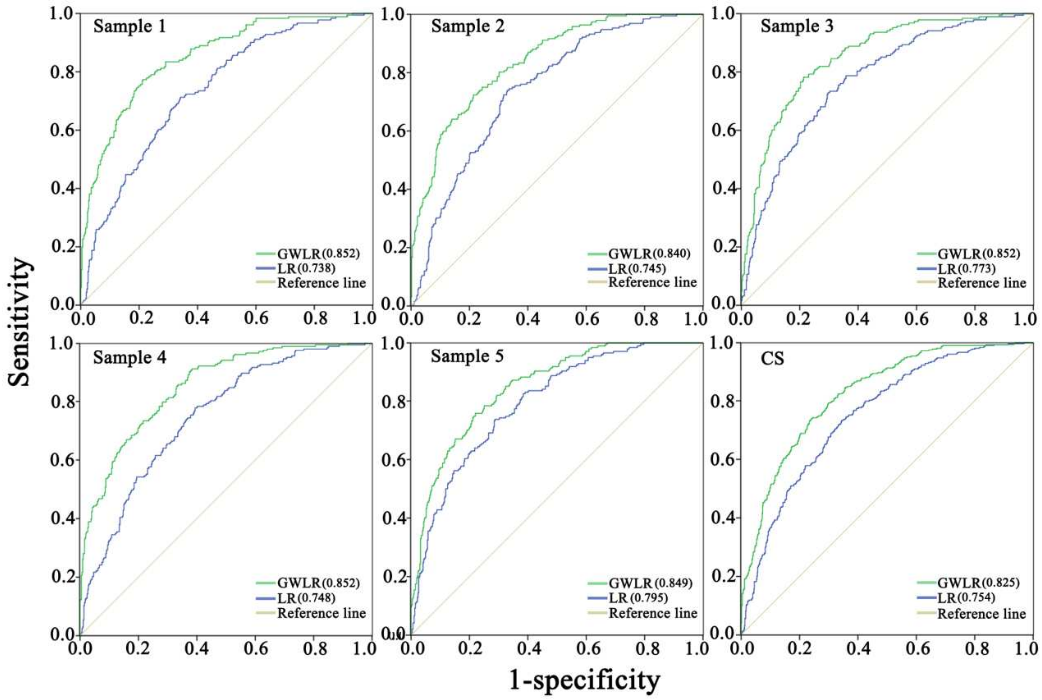

Fig. A2. ROC Curves of the LR and GWLR models for five samples and the complete dataset (CS) with all variables. The upper curve (GWLR) has greater area under the curve (AUC) than that of LR, indicating GWLR has a relatively higher model fitting ability. AUC values of the respective sample are given in parenthesis.

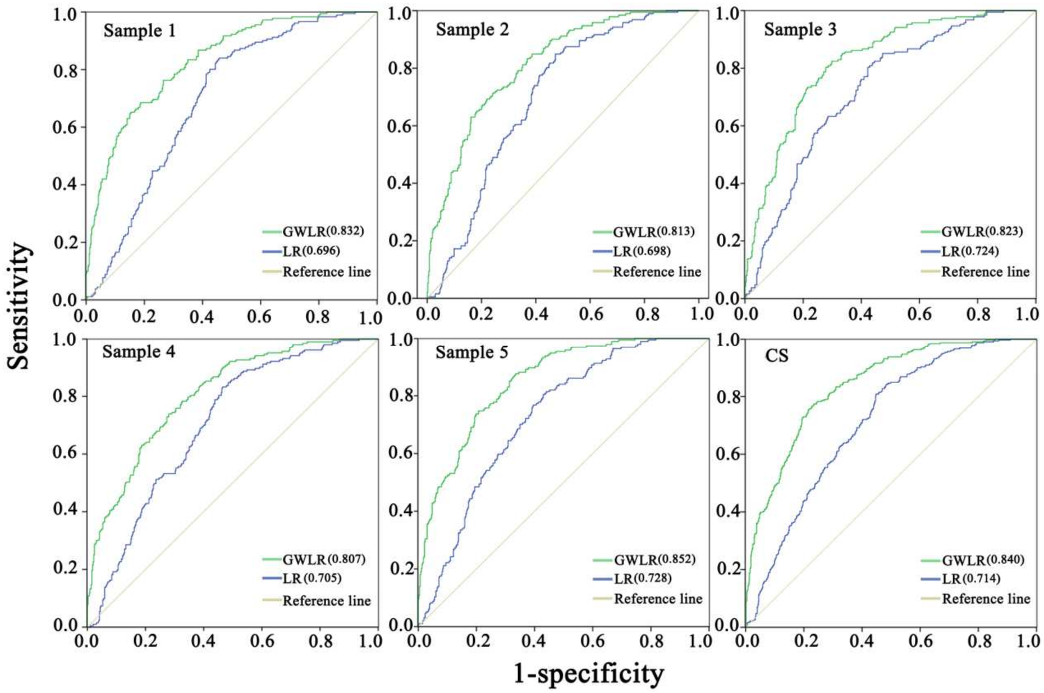

Fig. A3. ROC Curves of the LR and GWLR models for five samples and the complete dataset (CS) based on cross-validation. The upper curve (GWLR) has greater area under the curve (AUC) than that of LR, indicating GWLR has a relatively higher model fitting ability. AUC values of the respective sample are given in parenthesis.

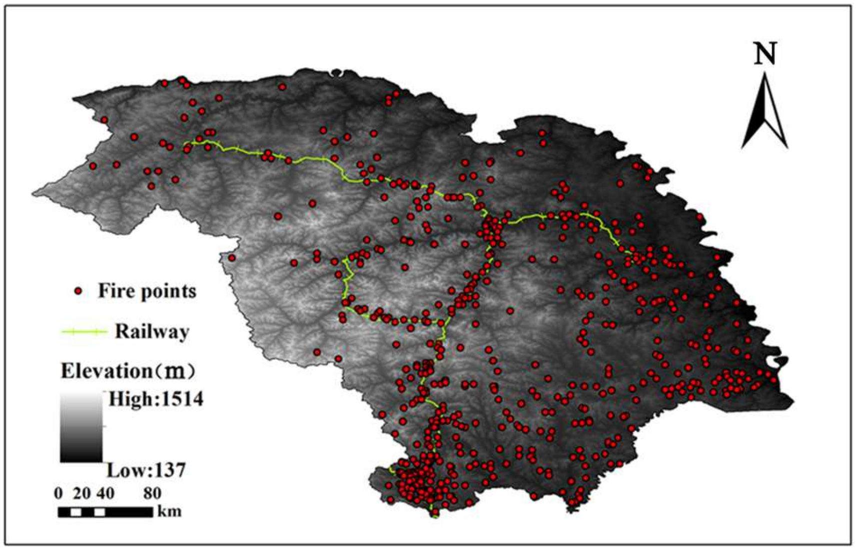

Fig. A4. The distribution of fire points and railway tracks in Daxing'an Mountains.