Environmental Modelling & Software 29 (2012) 44–50

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

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

Using model-based geostatistics to predict lightning-caused wildfires

C. Ordóñez [a,*], A. Saavedra [b], J.R. Rodríguez-Pérez [c], F. Castedo-Dorado [d], E. Covián [a]

[a] Polytechnic School of Mieres, University of Oviedo, Campus de Mieres, C/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Spain
[b] Department of Statistics, University of Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain
[c] Geomatics Engineering Group, University of Leon, Campus de Ponferrada, Avda. de Astorga s/n, 24400 Ponferrada, Spain
[d] Department of Enginneering and Agricultural Sciences, University of León, ESTIA, Avda. Astorga s/n, 24400 Ponferrada, Spain

ARTICLE INFO

Article history:
Received 11 January 2011
Received in revised form
28 July 2011
Accepted 14 October 2011
Available online 17 November 2011

Keywords:
Generalized linear spatial model
Markov-chain Monte Carlo
Spatial correlation
geostatistics
wildfire

ABSTRACT

The probability of fire in a particular area depends on a range of environmental and geographic variables. Fire prevention planning can be assisted by the construction of models to identify the variables that have a significant influence on the occurrence of fires and by building maps showing the spatial probability distribution for fires occurring in specific geographic areas. We used generalized spatial linear models to predict spatially distributed probabilities for fire occurrence in locations where storms featuring lightning occurred, on the basis of a set of variables related to climatology, orography, vegetation and lightning characteristics, and to assess the relative importance of these variables. A comparison of this model with simple logistic regression models used by other researchers to resolve similar problems demonstrates the importance of bearing in mind spatial correlation between variables.

1. Introduction

Lightning is the main natural cause of ignition in woodlands all over the world (Pyne et al., 1996) and is even the main cause of ignition, whether natural or otherwise, in boreal forests (Johnson, 1992). In Spain and other Mediterranean countries fires caused by humans are, by far, the most common kind of fire (Vázquez and Moreno, 1998; Loepfe et al., 2011) and this is probably the reason why not much attention has been paid to forest fires caused by lightning. Nonetheless, lightning-caused fires can burn larger areas of forest than human-caused fires because of remoteness and aggregation in time and space (Podur et al., 2003).

The probability of fire depends on a range of environmental and geographic variables. Díaz-Avalos et al. (2001) have indicated that topographic variables such as terrain altitude, gradient and orientation have a bearing on the probability of ignition. The type of vegetation is another factor that affects ignition (Dissing and Verbyla, 2003). Dlamini (2008) used a Bayesian belief network to predict wildfire from variables such as land cover, temperature, elevation, rainfall, aspect, slope, among others. Land cover, elevation, and climate (mean annual rainfall and mean annual temperature) were found to be strong predictors of wildfire occurrence, while aspect had the least influence on the wildfire occurrence.

Lightning characteristics (frequency of strikes, polarity, multiplicity, etc.) also affect the occurrence of fires in wooded areas (Flannigan and Wotton, 1991). Weather characteristics such as rainfall, relative humidity and temperature are other factors that have a bearing on ignition (Hely et al., 2001). Most of the variables that affect the occurrence of lightning-ignited fires are spatially correlated; consequently, the application of models reflecting the spatial component is likely to improve not only interpretation of the influence of known variables but also the prediction of the probability of a fire. However, most research into wildfire prediction relies on models that fail to take into account the spatial nature of the explanatory variables (Martell et al., 1989; Vega-García et al., 1995; Perestrello de Vasconcelos et al., 2001; Wotton and Martell, 2005; Catry et al., 2008; Martínez et al., 2009).

In this article we analyze the applicability of model-based geostatistics to assessing the probability of fires using the generalized linear spatial model (GLSM). The term 'model-based geostatistics' was coined by Diggle et al. (1998), who extended this geostatistical approach to situations in which stochastic variation in the data is known to be non-Gaussian. These authors described an approach to geostatistical problems based on formal statistical models and inferential procedures. Other authors that have studied

* Corresponding author. Tel.: +34 985 458 027; fax: +34 985 458 001.
E-mail addresses: ordonezcelestino@uniovi.es (C. Ordóñez), saavedra@uvigo.es (A. Saavedra), jr.rodriguez@unileon.es (J.R. Rodríguez-Pérez), fcasd@unileon.es (F. Castedo-Dorado), covianenrique@uniovi.es (E. Covián).

1364-8152/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2011.10.004

this model are Diggle et al. (1998), Christensen and Waagepetersen (2002); Diggle et al. (2002, 2003); and Zhang (2002, 2003). Christensen (2004) described algorithms for estimating the parameters involved in the GLSM using Markov-chain Monte Carlo (MCMC).

The article is laid out as follows: Section 2 provides a brief description of the GLSM and other statistical methods used in our research; Section 3 describes the study data and the formulated model; Section 4 covers the statistical results and analyses; finally, Section 5 contains our conclusions.

2. Statistical analysis methods

2.1. Generalized linear spatial models

The notion of the generalized linear model (GLM) was developed by Nelder and Wedderburn (1972) and discussed in detail in McCullagh and Nelder (1983). In the GLM, the response variables $Y_1, Y_2, \ldots Y_n$ are assumed to be mutually independent, with expectation related to a linear predictor $E[Y] = g^{-1}(dT\beta)$, where $\beta \in \Re^p$ is a vector of unknown regression parameters, d are known dependent covariates and g is a known function called the link function. The link function relates the linear predictor and the mean of the distribution function. Many different link functions are available, but one should be chosen that matches the domain of the link function to the range of the distribution function mean. Thus, if the response variable follows a normal distribution, identity is typically used as the link function: $E[Y] = d^T\beta$. For Poisson distributions, g is the ln function, hence

$$g(E[Y]) = \ln(E[Y]) = d^T\beta;\ E[Y] = \exp\left(d^T\beta\right) \tag{1}$$

The logit function is usually used for binomial distributions:

$$g(E[Y]) = \ln\left(\frac{E[Y]}{1-E[Y]}\right) = d^T\beta;\ E[Y] = \frac{\exp(d^T\beta)}{1+\exp(d^T\beta)} \tag{2}$$

An important extension of the GLM is the generalized linear mixed model (GLMM) (Breslow and Clayton, 1993), in which the response variables are assumed to be mutually independent conditionally on the realized values of a set of latent variables. The GLSM (Diggle and Ribeiro, 2007) is a GLMM in which the latent variables are derived from a spatial process. This leads to the following model specification:

Consider n distinct locations $\{x_1, \ldots x_n\} \subset I \subset \Re^2$ and suppose we observe a realization $y = (y_1, \ldots y_n)^T$ of $Y = (Y_1, \ldots Y_n)^T$, where $Y_i = Y(x_i)$. Let $S = \{S(x): x \in I\}$, denote a Gaussian random field with mean function $E[S(x)] = d(x)^T\beta$ and covariance $\mathrm{cov}(S(x), S(x')) = \sigma^2\rho(x, x'; \varphi) + \tau^2 1\{x = x'\}$, where $\beta \in \Re^p$ is, as in the GLM, a vector of unknown regression parameters, $d(x)$ are known location-dependent covariates, $\rho(x, x'; \varphi)$ is a correlation function in $\Re^2$, φ is a scale parameter directly related to the range of the correlation function that indicates the distances from which the spatial correlation can be considered null, and $\tau^2 \geq 0$ represents what is called the nugget effect in the geostatistical context.

The Gaussian process $S(x)$ can be understood as the residual spatial process remaining after eliminating from the response variable $Y(x)$ the effect of the explanatory variables. Thus, for a fixed $x \in I$ the one-dimensional variable $[Y(x)|S]$ has a distribution that only depends on the conditional mean $E[Y(x_i)|S(x_i)]$. Moreover, the response variable $\{Y(x), x \in I\}$, conditionally on the random field, S, is formed of random, mutually independent variables. Finally, $S(x)$ and the conditional mean are related by means of a known link function in such a way that $E[Y(x_i)|S(x_i)] = g^{-1}(S(x_i))$, where g is, as in the GLM, the link function.

We assume that the errors follow a binomial distribution. This kind of distribution has been previously used by Zhang (2002) and Diggle et al. (2002). A class of transformations that could be used as link functions for this distribution was described by De Oliveira et al. (1997).

In interpreting the regression parameters β of the GLSM, it should be remembered that a direct interpretation of these coefficients is only possible when the link function is identity and the error distribution $Y(x)|S$ is Gaussian. In the remaining cases, the parameters β have a conditional and non-marginal interpretation. See, for example, Diggle et al. (1994) for further details.

Since it is not possible to directly observe the process $S(x,)$ we cannot obtain a closed expression for the likelihood function from which to estimate the parameters. The implementation of algorithms based on MCMC, as suggested by Diggle et al. (1998), enables the parameters for the GLSM to be approximated within a Bayesian framework with vague priors. MCMC covers a range of simulation methods that are particularly useful for simulating observations from multivariate distributions or distributions whose density function is analytically complicated, as occurs with the likelihood function in the GLSM. These methods are based on establishing a Markov chain with the multivariate distribution of interest as the equilibrium distribution, in such a way that the probabilities of transition of the chain are analytically manageable. By performing a sufficiently large number of Markov chain iterations—called burn-in—we obtain a realization of the equilibrium distribution of interest. Continuing with the iterations, we can obtain a sample that is almost independent of the distribution, provided that the realizations of this sample are separated by a sufficiently large number of Markov chain iterations. The analysis described in this article is based on the work of Christensen (2004) and was performed using the geoRglm package that is freely available within the framework of the open-source R statistical system (R Development Core Team, 2010).

2.2. Model fit evaluation: ROC curves

A binary classification system can be used to calculate receiver operating characteristic (ROC) curves and determine the precision of a diagnostic test. ROC curves represent true positive versus false positive ratios for different values of the discrimination threshold (Zhou et al., 2002). By calculating the area under the ROC curve (AUC) associated with a binary classifier, we can compare models and select explanatory variables to be included in the model. This approach was used by Rodríguez-Álvarez et al. (2010), who described a bootstrap-based method for testing for a significant factor effect on the ROC curve. ROC curves can therefore be used to select an optimal model from among several GLMs with a binomial error distribution. The AUC can also be used to calculate the goodness-of-fit of the GLSM compared to similar models without a spatial component (i.e., GLMs).

3. Data description and model formulation

3.1. The study area and the geographic database

The study area, located in the province of Leon in northwest Spain (Fig. 1), measured 15,590 km^2. It was divided into cells measuring 3 × 3 km, resulting in a total of 1882 cells. The cell size was chosen on the basis of accuracy in the locations of the ignition and on the assumption that weather and fuel conditions would be homogeneous within a cell.

Data on the ignition locations (approximate UTM coordinates) for the years 2002 − 2007 were supplied by the Spanish Ministry of the Environment. The 1882 cells into which the province of Leon was divided were coded for occurrence of lightning-caused ignitions during the study period: 0 for non-occurrence and 1 for occurrence. Only major fires, i.e. those requiring intervention, were considered.

Lightning locations and data on intensity, polarity and date of occurrence were provided by the Spanish Meteorology Agency. A total of 78,256 flashes were recorded; fewer than 5% had negative polarity.

Topographic variables such as average slope, altitude and aspect were obtained from a digital elevation model with a pixel resolution of 50 m, obtained from the Spanish.

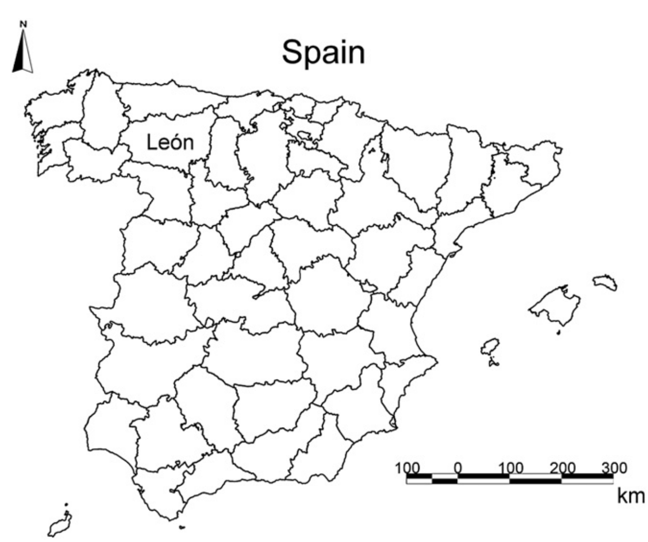

Fig. 1. Location of the study area: province of Leon (Spain).

Cartographic Database using the ArcGIS 9.2 spatial analysis module, a computer program developed by ESRI (Environmental System Research Institute). Data on land cover was obtained from the forestry map for 2003 for the province of Leon, supplied by the Spanish Ministry of the Environment. The different types of land cover were categorized in 7 classes.

Daily meteorological data for each cell were derived from raw data collected at 344 weather stations located in Leon and nearby provinces. The data, provided by the Spanish Meteorology Agency, consisted of daily observations of temperature, relative humidity and rainfall. Rainfall and relative humidity for each 3×3 km cell were estimated using universal kriging. Temperature for each cell was estimated by cokriging interpolation, with elevation used as a secondary variable given that it is highly correlated with temperature.

Spanish Meteorology Agency data were considered error free since this agency does not provide any additional information concerning the quality of the data. This is obviously a problem when estimating the goodness of the results. This problem becomes worse in the case of the variables estimated using kriging because observation errors are propagated in the interpolation process. A possible solution would be to perform a sensitivity analysis (Yang, 2011) assuming that the explanatory variables have associated some kind of uncertainty which can be model using random distribution functions. However, this is not affordable from a computational point of view given the time required to run the algorithm each time.

Kriging interpolation has been carried out in a classic way but in future research it will be done using a hierarchical spatial model (Banerjee et al., 2004). In this model, the spatial error component is added to the media, leaving in the covariance matrix the residual error (nugget).

3.2. Model formulation

The response variable, namely, the occurrence of fire, follows a binomial distribution with a spatial dependence structure that can be modeled using a GLSM. According to this model, it is assumed that conditionally on a stationary Gaussian process, $S(x)$, the response variables Y_i, $I = 1, \dots n$, are modeled as independent binomial variables. The parameters involved in this binary GLSM are $\theta = (\sigma^2, \varphi, \tau^2, \beta)$, with $\beta = (\beta_0, \dots \beta_p)$, where β_0 is the intercept and $\beta_1, \dots \beta_p$ are the regression coefficient corresponding to the i variables or factors.

It is assumed that $S(.)$ is a stationary Gaussian process that determines the spatial variation of the probability of a fire occurring in a specific location x, in such a way that $P(x) = E[Y(x)|S(x)]$. A logit function was used as the link function, as this kind of function is typical in binary GLSMs, as it establishes a relationship between $S(x)$ and the conditional mean $P(x)$; thus:

$$g(P(x)) = \ln\left(\frac{P(x)}{1 - P(x)}\right) = S(x) \tag{3}$$

To infer the parameters of the model we implemented a method based on MCMC. Within the Bayesian framework it can be considered that the vector of parameters, θ, is conditionally independent of Y given S. Considering a realization $y = (y_1, \dots y_n)^T$ of $(Y_1, \dots Y_n)^T$, the likelihood function to be maximized is as follows:

$$L(\theta) = f(y|\theta) = \int f(y|s)f(s|\theta)\,ds = \int \frac{f(y|s)f(s|\theta)}{\tilde{f}(y,s)}\tilde{f}(y,s)ds$$
$$\propto \int \frac{f(y|s)f(s|\theta)}{f(y|s)\tilde{f}(s)}\tilde{f}(s|y)ds = \tilde{E}\left[\frac{f(S|\theta)}{\tilde{f}(S)}|y\right] \approx \frac{1}{m}\sum_{j=1}^{m}\frac{f(s(j)|\theta)}{\tilde{f}(s(j))} \tag{4}$$

where $\tilde{f}(y,s) = f(y|s)\tilde{f}(s)$, $\tilde{f}(s) = f(s|\theta_0)$, θ_0 is a vector of initial parameters and $\tilde{f}(s|y) \propto f(y|s)\tilde{f}(s)$ and $\tilde{E}[\cdot|y]$ denote expectation with respect to $\tilde{f}(\cdot|y)$.

Briefly, the algorithm functions as follows:

1. Select an initial value for θ_0. If $S \equiv 0$ is assumed in this first step, the parameters estimated for the GLM can be used as θ_0.
2. Obtain values for $s^{(0)}(1), \dots s^{(0)}(m)$, sampled by an MCMC from the distribution $\tilde{f}(\cdot|y)$. The initial K samples are discarded as burn-in until it is judged that the equilibrium distribution of interest has been achieved. The subsequent realizations are used to obtain the $s^{(0)}(j)$. Furthermore, to build a sample that is almost independent of the distribution, rather than storing the first m samples of the simulated chain, each sample is taken at every Lth iteration.
3. Select the θ_1 that maximizes the following function:

$$L_m(\theta) = \frac{1}{m}\sum_{j=1}^{m}\frac{f(s^{(0)}(j)|\theta)}{\tilde{f}(s^{(0)}(j))} \tag{5}$$

4. Repeat steps 2 and 3 until $\theta_i \approx \theta_{i+1}$.

The likelihood function $L(\theta)$ can be approximated by using the MCMC to simulate samples from $\tilde{f}(\cdot|y)$, departing from an initial θ_0. The estimator for θ will be that which maximizes this approximation.

4. Results and discussion

Table 1 summarizes all the variables considered in regard to the construction of the model. The large number of explanatory variables considered prioritized finding a GLM with a high goodness-of-fit using a reduced subset of factors.

Taking the occurrence of fire as the response variable, binary GLM models were fitted to different groups of explanatory

Table 1
Explanatory variables observed in the studied area.

Category	Variable
Topography	Mean altitude (m)
	Mean slope (%)
	Mean aspect (degrees)
Vegetation cover	Percentage of woodland
	Percentage of scrubland
	Percentage of coniferous woodland
	Percentage of broadleaf woodland
	Percentage of mixed woodland
	Percentage of open woodland
	Percentage of forested areas
Lightning	Number of positive strikes
	Number of negative strikes
	Mean intensity of positive strikes (kA)
	Mean intensity of negative strikes (kA)
Meteorology	Number of dry-storm days
	Number of lower-than-average moisture days
	Number of lower-than-average-temperature days
Lightning-vegetation cover interaction	% of strikes in woodland
	% of strikes in scrubland
	% of strikes in coniferous woodland
	% of strikes in broadleaf woodland
	% of strikes in mixed woodland
	% of strikes in open woodland
	% of strikes in forested areas

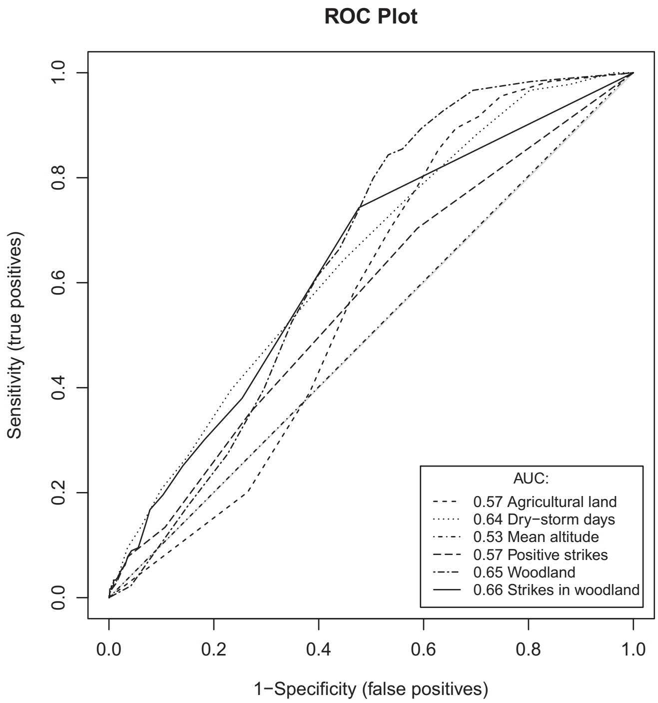

Fig. 2. ROC curves and AUC values for the binary GLM fitted for a single explanatory variable.

variables to locate the minimum number of variables providing a high AUC value. Given the correlation between the explanatory variables, the solution to this problem was not a single subset of factors; rather, several combinations of factors obtained similar AUC values. Finally, only 6 explanatory variables those with an AUC above 0.5 were considered, variables that were highly

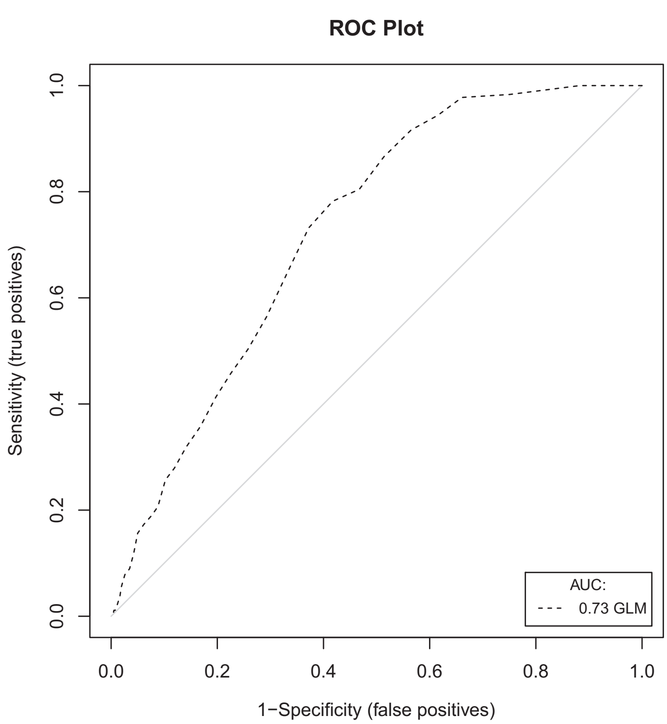

Fig. 3. ROC curve and AUC value for the GLM fitted for 6 explanatory variables.

Table 2
Estimated coefficients and 2.5% and 97.5% quantiles for the GLM model.

	Coefficient	2.5% quantile	97.5% quantile
β_0 (intercept)	−2.624	−4.699	−1.381
β_1 (agriculture land)	1.577E-05	−1.34E-03	2.45E-03
β_2 (dry-storm days)	1.013E-01	5.21E-02	1.68E-01
β_3 (mean altitude)	−1.924E-03	−2.47E-03	−1.29E-03
β_4 (positive strike)	1.816E-02	−1.28E-01	1.44E-01
β_5 (woodland area)	2.799E-03	1.42E-03	5.18E-03
β_6 (broadleaf wood)	9.906E-03	−3.20E-03	2.24E-02

correlated with these 6 variables were discarded. These variables were:

1. Percentage of agricultural land.
2. Number of dry-storm days (days with accumulated rainfall below 2.5 mm).
3. Mean altitude of the cell (m).
4. Number of positive strikes in the cell.
5. Percentage of woodland.
6. Number of strikes in broadleaf woodland.

Fig. 2 shows the ROC curves and corresponding AUC values for the binary GLM fitted for a single explanatory variable; Fig. 3 includes the ROC curve and AUC value of 0.73 for the GLM fitted for the 6 variables referred to above.

Once the explanatory variables were selected, a binary non-spatial GLM model was fitted using Bayesian methods and the MCMClogit function from the MCMCpack (R language). Table 2 shows the estimated coefficients, and the 2.5% and 97.5% quantiles for the fitted model. It can be observed that, for a significance level of $\alpha = 0.05$, only 3 variables are significant: number of dry-storm days, mean altitude and percentage of woodland area.

Bearing in mid the special nature of the variables, a GLSM was fitted keeping the 5 explanatory variables. The correlation function considered was of the type $\mathrm{cov}\,(S\,(x), S\,(x')) = \sigma^2 \rho\,(x, x'; \varphi) + \tau^2 1\ \{x = x'\}$, with $\rho\ (x, x';\ \varphi) = \exp\ (-//x{-}x'//\varphi)$ as the isotropic exponential model. The fit was performed using Bayesian inference implemented via MCMC algorithms. First, it is necessary to define options for the simulation of the samples using the MCMC algorithm. The first 10,000 samples, where convergence was judged to have occurred, were discarded, and all the subsequent samples were used to obtain the posterior distributions of the parameters of interest. To avoid bias, not every observation for the MCMC algorithm was taken into account, but the chain was sampled for each 100th iteration of 50,000 iterations, yielding a sample of 500 values (see Christensen (2004) for more details). Subsequently, we used the binom.krige.bayes function of the geoRglm package to obtain the parameters that maximize the likelihood approximation to function given in Equation (5).

Table 3
Estimated coefficients and 2.5% and 97.5% quantiles for the GLSM model.

	Coefficient	2.5% quantile	97.5% quantile
β_0 (intercept)	−3.336	−5.725	−1.435
β_1 (agriculture land)	5.858E-04	−1.80E-03	2.90E-03
β_2 (dry-storm days)	1.111E-01	3.79E-02	1.88E-01
β_3 (mean altitude)	−2.352E-03	−3.58E-03	−1.31E-03
β_4 (positive strike)	4.742E-02	−1.04E-01	1.95E-01
β_5 (woodland area)	3.443E-03	1.12E-03	5.59E-03
β_6 (broadleaf wood)	8.460E-03	−1.18E-02	2.64E-02

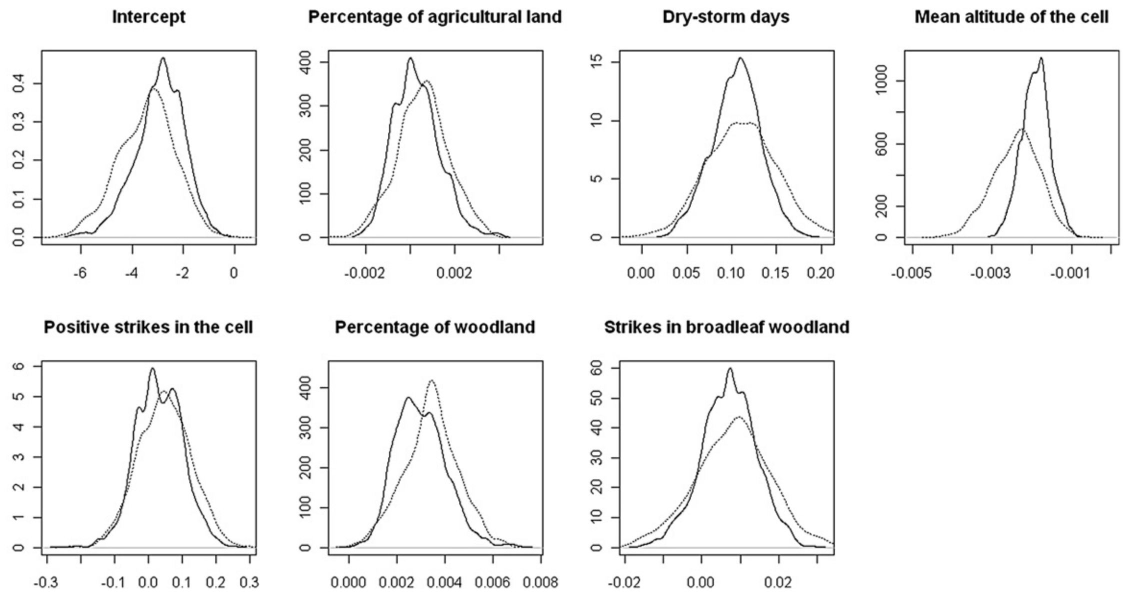

Fig. 4. Density plots for the bi coefficients estimated using the GLM (unbroken line) and the GLSM (broken line).

The estimated parameters were as follows: $\varphi = 6490$, $\sigma^2 = 0.173$, $\tau^2 = 0.732$ and $\beta = (-3.336$, 5.858e-04, 1.111e-01, -2.352e-03, 4.742e-02, 3.443e-03, 8.460e-03).

It is important to remember that β has a conditional rather than a marginal interpretation; furthermore, since the estimation approaches are essentially different, a direct comparison of the spatial and non-spatial GLM is not recommended.

Table 3 shows the values for the coefficients estimated for the GLSM and the 2.5% and 97.5% quantiles. Comparing Tables 2 and 3 it can be observed that, for a significance level of $\alpha = 0.05$, the same significant variables are maintained in the 2 models, namely, number of dry-storm days, mean altitude and percentage of woodland area.

In Fig. 4 we overplotted the density plots for the coefficients estimated using GLM and GLSM. It can be observed that the inclusion of the spatial component in the GLSM method seems to have little influence on the distribution of most of the coefficients.

The ROC curve and AUC value were calculated for the GLSM. The AUC value for the spatial binary model was 0.99, indicating a substantial improvement in precision for the GLSM. This improvement is reflected in Fig. 5, which shows ROC curves and the corresponding AUC values for the binary GLM and GLSM.

Fig. 6 shows the fire probability map obtained using the GLSM. The highest probabilities can be observed to occur in the east and north of the province.

Referring to the sign and significance of the coefficients for each explanatory variable obtained in the GLM and GLSM, it can be deduced from the data in Table 2 and Table 3 that altitude plays a negative role in the probability of lightning igniting a fire. This contrasts with the results obtained by other authors (Rivas et al., 2005). The other topographic variables considered were not associated with the occurrence of fires. The presence of wooded areas was positively related to fire occurrence in both models. This corroborates the results obtained by other researchers (Vázquez and Moreno, 1998).

The number of dry-storm days, significant in both models, was another variable that favored the occurrence of fire. This result corroborates those of other authors (Rorig et al., 2007). The explanation is likely to be that the occurrence of storms on low-rainfall days favors the occurrence and propagation of fires.

The presence of positive lightning strikes or of strikes in broadleaf woodland was not significant in either of the models. This contrasts with the hypothesis of authors such as Latham and Williams (2001), who consider that the presence of fine fuels, such as dead leaves, on the forest floor favor ignition.

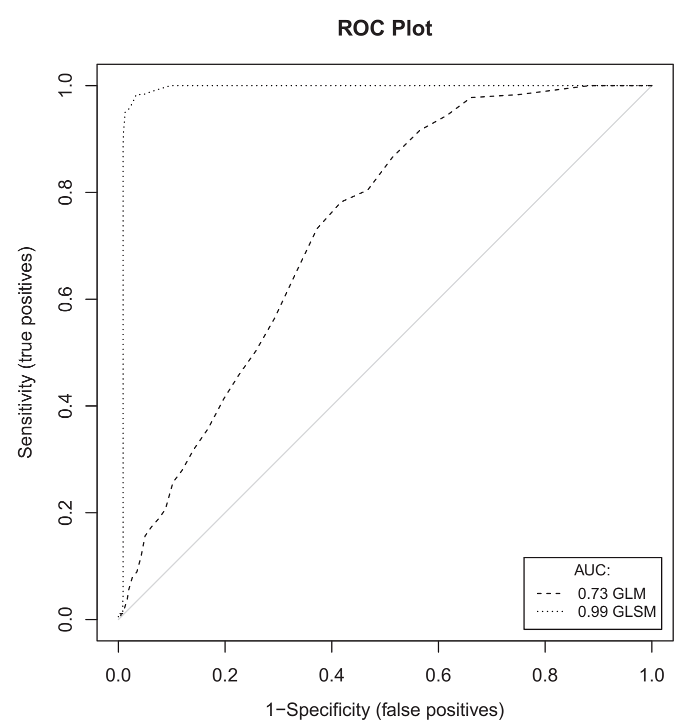

Fig. 5. ROC curves and corresponding AUC values for the binary GLM and GLSM.

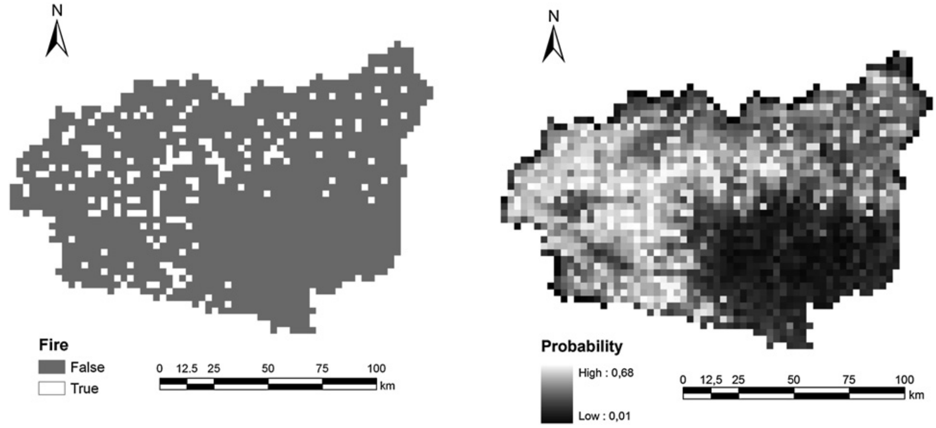

Fig. 6. Locations of fires (left) and fire probability map for lightning-caused fires in the province of Leon in Spain (right).

5. Conclusions

Most research to date into mathematical models that relate the occurrence of lightning-caused fires with physiographic and environmental variables fails to take into account the spatial structure of these variables. A general interpretation for the GLSM used in this work is that the spatial term represents the cumulative effect of unidentified and unobserved spatially structured covariates. In non-spatial GLM, the spatial correlation of residuals is not considered and this can significantly affect the quality of the statistical results.

It was shown that the GLSM fitted the data better than the standard GLM. It appears, therefore, that the GLSM is necessary to take account of any unexplained spatial variation not considered. This statement is justified by a comparison between the ROC curves and AUC values. In the specific problem considered in this article, the AUC for the GSLM is 35% larger than in the GLM model.

The analysis carried out above demonstrates the great potential of GLSM for analyzing spatial data. It also indicates the significant risk of drawing erroneous conclusions when non-spatial models are used for spatially structured data.

Acknowledgments

This work was partly funded by project 10PXIB300068PR of the Xunta de Galicia and by projects MTM2008-03129 and MTM2011-23204 of the Spanish Ministry of Science and Innovation.

References

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC, USA.

Breslow, N., Clayton, D., 1993. Approximate inference in generalized linear mixed models. Journal of American Statistical Association 88, 9–25.

Catry, F.X., Rego, F.C., Moreira, F., Bacao, F., 2008. Characterizing and modelling the spatial patterns of wildfire ignitions in Portugal: fire initiation and resulting burned area. In: de las Heras, J., Brebbia, C.A., Viegas, D., Leone, V. (Eds.), Modelling, Monitoring and Management of Forest Fires. WIT Press, Southampton (UK), pp. 213–224.

Christensen, O.F., Waagepetersen, R., 2002. Bayesian prediction of spatial Count data using generalized linear mixed models. Biometrics 58, 280–286.

Christensen, O.F., 2004. Monte Carlo Maximum likelihood in model-based geostatistics. Journal of Computational and Graphical Statistics 13 (3), 702–718.

De Oliveira, V., Kedem, B., Short, D.A., 1997. Bayesian prediction of Transformed Gaussian random Fields. Journal of the American Statistical Association 92, 1422–1433.

Díaz-Avalos, C., Peterson, D.L., Alvarado, E., Ferguson, S.A., Besag, J.E., 2001. Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. Canadian Journal of Forest Research 31 (9), 1579–1593.

Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 1994. The Analysis of Longitudinal Data. Oxford University Press, Oxford (UK).

Diggle, P.J., Moyeed, R.A., Rowlingson, B., Thomson, M., 2002. Childhood malaria in the Gambia: a case-study in model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 51 (4), 493–506.

Diggle, P.J., Ribeiro Jr., P.J., Christensen, O.F., 2003. An Introduction to model-based geostatistics. In: Møller, J. (Ed.), Spatial Statistics and Computational Methods. Springer Verlag, New York (USA), pp. 43–86.

Diggle, P.J., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer, New York (USA).

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 47 (3), 299–350.

Dissing, D., Verbyla, D.L., 2003. Spatial patterns of lightning strikes in interior Alaska and their relations to elevation and vegetation. Canadian Journal of Forest Research 33 (5), 770–782.

Dlamini, W., 2008. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. Environmental Modelling & Software 25 (2), 199–208.

Flannigan, M.D., Wotton, B.M., 1991. Lightning ignited forest fires in Northwestern Ontario. Canadian Journal of Forest Research 21, 277–287.

Hely, C., Flannigan, M.D., Bergeron, Y., McRae, D., 2001. Role of vegetation and weather on fire behaviour in the Canadian mixedwood boreal forest using two fire behaviour prediction systems. Canadian Journal of Forest Research 31, 430–441.

Johnson, E.A., 1992. Fire and Vegetation Dynamics: Studies from the North American Boreal Forest. Cambridge University Press, Cambridge (UK).

Latham, D., Williams, E., 2001. Lightning and forest fires. In: Johnson, E.A., Miyanishi, K. (Eds.), Forest Fires: Behavior and Ecological Effects. Academic Press, San Diego, California (USA), pp. 375–416.

Loepfe, L., Martínez-Vilalta, J., Piñol, J., 2011. An integrative model of human-influenced fire regimes and landscape dynamics. Environmental Modelling & Software 26 (8), 1028–1040.

Martínez, J., Vega-García, C., Chuvieco, E., 2009. Human-caused wildfire risk rating for prevention planning in Spain. Journal of Environmental Management 90 (2), 1241–1252.

McCullagh, P., Nelder, J.A., 1983. Generalized Linear Models. Chapman and Hall, London (UK).

Martell, D.L., Bevilacqua, E., Stocks, B.J., 1989. Modelling seasonal variation in daily people-caused forest fire occurrence. Canadian Journal of Forest Research 19, 1555–1563.

Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. Journal of the Royal Statistical Society: Series A (General) 135 (3), 370–384.

Perestrello de Vasconcelos, M.J., Silva, S., Tomé, M., Alvim, M., Cardoso, J.M., 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. Photogrammetric Engineering and Remote Sensing 67, 73–81.

Podur, J., Martell, D.L., Csillag, F., 2003. Spatial patterns of lightning-caused forest fires in Ontario, 1976-1998. Ecological Modelling 164, 1–20.

Pyne, S.J., Andrews, P.L., Laven, R.D., 1996. Fire behaviour. In: Pyne, S.J., Andrews, P.L., Laven, R.D. (Eds.), Introduction to Wildland Fire, second ed. John Wiley & Sons, Inc, New York (USA), pp. 46–127.

R Development Core Team, 2010. R: a Language and Environment for Statistical Computing [on line]. R Foundation for Statistical Computing, Vienna (Austria), ISBN 3-900051-07-0 [Ref. Jan 1st, 2011]. Available on: http://www.R-project.org.

Rivas, L., de Pablo, F., Tomás, C., 2005. Ten-year study of cloud-to-ground lightning activity in the Iberian Peninsula. Journal of Atmospheric and Solar-Terrestrial Physics 67 (16), 1632–1639.

Rodríguez-Álvarez, M.X., Roca-Pardiñas, J., Cadarso-Suárez, C., 2010. ROC curve and covariates: extending induced methodology to the non-parametric framework [on line]. In: Statistics and Computing. Springer, New York (USA) [Ref. Jan 1st, 2011]. Available on: http://www.springerlink.com.

Rorig, M.L., McKay, S.J., Ferguson, S.A., Werth, P., 2007. Model-generated predictions of dry thunderstorm potential. Journal of Applied Meteorology and Climatology 46, 605–614.
Vázquez, A., Moreno, J.M., 1998. Patterns of lightning and people caused fires in Peninsular Spain. International Journal of Wildland Fire 8, 103–115.
Vega-García, C., Woodard, P.M., Titus, S.J., Adamowicz, W.L., Lee, B.S., 1995. A logit model for predicting the daily occurrence of human caused forest-Fire. International Journal of Wildland Fire 5, 101–111.
Wotton, B.M., Martell, D.L., 2005. A lightning fire occurrence model for Ontario. Canadian Journal of Forest Research 35 (6), 1389–1401.
Yang, J., 2011. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environmental Modelling & Software 26 (4), 444–457.
Zhang, H., 2002. On estimation and prediction for spatial generalized linear mixed models. Biometrics 58, 129–136.
Zhang, H., 2003. Optimal interpolation and the Appropriateness of Cross-Validating Variogram in spatial generalized linear mixed models. Journal of Computational and Graphical Statistics 12, 698–713.
Zhou, X.H., McClish, D.M., Obuchowski, N.A., 2002. Statistical Methods in Diagnostic Medicine. John Wiley & Sons Inc., Hoboken, New Jersey (USA).