Applied Geography 48 (2014) 52–63

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression

M. Rodrigues [a,*], J. de la Riva [a], S. Fotheringham [b]

[a] GEOFOREST Group, IUCA, Department of Geography and Land Management, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain
[b] School of Geography and Geosciences, Irvine Building, University of St. Andrews, St. Andrews, Fife KY16 9AL, Scotland, UK

Keywords:
Fire risk
Human causality
Forest fires
GWR
Logistic regression
GIS modeling

A B S T R A C T

Forest fires are one of the main factors transforming landscapes and natural environments in a wide variety of ecosystems. The impacts of fire occur both on a global scale, with increasing emissions of greenhouse gases, and on a local scale, with land degradation, biodiversity loss, property damage, and loss of human lives. Improvements and innovations in fire risk assessment contribute to reducing these impacts. This study analyzes the spatial variation in the explanatory factors of human-caused wildfires in continental Spain using logistic regression techniques within the framework of geographically weighted regression models (GWR). GWR methods are used to model the varying spatial relationships between human-caused wildfires and their explanatory variables. Our results suggest that high fire occurrence rates are mainly linked to wildland–agricultural interfaces and wildland–urban interfaces. The mapping of explanatory factors also evidences the importance of other variables of linear deployment such as power lines, railroads, and forestry tracks. Finally, the GWLR model gives an improved calculation of the probabilities of wildfire occurrence, both in terms of accuracy and goodness of fit, compared to global regression models.

Introduction

Forest fires are an important factor in landscape transformation, vegetation succession, land degradation, and air quality. Although fire has been traditionally used as a land management tool, and many ecosystems are well adapted to fire cycles, recent changes in weather and social factors relating to wildfire could be modifying the historical fire regimes (Gonzalez, Neilson, Lenihan, & Drapek, 2010; San-Miguel Ayanz et al., 2012), possibly resulting in undesired effects. Indeed, the influence of climate change on an increase in fire frequency and intensity has been reported in several ecosystems (Kasischke & Turetsky, 2006; Westerling, Hidalgo, Cayan, & Swetnam, 2006). Climatic projections suggest worse conditions in future decades in tropical and boreal regions (Flannigan, Logan, Amiro, Skinner, & Stocks, 2005). In addition to these global effects, wildfires also have relevant local effects which are commonly associated with the frequency and intensity of fires, often implying soil loss and land degradation, loss of lives or biodiversity, and damage to property and infrastructure (Omi, 2005). On the other hand, human beings have a great impact on fire regimes because they alter ignition frequency and fuel fragmentation and suppress fires (Guyette, Muzika, & Dey, 2002). The dynamics of fire regimes in southern Europe are mainly related to human factors. In fact, humans are responsible for more than 95% of the fires in this region (San-Miguel Ayanz & Camiá, 2009). In the case of Spain, nearly 90% of wildfires are related to an anthropogenic source (Chuvieco et al., 2012; Martinez, Vega-Garcia, & Chuvieco, 2009). It is thus clear that human factors play an important role in fire ignition. Furthermore, determining the explanatory factors facilitates the development of future wildfire scenarios in the context of climate change. Therefore, a better comprehension of the local driving forces of fire ignition and of predicting where fires are likely to start are core elements in designing strategies to mitigate wildfire initiation and to identify areas at risk (Finney, 2005). In recent years, several methods for wildfire risk assessment have been developed using different methodological schemes, variables, and scales (Martínez-Vega, Echevarría, Ibarra, Echeverría, & Rodrigues, 2012). Without being exhaustive, some of the more recent efforts have included those by Amatulli, Rodrigues, Trombetti, and Lovreglio (2006), Chuvieco et al. (2010, 2012), Cooke et al. (2007), Loboda (2009),

* Corresponding author. Tel.: +34 876 554 058; fax: +34 976 761 506.
E-mail address: rmarcos@unizar.es (M. Rodrigues).

0143-6228/$ – see front matter
http://dx.doi.org/10.1016/j.apgeog.2014.01.011

Martínez, Chuvieco, and Koutsias (2013), Martinez et al. (2009), Padilla and Vega-Garcia (2011), and Romero-Calcerrada, Barrio-Parra, Millington, and Novillo (2010). Similar efforts have been invested in modeling fire occurrence (see Plucinksi, 2011 for an exhaustive review) and, particularly, to human-caused ignition (Martinez et al., 2009, 2013; Padilla & Vega-Garcia, 2011). The analysis of human factors in forest fires is widely recognized as critical for fire risk estimation (Kalabokidis, Gatzojannis, & Galatsidas, 2002; Martínez, Chuvieco, & Martín, 2004), however the literature on this topic is scarce and mainly site-specific (Krawchuk et al., 2009; Le Page, Oom, Silva, Jönsson, & Pereira, 2010; Martinez et al., 2009), perhaps due to the complexity of predicting human behavior, both in space and time. Currently, most fire risk models in use are based on physical parameters such as weather data or fuel moisture content – there is no global forest fire risk system that includes the human factors operationally, although some consider it in their components (San-Miguel Ayanz & Camiá, 2009). However, over recent years, the role of human factors in fire behavior modeling has been increasing, and several models now include an anthropogenic component in their assessments (Chuvieco et al., 2010, 2012; Loepfe, Martinez-Vilalta, & Piñol, 2011).

Additionally, the fit of statistical models of risk estimation, previously discussed for different regions of the Iberian Peninsula by Chuvieco et al. (2010), shows that the explanatory factors vary spatially in their significance and contribution. This finding is also supported by Padilla and Vega-Garcia (2011), who reported the existence of high spatial variation in the relationships between explanatory variables and historical human-caused fire occurrences. Accordingly, the use of global regression methods over wide areas, such as here, could be inappropriate due to the application of stationary coefficients over the whole study area, possibly masking local interactions with the explanatory factors. Hence, to better understand the causes of wildfires, the spatial variation of the human factors associated with wildfires must be properly analyzed. To overcome this limitation, in the present paper we use geographically weighted regression techniques (GWR) (Fotheringham, Brunsdon, & Charlton, 2002), which allow us to incorporate in the models the spatial variation of the explanatory variables, in a way similar to Martínez and Koutsias (2013) but focusing exclusively in the human influence on wildfire ignitions. Examples of the application of GWR to a number of subjects are found in Cardozo, García-Palomares, and Gutiérrez (2012), Chalkias et al. (2013), Chi, Grigsby-Toussaint, Bradford, and Choi (2013), Li, Heap, Potter, and Daniell (2011), Lu, Charlton, and Fotheringhama (2011), Tu (2011), Su and Zhang (2012), Wang, Zhang, and Li (2013) and Xiao et al. (2013); GWR is applied specifically to the occurrence of forest fires in Chuvieco et al. (2012), Koutsias, Chuvieco, and Allgöwer (2005), Martínez and Koutsias (2011), Martinez et al. (2013), and Rodrigues and de la Riva (2012). In this context, we apply binary logistic regression, commonly used for probabilistic explanation of human-caused occurrence (Chuvieco et al., 2010; Martínez, Chuvieco, et al., 2004; Vasconcelos, Silva, Tomé, Alvim, & Pereira, 2001; Vega-Garcia, Woodard, Titus, Adamowicz, & Lee, 1995), but within the framework of GWR models.

Therefore, the aim of this paper was to model and analyze, using GWLR techniques, the spatial variation in the human factors associated with forest fires. Our hypothesis is that the explanatory factors for human wildfires are not-stationary, rather their relationship with fires changes significantly over the space. The fit of GWLR models (geographically weighted logistic regression) required the statistical analysis and spatialization both of the historical occurrence (in the period 1988–2007) and of a large number of explanatory variables, selected based on experience of models at regional and national scales (Chuvieco et al., 2010; Martinez et al., 2009; Vilar del Hoyo, Martín Isabel, & Martínez Vega, 2008). Ignition data was retrieved from the General Statistics of Wildfires database (EGIF), one of the oldest 'complete' wildfire databases in Europe, beginning in 1968 (Vélez, 2001). The EGIF database registers information about several parameters related with fire ignition such as location, cause, date, size or affected vegetation. The explanatory variables were derived from spatial datasets and statistical data obtained from official data sources of the Spanish Government, later explained in detail. Model adjustment was carried out using a random sample of 60% of the ignition data, reserving the remaining 40% for the validation process. Additionally, an alternative validation sample constructed from the occurrence in the period 2008–2011 was used in the validation process to test the predictive capacity of the model.

This work was developed within the framework of the FIREGLOBE project (www.fireglobe.es, Chuvieco et al., 2011, 2012). In following sections, we describe the method used for modeling the spatial variation of the explanatory factors, the main results of the application of the methodology to peninsular Spain, the degree of fit of the model, and the results of the validation process. A comparison of the performance of GWR and global models, and of our work and similar studies is also conducted. Finally, we present our conclusions and suggestions for further research.

Materials and methods

The methodology for modeling human causality in forest fires is based on GWLR techniques. Specifically, we used the GWR 3.0 software developed by the NCG (Fotheringham et al., 2002). Like global logistic regression models (GLR), GWLR are statistical models that provide insights into the relationship between a qualitative dependent variable, dichotomous in our case, and one or more independent explanatory variables, whether qualitative or quantitative. Therefore, its development requires on the one hand a binary dependent variable, in this case the high/low occurrence of fires, and secondly a set of predictor variables, which are listed below. Fig. 1 shows a schematic of the workflow followed for modeling human causality.

Overview of GWLR

GWR techniques extend the traditional use of global regression models, allowing calculation of local regression parameters. Taking as a starting point the typical equation of the logistic regression:

$$y_i = \frac{e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})}}{1+e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})}} \tag{1}$$

the mathematical expression of its geographically weighted version is:

$$y_{(u_i,v_i)} = \frac{e^{(\beta_0(u_i,v_i)+\beta_1(u_i,v_i)x_{1i}+\cdots+\beta_k(u_{ki},v_{ki})x_{ki})}}{1+e^{(\beta_0(u_i,v_i)+\beta_1(u_i,v_i)x_{1i}+\cdots+\beta_k(u_{ki},v_{ki})x_{ki})}} \tag{2}$$

where (u_i,v_i) are the location coordinates in space of point i.

Accordingly, the use of GWLR models allows one to obtain regression coefficients whose values vary spatially, thus obtaining a different set of regression coefficients for each location in the study area. To do this, a regression model is adjusted for each point and its nearest neighbors. The influence of the points in this neighborhood varies according to the distance to the central point (Fotheringham et al., 2002). The optimum distance threshold (also known as the bandwidth) or the optimum number of neighbors is determined in two ways: by minimizing the square of the residuals (cross-Validation, Cleveland, 1979) or by minimizing the Akaike Information Criterion (AIC, adapted for GWR by Hurvich, Simonoff, & Tsai, 1998).