Advances in Forest Fire Research

Domingos Xavier Viegas
Editor

2014

Assessing the effect on fire risk modeling of the uncertainty in the location and cause of forest fires

Marcos Rodrigues[ac], Juan de la Riva[bc]

[a] *University of Zaragoza, Pedro Cerbuna 12 50009 Zaragoza (Spain), rmarcos@unizar.es*
[b] *University of Zaragoza, Pedro Cerbuna 12 50009 Zaragoza (Spain), delariva@unizar.es*
[c] *GEOFOREST-IUCA, University of Zaragoza, http://geoforest.unizar.es/*

Abstract
Wildfire risk assessments in Spain usually make little or no reference to the uncertainty of the results due to ignition data quality, or the implications that this potential uncertainty may have on wildfire management decisions. In Spain the autonomous regions have historically been the competent authorities in forest management and environmental protection as a result of the 1978 Constitution and, therefore, responsible on the operational application of the criteria defined for the country for wildfire classification and location. This competency framework has generated significant regional differences in the application of the criteria for wildfire classification among the different autonomous regions, arising potential uncertainty on wildfire assessments and fire risk models based on this historical series of data. This work explores six scenarios based on the classification of fire ignition causes and location data, reported in the General Statistics of Wildfires database (EGIF), to address the potential uncertainty from the point of view of the variability in predicted ignition probability and the changes in its spatial patterns. The analysis is focused on analyzing the effects on human-caused wildfires by using Random Forest algorithms to predict the ignition likelihood and cluster and outlier analysis (hot and cold spot) to detect changes in the spatial pattern of probability. Results suggest that there is significant uncertainty both in predicted human-caused ignition and spatial pattern related to the ignition source and location of fire events compiled in the EGIF database. The accuracy of the predictions ranges from AUC values of 0.90, when considering most of the records of the database, to around 0.76 in scenarios characterized by using only known-caused allocated fire events, probably due to differences in the proportions of unidentified and allocated fires within the mainland Spain.

***Keywords**: Uncertainty, wildfire, point location, ignition cause*

1. Introduction

During the last decades, the Spanish forest fire authorities have encouraged the investigation of fire causes, which is decisive to better understand patterns of fire occurrence and improve fire prevention measures (Martínez *et al*., 2009). However, the 29% of the fire causes remain unidentified in the period 1988-2007. According to Lovreglio *et al*. (2006), little is known about wildfire causes, which often are more diverse than what is assumed by the traditional classifications employed for statistical purposes. In face of the arising uncertainties, a better knowledge on spatial patterns of fire occurrence and their relationships with its underlying causes becomes a necessity to locate and make prevention efforts more efficient (Martínez *et al*., 2009). From a scientific perspective, improving decision quality in natural resource management begins with uncertainty management (Borchers, 2005). Uncertainty is essentially a lack of information; complete ignorance represents one end of the spectrum and perfect information (i.e., certainty) the other (Thompson and Calkin, 2011). However, viewing uncertainty as ‘information about information’ may be the first step in transforming a problem into knowledge (Bradshaw and Borchers, 2000).
The aim of this paper is to deal with the potential uncertainty linked to location and ignition cause of wildfires, with special attention to the human-caused fires in the mainland Spain. The analysis of human factors in forest fire is widely recognized as very critical for fire danger estimation (Kalabokidis *et al*., 2002; Martínez *et al*., 2009), especially in human-dominated landscapes where anthropogenic

ignitions widely surpass natural ignitions, like the peninsular Spain (Amatulli *et al*., 2007; Chuvieco *et al*., 2010; Chuvieco *et al*., 2012).

In Spain, fire events are recorded in the General Statistics of Wildfires database (EGIF). The EGIF database is one of the oldest 'complete' wildfire databases in Europe, beginning in 1968 (Vélez, 2001), though its data is not considered as completely reliable until 1988 (Martínez *et al*., 2009). The database is compiled by the Ministry of Environment, Rural and Marine affairs (MARM) using forest fire reports of the autonomous regions (Moreno *et al*., 2011). The autonomous regions have received competition in forest fires from the 1978 Constitution (article 148), and therefore are responsible of the application of the criteria and procedures defined for the entire national territory concerning wildfire classification and location. However, the fact that there is no single administration responsible on this topic has led to differences in the application of the criteria among the autonomous regions. A quick overview on the data collected in the historical database arises some inconsistences in the reported information. For instance, the proportion of unknown causes or the proportion of correctly located fire events (located with coordinates) differs from one region to another, becoming a potential source of uncertainty. This is especially important since research on forest fires in Spain is made from data collected in the EGIF database (Amatulli *et al*., 2007; Chuvieco *et al*., 2010; Chuvieco *et al*., 2012; de la Riva *et al*., 2004; Martinez *et al*., 2009; Padilla and Vega-García, 2011; Rodrigues *et al*., 2014; Rodrigues and de la Riva, 2014). Notwithstanding, the influence of uncertainty in historical fire data is scarcely considered (or at least not specifically addressed) and is mainly focused on location precision rather than ignition cause (Amatulli *et al*., 2006; Amatulli *et al*., 2007). Assessing the effects of uncertainty of Spanish ignition data is particularly interesting since it is a component of the wildfire information compiled European Forest Fires System Database (EFFIS), thus analyzing the effects of uncertainty at the Spanish level could be very helpful to understand wildfire patterns in the European scale, even more since Spain is the more fire-affected country within the European Union (Rodrigues *et al*., 2013).

In this work, we will explore six scenarios based on the classification of ignition causes and location data reported in the EGIF database to assess the potential uncertainty from the point of view of the variability in predicted ignition probability and changes in the spatial pattern of probability.

The occurrence probability will be calculated using Random Forest (RF) algorithms (Breiman, 2001) whereas the changes in the spatial probability patterns will be addresses through local Hot Spot analysis. RF algorithms have proved to be a useful tool for wildfire modeling (Bar Massada *et al*., 2012; Rodrigues and de la Riva, 2014), improving the performance of traditional regression techniques (e.g. logit Generalised Linear Models). The comparison of the proposed occurrence scenarios is conducted from the point of view of the accuracy in the classification based on a k-fold procedure (Fielding and Bell, 1997) and according to the variation in variable importance (Breiman, 2001). On the other hand, Hot Spot methods are one of the most adequate for the analysis of large-scale fire occurrence patterns (Allgöwer *et al*., 2005). The analysis of the changes in the predicted ignition probability patterns in each scenario is carried out by cluster and outlier analysis through the Anselin's Local Moran's I (Anselin, 1995).

2. Materials and methods

2.1. Study area and fire data

The study area covered the whole mainland Spain excluding the Balearic and Canary Islands as well as the autonomous cities of Ceuta and Melilla, due to the lack of data in those areas. Thus the total area of the study region was around 498 000 km^2. The fire events considered in this work are those occurred during the period 1988-2007.

2.2. An overview to the EGIF database

The EGIF database is compiled by the Ministry of Environment, Rural and Marine affairs (MARM) using the forest fire reports from the autonomous regions. The database classifies each fire event following a hierarchy of criteria which first differences between known (K) and supposed (S) cause and then into the most likely ignition source (natural or human). In turn, the ignition source is classified according to six categories: natural (lightning; L), human (negligence, accident or arson; H), restarted fires (R) and unknown or unidentified fires (U). Ideally, only K fires should be considered when developing any kind of fire analysis as they appear to be the most reliable. However, an insight into the classification of fire events in terms of number of fires in each category (Table 1) reveals that the proportion of fires with a S cause is more than 73 % of the total number of fires in the period 1988-2007. Hence, by excluding S fires the majority of fire events are being discarded (Figure 1).

Table 1. Classification of fire events according to its ignition causes (number of fires).

	Lightning	Human	Unknown	Restarted	All
Known	6775	35443	30952	1957	75127
Supposed	7931	228694	44706	2420	283751
Total	14706	264137	75658	4377	358878

This classification system also influences the proportion of fires according to its ignition source. Attending to K source, L fires represent the 9% of the occurrence whereas H fires are only the 47%. The remaining fires mostly correspond to U sources. This proportion changes drastically when S cause fires are accounted for, decreasing the proportion of L fires to 4% and raising H fires to a 73%. However, this 73 % of H fires is still far from the 90% value usually reported for Mediterranean European Countries (San-Miguel-Ayanz *et al.*, 2012; San-Miguel-Ayanz, 2009) and, particularly, for Spain (Martínez *et al.*, 2009). This fact suggests that there is great amount of U fires potentially related to H ignition factors and thus, when excluding unknown fires in human-caused wildfire assessments, a significant part of the human occurrence is not taken into account. However, while U fires are quite important attending to national overall values, mapping the spatial distribution of these proportions uncovers the existence of high spatial heterogeneity, increasing the uncertainty on the data (Figure 1). On the other hand, a second source of uncertainty is related to the location of fire events. In the EGIF database wildfires are located following to different procedures: (i) geocoding the location on the basis of a reference 10x10 km ICONA grid (used by the firefighting services for approximate location of fire events) and the municipality origin of the ignition; and (ii) georeferencing fire events using spatial coordinates. Again, the existence of coordinates should imply a precise allocation of the ignition points, however not all the fire events are georeferenced –only the 11% (Table 2)– and, as in the case of the ignition source, the proportion of fire events with coordinates varies from one region to another (Figure 2). This situation usually led to face the spatialization of the fire occurrence using geocoded location information (Amatulli *et al.*, 2007; Chuvieco *et al.*, 2010; Chuvieco *et al.*, 2012; de la Riva *et al.*, 2004; Martínez *et al.*, 2009). On top of this, sometimes the assigned coordinates are incorrect. For instance, 2267 fires are located outside Spain, 23 are assigned a wrong UTM zone and 757 are located in the exact intersection of the ICONA grid (Table 2). This means that the 7.6% of the forest fires with spatial coordinates are mistakenly allocated.

Table 2. Number of fires with coordinates and wrong located wildfires.

	Located	Outside	Wrong zone	Intersects Grid	Total incorrect
Known	16435	962	18	347	1327
Supposed	23581	1305	5	410	1720
Total	40016	2267	23	757	3047

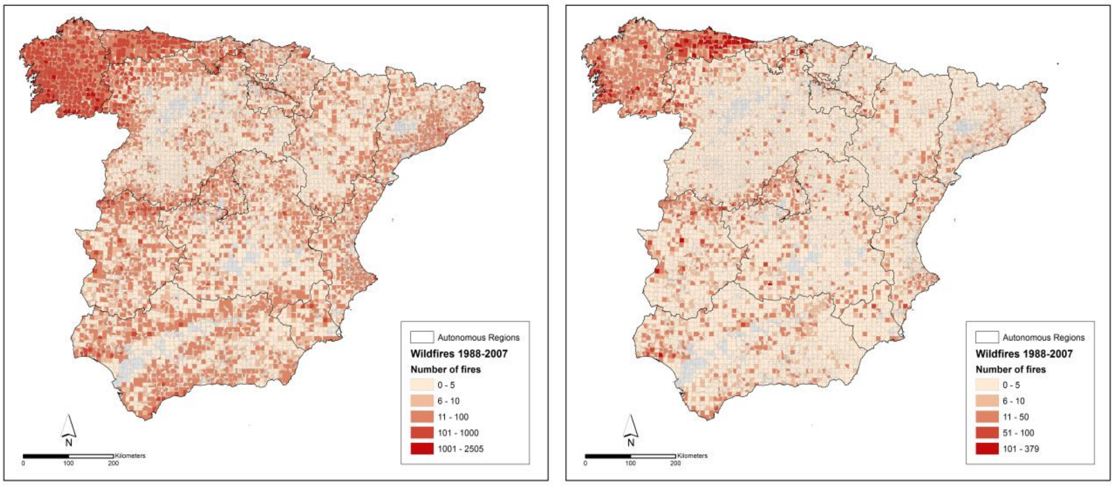

Figure 1. Spatial distribution of wildfires. Left total number of fires, right K fires.

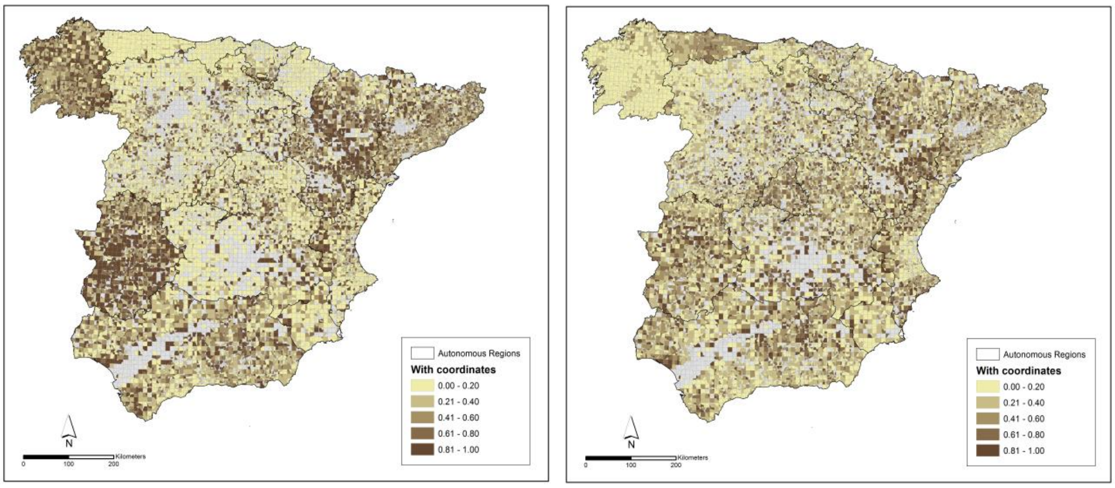

Figure 2. Spatial distribution of the proportion of points with coordinates. Total number of fires (left), fires with known cause (right).

2.3. EGIF scenarios

In this work, we explored six scenarios based on the classification of ignition causes and location data reported in the EGIF database. The proposed scenarios were constructed to simulate the most probable assumptions to select an occurrence sample for wildfire modeling purposes. The criteria followed to design the scenarios were based mainly in three parameters: certainty of the cause (known or supposed), certainty of the source (human or unknown) and presence of coordinates. Thus, the proposed scenarios are:

- Scenario 1: this scenario considers all human-caused fires, including both known and supposed cause, and a proportion of unknown fires according to the observed proportion of human-caused fires in the corresponding autonomous region.
- Scenario 2: this scenario considers all human-caused fire, including both known and supposed cause, excluding those fires with an unknown source.

- Scenario 3: this scenario considers all human-caused fire, but only those with known cause, excluding fires with a supposed cause, but including a proportion of unknown fires according to the observed proportion of human-caused fires in the corresponding autonomous region.
- Scenario 4: this scenario considers all human-caused fire, but only those with known cause, excluding fires with a supposed cause or an unknown source.
- Scenario 5: this scenario considers all human-caused fire, including both known and supposed cause located using coordinates, excluding fires with an unknown source or those which are wrongly located according to Table 2.
- Scenario 6: this scenario considers all human-caused fire, but only those with known cause and located using coordinates, excluding fires with a supposed cause, an unknown source or those which are wrongly located according to Table 2.

2.4. **Wildfire modelling**

The assessment of human-caused wildfire occurrence was carried out using RF algorithms an ensemble classifier which uses decision trees as base classifiers (Breiman, 2001).
The dependent variable for each scenario was constructed by selecting human-caused fires (e.g. negligence, accident or arson). Then wildfires were spatialized through the assignment of each fire to its respective combination of ICONA grid, municipality and forest perimeter (Amatulli *et al*. 2007; Chuvieco *et al*., 2010,2012; de la Riva *et al*., 2004; Rodrigues *et al*., 2014; Rodrigues and de la Riva, 2014). This allowed the calculation of fire density maps at a spatial resolution of 1 Km^2 by overlaying the random point cloud with the Spanish 1x1 Km UTM grid. The dependent variable was developed for each scenario by classifying the occurrence values into two categories: high occurrence (presence) in locations with two or more fires, and low occurrence (pseudo-absence or background) in locations with only one fire.
The explanatory variables were selected based on the experience of the authors in models at regional and national scales (Amatulli *et al*., 2007; Chuvieco *et al*., 2010, 2012; de la Riva *et al*., 2004; Martínez *et al*., 2009; Rodrigues *et al*., 2014; Rodrigues and de la Riva, 2014). The predictive variables considered were: wildland-agricultural interface (WAI), wildland-urban interface (WUI), density of agricultural machinery (DAM), changes in demographic potential 1991-2006 (CDP; Calvo and Pueyo, 2008), protected areas (PA), forestry area in public utility (FAPU), forestry tracks (TRCK), railroads (RRDS) , power lines (PWR) and land use change 1991-2006 (LUC).
The comparison of the outputs (predicted probability of occurrence) from each proposed scenario was conducted from the point of view of the accuracy in the classification based on a k-fold cross-validation procedure (Fielding and Bell, 1997) and according to the variation in the variable importance (Breiman, 2001). In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples (k=5 in this work). Each time, one of the k subsets is used as the test set and the other k -1 subsets are putted together to conform the training set. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds then can be averaged to produce single error estimation (Bar Massada *et al*., 2012).
Variable importance assessment was carried out by summarizing the influence of the explanatory variables according to the increase in mean square error (*IncMSE*) and the increase in node purity (*IncNP*). *IncSME* is defined as the increase in the mean of the error of a tree in the forest when the observed values of this variable are randomly permuted in the *out-of-bag* samples. *IncNP* is measured using the Gini criterion, from all the splits in the forest based on a particular variable (Breiman, 2001). The variability in variable importance was addressed through the fluctuations in the ranks obtained by ordering the explanatory variables from more to less importance according to *IncSME* and IncNP.

2.5. Spatial variation in the ignition probability patterns

Changes in the spatial probability patterns were addressed through local Hot Spot analysis, one of the most adequate for this purpose (Allgöwer *et al*., 2005). The assessment of changes in the spatial pattern of predicted probability was based on the assumption that one of the key factors in wildfire management was guiding governments or responsible authorities through prioritization across fires and resources at risk. We considered that the identification of areas with high values of occurrence probability (Hot Spot) is linked to the identification of priority intervention areas.

The assessment of the changes in the predicted spatial pattern at each scenario is carried out by cluster and outlier analysis through the Anselin's Local Moran's I (Cluster and Outlier Analysis). This kind of analysis allows identifying and allocating Hot Spot areas as well as characterizes its typology of cluster. Given a set of weighted features, the Cluster and Outlier Analysis tool identifies clusters of features with values similar in magnitude. The tool also identifies spatial outliers. To do this, the tool calculates a Local Moran's I value, a Z score, a p-value, and a code representing the cluster type for each feature. The Distance Band or Threshold established for the cluster detection was 10 km. The results were mapped according to the significant detected cluster typology: Hot Spot (HH), Hot Spot surrounded by Cold Spot (HL), Cold Spot (LL) and Cold Spot surrounded by Hot Spot (LH).

3. Results

3.1. Predicted probability of occurrence

There is high variability (and therefore uncertainty) in predicted probability values among the six scenarios (Figure 3). In general terms, scenarios characterized by the use of both K and S causes, mainly scenarios 1 and 2, show high performance with AUC values stand above 0.9 (McCune *et al*., 2002). Scenarios 2 and 3, where the occurrence used to construct the dependent variable only consider K causes are less accurate (AUC near 0.83) and values in the high probability interval (0.8 to 1) are almost inexistent. Scenarios where the ignition points are georeferenced using coordinates show the poorest accuracy and probability values are grouped in the first interval (0 to 0.2). In addition, the range of AUC values (difference between minimum and maximum value) shows a similar behavior, with lower values in scenarios 1 and 2, and increasing until scenarios 5 and 6. This means that the models fitted using a dependent variable constructed with both K and S causes are more stable and therefore more reliable. Table 3 summarizes the obtained AUC values. On the other hand, the same comportment is observed when considering the values of Max TPR+TNR. This parameter represents the best threshold to distinguish between presence/absence according to the maximum value of the kappa index i.e. the highest values of true positive rate (TPR) and true negative rate (TNR). In general terms, the higher the threshold the higher the accuracy of the model since it means that the model distinguish more efficiently between presence and background values.

The uncertainty observed in the probability values is also detected in the contribution of the explanatory variables for each scenario. Although the variability is higher in the importance ranks for *IncSME* than in *IncNP* (Table 4) there is a general tendency to promote always the same variables: DAM, CDP, WAI, PA and TRCK (the later only is observed in the *IncNP*). The rest of the variables are swapping ranks among the different scenarios.

Table 3. Summary of k-fold validation with k=5.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**
Max AUC	0.908	0.904	0.838	0.844	0.845	0.784
Min AUC	0.906	0.899	0.827	0.829	0.821	0.746
Max TPR+TNR	0.341	0.324	0.125	0.146	0.123	0.062

Table 4. Importance ranks for the explanatory variables. Top IncSME, bottom IncNP.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	scenario5	Scenario 6	ranks
DAM	1	1	1	1	1	1	0
CDP	2	2	2	2	2	3	1
WAI	3	3	4	5	6	5	4
PA	4	4	3	3	3	4	3
FAPM	5	5	6	7	4	2	5
WUI	6	6	5	4	5	6	3
RAIL	7	8	7	6	7	9	4
PWL	8	9	9	8	9	8	2
LUC	9	10	10	10	10	10	2
TRCK	10	7	8	9	8	7	4
DAM	1	1	1	1	1	1	0
CDP	2	2	2	2	2	2	0
WAI	3	3	3	3	3	3	0
PA	5	5	5	5	5	5	0
FAPM	6	6	7	7	8	7	3
WUI	10	10	10	10	10	10	3
RAIL	9	9	9	9	7	8	3
PWL	7	7	6	6	6	6	2
LUC	8	8	8	8	9	9	2
TRCK	4	4	4	4	4	4	0

3.2. **Variation in spatial patterns of probability**

Figure 4 shows the spatial distribution of the cluster characterization of the predicted probabilities. In the same way that occurs in the predicted probability of occurrence, there is high heterogeneity in the spatial pattern at each scenario. However, in this case a similar spatial pattern of cluster is observable among the six scenarios, with HH clusters in the northwest of the peninsula and the Mediterranean coast, HL clusters in Pyrenees and the central area of the peninsula and LH in the Cantabrian coast. However, the scenarios using known causes (scenarios 4 and 6) are presenting LL clusters in some regions of the Northwest of the peninsula which is not that would be expectable since this area presents the highest occurrence values (Figure 1).

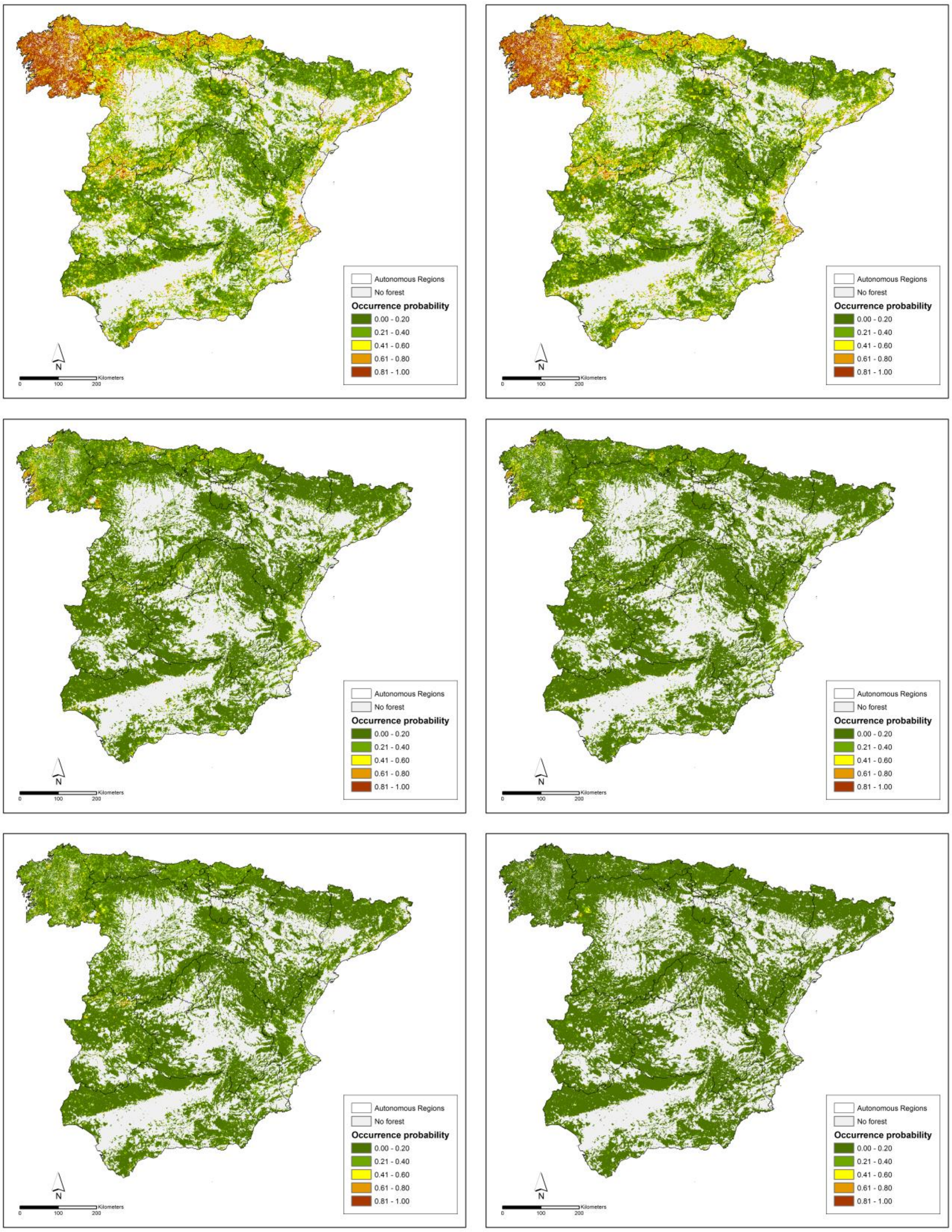

Figure 3. Spatial distribution of the predicted probability values. Scenarios are ordered consecutively left-right-top-bottom.

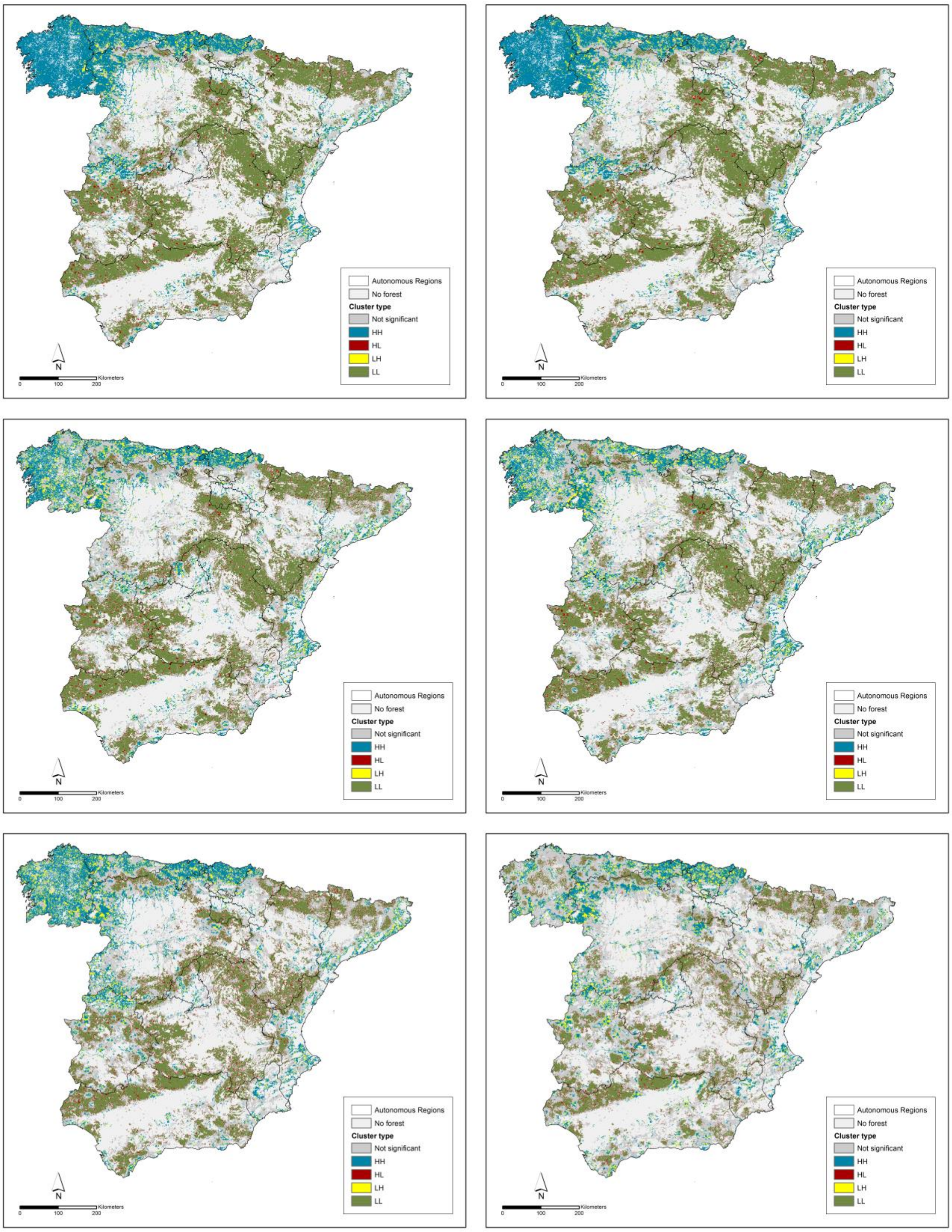

Figure 4. Spatial distribution of the cluster type. Scenarios are ordered consecutively left-right-top-bottom.

4. Discussion

Multiple sources of uncertainty remain with regard to modelling wildfire occurrence (Thompson and Calkin, 2011). Therefore there is a need to better understand how uncertainty and errors propagate through models (Sullivan, 2009). As little is known about wildfire causes (Lovreglio *et al*., 2006) many authors have chosen to deal globally with human-caused fires, avoiding uncertain specifications of causes, and have been able to derive useful recommendations for management (Stephens, 2005).

Nevertheless, using a coherent framework informs management authorities by facilitating the identification of potential sources of uncertainty and the quantification of their impact.

In Spain, near a 29% of the fire events in the period 1988-2007 have an unidentified cause and the remaining 71% are not fully reliable because the existence of certain degree of uncertainty regarding ignition source and location. This uncertainty is firstly detected while analyzing and mapping fire data; and secondly when occurrence data is used for wildfire modeling. Uncertainty is affecting both to the predicted probability values as well as the spatial pattern of probability.

According to the results may vary greatly depending on to the assumptions made when constructing the dependent variable. Results suggest that the scenarios based on the consideration of all causes (K and S) as well as a proportion of the fires with a U source are more accurate, with AUC values above 0.9. We believe that this is mainly because when considering the whole occurrence, the dependent variable is less 'spatially biased' since there is no partial criterion to leave out a particular set of fire events and, thereby, the spatial pattern should be closer to reality. It is expectable that the scenario with the less uncertainty in its occurrence data, i.e. a scenario with known causes and (scenario 6), would be the most accurate. However, the fact that there are differences in the proportions of unidentified and allocated fires within the Spanish peninsula is harming the quality of the data.

In addition, there is also uncertainty in the contribution/importance of the predictive variables. This might be a big issue in research works aiming to determine the factors that are explaining wildfire occurrence because the assumptions made when constructing the dependent variable are influencing the contribution of the explanatory variables.

Regarding to the predicted probability spatial pattern, although the variability is lower than the detected in the case of the predicted probability, it is still great. As in the case of the probability of occurrence, scenarios based on the consideration of all causes (K and S) including a proportion of the fires with an unidentified source seem to be the most realistic approach.

5. Conclusions

The lack of uniformity in the application of the criteria among the autonomous regions on forest fire management is a potential source of uncertainty for wildfire risk assessment which is affecting both the predicted probability values as well as the spatial pattern of probability. This is especially significant since research on forest fires in Spain is made from historical data collected in the EGIF database (Amatulli *et al*., 2007; Chuvieco *et al*., 2010, 2012; de la Riva *et al*., 2004; Martínez *et al*., 2009; Padilla and Vega-García, 2011), being more affected the older the data, as technological advances have greatly contributed to improve the quality of the data (such as GPS measurements or database administration capabilities) and, thus, reducing the uncertainty. In any case, some studies have been able to derive useful recommendations for management avoiding uncertain specifications of causes (Stephens, 2005), addressing arising uncertainty in occurrence data can help improve assessments.

The spatial distribution of wildfire ignition greatly varies depending on the assumptions made when considering the ignition cause and source, leading to different predictions. However, it is possible to determine the best scenarios for modeling wildfire occurrence or risk. According with our results the best choice is consider both K and S causes with a proportion of forest fires with unknown source. There is a big amount of unidentified fires potentially related to a human ignition source and thus, when excluding unknown fires in human-caused wildfire assessments, a significant portion of the occurrence is not accounted for. Considering this supposedly human-caused occurrence reduces the spatial biased conducting to more robust and reliable predictions.

Uncertainty is also affecting the contribution of the explanatory variables. Results suggest that DAM, CDP, WAI and PA are the least sensitive variables to variations in the spatial distribution of the occurrence.

6. Acknowledgements

This work was financed by the National Plan of I+D of the Spanish Ministry of Science and Innovation: FPI grant BES-2009-023728. The research was conducted within the framework of the projects FIREGLOBE: Analysis of fire risk scenarios at the national and global scales (CGL2008-01083/CLI) and PYRORAMA Modeling future scenarios of fire risk and land use changes at national scale (CGL2011-29619-C03-01) subproject 1 of the PYROSKENE project (CGL2011-29619-C03).

7. References

Allgöwer B, Camia A, Francesetti A, Koutsias N (2005) Fire hot spot areas in Southern Europe. Detection of large-scale wildland fire occurrence patterns by adaptive kernel density interpolation, In 'Proceedings of the 5th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management: Fire Effects Assessment'. (Eds J de la Riva, F Perez-Cabello, E Chuvieco) pp. 47-50 (EARSeL-UZ: Zaragoza).

Amatulli G, Rodrigues MJ, Trombetti M, Lovreglio R (2006) Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research* **111**, G04S05.

Amatulli G, Peréz-Cabello, F, de la Riva J (2007) Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling* **200**, 321-333.

Anselin L (1995) Local Indicators of Spatial Association-LISA. *Geographical Analysis* **27**, 93-115.

Bar Massada A, Syphard AD, Stewart SI, Radeloff VC (2012) Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire* **22**(2), 174-183.

Bradshaw GA, Borchers JG (2000) Uncertainty as information: narrowing the science-policy gap. *Conservation Ecology* **4**, 7.

Breiman L (2001) Random forests. *Machine Learning* **45**, 5-32.

Borchers JG (2005) Accepting uncertainty, assessing risk: Decision quality in managing wildfire, forest resource values, and new technology. *Forest Ecology and Management* **211**, 36-46.

Calvo JL, Pueyo A (2008) Atlas Nacional de España: Demografía (CNIG: Madrid).

Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Martín MP, Vilar L, Martínez J, Martín S, Ibarra P, de la Riva J, Baeza J, Rodríguez F, Molina JR, Herrera MA, Zamora R (2010) Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling* **221**, 46-58.

Chuvieco E, Aguado I, Jurdao S, Pettinari ML, Yebra M, Salas J, Hantson S, de la Riva J, Ibarra P, Rodrigues M, Echeverría M, Azqueta D, Román MV, Bastarrika A, Martínez S, Recondo C, Zapico E., Martínez-Vega FJ (2012) Integrating geospatial information into fire risk assessment. *International Journal of Wildland Fire*.

de la Riva J, Pérez-Cabello F, Lana-Renault N, Koutsias N (2004) Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment* **92**,363-369.

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24** (1), 38–49.

Kalabokidis KD, Gatzojannis S, Galatsidas S (2002) Introducing wildfire into forest management planning: towards a conceptual approach. *Forest Ecology and Management* **158**, 41-50.

Lovreglio R, Leone V, Giaquinto P, Notarnicola A (2006) New tools for the analysis of fire causes and their motivations: The Delphi technique. *Forest Ecology and Management* **234**, S18.

McCune B, Grace JB, Urban DL (2002) Analysis of ecological communities. MJM Software Design (Glenden Beach).

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

Moreno MV, Malamud BD, Chuvieco E (2011) Wildfire Frequency-Area Statistics in Spain. *Procedia Environmental Sciences* **7**, 182-187.

Padilla M, Vega-García 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. *International Journal of Wildland Fire* **20**, 46-58.

Rodrigues M, San Miguel J, Oliveira S, Moreira F, Camia A (2013). An insight into Spatial-Temporal Trends fo Fire Ignitions and Burned Areas in the European Mediterranean Countries. *Journal of Earth Science and Engineering* **3**, 497-505.

Rodrigues M, de la Riva J, Fotheringham S (2014) Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography* **48**, 52-63.

Rodrigues M, de la Riva (2014) An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software* In Press.

San-Miguel-Ayanz J (2009) Forest fires at a glance: facts, figures and trends in the EU. In 'Living with wildfires: what science can tell us. A contribution to the Science-Policy dialogue'. (Eds Y Birot) pp. 11-18 (European Forest Institute).

San-Miguel-Ayanz J, Rodrigues M, Oliveira S, Pacheco C, Moreira F, Duguy B, Camia A (2012) Land Cover Change and Fire Regime in the European Mediterranean Region. In 'Post-Fire Management and Restoration of Southern European Forests'. (Eds F Moreira, M Arianoutsou, P Corona, J de las Heras) pp. 21-43 (Springer: Netherlands).

Sullivan AL (2009) Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire* **18**, 349-368.

Stephens SL (2005) Forest fire causes and extent on United States Forest Service lands. *International Journal of Wildland Fire* **14**, 213-222.

Thompson MP, Calkin DE (2011) Uncertainty and risk in wildland fire management: A review. *Journal of Environmental Management* **92**,1895-1909.

Vélez R (2001) Fire Situation in Spain. In 'Global Forest Fire Assessment 1990-2001'. (Eds JG Goldammer, RW Mutch, P Pugliese). (FAO:Roma).