ARTICLES

Boletín de la A.G.E. N.º 47 - 2008, págs. 357-361

APPLICATION OF LOGISTIC REGRESSION TECHNIQUES TO OBTAIN HUMAN WILDFIRE RISK MODELS AT REGIONAL SCALE

Lara Vilar del Hoyo
Mª Pilar Martín Isabel
Javier Martínez Vega
Instituto de Economía y Geografía (IEG). Centro de Ciencias Humanas y Sociales. CSIC

I. INTRODUCTION

Wildfires in Spain are related to Mediterranean meteorological conditions but also to the human activity, which is responsible of more than 96% of fires (DGB, 2006). Traditionally, Mediterranean wildfires have had an ecological significance, to the point in which some ecosystems can be explained by wildfire occurrence. However, in the last decades the historical equilibrium that explains the relationship between fires and Mediterranean ecosystems has been broken (Martínez *et al*, 2004). Some authors have found that the fire occurrence is less severe in southern African Mediterranean countries than in the northern Mediterranean counterpart, even though the former have the same or worse climate conditions and the same kind of vegetation. Therefore, these differences could be explained because of the different socio-economic conditions (Estirado and Molina, 2005). In recent times major changes have taken place in the European area. Socioeconomic, cultural and political changes have brought important economic, production and social transformations in rural areas (Moyano, 2006). Although 52% of Spain's surface is forested (MAPA, 2004) its contribution to the national economy is only 0.15% (Vélez, 2005). In addition, the spread of residences into rural areas has increased the wildland urban interface. Besides, there are new uses for forested areas, such as recreational activities (Izquierdo, 2006). These changes imply new ecological problems and most of all (despite the investment in fire management) an increase in fire risk. Therefore, it is necessary to improve fire prevention tasks. For this purpose, it is essential to know the fire causes. Fire records, which have been compiled in Spain since 1968, show that over 90% of fires are human related. These fire records classify the typology of fire causes into *lightning*, *negligence*, *deliberate*, *others* and *unknown*. The motivation behind deliberate fires, which account for over 59% of the total from 1991 to 2004, is also recorded. In contrast, less than 4% of fires were due to natural causes (lightning) during the same period (DGB, 2006).

Given the serious consequences of wildfires, it is important to study the wildfire risk in order to improve the prevention systems or actions. One of the most complete approaches to the analysis of fire risk includes three components: ignition, behaviour and vulnerability (Chuvieco *et al*, 2004). This is the conceptual framework of the *Firemap* project 'Integrated Analysis of Wildland Fire with Remote Sensing and GIS' (CGL2004-06049-C04-02/CLI). The outcome of *Firemap* will be a fire risk index which will integrate the human and natural factors related to fire ignition. This paper focuses on the analysis of the human factors.

The human component in fire risk is difficult to model since it involves attempting to quantify and to map human behaviour. However, it is possible to make some relevant approximations based on the analysis of explanatory variables that allow us to represent socioeconomic factors. These factors have a direct or indirect influence on fire occurrence and they are related to (Leone *et al,* 2003):

— socioeconomic changes
— traditional activities in rural areas
— accidents or negligence
— fire prevention activities
— deliberate fires because of conflicts

A wildfire risk model can be generated by analysing these factors. First, it is necessary to obtain input explanatory variables from cartographic or statistical sources. GIS tools are widely used to integrate the spatial information for wildfire risk analysis, and various authors (Chuvieco and Salas, 1994, 1996; Castro and Chuvieco, 1998; Gouma and Chronopoulou-Sereli, 1998; Pew and Larsen, 2001; Cardille *et al.,* 2001) have applied GIS to work on predictive models that explain the phenomenon.

The present work examines how to obtain human wildfire risk models from socioeconomic explanatory variables. Logistic regression techniques are used to generate predictive models at 1 km^2 grid level in the regions of Madrid and Valencia. The aim is to integrate these models in a complex risk system that includes other factors (vegetation and climate) related to fire occurrence. The specific objectives are:

— to identify and to represent the explanatory variables from each factor
— to define the variable response (fire occurrence due to human causes)
— to apply a statistical exploratory analysis of the explanatory variables
— to propose, generate and validate a logistic regression model

II. THE STUDY AREAS

The study areas are the regions of Madrid and Valencia. The former is located in central Spain and although it represents only 1.6% of the nation's surface, it is one of the most densely populated areas with more than 6 million inhabitants. Urban areas have increased in the last decades, spreading into agricultural and forested areas. The contact boundary between urban and forest areas (wildland urban interface) is one of the main concerns for fire managers. Despite the large urban sprawl, Madrid still keeps an important network of protected areas which have an intensive recreational use, and are therefore, very vulnerable to forest fires. The main causes of fire in the region of Madrid from 1990 to 2004 were attributed to *unknown* causes followed by negligence.

The region of Valencia suffers wildland fires every year due to its extreme climate conditions, rugged terrain and human pressure. It has a Mediterranean climate where severe summer droughts are followed by a maximum rainfall during the autumn. For the fire phenomenon it is very important to take into account the wind factor (Ferrando, 2004). Wildfires in the last 20 years have decreased the tree formations which have been replaced by more inflammable shrubland. Society has been using fires as a traditional tool, mainly to dispose of agricultural waste. Tourism in this region is very important, and many visitors arrive every year. From 1990 to 2004 the main causes of fire were attributed to *negligence*, followed by *deliberate* and *lightning*. It is remarkable the low percentage of *unknown* fires, which are under 10% since 1995.

III. METHODS

To obtain the predictive models first of all it is necessary to generate the explanatory variables which represent the human factors in space. Explanatory variables are structural variables, related to permanent elements in the territory. For each human factor group, explanatory variables have been spatially mapped from cartographic or statistical sources using GIS tools, and represented on a 1km^2 UTM grid which has been considered appropriated by fire experts to be operationally useful for fire management at the regional level.

The response variable (fire occurrence caused by human activities) has been obtained from the Spanish fire database (DGB, Ministry of Environment). In this database, the spatial location of fire records are entered both at municipal and 10×10 km grid level. Therefore, since the exact position (x, y coordinates) of the fire ignition points remains unknown, a methodology has been applied in this work to reduce its inaccuracy. This method is based on the kernel density estimation approach applied by Amatulli et al (2007) to map lightning/human-caused wildfires under ignition point location uncertainty. The final goal here is to have an approximation on fire location at the 1 km^2 grid resolution selected for the model. An additional problem in the region of Madrid was the high percentage of unknown fires. To avoid the loss of information, unknown fires in Madrid were proportionally assigned to human or natural causes.

Logistic Regression techniques require a dichotomous response variable (0, 1 values). For that reason, the previous response variable has been transformed into a dichotomous one. The continuous variable has been ranged and divided into three groups with the same number of records. The records in the first group (low fire occurrence) have been assigned as zero, and the records in the last group (high fire occurrence) have been assigned as one.

The result of the logistic regression indicates the probability of fire occurrence and also the relationship between the response and the explanatory variables. It is defined as follows:

$$P_i = \frac{1}{1 + e^{-z}}$$

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_\rho X_\rho$$

Where P_i is the probability of a fire occurring in a grid cell, z is the linear combination of the explanatory variables weighted by their regression coefficients (β) and X the value of and independent variable for any cell (Afifi and Clark, 1990 in Pew and Larsen, 2001; McGrew and Monroe, 1993).

Logistic Regression techniques assume that, when irrelevant or highly intercorrelated explanatory variables are present, the model cannot distinguish which part of the response variable is explained by each explanatory variable (Villagarcía, 2006). This is what is known as multicolineality, which has been avoided by previously carrying out an explanatory analysis aimed at excluding unnecessary variables.

The *Wald step forward* Logistic Regression method was applied. A random sample of 60% of the grid cells was used to generate the model and the remaining 40 % was used to validate it. The model was then applied on 100% of the sample, thus obtaining the probability for each grid cell. Cells with more than 50% of urban area were not included in the analysis.

IV. RESULTS

Out of the 17 models obtained in the region of Madrid, number seven was selected for having the best balance between the level of complexity (number of independent variables) and the prediction ability. The percentage of correct predictions in the final model was 70.6%, of which 75.4% was for low fire occurrence and 65.7% for high fire occurrence. In the region of Valencia the total percentage of correct predictions in the model was 68.4%, where 79.4% was for low fire occurrence and 57.4% for high fire occurrence. In the region of Madrid the variables with the highest influence in the model were the *Wildland Urban Interface* and the *Natural Protected Areas*, followed by *unemployment rate* and *buffer of paths*. In the region of Valencia variables with the highest influence were the *variation of the population* and the *demographic potential*.

Over and underestimations on the prediction of fire occurrence were mapped. In the region of Madrid underestimations are located in the north, north-east and south-east (*Sierra de Madrid*, *Alcalá de Henares* and *Aranjuez*). The overestimation is located in the central and south-western areas. The highest probability values are located to the west (*Sierra de Madrid*) and the south-east (a Natural Protected Area called *Parque Regional del Sureste*). The lowest probability values are located in the central and eastern areas. In the region of Valencia the underestimation areas are in the North and in the South and also in some West areas. The high fire occurrence is well predicted in the South-East area close to *Alicante*. The highest probability values are located in the areas close to the Mediterranean Sea, matching up with the most inhabited areas. The lowest fire probability is located inside the region.

V. DISCUSSION AND CONCLUSIONS

The correct prediction percentage of the fire occurrence is near 70% in both study areas. Low fire occurrence is better predicted also in both Valencia (57.4%) and especially in Madrid (76.4%).

Experienced fire managers agree with results concerning the explanatory variables obtained from the model in Madrid. The fire phenomenon in this region is related to negligence or accidents that happen in the Wildland Urban Interface (WUI), recreational areas, roads, etc. The results of the model match the characteristics of these land uses. In the region of Valencia the explanatory variables that explain the phenomenon are related to the population. Other important factors are changes in the forested area index and the pasture-urban interface. In the region of Valencia, as well as in other Mediterranean areas, there has been a major urban sprawl mainly boosted by tourism, which is more intense during the summer, when extreme meteorological conditions favour fire ignition. The highest probability values were predicted in the most populated areas. However, the model does not consider the WUI variable to be relevant. The results of prediction are less accurate than in the region of Madrid, maybe due to the bigger size of the region of Valencia, as well as its greater socio-economic differences. It could be done regional sub models for this area.

By applying logistic regression, different kinds of explanatory variables can be included. The calculation of the response variable implies the uncertainty of the location of the fire ignition points and a continuous surface of fire density applied. This procedure could influence the final results. It would be interesting compare this information with the real fire ignition points and also to have more detailed spatial information about some explanatory variables. In any case it is useful to obtain results of fire probability in order to integrate them in a general risk system as the one proposed in the *Firemap* project.

Despite its limitations, this technique can be applied to different study areas so long as the socioeconomic variables fit the factors related to fire occurrence. Obtaining prediction models for fire occurrence could be very useful for the fire risk managers since it allows identifying high fire occurrence areas and what kinds of variables have an influence on the wildfire. This work shows the importance of land use distribution and how the fire phenomenon is influenced by human activity. Finally, it shows the relevance of including socioeconomic factors in fire risk prevention systems in general.