

## Article

# Modelling Fire Risk Exposure for France Using Machine Learning

Baptiste Gualdi <sup>1</sup>, Emma Binet-Stéphan <sup>1</sup>, André Bahabi <sup>1</sup>, Roxane Marchal <sup>2,\*</sup>  and David Moncoulon <sup>2</sup>

<sup>1</sup> EURIA EUro Institut d'Actuariat Brest, 29200 Brest, France; baptiste0310015361u.gualdi@etudiant.univ-brest.fr (B.G.); emma.binetste@gmail.com (E.B.-S.); andre.bah@outlook.com (A.B.)

<sup>2</sup> Caisse Centrale de Réassurance, Department R&D Cat and Agriculture Modelling, 75008 Paris, France; dmoncoulon@ccr.fr

\* Correspondence: rmarchal@ccr.fr; Tel.: +33-1-44-35-31-00

**Abstract:** Wildfires generating damage to assets are extremely rare in France. The peril is not covered by the French natural catastrophes insurance scheme (law of 13 July 1982). In the context of the changing climate, Caisse Centrale de Réassurance—the French state-owned reinsurance company involved in the Nat Cat insurance scheme—decided to develop its knowledge on the national exposure of France to wildfire risks. Current and future forest fires events have to be anticipated in case one of the events threatens buildings. The present work introduces the development of a catastrophe loss risk model (Cat model) for forest fires for the French metropolitan area. Cat models are the tools used by the (re)insurance sector to assess their portfolios' exposure to natural disasters. The open-source national Prométhée database focusing on the South of France for the period 1973–2019 was used as training data for the development of the hazard unit using machine learning-based methods. As a result, we observed an extension of the exposure to wildfire in northern areas, namely Landes, Pays-de-la-Loire, and Bretagne, under the RCP 4.5 scenario. The work highlighted the need to understand the multi-peril exposure of the French country and the related economic damage. This is the first study of this kind performed by a reinsurance company in collaboration with a scholarly institute, in this case EURIA Brest.

**Keywords:** forest fires; Cat model; climate change; disaster risk; machine learning; R programming language



**Citation:** Gualdi, B.; Binet-Stéphan, E.; Bahabi, A.; Marchal, R.; Moncoulon, D. Modelling Fire Risk Exposure for France Using Machine Learning. *Appl. Sci.* **2022**, *12*, 1635. <https://doi.org/10.3390/app12031635>

Academic Editor: Jason K. Levy

Received: 7 January 2022

Accepted: 1 February 2022

Published: 4 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

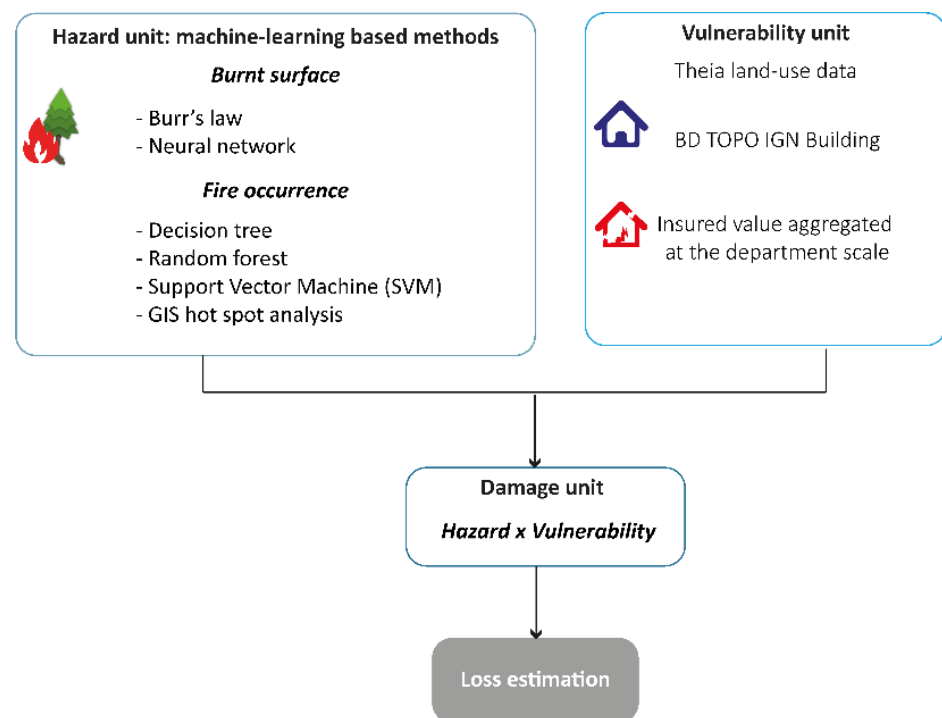
## 1. Introduction

In the world, large forest fire events are generating significant damage to natural ecosystems, human lives, and critical infrastructures [1]. In the last few years, large events occurred especially in the United States and in Australia [2]. In 2017 and 2018, in California, wildfire events were estimated, respectively, to have caused \$12 bn in damages for the Tubbs Fire and the CampFire. It has been estimated that wildfire caused \$150 bn damage globally, with \$27.7 bn for direct losses to buildings and houses, or 20% of the total [3–5]. Between 2011 and 2020, the average annual loss for the USA was \$4.7 bn for forest fires [3]. More recently, we had in mind last year's Black Summer in Australia, with the sad images of koalas and kangaroos burnt by the flames; in addition to the 10,000 people displaced, 25 people died, 5.5 million ha were destroyed, and 2448 homes were destroyed [6]. Those wildfires generated colossal economic losses. Periods of long and intense droughts elevated fire risk prediction, which is especially the case in Canada and the Western USA [7]. Nowadays, in early July 2021, the world watches, helplessly, the heat wave hitting Lytton (Canada), which recorded temperatures of 49.6 °C, with flames destroying the city [8], as well as the large events in Greece and Turkey due to the greatest heat wave in thirty four years [9].

Modelling wildfire is a complex task, as several parameters have to be defined (fire propagation, fuel, wind speed, terrain type, smoke, prevention actions and building sus-

ceptibility). This leads to the development of detailed models to assess the fire propagation, ignition and dynamic, as exemplified by the well-known models of the literature (FlamMap, FSPro, FARSITe, FIRETACTIC, FEPS, HYPPLIT, PHOENIX, and Minimum Travel Time) [1,4,10–14]. The insurance sector also developed models at the asset level, modelling the roofs, walls, and windows which were the most susceptible to burning with destruction functions [3,11,15–17]. In comparison to the US or Australia, southern Europe records less burnt surface, such as the 7.4 million ha burnt between 2000 and 2018 [18]. For France, from 1982 to 2017, 12 events were recorded for a total of 350,000 ha burnt in the Mediterranean area (EM-DAT data 2021, <https://public.emdat.be/>, (accessed on 2 August 2021)). Then, comparing these elements to the Prométhée database (Prométhée data, <https://www.promethee.com/>, (accessed on 12 October 2020)), France records an accumulation of small events (approximately 2000 events per year) with small-to-medium surfaces, with strong spatial and temporal fluctuations (approximately 7.3 ha burnt per year). At the end of August 2021, the South of France recorded a large wildfire event of 8100 ha in the natural park Plaine des Maures. The event was extreme in terms of its propagation speed of up to 8 km/h, destroying a dozen houses in Val de Gilly Griaud. It is the worst wildfire event occurring in France since 2–3 September 2003, when flames destroyed—in the same location—20,000 ha of forest, with three lives lost. It was demonstrated that the 1994 fire protection was successful [19]. On the contrary, for other countries, as explained by [20], there is a lack of international coordinated safety procedure for wildfires.

Nevertheless, considering climate change, it is important to anticipate the future exposure. Indeed, France is the fourth European country in terms of forest cover, with 17 million ha of forest, which means an increase of exposure in the next few years. Climate change affects the frequency of wildfires due to anomalous maximum temperatures, lower humidity, higher maximum wind speed, and fewer rainy days [5,21–23]. A study projected the fire danger due to climate change in Southern France [24]. In 2019, a lot of kermes oak trees died due to the heatwave, as the lethal temperature was reached with temperatures of more than 60 °C measured (French Ministry of Ecology, 2021, <https://www.ecologie.gouv.fr/prevention-des-feux-foret>, (accessed on 7 September 2021)). There is a clear need to develop a model to identify the exposed areas, and to protect them from significant losses (to the ecosystem and to the economy). Then, taking into account those elements, CCR experts in natural disaster modelling covered by the Nat Cat scheme raised interest in forest fires. In collaboration with EURIA Brest, we developed a Cat model from scratch within seven months, from data collection and hazard modelling through machine learning to exposure and damage estimates. A Cat model is the tool of the (re)insurance sector to estimate the consequences of natural disasters on their portfolios. It is composed of three submodules: hazard, vulnerability and damage units (Figure 1). We aimed to test the ability of machine learning-based methods to model the fire hazard, namely burnt surface and occurrence [25]. The hazards themselves, such as fire smoke and earth imagery, were not the target of this study. The outputs of the hazard and vulnerability units are combined into the damage unit in order to provide estimates of the amount of loss due to the natural events. A special interest was taken in the wildland–urban interface in order to consider the increasing number of houses in the French littoral at-risk areas [26]. Due to data availability, the RCP 4.5 IPCC scenario was used [27]. The model was developed using the R programming language. We combined meteorological data from spatially synchronized Safran daily weather data, building location BD TOPO IGN®, and insured values at the department scale.



**Figure 1.** Cat model structure used by the (re)insurance sector to estimate the amount of loss due to natural disasters, in this case wildfire events. For the study, machine learning-based methods were integrated into the model, namely into the hazard unit. The latest was based on the use of machine-learning methods to define the number of fire events and the burnt surface. Once the best model for each of the variables was defined, the outputs were combined with the damage unit. The vulnerability unit gathers all of the information about the insured portfolios with building locations and insured values. Then, the damage unit (risk assessment), allows the loss estimation for a wildfire event.

Cat modelling allows a probabilistic assessment of wildfire risk, examining key locations in order to determine the potential property losses; the model calculates risk by looking at a range of factors, in this case simplistic factors, in order to define the first exposure of France. The aim is to estimate whether the probability of fire occurrence and burnt surface per DFCI (DFCI, <https://www.geonov.fr/smartdata/carroyage/>, (accessed on 12 October 2020)) mesh according to the Safran data meteorological conditions of the mesh. This study was designed to demonstrate learning for residential exposure.

The paper is structured as follows. Section 2 relates the data collection process and the implemented machine learning methods for Cat model development. Section 3 presents the results of the Cat models from current to future exposure with potential losses. The paper ends with the discussions and conclusion.

## 2. Materials and Methods

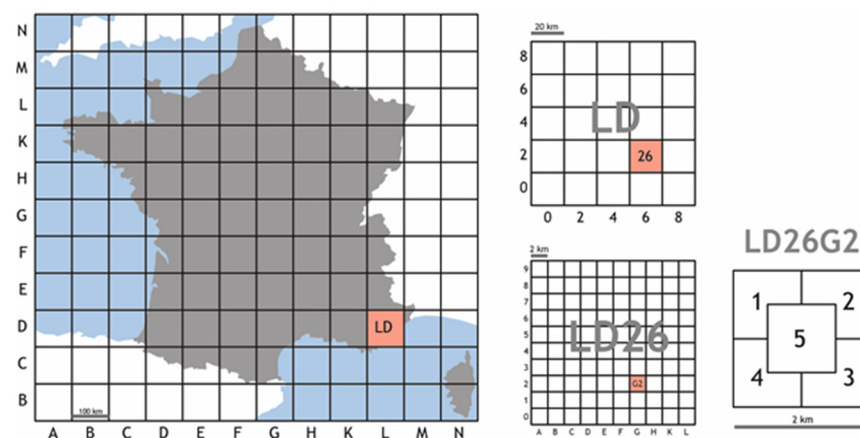
### 2.1. Fire and Meteorological Data as the Input Data for a Machine Learning-Based Hazard Unit

The historical patterns of wildfires in Southern France were based on the Prométhée dataset from 1973 to 2019. The database contains, for each DFCI mesh, several metrics (Table 1).

The DFCI geographical mesh system is used in France by actors for fire prevention from the 100-km to the 2-km resolution (Figure 2).

**Table 1.** Metrics of the Prométhée database used for the study.

Metrics	Details
Date	Date of occurrence of the wildfire
Number	Id of the fire
Type of fire	Unfilled variable
Department	Localization of the fire
INSEE ID	French ID for community
Community	Name of the community
DFCI mesh	Id of the DFCI mesh
Alert	The data and hour of the first fire alert
Origin of the alert	Policemen, population, aerial etc.
Burnt surface	In m <sup>2</sup>
Max_burnt_surf	Maximal burnt surface for each DFCI mesh



**Figure 2.** DFCI mesh, from the national to the local scale (<https://www.data.gouv.fr/fr/datasets/carroyage-dfci-2-km/>, (accessed on 12 October 2020)). The mesh has a value of 100 km with a letter code, exemplified with “LD”. Then, for the 20 km resolution, two figures are added—e.g., “LD26”—and for the 2 km a letter and a figure are added—e.g., “LD26G2”—providing a unique code for each mesh.

The Fire Weather Index (FWI) system was developed by the Canadian Forest Fire Danger Rating System (CFFDRS) in the seventies [28]. This indicator is used worldwide as a trustable indicator for the study of climate change effects in fire exposure [29]. The indicator is available on the EFFIS Copernicus website throughout Europe. The FWI is calculated daily by Météo-France for France via Arpège-Climat 4.6 over the period of reference: 1959–2007. The FWI data was downloaded from Drias’s Météo-France platform and for the period 1973–2007. The FWI is only available on a seasonal average from March to November. The higher the FWI is, the greater the probability of wildfire is. The winter season is not studied in this article.

The Safran data provide information about the temperature, humidity, wind and precipitation. The data were downloaded from the Drias’s Météo-France platform. The resolution is 8 km × 8 km; for different RCP scenarios, a total of 8602 points cover France. The evolution of the critical meteorological parameters was calculated for 1973–2005, and for the horizon 2050 RCP 4.5 seasonality of the parameters and a 20 × 20 km<sup>2</sup> analysis. The study focuses on a seasonal timescale in order to highlight the variation of the meteorological metrics. Climsec Météo-France data are available as a seasonal average for the entire year. We assume that the fire event within a DFCI mesh is uniquely determined by the mesh’s meteorological conditions.

EURO-CORDEX (Coordinated Downscaling Experiment) data are available daily by Safran point. The data were reanalyzed in order to obtain them for the same seasonality as the FWI and Climsec data (Table 2).

**Table 2.** EURO-CORDEX and Climsec data.

Metrics	Details
TASMIN	Daily minimal temperature at 2 m (altitude)
TASMAX	Daily maximal temperature at 2 m
TAS	Daily averaged temperature at 2 m
PR	Daily precipitations (mm)
SFCWIND	Wind speed at 10 m (altitude) (m/s)
SPI	Meteorological drought for 3 months
SSWI	Soil wetness index for 3 months

The meteorological Safran data were overlaid with the DFCI mesh. The resolution is 20 km × 20 km. The meteorological database is thus based on 1467 meshes (Appendix A).

## 2.2. Machine Learning-Based Methods for the Development of the Hazard Unit

Artificial intelligence and machine learning methods have been used in wildfire science since the 1990s. Within the sets of available tools suggested in Jain et al. [30], we focused only on the following: (i) decision trees, (ii) support vector machines, and (iii) artificial neural networks.

We used machine learning-based methods for the development of the hazard unit from the historical data collected in the Prométhée database. We focused only on the definition of the burnt surface and the occurrence of fire events. The meteorological data were integrated as indicators in order to assess their consequences on the area covered by the fire, and on the occurrence.

### 2.2.1. Burnt Surface

The first approach was to predict the burnt surface in each DFCI mesh, and to validate it for the real historical data. The training data were the total burnt surfaces and the maximal burnt surface. In order to solve the issue of extreme events and fires with low intensity, we focused only on the fires of 1 ha and 100 ha. The first tested method was the adjustment according to a law. Indeed, this makes it possible to study the data as a drawing of a random variable  $X$ , the law of which was known but the parameters of which were not. In order to do this, we must choose a known law of  $X$  that seems to be close to the distribution of our data. Then, by a method of optimization of the parameters of this law, we find the parameters that maximize the likelihood between the data and the density of this law. We can predict future data by randomly drawing this law as many times as necessary in a statistical approach. Linear regression establishes a linear relation between an explained variable and one or more explicative variables. The model was defined as follows, with  $Y$  being the explained variable,  $X_p$  being the  $p$  explicative variable, and the error and  $\beta$  being the parameters of the model.

We focused on the least-squares method, which minimizes the square deviation and the estimated regression. The the R packages used were fitdistr, MASS and fitdistrplus. After different tests, the most appropriate law was Burr's law.

The second process was to test the neural network. A neural network is composed of neurons distributed in several layers: input neurons, neurons in different hidden layers, and output neurons. The input data and our hidden layers will modulate these data by different weights and biases which provide output neurons and a value [31–33]. Then, the square error of the prediction was calculated by comparing the differences between the data and the predicted value. The neural network was then modified in order to minimize this error. Thus, by repeating this operation the neural network obtains accurate predictions while avoiding overlearning. The package used was nnet. The network length was 10, with 40 iterations.

### 2.2.2. Fire Occurrence

The objective was to predict the probability of a DFCI mesh being affected by a fire event, and to validate it with the real historical data. The training data were the number of fires recoded as a Boolean variable, in order to predict the occurrence of at least one fire event for the summer period. Three machine learning-based methods were tested, and are detailed below. In addition to the machine learning models developed with R, we tested the ArcGis<sup>®</sup> GIS-based machine learning hot-spot analysis.

A decision tree is a decision support tool that takes the form of a tree. At each node, a decision is taken according to a parameter, and we descend in the tree to a new node until we arrive at a leaf [34]. In a decision tree, in the case of a classification tree, the leaves contain qualitative variables (labels); in the case of a regression tree, the leaves contain quantitative variables. The package used was rpart. The maximal length of the trees was 3, with at least 50 individuals per terminal node. Between 1973 and 2005, during the summer season, 64% of the meshes recorded at least one fire; thus, the learning database will gather 50% of the database.

In a Random Forest, the main advantage of decision trees is their readability and speed of execution [35]. The package used was randomForest; 500 trees were chosen, with a height of five, and with at least 10 individuals per terminal node.

A Support Vector Machine (SVM) is a supervised learning technique. If the points (target values) are linearly separable in the space of explanatory variables, the SVM will search for the hyperplane boundary (the decision boundary) [36]. However, the points may be not separated by a hyperplane, and it is then possible to reconsider the problem in a higher dimensional space [37]. In order to deform the original space, we apply a kernel function; in this new space, it is then likely that there is a linear separation. The package used was e1017, and the duration of the calculation was a few minutes.

The space–time path is useful to visualize and understand the relationship between time and geography data. The geographical data are represented along the x and y axis, and the cube's height represents the time on the z axis [38]. The Prométhée dataset fits perfectly with the ArcGis<sup>®</sup> geovisualization tool, as there is information on the location and on the time series.

In order to perform the analysis, the Space Time Cube tool (ArcGIS Pro 2.8 online support: <https://pro.arcgis.com/fr/pro-app/latest/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm>, (accessed on 10 March 2021)) was used based on Prométhée data, in order to define the hot spots whilst considering their evolution over the timeline. The point data per year and community were integrated within the tool, and were aggregated considering space and time. The fire events were recapitulated within a hexagonal grid at yearly time steps, and then the spatial model provided the evolution in the time under a NetCDF format. We did not define the interval distance between the points, because they are the centroid of the community. Then, the Emerging Hot Spot Analysis tool was used to read the NetCDF file. It analyzes the area in which the fire events are statistically emergent or reduced in the area. The temporal interval is defined as a year.

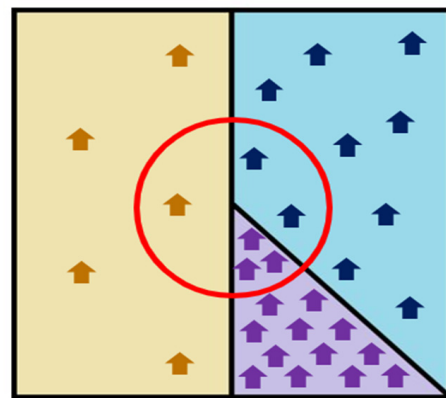
### 2.3. Vulnerability and Damage Modelling

For the development of the vulnerability unit, information about the portfolio exposure is required. First, the land-use type is needed; we used the Theia data at a 100-m resolution. Theia defined 23 land-use types; the data was re-categorized into 4 categories (Table 3). We assume the land use to be constant for the horizon 2050, as well as the number of buildings. The land-use data were overlaid with the DFCI meshes.

**Table 3.** New classification of the THEIA land-use data.

New Classification	THEIA Land Use Types
Urban	Dense buildings; diffuse buildings; industrial and commercial areas; roads
Agriculture	Colza; cereals; protein crops; soybeans; sunflowers; maize; rice; tubers/roots; grasslands; orchards; vineyards;
Forest	Wilderness deciduous forest; coniferous forest; grasslands; woody moors;
Other surfaces	Mineral; beaches; dunes; glaciers/snow; water;

Secondly, building location data were obtained from the vector building dataset BD TOPO IGN<sup>®</sup>. In this study, we only focused on the individual residential building, for which we have more detailed insurance data. For the damage model, we assume that when a fire crosses three departments with different building densities and insured values, that the three are proportionally damaged (Figure 3).



**Figure 3.** When a fire (in red) touches, proportionally, the department within the DFCI mesh, the density of the houses is applied, and we obtain the number of houses burnt and the damage costs.

In terms of insured damages, wildfires may have important consequences for the buildings. They can totally damage the infrastructure. Nevertheless, wildfire damage is not covered within the Nat Cat scheme, and events destroying residential assets are extremely rare in France. In French newspapers, there is information about the costs for firemen, but not for the insured losses. Thus, the wildfire-related insured damages are not available. In order to bridge the gap, we used confidential CCR portfolio data aggregated at the department scale. The leaflet R package was used to map the data. We consider, in the model, that if the fire touches a house it is totally destroyed within the damage function. As there are no data on fire-related claims, we considered the insured values as the claims. This is contrary to Australia or the USA, where buildings are destroyed and the destruction functions are then calibrated [11]. The price of a m<sup>2</sup> of building per department was downloaded (<https://www.meilleursagents.com/prix-immobilier/> (accessed on 1 December 2021)). The fictive model applies the maximal historical burnt surface of each DFCI mesh since 1973 to the existing urban areas. Then, we count the number of damaged assets,  $i_{house}$ , to which we apply the costs of a square meter of house per department,  $P€M2$ , and the insured values of the house and furniture at the department scale,  $P€fur$ .

$$Fori_{house} = \sum P€M2 + P€fur$$

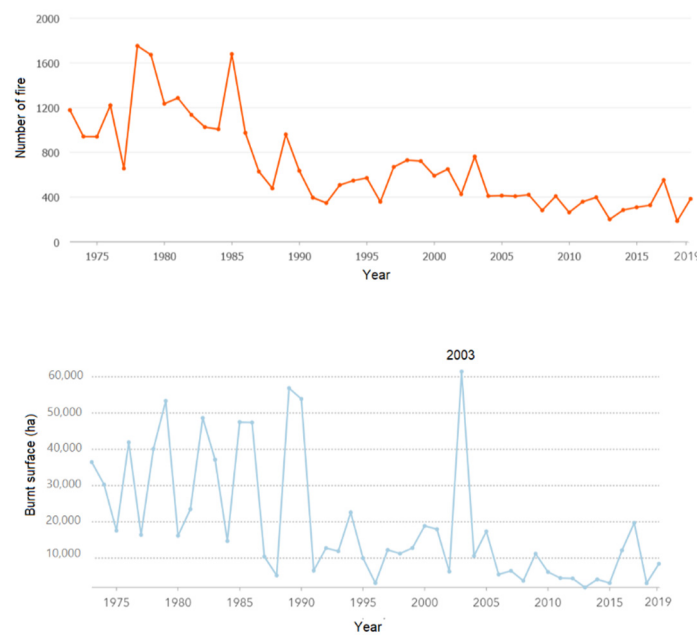
Third, the Wildland–Urban Interface (WUI), developed by the USA, describes areas where wildfires and urban areas interact, generating a potential loss of properties and life [10,39]. The WUI types intermix, and interface areas were applied to the entire French scale at the DFCI mesh of 20 × 20 km.

Concerning the damage model, we assume the application of the maximal surface of burnt areas of each DFCI mesh on the urban surface of the same mesh; then, by using the departmental insured values, we are able to calculate the potential damages.

### 3. Results

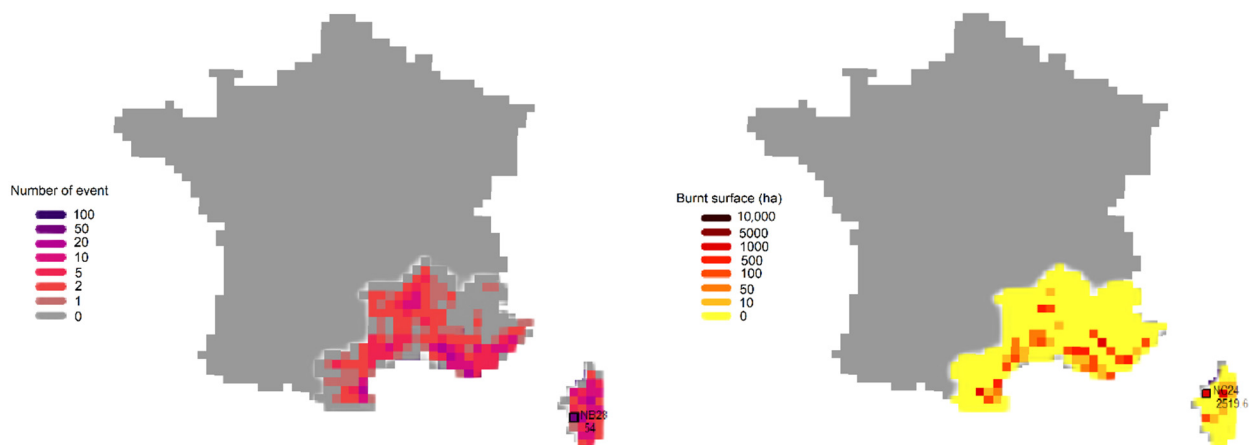
#### 3.1. Statistical Analysis

The statistical analysis of the data highlights the variations of the total burnt areas and fire occurrence per season per year in the Prométhée area. The seasonal variability is very important; it is correlated to the fact that if a lot of areas are burnt during the year  $n - 1$ , the probability of fire is decreased for the year 0 or year +1 (Figure 4).



**Figure 4.** The occurrence and surface area of fire events tend to decrease over time, especially since the 1990s, with the reinforcement of the preventive measures. In terms of burnt surface, the year 2003 is highlighted, and is well known regarding the high intensity of the heat wave.

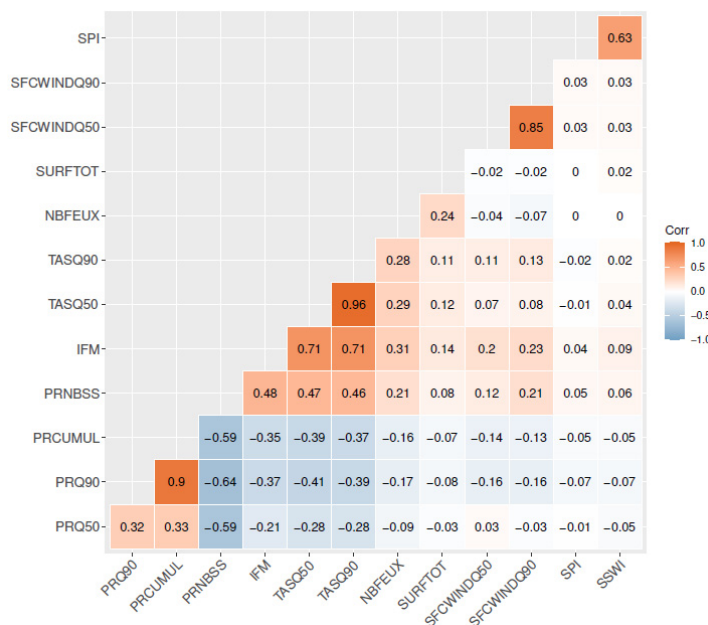
A large majority of the DFCI mesh has no or only two fire departures. On the contrary, some of the mesh has more than 100 fire departures over the entire studied period and over the years (Figure 5).



**Figure 5.** The two maps exemplify the diversity of the fire occurrence (left) and burnt surface (right) according to each DFCI mesh for the summer period in 2005.



For a daily analysis, the link between a high FWI and fire occurrence is important; nevertheless, when considering the average value of the FWI over three months, the link is not insured. The variability of the data poses an issue of extreme values, and adds complexity for the implementation of machine learning-based methods (see Section 3.2). In order to cope with the issue, the correlation matrix allows us to better understand the relationships and interdependence between the metrics. The number of fires is positively correlated with the FWI (0.31), the number of days without precipitation (0.21) and the mean temperature (0.29). The number of fires is, on the contrary, negatively correlated with the precipitation; in particular, the negative values are between  $-0.09$  and  $-0.17$ . This matrix also highlights that the SPI and SSWI indexes are not correlated with other variables, and especially with the targeted variables NBFEX and SURFTOT (values equal to 0). On the contrary, the variables representing the same data have an important coefficient of relation; for example, the variable for temperature TASQ50 and TASQ90 with a correlation of 0.96 (Figure 6).



**Figure 6.** The correlation matrix highlights the coefficient of correlation for the different possible pairs of variables. There is a positive correlation between the FWI and the Q50/Q90 of the daily temperature (0.71). We observed a correlation between the number of days without precipitation and the number of fires (0.21).

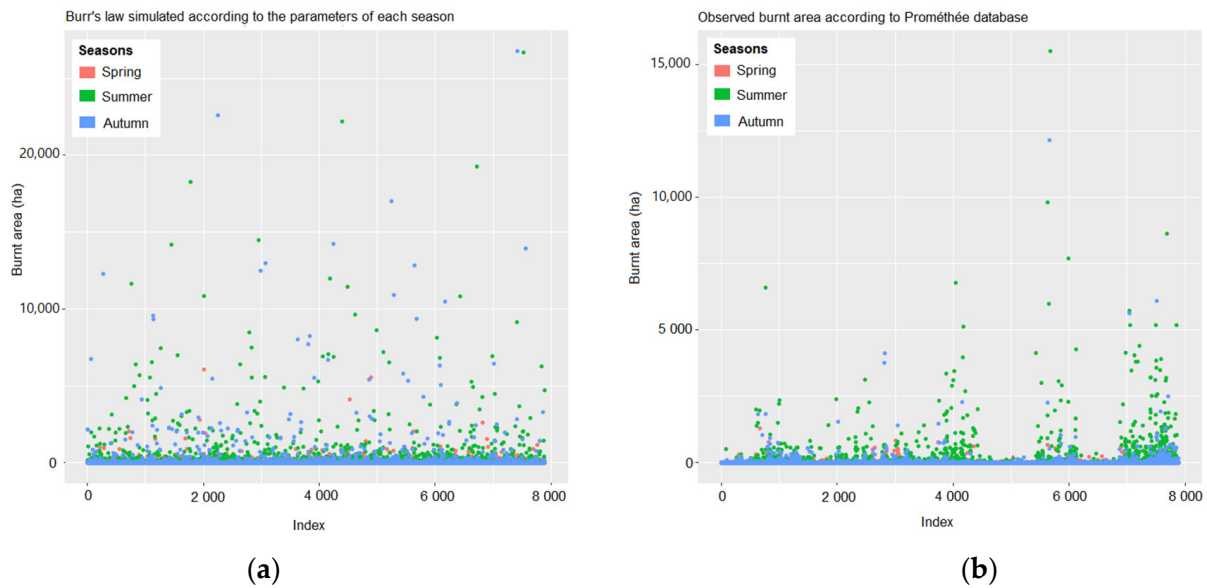
### 3.2. Hazard Unit

The results below show the comparison of the estimates from the machine learning-based methods between the real and modelled scenarios in order to predict the burnt surface and the probability of fire. We also integrated the projections at horizon 2050 under the RCP 4.5 IPCC Scenario.

#### 3.2.1. Burnt Surface

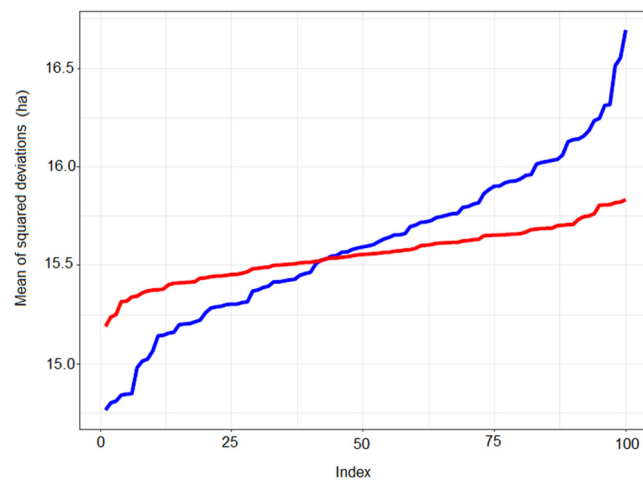
For this purpose, we started with the statistical approach based on Burr’s law. We obtained a similar distribution for the majority of the low-intensity fires, and for some of the extreme events. The quality of the simulation was determined using a quantile–quantile diagram. Figure 5 shows a soft Burr’s law overestimating the area of the burnt surface according to the seasons. The average error was 0.7. The densities are coherent with a very high probability of low fire intensity and a very low probability of extreme fire. The higher the fire, is the lower the probability is (Figure 7). It is a soft Burr’s law. We assumed an

interesting simulation, and we simulated with this law several thousand numbers in order to calculate the average burnt surface.



**Figure 7.** Burnt surface simulated with Burr’s law (a), and compared to the observed data from the Prométhée database (b).

In a second analysis, the neural network was tested for the estimation of the burnt surface. Figure 8 highlights the average absolute difference between our predicted values and the real ones. The average error on the learning base is in blue, and the error on the test base is given in red. The error on the test base is 15.6 ha. The use of the neural network highlights the need to focus on the prediction of the occurrence or absence of fire events.



**Figure 8.** Comparison of the squared deviation for the neural network based on the Prométhée database from 1973 to 2005 concerning the burnt surface.

Regarding the training data available and the machine learning outputs, we assume that Burr’s Law provides better results than the neural network concerning the estimation of the burnt surface.

### 3.2.2. Fire Occurrence

In order to model the fire occurrence, we tested three machine learning-based methods and one GIS-based approach.

First, we tried the decision tree. At the least, the learning base will contain 50% of the lines. We still obtained a high number of false negatives and badly classified results compared to the other machine learning methods (Figure 9). We could use a deeper tree, but the result wouldn't change much, and the tree would have too many leaves to be interpretable. The main variables are practically always the same, no matter which sample we take. The confusion matrix associated to the tree based on the learning base reveals 22.03% badly classified results; 40.01% false positives and 12.38% false negatives. The results are similar for the validation base, with 24.50% being badly classified, 43.56% false positives and 13.48% false negatives. Based on this method, the false negatives are too important and the results are insufficient.

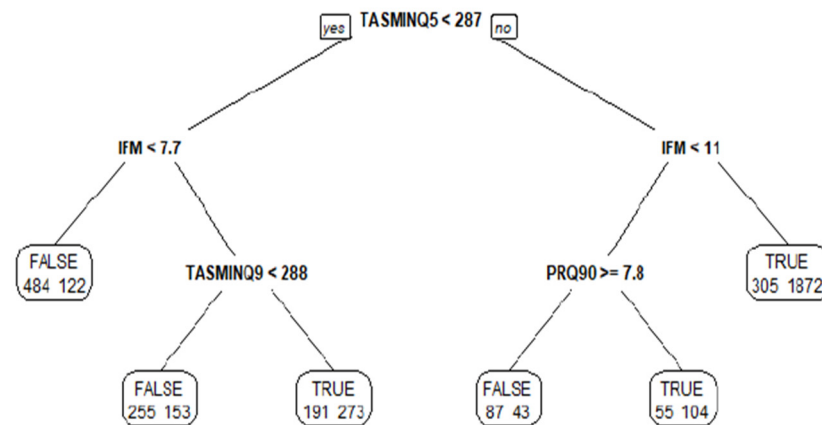


Figure 9. Extract of the decision tree developed for the study.

Next, we tested the Support Vector Machine (SVM). The objective was to split linearly the true and false results within the space of meteorological indexes (of dimension 14). The function svm of the package e1071 allows the realization of SVM by choosing among four nodal functions: linear, polynomial, radial or sigmoid. After different tests, the radial nodal function provides better results. Figure 10 provides the error rate on the learning base, and tests for a variation of the  $\gamma$  parameter. Among the three presented methods and related results, the three are similar in terms of performance. The more interesting rates to consider in the context of fire occurrence are the false negative and badly classified results, at around 14% (i.e., the number of fires that the model has not predicted). We assume that the models have good performance (Figure 10).

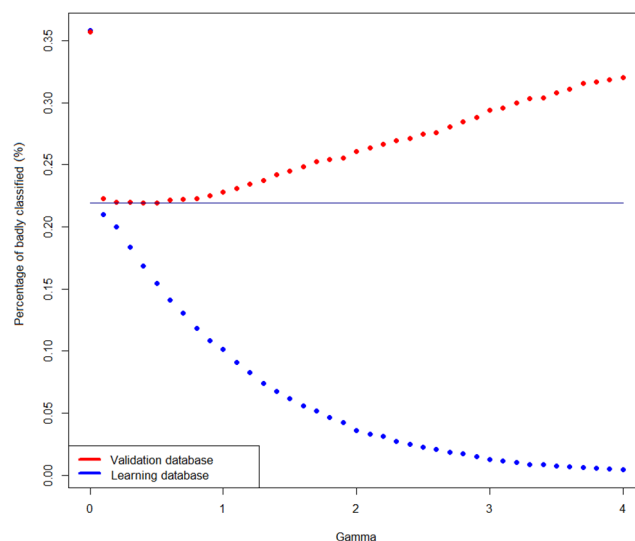
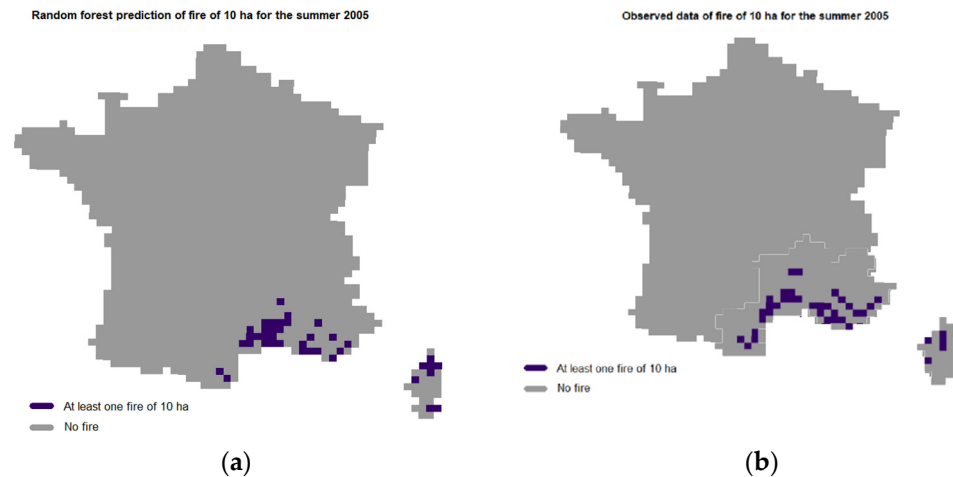


Figure 10. Error rates according to the parameter  $\gamma$  for a radial nodal.

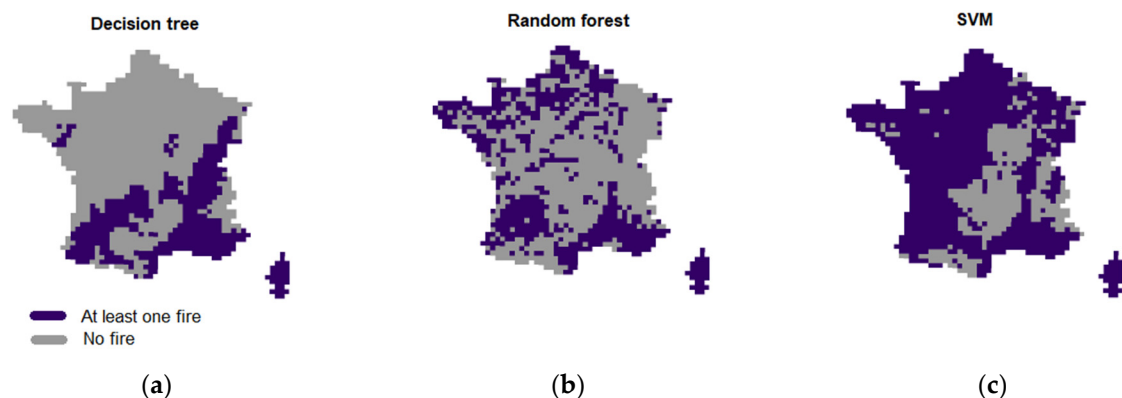
Then, in the continuity of the tests, we tested the random forest method. The implementation of the random forest of the learning base does not highlight false negatives and positives from phenomena of overlearning. Nevertheless, the random forest provides similar results from a dozen trees; the rate of bad classification was around 25% for a random forest predicting the occurrence of fires larger than 10 ha, the FWI, and the Q50 of the daily minimal temperature.

We also tried to estimate the prediction of a burnt area of more than 10 ha; the best model was a random forest (Figure 11). The results are coherent and applicable to the historical datasets and for the 2050 future climate.



**Figure 11.** The figure compares (a) the modelled prediction of fires of 10 ha for the summer of 2005 using a random forest, and observations from the same year (b).

We observe that the results of the projection are different according to the method of projection, as the models are calibrated on the Prométhée area, and are projected to the entire country (Figure 12). Each model has at least 20% badly classified results and 40% false positives.



**Figure 12.** Comparison of the results from the three machine learning tools regarding the probability of at least one fire within the 2 km DFCI mesh during summer 2050 RCP 4.5. (a) The decision tree projection is the more moderated model; we observed the exposure of Landes, Rhône valley and Vosges. (b) The random forest's projection highlights the Mediterranean and Corse areas' exposure, with an increasing exposure of the Rhône valley, Landes, Bretagne, Nord-Pas-de-Calais and Pays-de-Loire; this model seems to be the more coherent one. (c) SVM's projection is the more pessimistic model, generating fire events in the large majority of the territory; the results are more related to the RCP 8.5 pessimistic scenario.

The results are highly coherent for the Mediterranean area. We observe some similarities, especially for increasing exposure of Bretagne, Pays de la Loire, Centre Val de Loire, and the Atlantic coast (Landes forested areas), and an increasing exposure of the mediterranean area (Occitanie and Provence Alpes Côte d'Azur) [40].

Finally, the statistically significant hotspots and cold spots are represented on the map. The red areas indicate that, throughout the time, there is an aggregation of a high number of forest fires. The blue areas highlight a smaller number of fire events. Each hexagon is classified according to the timescale. The geographical analysis underlines the exposure of the areas nearby Béziers and Perpignan, which are areas of oscillating hot spots. The rest of the Mediterranean area is exposed in the same manner without a strong evolution in time (oscillating cold spots). For North Corse, at least 90% of the temporal intervals were statistically significant hotspots (Figure 13).

Emerging hot spot analysis – Prométhée database 1973–2019

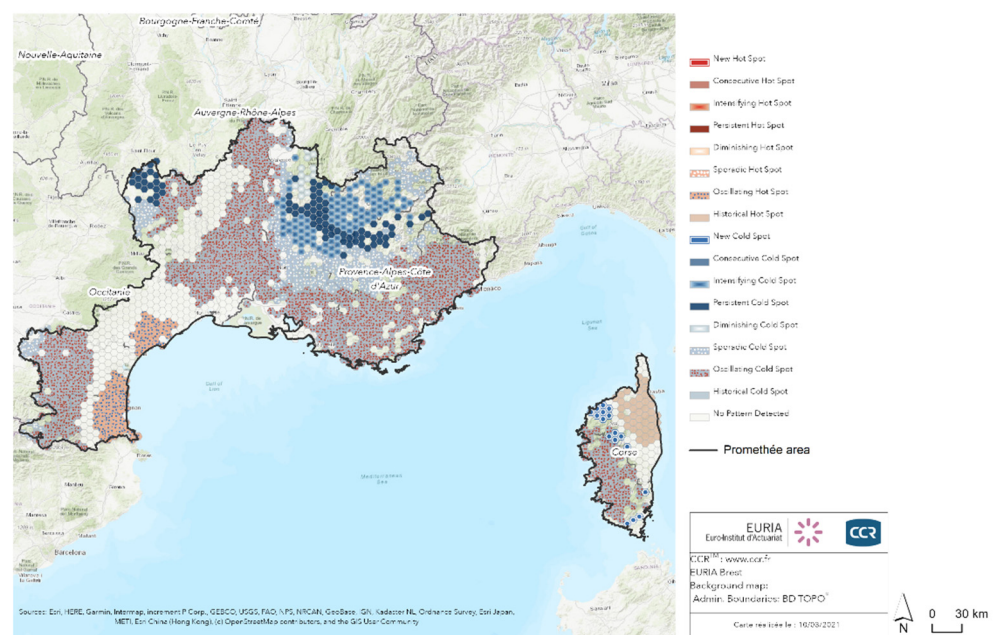


Figure 13. Emerging hot spot analysis for the Prométhée dataset 1973–2019.

### 3.3. Damage Model

The R code provides, for each year, a table with the potential costs of a fire for the DFCI. Within this model, we assume that the burnt surface is entirely inhabited. Under the hypothesis of the RCP 4.5 at horizon 2050 using a random forest model predicting the occurrence or not of fire greater than 10 ha, we estimate that the damage will be, in 2050, around 35 M€ on average for residential insured areas only. The evolution of the insured values is not considered, nor are the land-use changes. The spatial repartition of the future areas exposed to wildfire events further north in France highlights the increase of the economic exposure (Figure 10). This evolution can be compared to the results of Moncoulon et al. [26] on geotechnical drought and shrinking swelling clay (SSC), in that they reveal an increasing exposure of the southern communities, as well as those is the Atlantic area.

## 4. Discussion

Despite the many limitations associated with simulation modelling and machine learning-based methods in these experiments discussed above, the outputs from this work provide useful information on the exposure of France to wildfires. This work introduced the foundation of the Cat model for the assessment of forest fire exposure and projection to

horizon 2050. We acknowledge the limitations in the use of the model for the prediction of building exposure considering the spatial resolution chosen. First and foremost, our predictions were only applied to the RCP 4.5 scenarios. This study focused on wildfire disasters from an asset context, and we recognize that wildfires also threaten human lives and ecosystems, as well as cascading impacts on floods, landslides and potable water after severe fire events. Furthermore, the model could consider not only the physical damage but also the business interruption or critical infrastructure issues (highways, secondary roads, etc.). Furthermore, we did not consider the currently implemented fire prevention programs. We considered the exposure to be constant in the 2050 horizon. As the issue of population growth in littoral areas will be higher in the future, it will increase the exposure to fire events and greatly influence the WUI [26,41,42]. Regarding the damage model's development, wildfire-related claims data could support the determination of the damage functions; nonetheless, based on our knowledge, that kind of data is not available in France. Downscaling the model could also make the results and damage estimates more precise.

Although climate change is considered in the optimistic scenario RCP 4.5, the results of the future exposure are significant to start to raise and develop a risk culture in the future exposed areas (Bretagne, Alsace, etc.), and for the maintenance of the currently well-structured prevention processes in Southern France. There have been numerous studies that focused on the asset level and considered the fire conditions, landscape and properties' structures. Here, in order to obtain a global vision of the exposure of the entire area of France, we used simplistic models. Obtaining new data is challenging, as it requires waiting for future wildfires and potentially generating large losses, and ensuring that data of sufficient quality are collected. The relatively small number of fire events in terms of number or burned acres, and the very low number of burnt assets in France means that the applicability of the model developed overestimates the number, and the surface is not easily calibrated on the training data. Nevertheless, it offers a new visibility of France's exposure to that kind of natural disaster for the next few years. It provides elements for discussion on the issue of the underwriting of fire risk within the Nat Cat scheme.

## 5. Conclusions

This model synthesizes information for the French insurance sector, and contributes to understanding and reducing wildfire losses. CCR and Euria Brest developed a first-of-its-kind France Cat model projecting changes to wildfire potential under the RCP4.5 scenario at a granularity of about  $20 \times 20$  km. Finally, the best model for burnt-surface prediction is Burr's law, and the random forest for the fire occurrence.

In the future, this model could be combined with GIS analysis (within distance to vegetation, slopes, and fuel type) and with satellite imagery analysis in order to make the exposure analysis more precise.

New machine learning and remote sensing data could be used to develop specific damage curves for household buildings for the vulnerability models, such as those created for hurricane damage. This is the first time, as far as we know, that a reinsurance company developed, with an institute, a prototype model that links machine learning and insurance data, and applied these models to the estimation of the expected financial loss from wildfires. Likewise, the model could be applied to other countries, as well as the pessimistic scenario RCP 8.5. The different improvements will open the door to explore a wide range of exposure management in order to reduce the climate change impact, and to support the community for preventive measures. Strong decisions have to be taken in order to avoid making 2021 the last coolest year of the rest of our lives.

This paper offers new visibility for the improvement of the preparedness in future potentially affected areas. We hope that this work will support future potentially exposed areas to integrate the analysis within their disaster risk prevention and resilience plans. Robust cat models can improve the accuracy of the predictions of the locations of the greatest risks to assets, and could provide an indication of the implementation of preven-

tive measures. The proposed methodology could serve as a reference for wildfire risk assessment, and can be replicated elsewhere.

**Author Contributions:** Conceptualization, D.M. and R.M.; methodology, B.G., E.B.-S. and A.B.; software, B.G., E.B.-S. and A.B.; validation, D.M. and R.M.; formal analysis, B.G., E.B.-S. and A.B.; investigation, B.G., E.B.-S. and A.B.; resources, B.G., E.B.-S. and A.B.; data curation, B.G., E.B.-S. and A.B.; writing—original draft preparation, B.G., E.B.-S., A.B. and R.M.; writing—review and editing, R.M.; visualization, B.G., E.B.-S. and A.B.; supervision, D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Developed metrics from the overall database to be integrated within the machine learning tools.

Metrics	Details
NBFEUX	Number of fire departures
NBFEUX0_5HA	Number of fire >0.5 ha
NBFEUX1HA	Number of fire >1 ha
NBFEUX10HA	Number of fire >10 ha
NBFEUX100HA	Number of fire >100 ha
SURFQ50	Quantile 50 of the fire surface (ha)
SURFQ90	Quantile 90 of the fire surface (ha)
SURFMAX	Maximal surface of the fire (ha)
SURFTOT	Total burnt area (ha)
INBFEUX	Binear code: fire/no fire; whatever the burnt surface
INBFEUX0_5HA	Binear code: fire/no fire; burnt surface >0.5 ha
INBFEUX1HA	Binear code: fire/no fire; burnt surface >1 ha
INBFEUX10HA	Binear code: fire/no fire; burnt surface >10 ha
INBFEUX100HA	Binear code: fire/no fire; burnt surface >100 ha
TASMINQ50	Quantile 50 over 3 months of the minimal temperature
TASMINQ90	Quantile 90 over 3 months of the minimal temperature
TASMAXQ50	Quantile 50 over 3 months of the maximal temperature
TASMAXQ90	Quantile 90 over 3 months of the maximal temperature
TASQ50	Quantile 50 over 3 months of the mean temperature
TASQ90	Quantile 90 over 3 months of the mean temperature
PRQ50	Quantile 50 over 3 months of the daily precipitation
PRQ90	Quantile 90 over 3 months of the daily precipitation
PRCUMUL	Total precipitation over 3 months
PRNBSS	Number of days without precipitation
SFCWINQ50	Quantile 50 over 3 months of the wind speed
SFCWINQ90	Quantile 90 over 3 months of the wind speed
SPI	Meteorological drought for 3 months
SSWI	Soil wetness index for 3 months
FWI	Seasonal average from March to November

## References

1. Sfetos, A.; Giroud, F.; Clemencau, A.; Varela, V.; Freissinet, C.; Lecroart, J.; Vlachogiannis, D.; Politi, N.; Karozis, S.; Gkotsis, I.; et al. Assessing the Effects of Forest Fires on Interconnected Critical Infrastructures under Climate Change. Evidence from South France. *Infrastructures* **2021**, *6*, 16. [CrossRef]
2. Li, S.; Banerjee, T. Spatial and temporal pattern of wildfires in California from 2000 to 2019. *Sci. Rep.* **2021**, *11*, 1–17. [CrossRef]
3. Risk Management Solutions (RMS) Fundamentals of Wildfire Modeling. Virtual Online Meeting “Exceedance 2021”, May 2021. Available online: <https://www.rms.com/events/exceedance-2021> (accessed on 1 December 2021).

4. Ager, A.A.; Day, M.A.; Alcasena, F.J.; Evers, C.R.; Short, K.C.; Grenfell, I. Predicting Paradise: Modeling future wildfire disasters in the western US. *Sci. Total Environ.* **2021**, *784*, 147057. [CrossRef]
5. Wang, D.; Guan, D.; Zhu, S.; Mac Kinnon, M.; Geng, G.; Zhang, Q.; Zheng, H.; Lei, T.; Shao, S.; Gong, P.; et al. Economic footprint of California wildfires in 2018. *Nat. Sustain.* **2020**, *4*, 252–260. [CrossRef]
6. Whittaker, J.; Haynes, K.; Wilkinson, C.; Tofa, M.; Samson, S.; Dilworth, T.; Collins, J.; Tait, L. *Lessons from Black Summer: How People Experienced the 2019/20 NSW Fire Season; Hazard Note; Bushfire and Natural Hazards CRC: Merlbourne, VIC, Australia, 2021; p. 4.*
7. Phillips, C. How Wildfires Affect Climate Change—And Vice Versa. 2021. Available online: <https://theconversation.com/how-wildfires-affect-climate-change-and-vice-versa-158688> (accessed on 1 December 2021).
8. Canicule historique au Canada le village aux 49,6 °C détruit par le feu. *Le Monde.fr* 2021. Available online: [https://www.lemonde.fr/climat/article/2021/07/02/canicule-historique-au-canada-le-village-aux-49-6-c-detruit-par-un-incendie\\_6086626\\_1652612.html#:~:text=Canada-,Canicule%20historique%20au%20Canada%20%3A%20le%20village%20aux%2049%2C6%20%20%20ce%20record%20mardi%2029%20juin.&text=A%20260%20kilom%C3%A8tres%20de%20Vancouver,du%20nord%20Duouest%20du%20Canada](https://www.lemonde.fr/climat/article/2021/07/02/canicule-historique-au-canada-le-village-aux-49-6-c-detruit-par-un-incendie_6086626_1652612.html#:~:text=Canada-,Canicule%20historique%20au%20Canada%20%3A%20le%20village%20aux%2049%2C6%20%20%20ce%20record%20mardi%2029%20juin.&text=A%20260%20kilom%C3%A8tres%20de%20Vancouver,du%20nord%20Duouest%20du%20Canada). (accessed on 1 December 2021).
9. En Turquie et en Grèce, une canicule et des feux d’une ampleur historique. *Le Monde.fr* **2021**.
10. Mahmoud, H.; Chulahwat, A. A Probabilistic Cellular Automata Framework for Assessing the Impact of WUI Fires on Communities. *Procedia Eng.* **2017**, *198*, 1111–1122. [CrossRef]
11. Duff, T.J.; Penman, T.D. Determining the likelihood of asset destruction during wildfires: Modelling house destruction with fire simulator outputs and local-scale landscape properties. *Saf. Sci.* **2021**, *139*, 105196. [CrossRef]
12. Salis, M.; Arca, B.; Del Giudice, L.; Palaiologou, P.; Alcasena-Urdiroz, F.; Ager, A.; Fiori, M.; Pellizzaro, G.; Scarpa, C.; Schirru, M.; et al. Application of simulation modeling for wildfire exposure and transmission assessment in Sardinia, Italy. *Int. J. Disaster Risk Reduct.* **2021**, *58*, 102189. [CrossRef]
13. Jarry, M.; Campet, P. *Modélisation d’un Feu de Forêt; Master pro de didactique de mathématiques Mémoire de modélisation; Université Paris 7: Paris, France, 2010; p. 28.*
14. De Gennaro, M. *Modélisation de La Propagation Des Grands Incendies de Forêts et Élaboration d’un Outil Opérationnel d’aide à La Lutte Tactique. Ph.D. Thesis, Aix-Marseille University, Marseille, France, 2017.*
15. Badal-Valero, E.; Coll-Serrano, V.; Segura-Gisbert, J. Fire Risk Sub-Module Assessment under Solvency II. Calculating the Highest Risk Exposure. *Mathematics* **2021**, *9*, 1279. [CrossRef]
16. *IFRC Shelter Safety Handbook Some Important Information on How to Build Safer; IFRC: Geneva, Switzerland, 2011; p. 62.*
17. *The Zurich Flood Resilience Alliance PERC A Burning Issue: Insights for Resilience from Three Wildfire Events; Institute for Social and Environmental Transition—International: Boulder, CO, USA, 2020; p. 28.*
18. San-Miguel-Ayanz, J.; Durrant, T.; Boca, R.; Libertà, G.; Branco, A.; de Rigo, D.; Ferrari, D.; Maianti, P.; Vivancos, T.A.; Oom, D.; et al. *Forest Fires in Europe, Middle East and North Africa 2017 2018; European Commission: Brussels, Belgium, 2019; ISBN 978-92-76-11234-1.*
19. Evin, G.; Curt, T.; Eckert, N. Has fire policy decreased the return period of the largest wildfire events in France? A Bayesian assessment based on extreme value theory. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2641–2651. [CrossRef]
20. Vuorio, A.; Kovanen, P.T.; Budowle, B.; Sajantila, A.; Palo, J.U.; Stoop, J. Wildfire-Related Catastrophes: The Need for a Modern International Safety Investigation Procedure. *Front. Clim.* **2021**, *3*. [CrossRef]
21. Jolly, W.M.; Cochrane, M.A.; Freeborn, P.H.; Holden, Z.A.; Brown, T.J.; Williamson, G.J.; Bowman, D.M.J.S. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* **2015**, *6*, 7537. [CrossRef]
22. Norman, B.; Newman, P.; Steffen, W. Apocalypse now: Australian bushfires and the future of urban settlements. *npj Urban Sustain.* **2021**, *1*, 1–9. [CrossRef]
23. Xu, R.; Yu, P.; Abramson, M.J.; Johnston, F.H.; Samet, J.M.; Bell, M.L.; Haines, A.; Ebi, K.L.; Li, S.; Guo, Y. Wildfires, Global Climate Change, and Human Health. *N. Engl. J. Med.* **2020**, *383*, 2173–2181. [CrossRef]
24. Varela, V.; Vlachogiannis, D.; Sfetos, A.; Karozis, S.; Politi, N.; Giroud, F. Projection of Forest Fire Danger due to Climate Change in the French Mediterranean Region. *Sustainability* **2019**, *11*, 4284. [CrossRef]
25. Murphy, K.P. *Machine Learning: A Probabilistic Perspective; MIT Press: Cambridge, MA, USA, 2012; ISBN 978-0-262-30432-0.*
26. Moncoulon, D.; Desarthe, J.; Naulin, J.-P.; Onfroy, T.; Tinard, P.; Wang, Z.-X.; Hajji, C.; Veysseire, M.; Dequé, M.; Régimbeau, F. *Conséquences du changement climatique sur le coût des catastrophes naturelles en France à l’horizon 2050; Caisse Centrale de Réassurance & Météo France: Paris, France, 2018; p. 31.*
27. Core Writing Team; Pachauri, R.K.; Meyer, L.A. (Eds.) *IPCC Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014.*
28. Van Wagner, C.E. Development and Structure of the Canadian Forest Fire Weather Index System. *Can. For. Serv.* **1987**, *48*. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.460.3231> (accessed on 1 December 2021)
29. Bergeron, Y.; Flannigan, M.D. Predicting the effects of climate change on fire frequency in the southeastern Canadian boreal forest. *Water, Air, Soil Pollut.* **1995**, *82*, 437–444. [CrossRef]
30. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.W.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]



31. Touzet, C. LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME; Collection de l'EERIE; Marseille, France, EC2, 1992. Available online: <https://hal-amu.archives-ouvertes.fr/hal-01338010/document> (accessed on 1 December 2021).
32. Zhang, W.; Li, H.; Li, Y.; Liu, H.; Chen, Y.; Ding, X. Application of deep learning algorithms in geotechnical engineering: A short critical review. *Artif. Intell. Rev.* **2021**, *54*, 5633–5673. [[CrossRef](#)]
33. Zhang, W.; Phoon, K.-K. Editorial for Advances and applications of deep learning and soft computing in geotechnical underground engineering. *J. Rock Mech. Geotech. Eng.* **2022**. [[CrossRef](#)]
34. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: Abingdon, UK, 2017; ISBN 9781315139470.
35. Breiman, L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat. Sci.* **2001**, *16*, 199–231. [[CrossRef](#)]
36. St, J.P.O.E.D.; Dumais, S.; Osuna, E.; Platt, J.; Schölkopf, B. Support vector machines. *IEEE Intell. Syst. Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
37. Wang, J.; Chen, Q.; Chen, Y. RBF Kernel Based Support Vector Machine with Universal Approximation and Its Application. In *Advances in Neural Networks—ISNN 2004*; Yin, F.L., Wang, J., Guo, C., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3173.
38. Kraak, M. The Space-Time Cube Revisited from a Geovisualization Perspective. In Proceedings of the 21st International Cartographic Conference (ICC), Durban, South Africa, 10–16 August 2003.
39. Westhaver, A.; Taylor, S.W. *Developing a Method for Conducting Wildland/Urban Interface Fire Case Study Research*; Institute for Catastrophic Loss Reduction: Toronto, ON, Canada, 2020; p. 86.
40. Chatry, C.; Le Quentrec, M.; Laurens, D.; Le Gallou, J.-Y.; Lafitte, J.-J.; Creuchet, B. *Changement Climatique et Extension Des Zones Sensibles Aux Feux de Forêts*; Ministère de l'Alimentation, de l'Agriculture et de la Pêche, Ministère de l'Intérieur, de l'Outre-mer et des collectivités territoriales; Conseil Général de l'Environnement et du Développement Durable: Paris, France, 2010; p. 190.
41. Wu, C.; Venevsky, S.; Sitch, S.; Mercado, L.M.; Huntingford, C.; Staver, A.C. Historical and future global burned area with changing climate and human demography. *One Earth* **2021**, *4*, 517–530. [[CrossRef](#)]
42. Wu, Z.; Wei, R.; Chu, Z.; Liu, Q. Real-time rock mass condition prediction with TBM tunneling big data using a novel rock-machine mutual feedback perception method. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 1311–1325. [[CrossRef](#)]