# ADVANCES IN FOREST FIRE RESEARCH

## 2022

Edited by
**DOMINGOS XAVIER VIEGAS**
**LUÍS MÁRIO RIBEIRO**

# Fire cause classification of undetermined fires in southeastern France

Christos Bountzouklis*[1]; Dennis Michael Fox[1]; Elena Di Bernardino[2]

[1]*Univerisity of Côte d'Azur, UMR CNRS 7300 ESPACE, Nice, 06204, France, {christos.bountzouklis@univ-cotedazur.fr; dennis.fox@univ-cotedazur.fr}*
[2] *Univerisity of Côte d'Azur, UMR CNRS 7351 LJAD, Nice, 06108, France, {elena.di_bernardino@univ-cotedazur.fr}*

**Corresponding author*

**Keywords**

ignition causes, machine learning, arson fires, southeastern France

**Abstract**

Knowledge about fire ignition causes and their spatiotemporal patterns can greatly enhance the efficiency of fire management and fire strategies. In France, the majority of forest fire research is based on a 2x2 km gridded database that provides amongst other information, the cause of fire ignition. According to the same database however, approximately 70% of all fires between 1973 and 2020 were recorded without a cause of ignition. Therefore, information on fire causes for a very large part of the fires that were recorded in the last 50 years is not taken into consideration and can potentially provide significant evidence on fire ignition patterns. As arson fires are of particular interest in southeastern France since they are the most frequent and account for the largest volume of burned area, this study aimed to exploit a fire ignition point geodatabase and machine learning methodologies, to create and evaluate a model that can identify whether unknown caused fires can be classified as arson or non-arson fires, based on numerous environmental and anthropogenic factors. The results of the study suggest that cause identification can be adequately accurate using such a model, although a larger fire database would increase overall performance.

## 1. Introduction

Fire ignition patterns can vary significantly both temporally and spatially depending on the cause of ignition (Curt et al., 2016) and can be impacted by a plethora of environmental and anthropogenic drivers (Catry et al., 2009; Syphard et al., 2008; Syphard and Keeley, 2015). Some studies have demonstrated that arson fires can potentially be predicted both spatially and temporally (Gonzalez-Olabarria et al., 2012; Penman et al., 2013). In SE France, arson (particularly pyromania and conflict/interest) is the most frequent ignition cause for large fires (100>ha) (Ganteaume and Jappiot, 2013). Recording fire causes and studying their spatiotemporal patterns is important for establishing useful fire policies (Rodrigues et al., 2014) since a better understanding can enhance the efficacy of fire prevention measures (Oliveira et al., 2012). In France according to the national fire database (Prométhée) that contains records of fire ignition causes, approximately 70% of all fires between 1973 and 2020 were ignited by an unknown cause. The percentage of non-identified causes is high, and it is an additional constraint to the already limited research conducted on fire ignition causes. As in many other disciplines, applications of machine learning methodologies have seen a significant increase in wildfire science over the past years (Jain et al., 2020). Thus, this study aims to examine whether a fire ignition point dataset coupled with machine learning methods can be used to identify the source (arson or non-arson) of unknown caused fires and evaluate the importance of several environmental and anthropogenic factors in determining the ignition source.

## 2. Data & Methodology

### 2.1. Study area

The study area covers the administrative department of "Bouches-du-Rhône", which according to the official forest fire database in France (Prométhée), ranks second in terms of burned area and fire frequency in mainland France (Figure 1, Table 1). The department is characterized by gentle slopes and low to intermediate altitudes that increase when moving eastwards. Population density (388.8 people/km$^2$) is higher in the eastern half of the department since that is where the second most populated city in France (Marseille) is found and because the westernmost parts are covered by wetlands and a national park. Therefore, the westernmost section has a low potential for fire ignition and propagation but increases when moving towards the eastern half of the department.

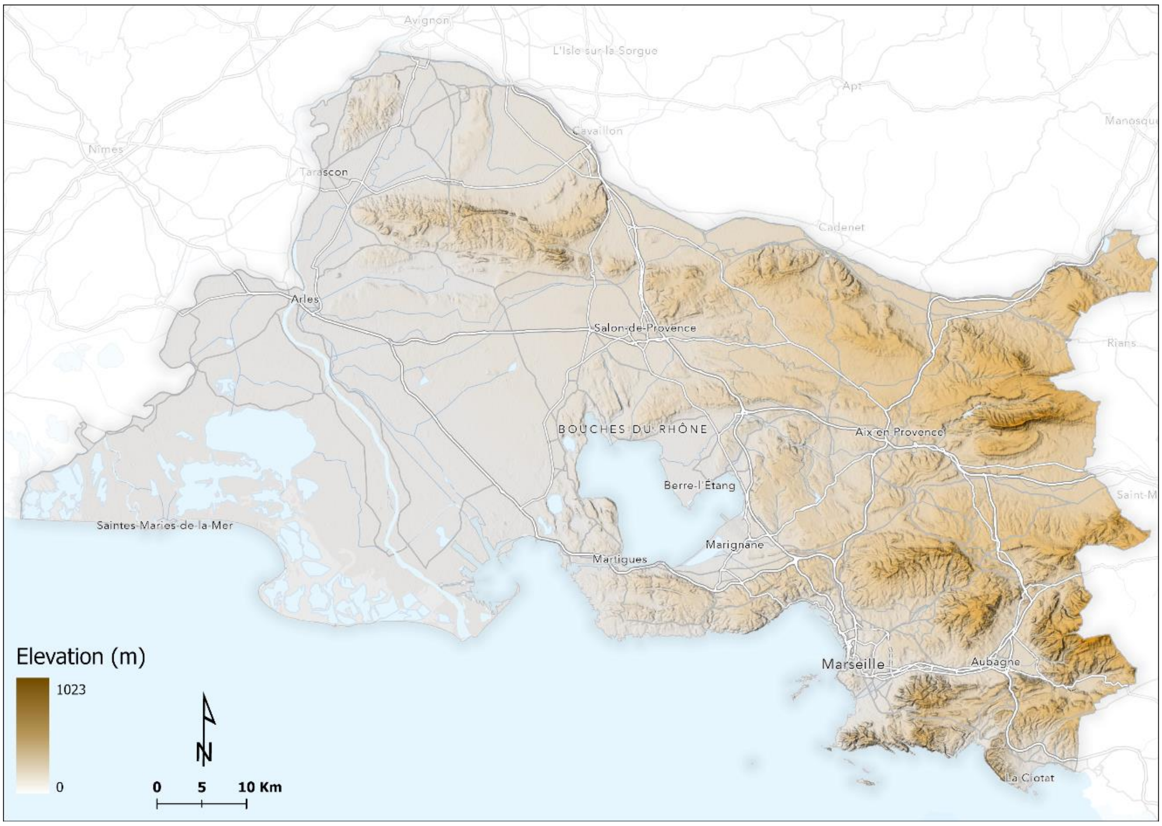


***Figure 1 Departmental limits of Bouches-du-Rhônes overlaid on a 25 m Digital Elevation Model.***

***Table 1 Environmental characteristics of Bouches-du-Rhône.***

| | |
|---|---|
| Total area (km²) | 3456 |
| Forested area (km²) | 1530 |
| Mean slope (°) | 8.8 |
| Median slope (°) | 5.7 |
| Mean elevation (m) | 142 |
| Median elevation (m) | 89 |

When considering only fires with a known cause, 63 % of the total burned area and approximately half (51 %) of all fire ignitions in the department are due to arson fires, according to Prométhée (Table 2). In addition, most of the large fires (>100 ha) in the study area are caused by arsonists (Figure 2). Even though negligence (professional & personal) is the second most frequent cause of fire ignition, it does not cause a proportionate volume of burned area.

***Table 2 Number of fires and volume of burned area per ignition cause from 1973 to 2020 in Bouches-du-Rhônes.***

| **Fire Ignitions (#)** | **Percentage (%)** | **Burned Area (ha)** | **Percentage (%)** | **Cause** |
|---|---|---|---|---|
| 349 | 3.3 | 3,020 | 3.3 | Accidental |
| 161 | 1.5 | 335 | 0.4 | Natural |
| 1,034 | 9.7 | 5,223 | 5.7 | Negligence |
| 1,556 | 14.7 | 14,493 | 15.7 | Arson |
| 7,524 | 70.8 | 69,105 | 75 | Unknown |
| 10,624 | - | 92,176 | - | Total |

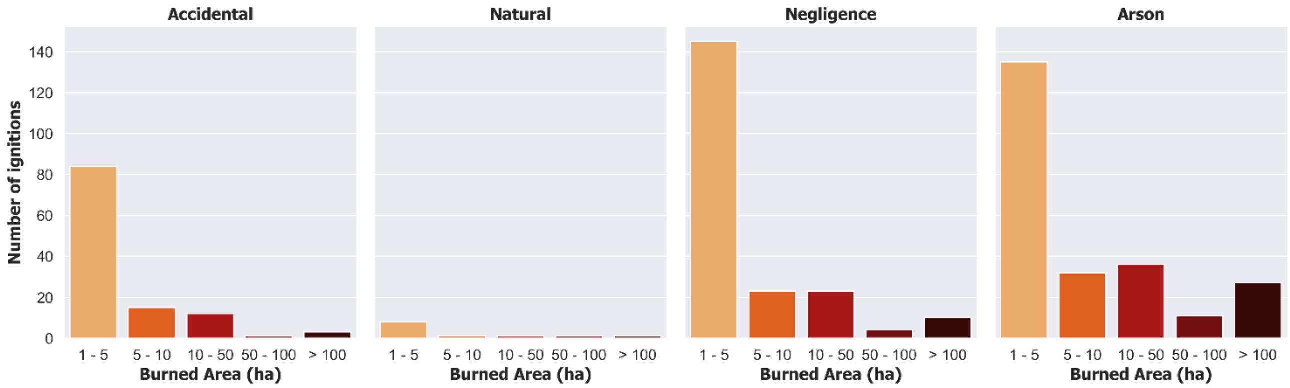


***Figure 2 Number of fires per volume category of burned area (> 1ha) and ignition cause from 1973 to 2020.***

## 2.2. Data

### *2.2.1. Fire database*

In France, the majority of forest fire research is based on Prométhée, the national database for forest fires. The specific fire database holds records of fires starting from 1973 and it includes information such as burned area, cause of ignition, date, and approximate location (within a 2x2 km grid) for each fire. In the current study, we used a geographic database that contains exact coordinates of fire ignitions that is provided by the National Forestry Office (Office National des Forêts, ONF), which to the best of our knowledge is the second time being utilized after Ganteaume and Long-Fournel, 2015. The dataset consists of 3,234 fire ignition points ranging from 1960 to 2012, which however does not contain information on the cause of ignition. To enrich the ONF point database with the cause of ignition, two additional databases were used (Figure 3). Information from the Prométhée database was firstly merged with a polygon fire geodatabase that is described in Bountzouklis et al., 2022 and subsequently spatially joined with the point geodatabase resulting in a combined dataset that contains ignition coordinates, burn scars, and cause of ignition.

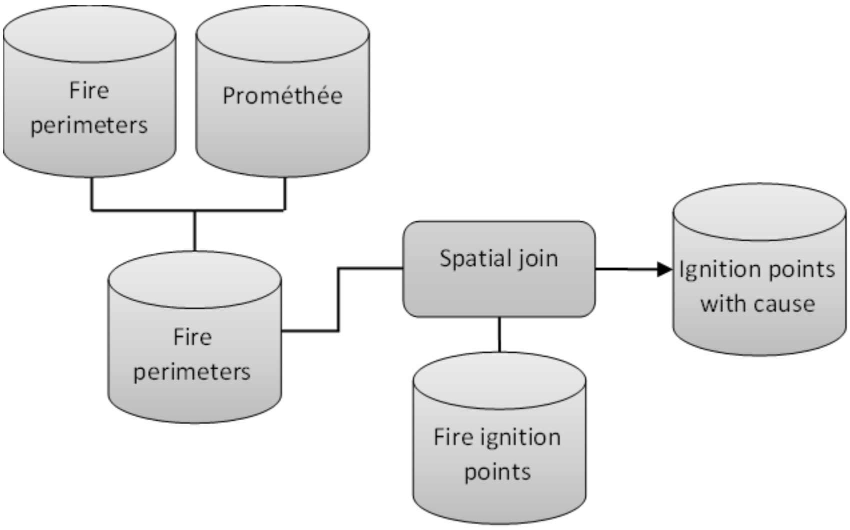


***Figure 3Flow chart depicting the processing steps to generate the final dataset.***

As earlier records on fire causes are considered less reliable (Ganteaume and Jappiot, 2013) only fires from 1996 to 2012 were considered, resulting in 323 fires (Figure 4). it was deemed best to classify causes into two major categories, arson and non-arson, due to the limited recorded number of fires caused by accident, negligence and lightning strikes but also due to the significance of arson fires in the specific area.

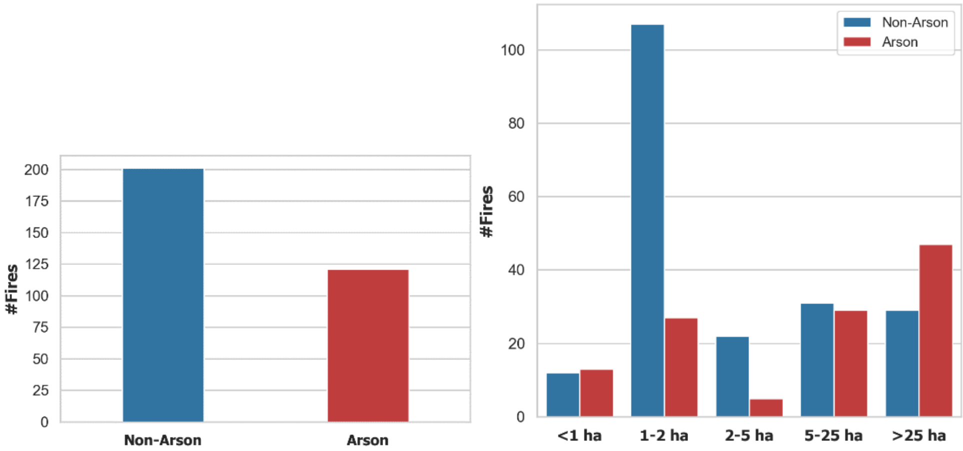


***Figure 4 Number of fires per cause and burned area size.***

### *2.2.2. Explanatory variables*

Multiple environmental and socioeconomic factors (Table 3 & Figure 5) that are known to be associated with forest fires, were acquired by a combination of European and national databases, in order to train a model that can identify the ignition cause of a fire. To account for any potential geometric errors of the ignition points and more importantly to include contextual geographic information, a circular buffer zone (500 m) was created around each fire ignition point to extract relative information.

***Table 3 List of environmental, anthropogenic and spatiotemporal variables considered.***

| **Type** | **Name** | **Description** | **Source** |
|---|---|---|---|
| **Land Cover** | Artificial surfaces | Percentage of cover within each zone | Corine Land Cover - 2006 |
| | Agriculture | | |
| | Vegetation | | |
| | Water/ Wetlands | | |
| **Topographic** | Slope inclination | Mean inclination (º) | National Geographic Institute - 5 m spatial resolution |
| | Slope aspect | Percentage of cover within each zone (Flat, N, E, S, W) | |
| | Elevation | Mean elevation (m) above sea level | |
| | Topographic wetness index | Topographic-driven control on soil moisture | |
| **Vegetation type** | Coniferous | Percentage of cover within each zone | National Geographic Institute - 2006 |
| | Coppice | | |
| | Hardwood | | |
| | Open Forest | | |
| | Shrublands | | |
| **Anthropogenic** | Population density | Individuals per sq. m. | National Institute of Statistics and Economic Studies - Sub-municipal level - 2006 |
| | Gini index | Inequality index | |
| | Income | Mean taxable income (€) | |
| | Unemployment | Unemployment rate (%) | |
| | Primary road distance | Euclidean distance from fire location (m) | National Geographic Institute - 2008 |
| | Secondary road distance | | |
| | Power pylons distance | | |
| | Railway distance | | |
| | Wildland Urban Interface distance | | |
| | House density | Buildings per sq. m. | |
| **Spatiotemporal characteristics** | Season | Winter, spring, summer, autumn | Fire database |
| | Burned area size | <1 ha, 1-2 ha, 1-2 ha, 2-5 ha, 5-25 ha, >25 ha | |
| | Coordinates | XY coordinates of each fire | |

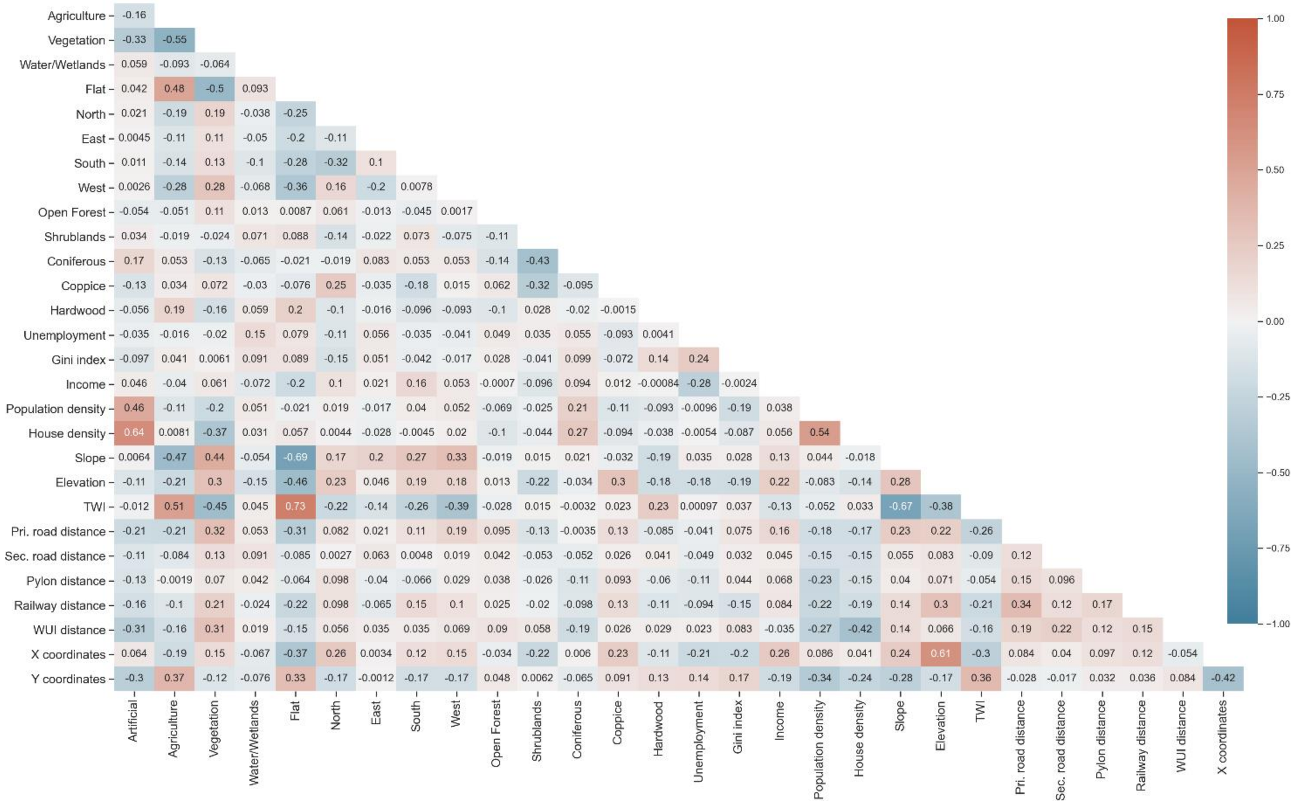


***Figure 5 Kendall's Tau rank correlation coefficient heatmap of explanatory variables.***

### 2.3. Methods

The analysis of the data is based on Random Forests (Breiman, 2001), a well-established machine learning algorithm in many disciplines but also in wildfire science (Jain et al., 2020). In order to train the model, 70% of the original dataset was utilized, while the remaining 30% was used for testing the accuracy in predicting the cause of a fire. Due to the limited number of observations, the process of splitting the data (using the same ratio) and executing the model was iterated (n=300) to have a more consistent perception of the accuracy of the model. Additionally, the processing chain included tuning the algorithm's hyperparameters as well as calculating the feature importance score for all variables. Feature importance was based on the Gini impurity method which can be used as a diagnostic which contributes to understanding which variables are driving the results of a model and which ones can be discarded.

## 3. Results

The accuracy of the model is illustrated in the form of a boxplot (Figure 6) that represents accuracy scores from all iterations of the model. The accuracy of the model to classify unknown caused fires can vary substantially ranging from 56% to 76% (median value 67%), due to the small size of the dataset.

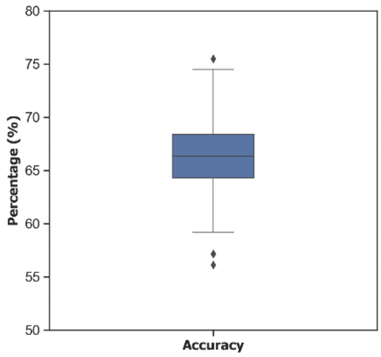


***Figure 6 Boxplot representing Random Forest's classification accuracy for all iterations (n=300). (i) Bar within the box is the median value, (ii) bottom part of the box is the first quartile, and (iii) top part of the box is the third quartile. Whiskers represent observations outside the middle 50 % and points represent outliers.***

Figure 7 shows the importance values of each explanatory variable used in the model. Overall, anthropogenic features (in blue) appear to surpass in importance the rest. Moreover, topographic factors (in brown) seem to be more important than land cover and vegetation type while spatiotemporal variables except for XY coordinates hold the lowest importance. More specifically, Secondary road distance, Shrublands and Unemployment rate are the three variables displaying the highest importance. Finally, several variables mainly related to spatiotemporal characteristics and others such as certain vegetation types (Hardwood & Coppice) as well as Water can be excluded to decrease model complexity and execution time.

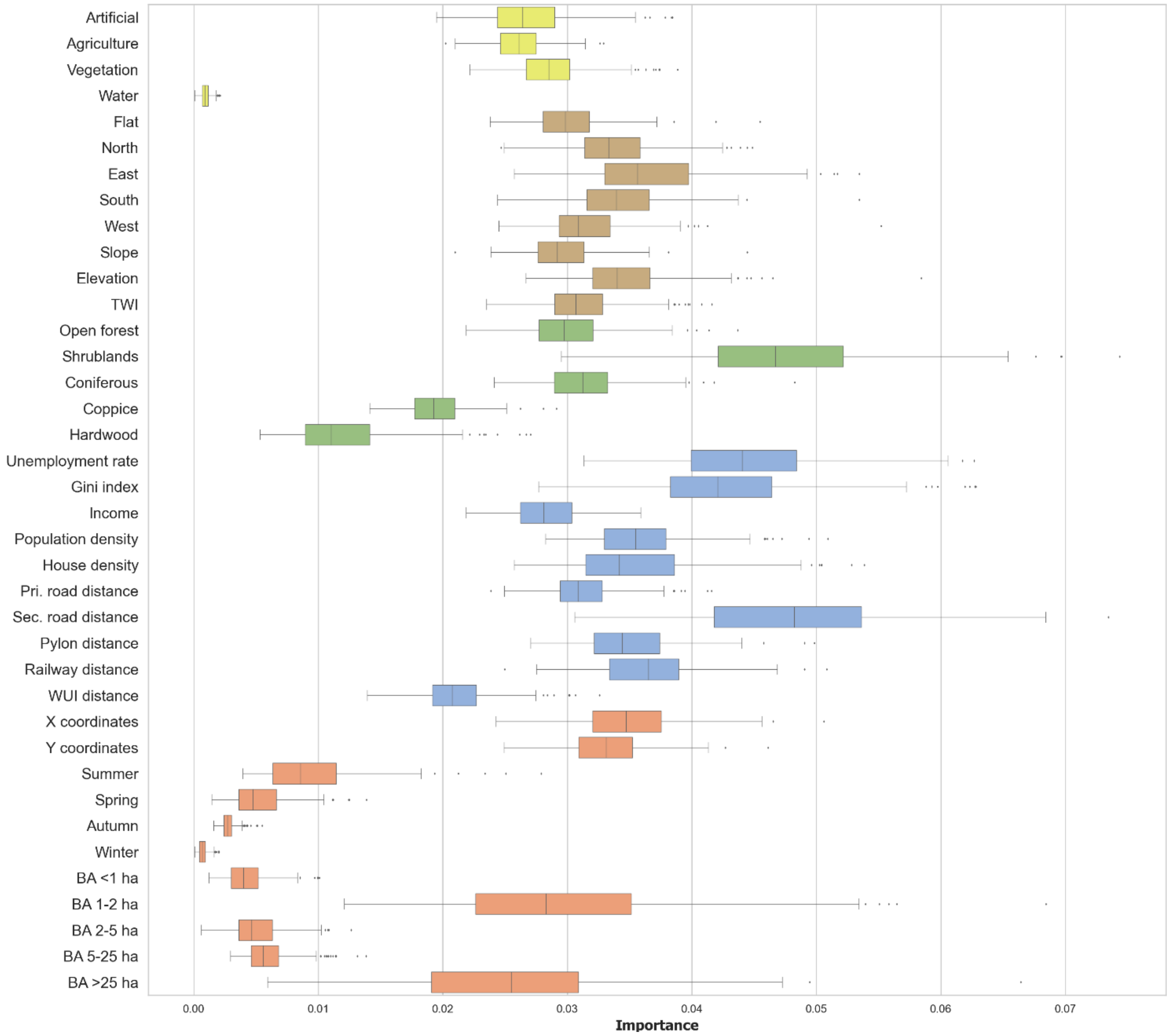


***Figure 7 Distribution of the variable importance values for all iterations (n=300).***

## 4. Conclusion

The results of the study suggest that the source of unknown caused fires can be identified at an acceptable level of accuracy even with a limited number of fires. Anthropogenic drivers such as distance to secondary roads and unemployment rate, along with higher volumes of shrublands around ignition points are the most important features in determining the classification of unknown caused fires for the specific area.

Overall performance of such models would most likely greatly benefit from the exploitation of larger datasets as well as from the inclusion of weather-related variables. Finally, as location holds particular importance over certain fire causes, spatial extensions of machine learning algorithms such as Geographic random forests

(Georganos et al., 2021) and Geographically weighted neural networks (Hagenauer and Helbich, 2022) could provide significant enhancements over the original algorithms.

## 5. References

Bountzouklis, C., Fox, D. M. and Di Bernardino, E.: Environmental factors affecting wildfire-burned areas in southeastern France, 1970–2019, Nat. Hazards Earth Syst. Sci., 22(4), 1181–1200, doi:10.5194/nhess-22-1181-2022, 2022.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, doi:https://doi.org/10.1023/A:1010933404324, 2001.

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

Curt, T., Fréjaville, T. and Lahaye, S.: Modelling the spatial patterns of ignition causes and fire regime features in southern France: Implications for fire prevention policy, Int. J. Wildl. Fire, 25(7), 785–796, doi:10.1071/WF15205, 2016.

Ganteaume, A. and Jappiot, M.: What causes large fires in Southern France, For. Ecol. Manage., 294, 76–85, doi:10.1016/j.foreco.2012.06.055, 2013.

Ganteaume, A. and Long-Fournel, M.: Driving factors of fire density can spatially vary at the local scale in south-eastern France, Int. J. Wildl. Fire, 24(5), doi:10.1071/WF13209, 2015.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E. and Kalogirou, S.: Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling, Geocarto Int., 36(2), 121–136, doi:10.1080/10106049.2019.1595177, 2021.

Gonzalez-Olabarria, J. R., Brotons, L., Gritten, D., Tudela, A. and Teres, J. A.: Identifying location and causality of fire ignition hotspots in a Mediterranean region, Int. J. Wildl. Fire, 21(7), 905, doi:10.1071/WF11039, 2012.

Hagenauer, J. and Helbich, M.: A geographically weighted artificial neural network, Int. J. Geogr. Inf. Sci., 36(2), 215–235, doi:10.1080/13658816.2021.1871618, 2022.

Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S. and Flannigan, M. D.: A review of machine learning applications in wildfire science and management, , doi:10.1139/er-2020-0019, 2020.

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

Penman, T. D., Bradstock, R. A. and Price, O.: Modelling the determinants of ignition in the Sydney Basin, Australia: Implications for future management, Int. J. Wildl. Fire, 22(4), 469–478, doi:10.1071/WF12027, 2013.

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

Syphard, A. D. and Keeley, J. E.: Location, timing and extent of wildfire vary by cause of ignition, Int. J. Wildl. Fire, 24(1), 37–47, doi:10.1071/WF14024, 2015.

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