CSIRO PUBLISHING

RESEARCH PAPER
https://doi.org/10.1071/WF23162

International Journal of WILDLAND FIRE
International Association of Wildland Fire

Pattern recognition and modelling of virulent wildfires in Spain

María Bugallo[A,*], María Dolores Esteban[A], Manuel Francisco Marey-Pérez[B] and Domingo Morales[A]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
María Bugallo
Center of Operations Research, Miguel Hernández University of Elche, Elche, Alicante, Spain
Email: mbugallo@umh.es

Received: 10 October 2023
Accepted: 28 January 2025
Published: 24 February 2025

Cite this: Bugallo M *et al.* (2025) Pattern recognition and modelling of virulent wildfires in Spain. *International Journal of Wildland Fire* **34**, WF23162. doi:10.1071/WF23162

© 2025 The Author(s) (or their employer(s)). Published by CSIRO Publishing on behalf of IAWF.
This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND)

OPEN ACCESS

ABSTRACT

Background. Driven by global warming and land use, fire dynamics are changing worldwide, increasing fire activity and its impacts on ecosystems, livelihoods and human settlements. In southern European countries, large forest fires are increasing annually and a high spatio-temporal concentration has been shown to overwhelm fire suppression systems. **Aims.** Propose sound statistical criteria for grouping fire ignitions based on auxiliary information. **Methods.** Our methodology combines pattern recognition and complex statistical modelling that, by capturing underlying dependencies and adapting to the nature of the data, provides a reliable statistical basis for formulating firefighting strategies. **Key results.** Weekly and provincial data from 2007 to 2015 are used to illustrate the performance of the procedures. **Conclusions.** Research indicates that weather conditions, the simultaneous occurrence of events and the proximity of urban areas contribute to making the Spanish landscape highly susceptible to wildfires in summer. **Implications.** Risk mapping is key to fire management and efficient resource allocation.

Keywords: fire management, megafire, pattern recognition, risk map, Spanish data, statistical modelling, wildfire forecasting, zero-inflated gamma mixed model.

Introduction

Many studies have documented that wildfires are becoming larger and more destructive around the world (Singleton *et al.* 2019; Duane *et al.* 2021; Kraus *et al.* 2022; Stoof *et al.* 2024). In recent years, the term virulent fire has been widely used to refer to wildfires that are extremely intense and difficult to control (Jain *et al.* 2022; Mansoor *et al.* 2022). These fires are often fuelled by hot, dry and windy weather conditions, making them more dangerous and difficult to fight (Xanthopoulos *et al.* 2020; Castel-Clavera *et al.* 2022; Sakellariou *et al.* 2023). Moreover, the expected increase in their abundance owing to climate change (Senande-Rivera *et al.* 2022; Grünig *et al.* 2023) and the abandonment of traditional land uses, such as extensive livestock farming (Lasanta *et al.* 2022), severely limit forest management (Marey-Pérez *et al.* 2021; Rodríguez *et al.* 2022; Arellano-del-Verbo *et al.* 2023). The abandonment of rural areas has led to an increase in dry vegetation, and the extreme conditions of high maximum temperatures and atmospheric stability have also led to the appearance of pyroconvection phenomena (Ribau *et al.* 2024). The latter imply a new dimension in the study of large wildfires, both from the point of view of behavioural research (Duane *et al.* 2021; Castellnou *et al.* 2022) and extinguishing difficulty (Eghdami *et al.* 2023; Lee *et al.* 2023).

Mediterranean countries have a high risk of large fires, with Portugal, Greece, Spain and Italy at the top of the list (Meier *et al.* 2022). As a result, there are many studies focusing on different regions of the Iberian Peninsula or on Spain as a whole. Among these, we mention some of particular interest. First of all, Poisson mixed models provide predictions of the number of fires by forest area (Boubeta *et al.* 2015, 2016) and logistic regression models deal with the presence or absence of forest fires (Martínez-Fernández *et al.* 2013; Rodrígues *et al.* 2014; Ríos-Pena *et al.* 2015). Generalised linear and additive models allow the study of spatial variation and human factors influencing the occurrence and nature of wildfires (Reid *et al.* 2015). Spatial correlation models and pattern recognition techniques can be used to model burned surfaces (Pereira *et al.* 2015; Kale

et al. 2017; Moanga *et al.* 2020). Last but not least, zero-inflated negative binomial mixed models deal with excess zeros and have already been successfully used in several studies to analyse fire data in Spain (Viedma *et al.* 2018; Bugallo *et al.* 2023).

A key factor in the occurrence and evolution of current and future wildfires is the presence of villages, towns and buildings (Bento-Gonçalves and Vieira 2020). This is where the wildland–urban interface (WUI) comes into play, both in its various original definitions (Radeloff *et al.* 2005; Stewart *et al.* 2007) and in more recent ones adapted to Mediterranean conditions (Alcasena Urdíroz *et al.* 2018; Badia *et al.* 2019). Indeed, it is well known that the presence of buildings in the vicinity of fires leads to a bias in firefighting guidelines. Recent studies have highlighted the global expansion of the WUI (Chen *et al.* 2024) and the increasing proximity of wildfires to buildings in regions such as California and Spain (Marey-Pérez *et al.* 2024). These fires exhibit different behaviour compared with those in forested areas (Purnomo *et al.* 2024), requiring the development of specialised protocols for firefighting operations. As a result, forest fires that are more dangerous for urban areas are prioritised over those in more isolated regions (Cruz and Alexander *et al.* 2019; Kuligowski *et al.* 2022).

Broadly speaking, the WUI is the area where human structures intersect with wild vegetation. In this context, the presence of buildings near or within forests, and the size, speed of spread and difficulty of eradication of wildfires justify structural changes in these areas (Prichard *et al.* 2021; Marey-Pérez *et al.* 2021; Wolpert *et al.* 2022; Pulido *et al.* 2023), or in the behaviour of the population (Fernández-Blanco *et al.* 2022; Lim 2022; Iglesias *et al.* 2022*b*; Xanthopoulos *et al.* 2022) and also in both (Copes-Gerbitz *et al.* 2022; Canadas *et al.* 2023). Furthermore, the simultaneity of events is crucial in the allocation of firefighting resources (Papathoma-Köhle *et al.* 2022), and all this leads to an increased likelihood of larger wildfires (Turco *et al.* 2018; Marey-Pérez *et al.* 2022; Meier *et al.* 2023).

The ecological implications of wildfires remain a topic of ongoing debate (Bargali *et al.* 2024; Sánchez-Pinillos *et al.* 2024), with further discussions on how these ecological consequences extend into the economic and social realms (Lloret *et al.* 2024). It is well established that wildfires significantly impact these areas (Poduška and Stajić 2024). Additionally, an increasing number of extreme wildfires are highlighting the inherent physical limitations of fire suppression capabilities and demonstrating that fire suppression-focused strategies are not effective in addressing these new challenges. Therefore, there is a broad consensus among policymakers and the scientific community on the need to shift the focus from fire suppression to fire prevention (Andrade and Hulse 2022; Granville *et al.* 2022), which is embedded in the concept of *FireSmart* (Iglesias *et al.* 2022*a*; Meacham and McNamee 2022; Campos *et al.* 2022), i.e. the intelligent management of forest fires.

A lack of foresight and capacity to respond to large fires in the near future may lead to severe deforestation problems, which, in combination with climate change, may lead to the collapse of natural ecosystems. Catastrophic events require accurate multidimensional prediction models (McColl-Gausden *jet al.* 2022) with different time scales (Gannon *et al.* 2020; Palaiologou *et al.* 2021) and different techniques (Bot and Borges 2022; Guyeux *et al.* 2022; Nasiri *et al.* 2022) at regional (Khanmohammadi *et al.* 2023) and national scales (Chew *et al.* 2022; Tedim *et al.* 2022; Al-Khuraisat 2023). Given the difficulty of rapidly deploying resources, it is becoming increasingly popular to implement new interventions at a larger scale of response, from regional to national and ultimately European scale.

As a follow-up to the study by Bugallo *et al.* (2023), the present research aims to develop a virulent fire prediction model at a provincial spatial and weekly temporal scale, and also plans to define a risk scale to map the results. For this purpose, data for all forest fires recorded in Spain at the provincial level were available for all weeks of the years 2002–2015 inclusive. Therefore, aggregated data are used, i.e. totals and averages in crosses between provinces and weeks. In this context, Tedim *et al.* (2020) introduce the concept of the War on Fire, which refers to all activities – including fire management and suppression – undertaken by modern societies to prevent fires from occurring or, if they do occur, to prevent them from overwhelming the system. The concept of 'collapse' refers to a situation where fire suppression efforts exceed society's capacity to manage the fire over a prolonged period. Specifically, the goal is to characterise the potential spread of a forest fire and establish statistical thresholds to identify when the system might fail.

Data and problem of interest

Spain is a country on the Iberian Peninsula in the southwest of Europe, sharing a maritime border with North Africa and partly surrounded by the Mediterranean Sea. It has two archipelagos, the Canary Islands and the Balearic Islands, and two autonomous cities, Ceuta and Melilla (Fig. 1). Owing to their respective insular and urban conditions, it is common to treat the archipelagos separately and to exclude both cities from forestry studies. The works of Linley *et al.* (2022) and Stoof *et al.* (2024) establish broad definitions and interpretations of what is currently considered a 'megafire' by applying terms of size, behaviour or intensity. In the case of Spain, and for the purposes of this research, we focus specifically on the size of the wildfire. Specifically, a megafire is defined as any wildfire that burns 500 ha of forest area in the peninsula, and 250 ha in the archipelagos (MITECO 2023). Megafires are defined at the unit level, although we work with average wildfires at annual crossings between provinces and weeks, which means smoother results requiring lower thresholds. In territorial terms, the province is an administrative demarcation

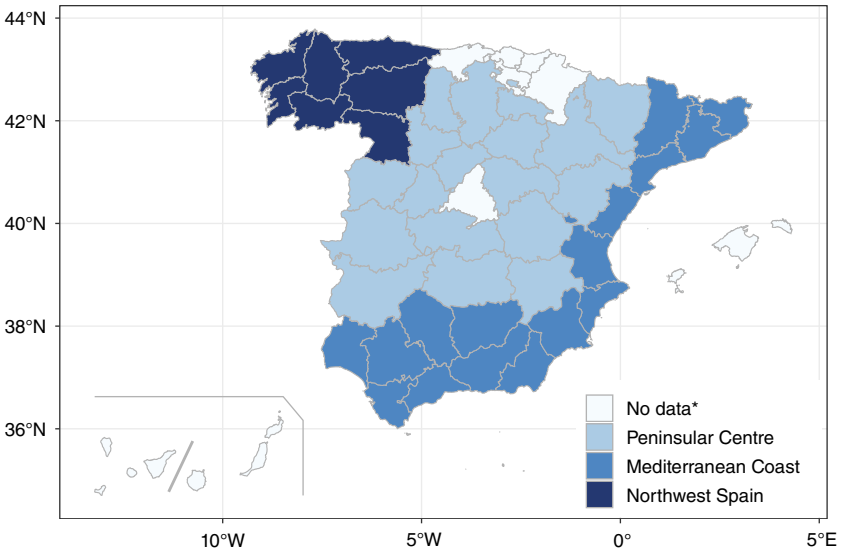

Fig. 1. Political map of Spain at the provincial level, including the Iberian Peninsula and the Canary and Balearic Islands. Territorial divisions are taken from MITECO (2023). The 'No data' category refers to provinces that do not provide data on the distance between fires and human buildings up to the last years of our study or they are geographically imprecise. These provinces have been left out.

with competences in environment and forest fire management so it is the unit of spatial aggregation in this paper. Each province can be classified into one of three main regions on the basis of its fire regime: Northwest Spain, Mediterranean Coast and Peninsular Centre (MITECO 2023).

Target variables and aggregation structure

Data are from the General Forest Fire Statistics (GFFS 2023), from the national agency of the European Forest Fire Information System, which provides services related to wildfires in Spain. GFFS data after 2015 are being cleaned and not available for research purposes. In terms of data recording, it is important to note that a wildfire is located in the province where it originated and therefore the ignition point is crucial for the distribution of events between provinces. The same applies to weekly distribution.

Some basic notation is introduced below. The aim is to model the total burned area y_{ijk}, measured in hectares (ha), for year $i \in \{1, \ldots, I\}$, week $j \in \{1, \ldots, J\}$ and province $k \in \{1, \ldots, K\}$. That is a non-negative continuous variable that ranges over $[0, \infty]$ and is defined as the sum of the areas burned in the n_{ijk} wildfires that took place in year i, week j and province k. In addition, we are interested in inferring the average of the burned areas $\bar{y}_{ijk} = y_{ijk}/n_{ijk}$ for year i, week j and province k. It is said that $\bar{y}_{ijk}$ denotes the average forest fire. Let $D = IJK$ be the total number of crosses. Compared with other studies, our target variables are weekly and provincial totals and averages. The estimation domains are the crosses between years, weeks and provinces. The latter has both advantages and disadvantages. On the one hand, it provides general guidelines for action and measures of risk at an appropriate scale. On the other hand, the level of aggregation makes the results applicable only to 'expected average fires' in a given domain. Large isolated wildfires can therefore go undetected.

For weeks, we use the standard ISO 8601 (2019) notation, which states that the first week of the year begins on the Monday of the week containing 4 January. Therefore, all weeks have 7 days, and there may be years with 52 or 53 weeks (not necessarily leap years). Given the seasonal nature of forest fires, the study covers July, August, September and October, focusing on data recorded between weeks 27 and 44 inclusive. The problem is first posed for $D = IJK = 12{,}600$ domains, defined by the crosses of years ($I = 14$), weeks ($J = 18$) and provinces ($K = 50$). Ceuta and Melilla are excluded.

Domain-level auxiliary information

Broadly speaking, weekly and provincial weather conditions and, to some extent, the socio-economic situation of a country and its investment in firefighting resources and support are critical factors in identifying patterns in the study of wildfires. In this context, a model-based approach uses data from other domains and underlying correlation structures between the available auxiliary variables to model and predict the target variable. Supplementary Appendix S2 describes the area-level model that is fitted to our data.

Starting with the climatological variables, their selection process was divided into three steps. Firstly, the most representative automatic meteorological stations were filtered from the AEMET database (https://www.aemet.es/es/datos_abiertos/AEMET_OpenData). A pre-processing task was then carried out to ensure that the dataset was homogeneous in terms of location and scale. Daily averages were calculated, taking into account that more than one meteorological station might be available. Weekly averages were then calculated. Finally, missing values were replaced by weekly and provincial medians.

The selection of the candidate variables for inclusion in the models is consistent with previous findings documented

in wildfire research. Giannaros and Papavasileiou (2023) studied which atmospheric conditions explain the causality of current megafires in Europe, with special attention to Mediterranean Europe. Their findings, also based on the wildfire literature, indicate that there is huge regional variability with a pattern that is the entry of an approaching upper-air trough ahead of a collapsing upper-air ridge, as occurred in the megafires of Portugal in 2017 (Ramos *et al.* 2023), associated with the passage of Hurricane Ophelia. In addition, Raposo *et al.* (2023) investigated how drought influenced new megafires. Specifically in Europe, there are numerous publications that analyse how drought, sometimes associated with seawater temperature, portends large wildfires. In addition, wind speed and heat waves are crucial because they are directly related to the formation of pyrocumulus (Werth *et al.* 2016). However, there are other conditions that are decisive in the spread of flames over a large area, which has led to the consideration of a wide set of auxiliary variables. One would expect that wind direction and temperatures could also alter the evolution of a forest fire (Keeley and Syphard 2019; Prichard *et al.* 2020).

Table 1 summarises the final set of area-level climatological variables, including their correlations with $\log\bar{y}_{ijk}$, denoted by $\hat{\rho}$, and the confidence intervals (CI) for ρ. We calculate the correlation with $\log\bar{y}_{ijk}$ because the link function in the zero-inflated models to be fitted is logarithmic (see Supplementary Appendix S2). Moreover, this is consistent with the nature of our response variables (average and total burned areas by domains), as they are non-negative and with right-weighted tails. We exclude null counts (domains where no forest fires have been observed), because the logarithm function is not well defined at zero.

Accounting for domains with a positive total burned area, we found statistically significant linear dependencies at 95% for all variables in Table 1 and the logarithm of the average burned area. Pearson's correlation coefficients and asymptotic CIs are included.

The sign of the correlation coefficients associated with prec and hr is worth commenting on. The higher the rainfall and humidity, the smaller the average fire size. The opposite applies for all other climatological auxiliary variables. Although these variables were initially of count type, they are interpreted as measures scaled to the mean and are unitless.

Secondly, we had information on provincial and weekly average distances between fires and human buildings. Therefore, we decided to work with two variables, bui1 and bui10, as described in Table 2. The Global Human Settlement Layer (GHSL 2014) was used to obtain these. Where no fire was observed, distances were missing. In this situation, it was decided to assign a default value of 0 because, if there are no fires, it can be interpreted that there is no physical distance between them. At this stage, we had to remove some areas to have precise coordinates, as these regions do not provide them until the last years of our study or they are geographically imprecise. Fig. 1 shows the political map of Spain at the provincial level, with the deleted provinces highlighted. To include data from Galicia, a region particularly affected by forest fires, the first year considered was 2007. As a result, our research is finally posed for $D = IJK = 6642$ domains, defined by the crosses of years ($I = 9$), weeks ($J = 18$) and provinces ($K = 41$).

Finally, there are the variables of firefighting means. As for their pre-processing, it was decided to standardise these in order to avoid problems of location and scale. The same transformation was applied to the variable n.fir, which counts the domain number of forest fires. Thus, although these variables were initially of count type, they are interpreted as measures scaled to the mean and are unitless. This ensures that the fitting algorithm of Supplementary Appendix S2 is not affected by convergence problems. Table 2 describes the non-climatological explanatory variables and their grouping into five categories: distance (2), firefighting staff (4), land-based machinery (4), aerial machinery (4) and fire count (1).

Table 2 also shows the Pearson correlation coefficients between the non-climatological explanatory variables and the logarithm of the average burned area for domains with positive outcomes. CIs at 95% confidence level are also

Table 1. Climatological auxiliary variables used in the application to the 2007–2015 GFFS weekly data.

Variable	Description	Units	$\hat{\rho}$	CI_ρ
dir	Direction of maximum wind speed	°∠	0.063	0.035, 0.091
prec	Accumulated average precipitation	mm	-0.107	-0.135, -0.080
sol	Duration of insolation	h	0.218	0.191, 0.245
tmax	Average maximum temperature	°C	0.190	0.163, 0.217
tmed	Average mean temperature	°C	0.127	0.099, 0.155
tmin	Average minimum temperature	°C	0.038	0.010, 0.066
wmed	Average speed of 07, 13, 18 UTC	m/s	0.078	0.050, 0.106
hr	Average relative humidity	Tenths of mm	-0.259	-0.285, -0.232

Description, units and Pearson's correlation coefficients with $\log\bar{y}_{ijk}$.

Table 2. Non-climatological auxiliary variables used in the application to the 2007–2015 GFFS weekly data.

Variable	Group and description	Units	$\hat{\rho}$	CI_ρ
bui1	Distance between the fire and the nearest human building	km	0.123	0.094, 0.150
bui10	Average distance between the fire and the 10 nearest human buildings	km	0.130	0.102, 0.158
tch.st	Firefighting staff – technicians and/or forestry agents		0.403	0.379, 0.427
brg.st	Firefighting staff – brigade personnel		0.381	0.357, 0.405
sec.st	Firefighting staff – state security forces		0.286	0.260, 0.311
oth.st	Firefighting staff – others, e.g. volunteers		0.216	0.190, 0.243
atb.eq	Land-based machinery – no. of fire engines		0.317	0.291, 0.342
bll.eq	Land-based machinery – no. of bulldozers		0.391	0.367, 0.414
trc.eq	Land-based machinery – no. of tractors		0.252	0.226, 0.278
oth.eq	Land-based machinery – no. of other machines		0.207	0.180, 0.234
amp.air	Aerial machinery – no. of amphibious aircraft		0.366	0.342, 0.390
car.air	Aerial machinery – no. of cargo aircraft		0.249	0.223, 0.276
ext.air	Aerial machinery – no. of firefighting helicopters		0.209	0.182, 0.236
tra.air	Aerial machinery – no. of transport helicopters		0.458	0.436, 0.480
n.fir	Fire count – no. of fires		0.212	0.185, 0.239

Description, units and Pearson's correlation coefficients with $\log\bar{y}_{ijk}$.

given. The linear relationship is significant for the two distance variables and for the standardised number of fires. In addition, there is a significant positive linear correlation between all variables related to firefighting equipment and the average burned area. Further, these statements will be reaffirmed by pattern recognition and burned area modelling. However, the Pearson correlations are of a marginal nature. Therefore, these estimates provide indicative results, but do not necessarily translate into significant dependencies in the data analysis.

Data analysis and statistical results

Keeping practicality first, this section applies the proposed methodology, focusing on the results and consigning the mathematical developments to Supplementary Appendices S1 and S2. Supplementary Appendix S1 describes the clustering algorithm, the *K*-means algorithm. Supplementary Appendix S2 describes the area-level zero-inflated Gamma (aZIG) mixed model, uses the Laplace approximation algorithm to calculate the maximum likelihood (ML) estimators of the model parameters, gives predictors of the target quantities and develops bootstrap procedures to estimate mean squared errors (MSEs). The construction of bootstrap percentile CIs of the model parameters is also provided. The data analysis is thus divided into two distinct branches, which are in fact two sides of the same coin. Pattern recognition has its origins in engineering and tends to be more applied in its application. However, the model-based approach has a solid statistical basis for modelling and prediction that successfully supports the findings.

Pattern recognition and clustering

The clustering algorithm in Supplementary Appendix S1 allows the identification of domains with a markedly anomalous profile. Specifically, representing each cluster by a single mean vector (cluster centre), the *K*-means algorithm is used. Table 2 shows the centre, size and distribution of the average megafires in the clusters. To find a favourable trade-off between interpretability, complexity and variability (measured by comparing within-cluster and between-cluster sums of squares), three clusters were considered (Forgy 1965). As recommended, random sets of different observations were repeatedly selected as initial centres.

The clustering algorithm identifies domains with a markedly anomalous profile. There is one cluster consisting of only five domains (0.07%). A slightly larger cluster consists of 34 observations (0.50%). The others form a large cluster with 6603 observations (99.41%). The average burned area in cluster 1 exceeds the threshold of 500 ha. The same is true for 48% of the observations in cluster 2 and for none in cluster 3. Results show that 0.07 + 0.50 = 0.57% of the studied events comprise the problem and it is on these that modelling should focus. Thus, the *K*-means algorithm successfully detects average megafires and supports the need for models capable of quantifying and forecasting the spatio-

temporal risk of extreme events. From Table 2, average megafires could be summarised as follows:

- **Weather conditions** – low rainfall and humidity, plenty of sunshine, higher than expected temperatures and strong winds. Wind direction is not discriminating.
- **Distances** – proximity to urban settlements more influential than isolated human buildings.
- **Firefighting resources (standardised variables)** – need for more personnel, ground equipment and air support. The differences with the average of the extinguishing systems are higher for megafires. Indeed, they skew the mean – not robust enough – and force it to be slightly negative for cluster 3, which captures fires with 'more common' patterns.
- **Simultaneity of fires (standardised variable)** – the more forest fires, the more virulent they are. In cluster 1, large forest fires occur in a period with fewer events than average, but the other fires in cluster 2 are generally associated with higher simultaneity.

Zero inflation

So far, we have discussed the need for auxiliary information, but we have not addressed the problem of excess zeros (Zuur *et al.* 2009; Michael and Thomas 2016). In this context, biologists, ecologists and engineers, to name just a few, often deal with target variables that contain more zeros than would be expected if the process generating their data came purely from a standard probability distribution (Feng and Dean 2012; Feng *et al.* 2020). Sometimes the zeros are not really zeros, but very small values that were not observed. Zero-inflated outcomes are very common in forestry databases, where excess zeros indicate undetectable events. For example, very small forest fires are not recorded in government databases. But there are also plausible scenarios where the observed values are very close to zero, for example, the total burned area in a week of very low insolation, high rainfall and unusually low thermal amplitude. Some engineers and statisticians have therefore modelled this peculiarity of the data. Recent related work includes Viedma *et al.* (2018) and Bugallo *et al.* (2023) for wildfires in Spain, and Tan *et al.* (2020) for an Indonesian scenario.

Our model-based approach, detailed in Supplementary Appendix S2, involves modelling a zero-inflation parameter that represents the probability of a given zero originating from the main distribution or being an excess zero, using generalised linear mixed models (GLMMs). Tables 3, 4 are included to illustrate the suitability of using a zero-inflated mixed model.

Table 4 shows the provincial percentiles of the total and average burned area for weeks 27–44 and years 2007–2015. As can be seen, at least 20% of the data are equal to 0, which is an extremely high percentage for continuous distributions, such as the Gamma distribution. This is a strong motivation for modelling burned forest areas by also including latent variables to account for excess zeros. Moreover, the total burned area has much greater dispersion than its average from the number of fires, as expected. Table 5 depicts the proportion of zeros per year. The percentage of zeros per year is close to 27% and seems to level off. As a result, it follows that the zero-inflated structure is stable over the years.

Model-based statistical analysis

Statistical modelling helps us to analyse what triggers an increase in the magnitude of wildfires and to understand recent changes in fire patterns. As a follow-up to the study by Bugallo *et al.* (2023), the aim is to model virulent fires at provincial spatial and weekly time scales and to define risk measures. Based on the area-level zero-inflated Gamma (aZIG) mixed model described in Supplementary Appendix S2, this section analyses the variables total burned area, y_{ijk} and average burned area, $\bar{y}_{ijk}$, in year $i \in \{1, \ldots, I\}$, week $j \in \{1, \ldots, J\}$ and province $k \in \{1, \ldots, K\}$, accounting for excess zeros.

Broadly speaking, the aZIG model has two submodels with Bernoulli and Gamma distributions, called the BE submodel and the GA submodel, respectively. The first submodel is a GLMM that focuses on the probabilities of zero or at least one wildfire, whereas the other submodel deals with burned area modelling. Because of aggregation, $u_{1,jk}$ and $u_{2,jk}$, $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$ are considered as random effects dependent on week and province crosses. Therefore, the dependence structure implies $2JK = 1476$ random effects.

For ease of presentation, the aZIG models for totals and averages are denoted aZIGT and aZIGA, respectively. Starting from the model fitting results, the ML parameter estimates of the regression parameters β_1 and ϕ_1 for the BE submodel are shown in Table 5, and the ML parameter estimates of the regression parameters β_2 and ϕ_2 for the GA submodels in Table 5. They also display the P values to test $H_0{:}\beta_{t\ell} = 0$, $t = 1, 2$, $\ell = 1, \ldots, q_t$ and $H_0{:}\phi_t = 0$, $t = 1, 2$, and the normal-asymptotic and bootstrap CIs at the 95% confidence level. For convenience, their lower (LB) and upper (UB) bounds are provided. Normal-asymptotic CIs and bootstrap CIs are discussed in Supplementary Appendix S2.

The models are fitted to data from 2007 to 2014, keeping 2015 to test near-future retroconditions. The final models incorporate only those variables that are significant at 1%. As a result, the BE submodel, fitted with the climatological variables of Table 1, contains $q_1 = 5$ covariables: $x_{1,1}$ = prec, $x_{1,2}$ = tmax, $x_{1,3}$ = tmed, $x_{1,4}$ = wmed, $x_{1,5}$ = hr. The GA submodel of model aZIGT contains $q_1 = 11$ covariates: $x_{2,1} =$, $x_{2,2}$ = bll.eq, $x_{2,3}$ = trc.eq, $x_{2,4}$ = amp.air, $x_{2,5}$ = car.-air, $x_{2,6}$ = tra.air, $x_{2,7}$ = prec, $x_{2,8}$ = tmax, $x_{2,9}$ = tmed, $x_{2,10}$ = wmed, $x_{2,11}$ = hr. To fit this model, we give more value to the resource variables and omit distances. The

Table 3. Results of the *K*-means algorithm described in Supplementary Appendix S1.

	dir	prec	sol	tmax	tmed	tmin	wmed	hr	bui1	bui10	tch.st	brg.st	sec.st
1	16.360	1.702	9.660	30.976	24.886	18.808	3.188	53.000	1.292	2.530	12.456	8.700	19.366
2	20.795	1.561	11.311	32.066	24.295	16.530	2.776	44.206	1.884	3.673	4.219	3.716	4.910
3	9.546	8.705	8.912	26.756	20.337	13.918	2.554	57.896	1.086	2.266	-0.031	-0.026	-0.040
	oth.st	atb.eq	bll.eq	trc.eq	oth.eq	amp.air	car.air	ext.air	tra.air	n.fir	$\bar{y}_{ijk}$	size	count
1	2.272	6.321	7.118	4.139	8.015	8.873	4.711	2.134	10.235	-0.209	3710.843	5	5
2	2.521	3.759	6.064	1.803	1.877	4.560	2.014	2.445	3.681	0.026	656.026	34	16
3	-0.015	-0.024	-0.037	-0.012	-0.016	-0.030	-0.014	-0.014	-0.027	0.000	5.919	6603	0

Cluster centres and sizes (size) and count (count) of average megafires for the three-group case

Table 4. Percentiles of total (top) and average (bottom) monthly and provincial burned areas for weeks 27–44 of the years 2007–2015.

	q_0	$q_{0.1}$	$q_{0.2}$	$q_{0.3}$	$q_{0.4}$	$q_{0.5}$	$q_{0.6}$	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	q_1
Total	0	0	0	0.200	0.900	2.200	5.080	11.307	27.048	80.092	15,256.210
Average	0	0	0	0.190	0.500	0.932	1.600	2.800	5.073	12.305	4674.110

Table 5. Proportion of zeros per year, i.e. domains with no reported forest fires.

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015
Proportion	0.257	0.295	0.203	0.276	0.168	0.324	0.251	0.339	0.314

Domains are defined as crossings between years, weeks and provinces.

Table 6. Maximum likelihood estimates of the regression parameters for the two Bernoulli (BE) submodels of the two area-level zero-inflated Gamma (aZIG) mixed models.

		BE submodel					
		β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	ϕ_1
	Estimate	0.015	-0.285	0.235	-0.447	0.041	1.141
	P value	0.000	0.000	0.000	0.000	0.000	0.000
Asymp	LB 95%	0.010	-0.339	0.167	-0.559	0.035	1.020
	UB 95%	0.020	-0.232	0.302	-0.335	0.046	1.275
Boot	LB 95%	0.011	-0.351	0.183	-0.555	0.036	1.015
	UB 95%	0.020	-0.248	0.317	-0.330	0.047	1.267

Models fitted with 2007–2014 data aggregated by province and week.

opposite is true for the aZIG model for domain averages. The GA submodel of model aZIGA contains $q_2 = 7$ covariates: $x_{2,1}$ = n.fir, $x_{2,2}$ = bui1, $x_{2,3}$ = prec, $x_{2,4}$ = tmax, $x_{2,5}$ = tmed, $x_{2,6}$ = wmed, $x_{2,7}$ = hr. The estimates of the scale parameters are $\nu_T = 0.808$ for the GA submodel of model aZIGT and $\nu_A = 0.805$ for the GA submodel of model aZIGA. There is no fixed intercept in either of the three submodels and the climatological variables of the two GA submodels are the same.

In general, the LBs, UBs and widths of the bootstrap CIs in Tables 6 and 7 are similar to those of the asymptotic CIs, suggesting that the theoretical distribution is close to the asymptotic one. Both the BE submodel and the GA submodels incorporate aggregated information on precipitation, high and average temperatures, wind and humidity. Although the interpretation of the estimated regression parameters is generally reasonable, it is important to note that both their sign and magnitude can change – sometimes significantly. Therefore, the results should be understood as being contingent on the proposed model. As the mathematical model is based on the current data, the inclusion of additional variables, different time periods, or changes in the dataset could lead to different outcomes, both in terms of sign and relevance.

It can be seen in Table 6 that for both y_{ijk} and $\bar{y}_{ijk}$, the Laplace approximation algorithm returns the same estimates for the BE submodel. The reason is that the objective

Table 7. Maximum likelihood estimates of the regression parameters for the two Gamma submodels of the two area-level zero-inflated Gamma (aZIG) mixed models.

		Total burned area – model aZIGT											
		β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}	β_{28}	β_{29}	β_{210}	β_{211}	$\phi_{2,T}$
	Estimate	0.316	0.384	0.274	0.267	0.171	0.336	0.008	0.356	-0.347	0.301	-0.016	1.347
	P value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Asymp	LB 95%	0.236	0.305	0.221	0.192	0.104	0.266	0.003	0.316	-0.400	0.226	-0.020	1.266
	UB 95%	0.397	0.462	0.326	0.342	0.239	0.407	0.012	0.396	-0.294	0.377	-0.012	1.433
Boot	LB 95%	0.290	0.334	0.217	0.188	0.107	0.295	0.002	0.308	-0.395	0.218	-0.019	1.261
	UB 95%	0.407	0.437	0.305	0.311	0.225	0.414	0.011	0.388	-0.285	0.369	-0.012	1.416

		Average burned area – model aZIGA							
		β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}	$\phi_{2,A}$
	Estimate	0.243	0.237	0.006	0.115	-0.080	0.331	-0.021	1.023
	P value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Asymp	LB 95%	0.191	0.188	0.002	0.079	-0.128	0.257	-0.025	0.959
	UB 95%	0.296	0.287	0.010	0.152	-0.032	0.406	-0.017	1.091
Boot	LB 95%	0.207	0.173	0.002	0.083	-0.121	0.253	-0.026	0.926
	UB 95%	0.310	0.252	0.011	0.152	-0.027	0.393	-0.018	1.077

Models fitted with 2007–2014 data aggregated by province and week.

function to be maximised is additively separable and the optimisation of the BE submodel summand does not depend on y_{ijk} or $\bar{y}_{ijk}$. Furthermore, the BE submodel is an area-level BE mixed model (aBE) with random effects that depend on week and province crosses and are stable over years. If we fit the aBE model directly to the binary target variables indicating zero or more than one wildfire event and apply the Laplace approximation algorithm, we obtain the same estimates as for the completed models aZIGT and aZIGA.

At the same time, we have modelled the target variables taking into account that they are semi-continuous, i.e. with many exact zeros and continuous positive outcomes. Linear models for normally distributed variables are the simplest and most commonly used statistical models. However, linear models are not appropriate for positive variables with asymmetric distributions. Here are some reasons for choosing a Gamma mixed model for the conditional target data $y_{ijk}|n_{ijk} > 0$ and $\bar{y}_{ijk}|n_{ijk} > 0$. If the response variable is positively skewed, a model based on the normal distribution does not take place. The Gamma distribution is appropriate when the response variables take values in $(0, \infty)$, where small values are expected to have small variability and large values are expected to have large variability. The link function of the Gamma GLMM is logarithmic. One reason is that these models assume multiplicative effects of the predictors on the original outcome and are easier to interpret. This is not the case with the canonical link of the Gamma GLMM (Lee *et al.* 2010). In addition, it is suitable to reduce the variability of positive variables with some small and some unusually high values. Finally, the usual Gamma GLMM assumes that the shape parameter is constant, just as the normal linear model assumes a constant variance. Likewise, we have maintained this assumption.

Influence and diagnostic analysis

Influence and diagnostic analysis are appropriate tools to validate a model, but also to provide valuable information for interpreting the underlying relationships in a dataset. Model residuals are defined for these purposes. As a reminder of some of the most important elements, we have one aBE model and two aZIG models, but we will restrict ourselves to the second. Up to this point, it is important to make the following comments.

Model aZIGT

Because of the range of the response variable, calculations are limited to the logarithmic scale. Consequently, raw residuals are the differences between the log-observed values and the log-fitted values. For standardised residuals, the sample mean is subtracted and divided by the sample standard deviation. At this point, there is a problem with null counts because the corresponding residuals are not defined. They must be omitted or an artificial value must be assigned to log(0). As we are interested in zero inflation, both options are problematic. This drawback provides strong arguments against this model.

Model aZIGA

Calculations can be performed at the original scale and both raw and standardised residuals are well defined for all domains. Values are coherent and interpretable.

Owing to the aforementioned advantages and limitations, we pay more attention to model aZIGA, whose estimated parameters are provided in Tables 6 and 7 (bottom). The aim is to investigate whether there are unfavourable cases with established tolerance limits and to link them to average megafires. Namely, of the 5904 observations, only 18 are average megafires, but they are particularly interesting and severely distort the residual plots. They must therefore be analysed separately.

Fig. 2 plots the standardised residuals of the aZIGA model, stressing their magnitude and with average megafires coloured in dark blue. As expected, it is challenging to accurately fit such events, which adds great value to our research. Indeed, average megafires are more volatile and their residuals skyrocket. Consequently, the model underpredicts them. In contrast, residuals belonging to null observations are conveniently close to zero. In addition, the majority of the standardised residuals are in the interval $[-3, 3]$. We conclude that the larger the response variable is, the larger the residuals of the aZIGA model are. Finally, no patterns are observed over the years.

Concerning the predicted random effects, most of these lie in the interval $[-3, 3]$ and do not differ from the $N(0, 1)$ distribution, although some atypical values were found. Moreover, their mean is close to zero and the estimated variances $\hat{\phi}_1$ and $\hat{\phi}_2$, as mentioned above, are significantly positive, which is further evidence to consider GLMMs. Even more, including week per province random effects helps to better explain the underlying variability between areas that is not explained by fixed effects.

2015 Spanish retroconditions

For a given forest fire, its proximity to the nearest human building is known and a sufficiently accurate short-term (e.g. weekly) weather forecast is feasible. It is also possible to plan the number of firefighting resources available to extinguish it and the sum of the remaining wildfires that are active at that time. Once the week and the province have been set, and the explanatory variables are known, the proposed models can be used to predict the total and average burned area in different provinces. For this purpose, we use the plug-in predictor defined in Supplementary Appendix S2.

There are two approaches. One is to use the fitting period of the data to assess the goodness-of-fit of the model. Another option is to work under hypothetical scenarios, i.e. artificial values of the explanatory variables and simulated target quantities. However, it is quite difficult to reproduce the variability of the real process, so we would have no guarantee that the results would be close to a realistic background. For this reason, we decided to set aside 2015 ($I = 9$) and use the available information to make retroconditions for the near future. Accordingly, we took the actual scenario of 2015 as the forecast scenario.

Fig. 3 shows line charts of the observed area-level values and retroconditions of the 2015 forest fires, based on the aZIGT and aZIGA models. The results have been averaged on a provincial basis according to the territorial division of Spain shown in Fig. 1. Thus, what is shown is the average behaviour of the provinces over time, as a summary measure of the three specified regions (Northwest Spain, Peninsular Centre and Mediterranean Coast). As a result, we conclude that the averaged observations and retroconditions follow similar patterns over time. Fig. 3 also indicates that in 2015, the month of July was, on average, the period with the highest number of reported medium to large fires, with a decreasing trend until the last month considered, October. We do not include a non-aggregated plot because the cardinality of the set of crosses between weeks and provinces clouds the interpretation.

We also provide estimates of the MSE of the plug-in predictor for the 2015 retroconditions. To do this, we use the resampling method described in Algorithm 3 and we consider both aZIGT and aZIGA models. Fig. 4 shows the relative root-MSE (RRMSE) estimates for domains of model

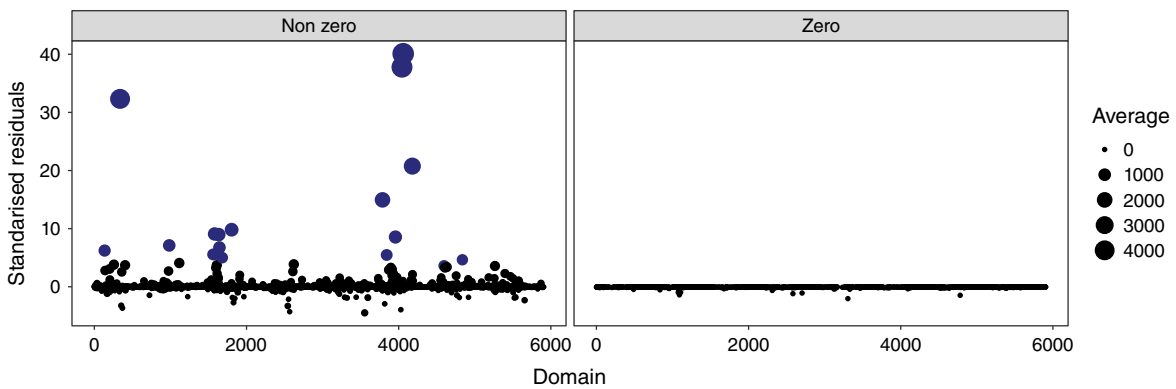

Fig. 2. Standardised residuals of the area-level zero-inflated Gamma mixed model for averages (aZIGA) by domain, for non-zero (left) and zero (right) outcomes. Average megafires are highlighted in blue and the size of the dots is proportional to the average burned area, in hectares.

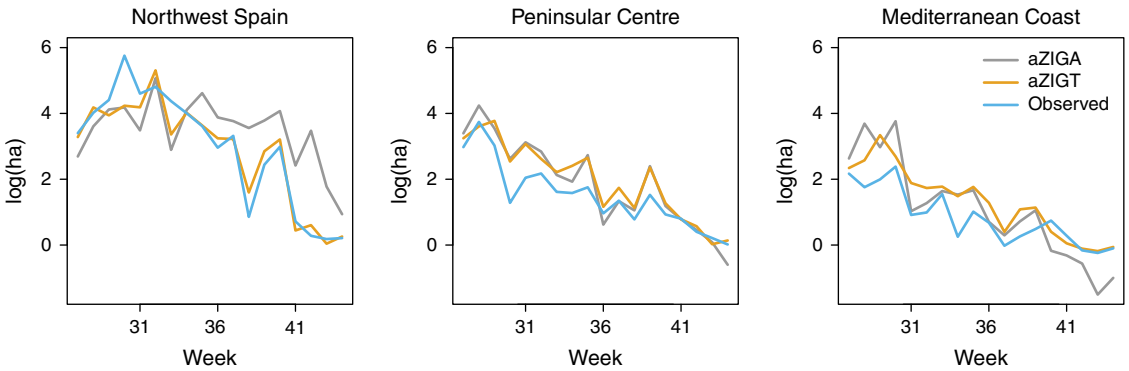

Fig. 3. Spanish forest fire plug-in retroconditions for 2015 over the study weeks (from July to October) and aggregated by the three predefined regions of the Iberian Peninsula.

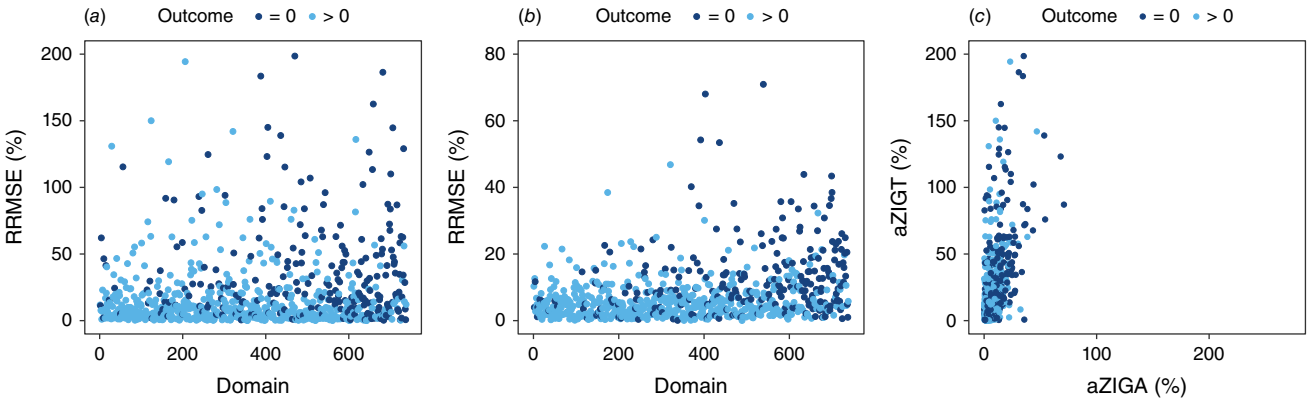

Fig. 4. Bootstrap estimates of the relative root-mean squared error (RRMSE) for the 2015 Spanish forest fire plug-in retroconditions by domain, i.e. by year, week and province: (*a*) model aZIGT, (*b*) model aZIGT, and (*c*) compared with each other.

aZIGT (left), model aZIGA (centre) and one model versus the other (right), all in percentage. To analyse what happens, a distinction is made between zero and non-zero observations. The aZIGT model performs worse, which is in line with the problems of fitting and treatment of large outcomes mentioned in the data analysis. For totals, the main quartiles are $q_0 = 0.036$, $q_{0.25} = 3.752$, $q_{0.5} = 10.085$, $q_{0.75} = 25.342$ and $q_1 = 198.558$. In contrast, the aZIGA model is much better, with considerably lower RRMSE estimates. For averages, the main quartiles are $q_0 = 0.039$, $q_{0.25} = 2.564$, $q_{0.5} = 5.386$, $q_{0.75} = 10.187$ and $q_1 = 70.940$. As can be inferred from Fig. 4 (right), domains with higher RRMSE estimates match for both models, although they are higher for totals. These are either domains with no records or very small aggregated values at weekly and provincial scales, both in quantity and magnitude. This is mainly explained by the relative nature of the RRMSE. In addition, it is difficult to fit a model when the distribution of the data across provinces and weeks is so dissimilar. However, our results are reassuring: it is desirable to predict better in those domains with larger average forest fires, as the severity of the environmental problem is of much more concern.

Finally, because one of our goals is to quantify the risk of large forest fires associated with both adverse weather conditions and the location and confluence of events, we devised a measure of risk. In this regard, we modelled total and average burned areas, which makes it difficult to detect extreme phenomena. Unfortunately, it is impossible to predict the presence or absence of average megafires. Indeed, only three domains in 2015 recorded average forest fires that exceeded the threshold of 500 ha, rising to 15 if we relax the threshold to 100 ha, and to 30 for 50 ha. Nevertheless, these values are consistent: for weekly and provincial averages, the results are smoothed. Even if several megafires are registered in a certain week and province, the weekly and provincial averages are also affected by other wildfires registered in that time.

Table 6 shows the weekly distribution of provinces where fires averaged more than 50 ha of burned area in 2015. Clearly, July stands out (weeks 27–30) and that is why we will illustrate our findings for this month. In contrast, the count of fires exceeding 50 ha burnt on average is practically null in all weeks of September and October because of seasonality. As for the provinces, we should mention Cáceres (3), Ciudad Real (2), León (2), Ourense (2), Palencia (2) and Zamora (2). No province in south-eastern Spain is listed.

For province i and week j, $i = 1, \ldots, I$, $j = 1, \ldots, J$, we develop a risk scale associated with the event that a wildfire burns more than 50 ha in 2015. So the process is based on 2015 forecasts, but can be applied to any year with available auxiliary information. As a starting point, bootstrap estimates of the quantiles of the predicted average burned areas, $\hat{q}_{ijk,\alpha}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, are calculated for $\alpha \in \{0.05, 0.15, 0.2, 0.3\}$, according to Algorithm 3. For each $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, we set 50 ha as a threshold to introduce the risk scale: *medium–low* if $\hat{q}_{ijk,0.30} \leq 50$ (<70%), *moderate* if $\hat{q}_{ijk,0.20} \leq 50 < \hat{q}_{ijk,0.30}$ (70–80%), *high* if $\hat{q}_{ijk,0.15} \leq 50 < \hat{q}_{ijk,0.20}$ (80–85%), *very high* if $\hat{q}_{ijk,0.05} \leq 50 < \hat{q}_{ijk,0.15}$ (80–85%) and *extreme* if $\hat{q}_{ijk,0.05} > 50$ (>95%). The risk conditions establish a hierarchy over the Spanish provinces, providing weekly horizon prospects.

Fig. 5 shows the risk maps for weeks 28, 29 and 30 of 2015 (when more average large fires were observed according to Table 8). For the study period, the presence of constant provincial patterns over time is quite evident, highlighting the high risk associated with the provinces of Galicia and Extremadura, as well as some provinces of Castilla y León. In these regions, the detection of a forest fire is an alarm signal, because there is a high probability that it will grow considerably and, in particular, that it will burn more than 50 ha. The opposite is true for the provinces on the Mediterranean Coast and the whole of Andalucía. Thus, our results are consistent with the facts: north-western Spain is the region most affected by the frequency and intensity of forest fires, owing to its large forest mass. It is important to note that the level of aggregation (province and week) makes the results applicable only to 'expected average fires' in a given domain (province, week and year). Therefore, this approach does not allow for the detection of individual events.

Fig. 6 measures the accuracy of our results by reporting the RRMSEs, defined in Step 3 of Algorithm 3, for each province and week. Graphically and analytically, we confirm that the aZIGA model is quite accurate: ~90% of the provinces have RRMSE estimates below 10% for weeks 28, 29 and 30 of 2015. Moreover, the relative error is particularly small for the higher-risk provinces. However, it is important to note that we are providing expected risk probabilities based on foreseen scenarios, and the variability in the data is fairly high.

Discussion

The goal of reducing forest fire devastation depends on accurate management-level information. A key challenge is to fully understand the conditions that lead to uncontrollable wildfires and to develop prevention strategies tailored to specific circumstances. Our approach provides solid evidence for defining risk measures by week and province and offers an initial approximation for describing megafires using readily available, unspecific data. Thus, our proposal represents a competitive contribution to state-of-the-art spatio-temporal fire prediction models, many of which rely on Machine Learning (Bot and Borges 2022; Chen 2022; Khanmohammadi *et al.* 2022; Sadatrazavi *et al.* 2022; Shao *et al.* 2022; Shmuel and Heifetz 2022; Yu *et al.* 2022; Abdollahi and Pradhan 2023) or other statistical techniques (Rodríguez and Marey-Pérez 2022; Marey-Pérez *et al.* 2022; Pozo *et al.* 2022; Ujjwal *et al.* 2022; Koh *et al.* 2023; Troumbis *et al.* 2023; Bugallo *et al.* 2023).

The interpretation of the fitted models depends on the available data, time periods and chosen modelling domains. Although performance was reasonable, it may vary as the model does. The same applies to the clustering results, as the analysis reflects the current situation. Therefore, careful selection of auxiliary information is crucial for effective mathematical modelling (Humi 2017). Cluster analysis has been widely used to identify forest fire patterns. For example, Pereira *et al.* (2015) and Sousa *et al.* (2021) applied these techniques to evaluate fire spatio-temporal aggregation in Portugal. Our results show that 0.50% of the events studied account for the majority of the problem, with modelling concentrated on these cases.

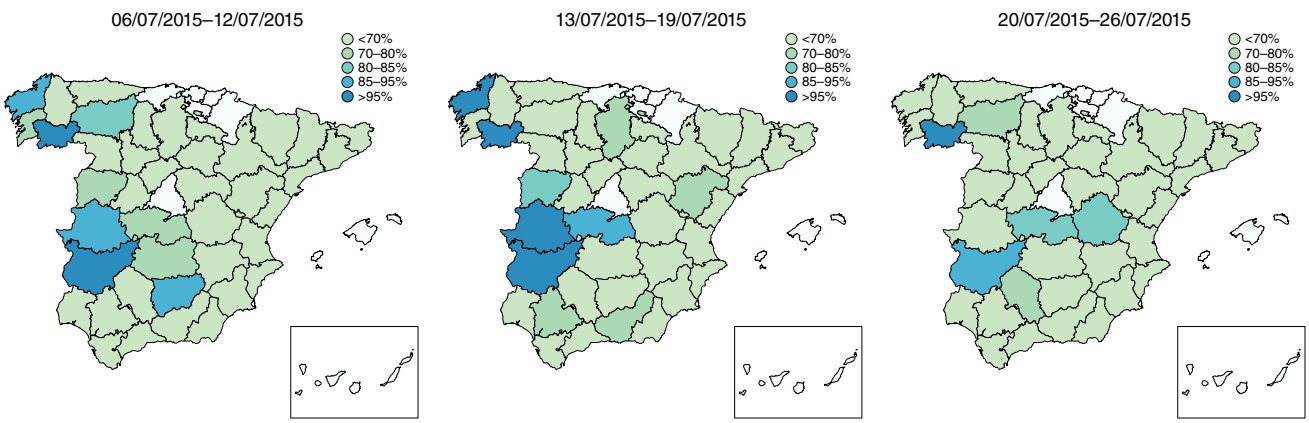

Fig. 5. Risk maps for weeks 28, 29 and 30 of the year 2015. The mapped values are calculated according to Algorithm 3 described in Supplementary Appendix S2.

Table 8. Weekly distribution of wildfires that, on average, burned more than 50 ha.

Week	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
No of provinces	5	7	4	4	1	3	1	1	3	0	1	0	0	0	0	0	0	0

Domains are identified as crosses of provinces, months and years. Then average wildfires are defined.

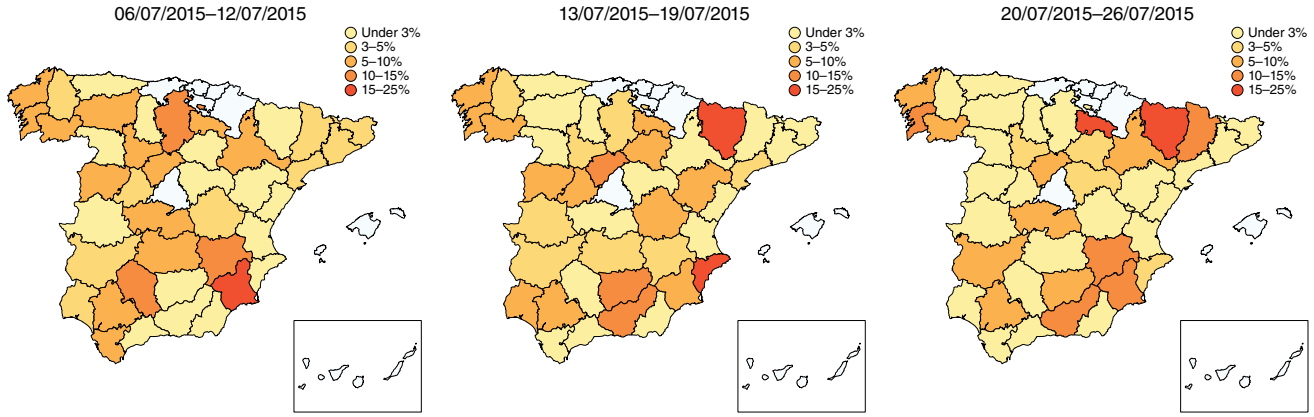

Fig. 6. Bootstrap estimates of the relative root-mean squared error (RRMSE) for the 2015 Spanish forest fire plug-in retroconditions and the first 3 weeks of July: 28, 29 and 30.

Against this background, there is no shortage of evidence that average global temperatures are rising, which our models suggest will have a significant impact on future wildfire patterns. Haque *et al.* (2023) note that one of the main drivers of the 2019–2020 'Black Summer' in Australia was the increase in average temperature over the Australian continent. Furthermore, Wu *et al.* (2022) point out that mean temperature is the most important variable among climatic factors. Although it was clear that certain temperature (tmax, tmed, tmin) and humidity (prec, sol, hr) conditions are necessary for fires to start, they do not explain the origin of arson fires. Nevertheless, different authors have included climatological variables in their models (Ríos-Pena *et al.* 2018; Jiménez-Ruano *et al.* 2019; Boubeta *et al.* 2019; Rodrígues *et al.* 2020; Rodríguez-García *et al.* 2022; Bugallo *et al.* 2023) and, recently, for the case of virulent wildfires (Rodrígues *et al.* 2023). Traditionally, most forecasting models take into account the maximum temperature (Chou 1992; Reid *et al.* 2010) and the level of vegetation moisture, which is crucial in the increase of fire intensity (Carmo *et al.* 2022; Kirkland *et al.* 2023), to which must be added the difficulty of extinguishment (Ni and Gernay 2022) and the high risk (Peña *et al.* 2022; Leone *et al.* 2023). Another typical element is wind (Moritz *et al.* 2005). Therefore, we include the direction of the maximum wind speed (dir) and the average speed at 07, 13, 18 UTC (wmed) per week and province, associated with continuous situations of intense wind.

Finally, many authors have studied the spatio-temporal aggregation of forest fires (Moreira *et al.* 2023; Crist 2023) and concluded that there is usually not a single large fire, but several simultaneous fires. For the study area, the publications related to August 2006 (Merino de Miguel *et al.* 2010; Barreiro and Hermosilla 2013; Ríos-Pena *et al.* 2017, 2018) and October 2017 (Sobrino *et al.* 2019; Chas-Amil *et al.* 2020; Llorens *et al.* 2021) stand out. These studies conclude that large fires cause secondary fires, leading to a collapse of firefighting resources. Other papers (Zambon *et al.* 2019; Colantoni *et al.* 2020; D'Este *et al.* 2020; Soto *et al.* 2022) have shown that many of these fires start close to cities. Research in these WUI areas (Glickman and Babbitt 2001) is a hot topic (Winandy *et al.* 2022; Zong and Tian 2022; Bar-Massada *et al.* 2023). In this sense, our results show how the distance between the fire and the nearest human building (bui1) contributes to the average burned area. This influence can be explained in the context of the difficulty of extinguishing fires: those closer to human settlements are more difficult to extinguish and require more effort to do so. This in turn means that wildfires themselves increase the area affected because they are not fought, as also noted by Gaudet *et al.* (2020) and Bachantourian *et al.* (2023). Our results are consistent with the Fire Paradox concept defined by Hudson (2011), where efforts in fire planning and suppression in Spain, and other Western countries, lead to biomass accumulation, increasing fire risk. As recent studies (Martínez *et al.* 2023; Aslan *et al.* 2024) have shown, this has contributed to changes in fire regimes worldwide.

Conclusions

This paper aims to contribute to the weekly prediction of virulent wildfires at the provincial level. Both climatological and non-climatological variables were tested. The modelling

of burned forest area demonstrated the suitability of the Gamma distribution to fit continuous and asymmetric outcomes. The presence of zeros, i.e. the non-observation of forest fires, led us to propose a model that combines Gamma and Bernoulli distributions. Further, it presents solid and accurate results for the following reasons: (1) wildfire data are highly affected by uncertainty and opportunism; (2) short-term (weekly) fire modelling is key to managing resources in potentially vulnerable locations and planning future firefighting; (3) *easy-to-predict* meteorological data have been used, which simplifies forecasting tasks; (4) the error inherent in modelling large events is conditioned by simplified representations of the underlying physical processes. Apart from a few specific domains, the RRMSE estimates are appropriate, with 90% of the domains below 10% relative errors for the aZIGA model. Therefore, the research objectives were satisfactorily met, being able to predict the number and magnitude of fires for the proposed spatio-temporal units.

In the future, our efforts will focus on decision support for fighting large forest fires. The main limitations so far have been the lack of good vegetation mapping and the availability of information about the arrival of firefighting resources at forest fires. We expect to present results for the region of Galicia in the near future with improved models using this information.

Supplementary material

Supplementary material is available online.

References

Abdollahi A, Pradhan B (2023) Explainable Artificial Intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. *Science of The Total Environment* **879**, 163004. doi:10.1016/j.scitotenv.2023.163004

Alcasena Urdíroz FJ, Evers CR, Vega García C (2018) The wildland–urban interface raster dataset of Catalonia. *Data in Brief* **17**, 124–128. doi:10.1016/j.dib.2017.12.066

Al-Khuraisat S (2023). Using decision-tree analysis for predicting wildfires impact on vegetation in Australia. Master's Thesis, Rochester Institute of Technology, NY USA. Available at https://repository.rit.edu/theses/11405/

Andrade SR, Hulse DE (2022) Evaluation and improvement of system-of-systems resilience in a simulation of wildfire emergency response. *IEEE Systems Journal* **17**, 1877–1888.

Arellano-del-Verbo G, Urbieta IR, Moreno JM (2023) Large fire ignitions are higher in protected areas than outside them in west-central Spain. *Fire* **6**(1), 28. doi:10.3390/fire6010028

Aslan C, Tarver R, Brunson M, Veloz S, Sikes B, Epanchin-Niell R (2024) Experiences with wildfire are associated with private landowners management decisions, relationships, and perceptions of risk. *Landscape and Urban Planning* **247**, 105067. doi:10.1016/j.landurbplan.2024.105067

Bachantourian M, Kalabokidis K, Palaiologou P, Chaleplis K (2023) Optimizing fuel treatments allocation to protect the wildland–urban interface from large-scale wildfires in Greece. *Fire* **6**(2), 75. doi:10.3390/fire6020075

Badia A, Pallares-Barbera M, Valldeperas N, Gisbert M (2019) Wildfires in the wildland–urban interface in Catalonia: vulnerability analysis based on land use and land cover change. *Science of The Total Environment* **673**, 184–196. doi:10.1016/j.scitotenv.2019.04.012

Bar-Massada A, Alcasena F, Schug F, Radeloff VC (2023) The wildland–urban interface in Europe: spatial patterns and associations with socioeconomic and demographic variables. *Landscape and Urban Planning* **235**, 104759. doi:10.1016/j.landurbplan.2023.104759

Bargali H, Pandey A, Bhatt D, Sundriyal RC, Uniyal VP (2024) Forest fire management, funding dynamics, and research in the burning frontier: a comprehensive review. *Trees, Forests and People* **16**, 100526. doi:10.1016/j.tfp.2024.100526

Barreiro JB, Hermosilla T (2013) Socio-geographic analysis of the causes of the 2006's wildfires in Galicia (Spain). *Forest Systems* **22**(3), 497–509. doi:10.5424/fs/2013223-04165

Bento-Gonçalves A, Vieira A (2020) Wildfires in the Wildland-Urban Interface: Key concepts and evaluation methodologies. *Science of The Total Environment* **707**, 135592. doi:10.1016/j.scitotenv.2019.135592

Bot K, Borges JG (2022) A systematic review of applications of Machine Learning techniques for wildfire management decision support. *Inventions* **7**(1), 15. doi:10.3390/inventions7010015

Boubeta M, Lombardía MJ, Marey-Pérez MF, Morales D (2015) Prediction of forest fires occurrences with area-level Poisson mixed models. *Journal of Environmental Management* **154**, 151–158. doi:10.1016/j.jenvman.2015.02.009

Boubeta M, Lombardía MJ, González-Manteiga W, Marey-Pérez MF (2016) Burned area prediction with semiparametric models. *International Journal of Wildland Fire* **25**(6), 669–678. doi:10.1071/WF15125

Boubeta M, Lombardía MJ, Marey-Pérez MF, Morales D (2019) Poisson mixed models for predicting number of fires. *International Journal of Wildland Fire* **28**(3), 237–253. doi:10.1071/WF17037

Bugallo M, Esteban MD, Marey-Pérez MF, Morales D (2023) Wildfire prediction using zero-inflated negative binomial mixed models: application to Spain. *Journal of Environmental Management* **328**, 116788. doi:10.1016/j.jenvman.2022.116788

Campos JC, Rodrigues S, Sil Â, *et al.* (2022) Climate regulation ecosystem services and biodiversity conservation are enhanced differently by climate-and fire-smart landscape management. *Environmental Research Letters* **17**(5), 054014. doi:10.1088/1748-9326/ac64b5

Castel-Clavera J, Pimont F, Opitz T, Ruffault J, Rivière M, Dupuy JL (2022) Disentangling the factors of spatio-temporal patterns of wildfire activity in south-eastern France. *International Journal of Wildland Fire* **32**(1), 15–28. doi:10.1071/WF22086

Castellnou M, Bachfischer M, Miralles M, Ruiz B, Stoof CR, Vilà-Guerau de Arellano J (2022) Pyroconvection classification based on atmospheric vertical profiling correlation with extreme fire spread observations. *Journal of Geophysical Research: Atmospheres* **127**(22), e2022JD036920. doi:10.1029/2022JD036920

Canadas MJ, Leal M, Soares F, *et al.* (2023) Wildfire mitigation and adaptation: two locally independent actions supported by different policy domains. *Land Use Policy* **124**, 106444. doi:10.1016/j.landusepol.2022.106444

Carmo M, Ferreira J, Mendes M, *et al.* (2022) The climatology of extreme wildfires in Portugal, 1980-2018: contributions to forecasting and preparedness. *International Journal of Climatology* **42**(5), 3123–3146. doi:10.1002/joc.7411

Chas-Amil ML, García-Martínez E, Touza J (2020) Iberian Peninsula October 2017 wildfires: burned area and population exposure in Galicia (NW of Spain). *International Journal of Disaster Risk Reduction* **48**, 101623. doi:10.1016/j.ijdrr.2020.101623

Chen A (2022) Evaluating the relationships between wildfires and drought using Machine Learning. *International Journal of Wildland Fire* **31**(3), 230–239. doi:10.1071/WF21145

Chen B, Wu S, Jin Y, Song Y, Wu C, Venevsky S, *et al.* (2024) Wildfire risk for global wildland–urban interface areas. *Nature Sustainability* **7**(4), 474–484. doi:10.1038/s41893-024-01291-0

Chew YJ, Ooi SY, Pang YH, Wong KS (2022) A review of forest fire combating efforts, challenges and future directions in Malaysia, Sabah, and Sarawak. *Forests* **13**(9), 1405. doi:10.3390/f13091405

Chou YH (1992) Management of wildfires with a geographical information system. *International Journal of Geographical Information Systems* **6**(2), 123–140. doi:10.1080/02693799208901900

Colantoni A, Egidi G, Quaranta G, DAlessandro R, Vinci S, Turco R, Salvati L (2020) Sustainable land management, wildfire risk and the role of grazing in Mediterranean urban–rural interfaces: a regional approach from Greece. *Land* **9**(1), 21. doi:10.3390/land9010021

Copes-Gerbitz K, Dickson-Hoyle S, Ravensbergen SL, *et al.* (2022) Community engagement with proactive wildfire management in British Columbia, Canada: perceptions, preferences and barriers to action. *Frontiers in Forests and Global Change* **5**, 829125. doi:10.3389/ffgc.2022.829125

Crist MR (2023) Rethinking the focus on forest fires in federal wildland fire management: landscape patterns and trends of non-forest and forest burned area. *Journal of Environmental Management* **327**, 116718. doi:10.1016/j.jenvman.2022.116718

Cruz MG, Alexander ME (2019) The 10% wind speed rule of thumb for estimating a wildfire forward rate of spread in forests and shrublands. *Annals of Forest Science* **76**(2), 44. doi:10.1007/s13595-019-0829-8

D'Este M, Ganga A, Elia M, *et al.* (2020) Modeling fire ignition probability and frequency using Hurdle models: a cross-regional study in Southern Europe. *Ecological Processes* **9**, 54. doi:10.1186/s13717-020-00263-4

Duane A, Castellnou M, Brotons L (2021) Towards a comprehensive look at global drivers of novel extreme wildfire events. *Climatic Change* **165**, 43. doi:10.1007/s10584-021-03066-4

Eghdami M, Juliano TW, Jiménez PA, Kosovic B, Castellnou M, Kumar R, Vila-Guerau de Arellano J (2023) Characterizing the role of moisture and smoke on the 2021 Santa Coloma de Queralt pyroconvective event using WRF-Fire. *Journal of Advances in Modeling Earth Systems* **15**(5), e2022MS003288. doi:10.1029/2022MS003288

Feng CX, Dean CB (2012) Joint analysis of multivariate spatial count and zero-heavy count outcomes using common spatial factor models. *Environmetrics* **23**(6), 493–508. doi:10.1002/env.2158

Feng CX, Li L, Sadeghpour A (2020) A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology* **20**, 175. doi:10.1186/s12874-020-01055-2

Fernández-Blanco CR, Górriz-Mifsud E, Prokofieva I, Muys B, Novoa CP (2022) Blazing the trail: social innovation supporting wildfire-resilient territories in Catalonia (Spain). *Forest Policy and Economics* **138**, 102719. doi:10.1016/j.forpol.2022.102719

Forgy EW (1965) Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics* **21**, 768–769.

Gannon BM, Thompson MP, Deming KZ, *et al.* (2020) A geospatial framework to assess fireline effectiveness for large wildfires in the western USA. *Fire* **3**(3), 43. doi:10.3390/fire3030043

Gaudet B, Simeoni A, Gwynne S, Kuligowski E, Benichou N (2020) A review of post-incident studies for wildland–urban interface fires. *Journal of Safety Science and Resilience* **1**(1), 59–65. doi:10.1016/j.jnlssr.2020.06.010

GFFS (2023). 'General Forest Fire Statistics. Ministry for the Ecological Transition and the Demographic Challenge.' (Government of Spain)

GHSL (2014). 'Global Human Settlement Layer.' (European Commission)

Giannaros TM, Papavasileiou G (2023) Changes in European fire weather extremes and related atmospheric drivers. *Agricultural and Forest Meteorology* **342**, 109749. doi:10.1016/j.agrformet.2023.109749

Glickman D, Babbitt B (2001) Urban–wildland interface communities within the vicinity of federal lands that are at high risk from wildfire. *Federal Register* **66**(3), 751–777.

Granville K, Woolford DG, Dean CB, McFayden CB (2022) Wildland fire prevention: the impact of the modifying industrial operations protocol on the growth of industrial forestry-caused wildland fires in Ontario, Canada. *International Journal of Wildland Fire* **31**(9), 825–834. doi:10.1071/WF22074

Grünig M, Seidl R, Senf C (2023) Increasing aridity causes larger and more severe forest fires across Europe. *Global Change Biology* **29**(6), 1648–1659. doi:10.1111/gcb.16547

Guyeux C, Makhoul A, Bahi JM (2022) How to build an optimal and operational knowledge base to predict firefighters interventions. In 'Proceedings of SAI Intelligent Systems Conference, vol. 1' (Eds C Guyeux, A Makhoul, JM Bahi) pp. 558–572. (Springer)

Haque KMS, Uddin M, Ampah JD, *et al.* (2023) Wildfires in Australia: a bibliometric analysis and a glimpse on 'Black Summer'(2019/20) Disaster. *Environmental Science and Pollution Research* **30**, 73061–73086. doi:10.1007/s11356-023-27423-1

Hudson M (2011) 'Fire management in the American West: forest politics and the rise of megafires.' (University Press of Colorado)

Humi M (2017) 'Introduction to Mathematical Modelling.' (CRC Press)

Iglesias MC, Hermoso V, Campos JC, , *et al* (2022*a*) Climate and fire smart landscape scenarios call for redesigning protection regimes to achieve multiple management goals. *Journal of Environmental Management* **322**, 116045. doi:10.1016/j.jenvman.2022.116045

Iglesias V, Stavros N, Balch JK, , *et al* (2022*b*) Fires that matter: reconceptualizing fire risk to include interactions between humans and the natural environment. *Environmental Research Letters* **17**(4), 045014. doi:10.1088/1748-9326/ac5c0c

ISO 8601 (2019) 'Date and time – Representations for information interchange.' (International Organization for Standardization)

Jain P, Castellanos-Acuna D, Coogan SC, Abatzoglou JT, Flannigan MD (2022) Observed increases in extreme fire weather driven by atmospheric humidity and temperature. *Nature Climate Change* **12**(1), 63–70. doi:10.1038/s41558-021-01224-1

Jiménez-Ruano A, Mimbrero MR, Jolly WM, De la Riva Fernández J (2019) The role of short-term weather conditions in temporal dynamics of fire regime features in mainland Spain. *Journal of Environmental Management* **241**, 575–586. doi:10.1016/j.jenvman.2018.09.107

Kale MP, Ramachandran RM, Pardeshi SN, *et al.* (2017) Are climate extremities changing forest-fire regimes in India? An analysis using MODIS fire locations of 2003-2013 and gridded climate data of India meteorological department. *Proceedings of the National Academy of Sciences* **87**, 827–843. doi:10.1007/s40010-017-0452-8

Keeley JE, Syphard AD (2019) Twenty-first century California, USA, wildfires: fuel-dominated vs. wind-dominated fires. *Fire Ecology* **15**(1), 24. doi:10.1186/s42408-019-0041-0

Khanmohammadi S, Arashpour M, Mohammadi Golafshani E, Cruz MG, Rajabifard A, Bai Y (2022) Prediction of wildfire rate of spread in grasslands using Machine Learning methods. *Environmental Modelling and Software* **156**, 105507. doi:10.1016/j.envsoft.2022.105507

Khanmohammadi S, Arashpour M, Golafshani EM, Cruz MG, Rajabifard A (2023) An artificial intelligence framework for predicting fire spread sustainability in semiarid shrublands. *International Journal of Wildland Fire* **32**(4), 636–649. doi:10.1071/WF22216

Kirkland M, Atkinson PW, Pearce-Higgins JW, *et al.* (2023) Landscape fires disproportionally affect high conservation value temperate peatlands, meadows, and deciduous forests, but only under low moisture conditions. *Science of The Total Environment* **884**, 163849. doi:10.1016/j.scitotenv.2023.163849

Koh J, Pimont F, Dupuy JL, Opitz T (2023) Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics* **17**(1), 560–582. doi:10.1214/22-AOAS1642

Kraus D, Wohlgemuth T, Castellnou M, Conedera, M (2022) Fire in forest ecosystems: processes and management strategies. In 'Disturbance ecology, Vol. 32.' (Eds T Wohlgemuth, A Jentsch, R Seidl) pp. 143–171. (Springer)

Kuligowski ED, Zhao X, Lovreglio R, *et al.* (2022) Modeling evacuation decisions in the 2019 Kincade fire in California. *Safety Science* **146**, 105541. doi:10.1016/j.ssci.2021.105541

Lasanta T, Cortijos-López M, Errea MP, Khorchani M, Nadal-Romero E (2022) An environmental management experience to control wildfires in the mid-mountain mediterranean area: shrub clearing to generate mosaic landscapes. *Land Use Policy* **118**, 106147. doi:10.1016/j.landusepol.2022.106147

Lee AH, Zhao Y, Yau KKW, Xiang L (2010) How to analyze longitudinal multilevel physical activity data with many zeros? *Preventive Medicine* **51**(6), 476–481. doi:10.1016/j.ypmed.2010.09.012

Lee JM, Mirocha JD, Lareau NP, Whitney T, To W, Kochanski A, Lassman W (2023) Sensitivity of pyrocumulus convection to tree mortality during the 2020 Creek fire in California. *Geophysical Research Letters* **50**(16), e2023GL104193. doi:10.1029/2023GL104193

Leone V, Elia M, Lovreglio R, Correia F, Tedim F (2023) The 2017 extreme wildfires events in Portugal through the perceptions of volunteer and professional firefighters. *Fire* **6**(4), 133. doi:10.3390/fire6040133

Lim JR (2022) Why people adopt climate change adaptation and disaster risk reduction behaviors: integrated model of risk communication and results from hurricanes, floods, and wildfires. *Bulletin of the*

American Meteorological Society **103**(10), 2440–2469. doi:10.1175/BAMS-D-21-0087.1
Linley GD, Jolly CJ, Doherty TS, Geary WL, Armenteras D, Belcher CM, *et al.* (2022) What do you mean, 'megafire'? *Global Ecology and Biogeography* **31**(10), 1906–1922. doi:10.1111/geb.13499
Llorens R, Sobrino JA, Fernández C, Fernández-Alonso JM, Vega JA (2021) A methodology to estimate forest fires burned areas and burn severity degrees using Sentinel-2 data. Application to the October 2017 Fires in the Iberian Peninsula. *International Journal of Applied Earth Observation and Geoinformation* **95**, 102243. doi:10.1016/j.jag.2020.102243
Lloret F, Escudero A, Lloret J, Valladares F (2024) An ecological perspective for analysing rural depopulation and abandonment. *People and Nature* **6**(2), 490–506. doi:10.1002/pan3.10606
Mansoor S, Farooq I, Kachroo MM, *et al.* (2022) Elevation in wildfire frequencies with respect to the climate change. *Journal of Environmental Management* **301**, 113769. doi:10.1016/j.jenvman.2021.113769
Marey-Pérez MF, Loureiro X, Corbelle-Rico EJ, Fernández-Filgueira C (2021) Different strategies for resilience to wildfires: the experience of collective land ownership in Galicia (northwest Spain). *Sustainability* **13**(9), 4761. doi:10.3390/su13094761
Marey-Pérez MF, Fuentes-Santos I, Saavera-Nieves P, González-Manteiga W (2022) Non-parametric comparative analysis of the spatiotemporal pattern of human-caused and natural wildfires in Galicia. *International Journal of Wildland Fire* **32**(2), 178–194. doi:10.1071/WF22030
Marey-Pérez M, López-Álvarez Ó, Franco-Vázquez L (2024) Analysis of trends in the distance of wildfires from built-up areas in Spain and California (USA): 2007–2015. *Forests* **15**(5), 788. doi:10.3390/f15050788
Martínez DJ, Seraphin B, Marks-Block T, Nelson P, Vinyeta K (2023) Indigenous fire futures: anticolonial approaches to shifting fire relations in California. *Environment and Society* **14**(1), 142–161. doi:10.3167/ares.2023.140109
Martínez-Fernández J, Chuvieco E, Koutsias N (2013) Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Sciences* **13**(2), 139–153.
McColl-Gausden SC, Bennett LT, Clarke HG, Ababei DA, Penman TD (2022) The fuel climate fire conundrum: how will fire regimes change in temperate eucalypt forests under climate change? *Global Change Biology* **28**(17), 5211–5226. doi:10.1111/gcb.16283
Meacham BJ, McNamee M (2022) Sustainable and fire resilient built environment. In 'Handbook of fire and the environment'. (Eds BJ Meacham, M McNamee) pp. 421–456. (Springer)
Meier S, Strobl E, Elliott RJR, Kettridge N (2022) Cross-country risk quantification of extreme wildfires in Mediterranean Europe. *Risk Analysis* **43**, 1745–1762. doi:10.1111/risa.14075
Meier S, Elliott RJ, Strobl E (2023) The regional economic impact of wildfires: evidence from southern Europe. *Journal of Environmental Economics and Management* **118**(2), 102787. doi:10.1016/j.jeem.2023.102787
Merino de Miguel S, Huesca M, González-Alonso F (2010) MODIS reflectance and active fire data for burn mapping and assessment at regional level. *Ecological Modelling* **221**(1), 67–74. doi:10.1016/j.ecolmodel.2009.09.015
Michael F, Thomas D (2016) 'Discrete data analysis with R: visualization and modeling techniques for categorical and count data.' (Chapman and Hall)
MITECO (2023) Common basic guidelines for sustainable forest management. Spanish Ministry for the Ecological Transition and the Demographic Challenge (MITECO).
Moanga D, Biging G, Radke J, Butsic V (2020) The space–time cube as an approach to quantifying future wildfires in California. *International Journal of Wildland Fire* **30**, 311–327. doi:10.1071/WF19062
Moreira F, Leal M, Bergonse R, *et al.* (2023) Recent trends in fire regimes and associated territorial features in a fire-prone Mediterranean Region. *Fire* **6**(2), 60. doi:10.3390/fire6020060
Moritz MA, Morais ME, Summerell LA, Carlson JM, Doyle J (2005) Wildfires, complexity, and highly optimized tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **102**(50), 17912–7. doi:10.1073/pnas.0508985102
Nasiri V, Sadeghi SMM, Bagherabadi R, Moradi F, Deljouei A, Borz SA (2022) Modeling wildfire risk in western Iran based on the integration of AHP and GIS. *Environmental Monitoring and Assessment* **194**(9), 644. doi:10.1007/s10661-022-10318-y
Ni S, Gernay T (2022) On the effect of exposed timber on the severity of structural fires in a compartment and required firefighting resources. *Fire Technology* **58**(5), 2691–2725. doi:10.1007/s10694-022-01254-x
Palaiologou P, Kalabokidis K, Ager AA, Galatsidas S, Papalampros L, Day MA (2021) Spatial optimization and tradeoffs of alternative forest management scenarios in Macedonia, Greece. *Forests* **12**(6), 697. doi:10.3390/f12060697
Papathoma-Köhle M, Schlögl M, Garlichs C, Diakakis M, Mavroulis S, Fuchs S (2022) A wildfire vulnerability index for buildings. *Scientific Reports* **12**(1), 6378. doi:10.1038/s41598-022-10479-3
Peña PF, Ragab AR, Luna MA, Isaac MSA, Campoy P (2022) Wild hopper: a heavy-duty UAV for day and night firefighting operations. *Heliyon* **8**(6), 09588. doi:10.1016/j.heliyon.2022.e09588
Pereira MG, Caramelo L, Vega-Orozco C, Costa R, Tonini M (2015) Space–time clustering analysis performance of an aggregated dataset: The case of wildfires in Portugal. *Environmental Modelling and Software* **72**, 239–249. doi:10.1016/j.envsoft.2015.05.016
Poduška Z, Stajić S (2024) The cost of forest fires: a socioeconomic analysis. In 'Fire hazards: socio-economic and regional issues' (Eds J Rodrigo-Comino, L Salvati) pp. 123–135. (Springer)
Pozo RA, Galleguillos M, González ME, Vásquez F, Arriagada R (2022) Assessing the socio-economic and land-cover drivers of wildfire activity and its spatiotemporal distribution in south-central Chile. *Science of The Total Environment* **810**, 152002. doi:10.1016/j.scitotenv.2021.152002
Prichard SJ, Povak NA, Kennedy MC, *et al.* (2020) Fuel treatment effectiveness in the context of landform, vegetation, and large, wind-driven wildfires. *Ecological Applications* **30**(5), e02104. doi:10.1002/eap.2104
Prichard SJ, Hessburg PF, Hagmann RK, *et al.* (2021) Adapting western North American forests to climate change and wildfires: 10 common questions. *Ecological Applications* **31**(8), 02433. doi:10.1002/eap.2433
Pulido F, Corbacho J, Bertomeu M, *et al.* (2023) Fire smart territories: a proof of concept based on Mosaico approach. *Landscape Ecology* **38**, 3353–3370. doi:10.1007/s10980-023-01618-w
Purnomo DM, Qin Y, Theodori M, Zamanialaei M, Lautenberger C, Trouvé A, Gollner MJ (2024) Integrating an urban fire model into an operational wildland fire model to simulate one dimensional wildland–urban interface fires: a parametric study. *International Journal of Wildland Fire* **33**(10), WF24102. doi:10.1071/WF24102
Radeloff VC, Hammer RB, Stewart SI, *et al.* (2005) The wildland–urban interface in the United States. *Ecological Applications* **15**(3), 799–805. doi:10.1890/04-1413
Ramos AM, Russo A, DaCamara CC, Nunes S, Sousa P, Soares PMM, *et al.* (2023) The compound event that triggered the destructive fires of October 2017 in Portugal. *IScience* **26**, **3**, 106141. doi:10.1016/j.isci.2023.106141
Raposo VDMB, Costa VAF, Rodrigues AF (2023) A review of recent developments on drought characterization, propagation, and influential factors. *Science of The Total Environment* **898**, 165550. doi:10.1016/j.scitotenv.2023.165550
Reid AM, Fuhlendorf SD, Weir JR (2010) Weather variables affecting Oklahoma wildfires. *Rangeland Ecology and Management* **63**(5), 599–603. doi:10.2111/REM-D-09-00132.1
Reid CE, Jerrett M, Petersen ML, *et al.* (2015) Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using Machine Learning. *Environmental Science and Technology* **49**(6), 3887–3896. doi:10.1021/es505846r
Ribau MC, Bachfischer M, Bover MM, de Arellano JVG (2024) Dry pyroclouds promote longer-lasting extreme wildfire events intensification [Preprint]. Research Square. doi:10.21203/rs.3.rs-4053550/v1
Ríos-Pena L, Cadarso-Suárez C, Kneib T, Marey-Pérez MF (2015) Applying binary structured additive regression for predicting wildfire in Galicia, Spain. *Procedia Environmental Sciences* **27**, 123–126. doi:10.1016/j.proenv.2015.07.121
Ríos-Pena L, Kneib T, Cadarso-Suárez C, Klein N, Marey-Pérez MF (2018) Studying the occurrence and burnt area of wildfires using zero-one-inflated structured additive beta regression. *Environmental Modelling and Software* **110**, 107–118. doi:10.1016/j.envsoft.2018.03.008

Rodrígues 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. doi:10.1016/j.apgeog.2014.01.011

Rodrígues M, Alcasena F, Gelabert P, Vega-García C (2020) Geospatial modeling of containment probability for escaped wildfires in a Mediterranean region. *Risk Analysis* **40**(9), 1762–1779. doi:10.1111/risa.13524

Rodrígues M, Cunill Camprubí À, Balaguer-Romano R, *et al.* (2023) Drivers and implications of the extreme 2022 wildfire season in Southwest Europe. *Science of The Total Environment* **859**, 160320. doi:10.1016/j.scitotenv.2022.160320

Rodríguez GL, Marey-Pérez M F (2022) Chromatic coding for the detection of spatial-temporal patterns in collective lands in Galicia (Spain). *Fire* **5**(6), 179. doi:10.3390/fire5060179

Rodríguez GL, Vicente VR, Marey-Pérez MF (2022) Influence of the declaration of protected natural areas on the evolution of forest fires in collective lands in Galicia (Spain). *Forests* **13**(8), 1161. doi:10.3390/f13081161

Rodríguez-García E, Santana VM, Alloza JA, Vallejo VR (2022) Predicting natural hyperdense regeneration after wildfires in *Pinus halepensis* forests using prefire site factors, forest structure and fire severity. *Forest Ecology and Management* **512**, 120164. doi:10.1016/j.foreco.2022.120164

Sadatrazavi A, Motlagh MS, Noorpoor A, Ehsani AH (2022) Predicting wildfires occurrences using meteorological parameters. *International Journal of Environmental Research* **16**(6), 106. doi:10.1007/s41742-022-00460-3

Sakellariou S, Sfougaris A, Christopoulou O, Tampekis S (2023) Spatial resilience to wildfires through the optimal deployment of firefighting resources: Impact of topography on initial attack effectiveness. *International Journal of Disaster Risk Science* **14**, 98–112. doi:10.1007/s13753-023-00464-3

Sánchez-Pinillos M, Dakos V, Kéfi S (2024) Ecological dynamic regimes: a key concept for assessing ecological resilience. *Biological Conservation* **289**, 110409. doi:10.1016/j.biocon.2023.110409

Senande-Rivera M, Insua-Costa D, Miguez-Macho G (2022) Spatial and temporal expansion of global wildland fire activity in response to climate change. *Nature Communications* **13**(1), 1208. doi:10.1038/s41467-022-28835-2

Shao Y, Feng Z, Sun L, *et al.* (2022) Mapping China's forest fire risks with Machine Learning. *Forests* **13**(6), 856. doi:10.3390/f13060856

Shmuel A, Heifetz E (2022) Global wildfire susceptibility mapping based on Machine Learning models. *Forests* **13**(7), 1050. doi:10.3390/f13071050

Singleton MP, Thode AE, Meador AJS, Iniguez JM (2019) Increasing trends in high-severity fire in the south western USA from 1984 to 2015. *Forest Ecology and Management* **433**, 709–719. doi:10.1016/j.foreco.2018.11.039

Sobrino JA, Llorens R, Fernández C, Fernández-Alonso JM, Vega JA (2019) Relationship between soil burn severity in forest fires measured *in situ* and through spectral indices of remote detection. *Forests* **10**(5), 457. doi:10.3390/f10050457

Soto C, Martínez M, Bonilla S, García M (2022) Calculating minimum safety distance against wildfires at the wildland–urban interface in Chile and Spain. *Heliyon* **8**(11), e11238. doi:10.1016/j.heliyon.2022.e11238

Sousa D, Cruz-Jesus F, Sousa A, Painho M (2021) A multivariate approach to assess the structural determinants of large wildfires: evidence from a Mediterranean country. *International Journal of Wildland Fire* **30**(4), 241–254. doi:10.1071/WF20119

Stewart SI, Radeloff VC, Hammer RB, Hawbaker TJ (2007) Defining the wildland–urban interface. *Journal of Forestry* **105**(4), 201–207.

Stoof CR, de Vries JR, Castellnou Ribau M, *et al.* (2024) Megafire: an ambiguous and emotive term best avoided by science. *Global Ecology and Biogeography* **33**(2), 341–351. doi:10.1111/geb.13791

Tan ZD, Carrasco LR, Taylor D (2020) Spatial correlates of forest and land fires in Indonesia. *International Journal of Wildland Fire* **29**(12), 1088–1099. doi:10.1071/WF20036

Tedim F, McCaffrey S, Leone V, Delogu GM, Castelnou M, McGee TK, Aranha J (2020) What can we do differently about the extreme wildfire problem: an overview. In 'Extreme wildfire events and disasters'. (Eds F Tedim, V Leone, TK McGee) pp. 233–263. (Elsevier) doi:10.1016/B978-0-12-815721-3.00013-8

Tedim F, Leone V, Lovreglio R, *et al.* (2022) Forest fire causes and motivations in the southern and south-eastern Europe through experts perception and applications to current policies. *Forests* **13**(4), 562. doi:10.3390/f13040562

Troumbis AY, Gaganis CM, Sideropoulos H (2023) Probabilistic wildfire risk assessment and modernization transitions: the case of Greece. *Fire* **6**(4), 158. doi:10.3390/fire6040158

Turco M, Jerez S, Doblas-Reyes FJ, *et al.* (2018) Skilful forecasting of global fire activity using seasonal climate predictions. *Nature Communications* **9**, 2718. doi:10.1038/s41467-018-05250-0

Ujjwal KC, Hilton J, Garg S, Aryal J (2022) A probability-based risk metric for operational wildfire risk management. *Environmental Modelling and Software* **148**, 105286. doi:10.1016/j.envsoft.2021.105286

Viedma O, Urbieta IR, Moreno JM (2018) Wildfres and the role of their drivers are changing over time in a large rural area of west-central Spain. *Scientific Reports* **8**, 17797. doi:10.1038/s41598-018-36134-4

Werth PA, Potter, BE, Alexander, ME, *et al.* (2016) Synthesis of knowledge of extreme fire behavior: Volume 2 for fire behavior specialists, researchers, and meteorologists. (USDA Forest Service, Pacific Northwest Research Station)

Winandy JE, Wiesner F, Hassan B, Morrell JJ (2022) Fire performance of timber: review for use in wildland–urban interfaces. *Holzforschung* **76**, 679–698. doi:10.1515/hf-2022-0038

Wolpert F, Quintas-Soriano C, Pulido F, Huntsinger L, Plieninger T (2022) Collaborative agroforestry to mitigate wildfires in Extremadura, Spain: land manager motivations and perceptions of outcomes, benefits, and policy needs. *Agroforestry Systems* **96**(8), 1135–1149. doi:10.1007/s10457-022-00771-6

Wu Z, Wang B, Li M, Tian Y, Quan Y, Liu J (2022) Simulation of forest fire spread based on artificial intelligence. *Ecological Indicators* **136**, 108653. doi:10.1016/j.ecolind.2022.108653

Xanthopoulos G, Delogu GM, Leone V, Correia FJ, Magalhães CG (2020) Firefighting approaches and extreme wildfires. In 'Extreme wildfire events and disasters'. (Eds F Tedim, V Leone, TK McGee) pp. 117–132. (Elsevier)

Xanthopoulos G, Athanasiou M, Nikiforaki A, *et al.* (2022) Innovative action for forest fire prevention in Kythira Island, Greece, through mobilization and cooperation of the population: methodology and challenges. *Sustainability* **14**(2), 594. doi:10.3390/su14020594

Yu Y, Mao J, Wullschleger SD, *et al.* (2022) Machine Learning-based observation-constrained projections reveal elevated global socio-economic risks from wildfire. *Nature Communications* **13**(1), 1250. doi:10.1038/s41467-022-28853-0

Zambon I, Cerdà A, Cudlin P, Serra P, Pili S, Salvati L (2019) Road network and the spatial distribution of wildfires in the Valencian Community (1993-2015). *Agriculture* **9**(5), 100. doi:10.3390/agriculture9050100

Zong X, Tian X (2022) Bibliometric analysis of fires on Wildland-Urban Interfaces. *Natural Hazards Research* **2**(3), 147–153. doi:10.1016/j.nhres.2022.08.003

Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Zero-Truncated and Zero-Inflated Models for Count Data. In 'Mixed effects models and extensions in ecology with R', Ch. 11. pp. 261–293. (Springer) doi:10.1007/978-0-387-87458-6

Data availability. The fitting and analysis of the model were completed using open source software shared on GitHub that includes code and data. The user manual README.pdf gives a detailed explanation of our repository, indicating which files correspond to the main code of the paper and which correspond to the data used. The authors' experimental set-up was as follows: software name: aZIG11Fires. Developer: María Bugallo Porto [aut, cre]. Contact information: mbugallo@umh.es. First year available: 2023. Program language: R (version 4.1.3). Software availability: https://github.com/mbugallo/aZIG11Fires.git. License: GNU Affero GPL v3.0.

Conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of funding. This work was supported by the Spanish Ministry of Universities, through projects PGC2018-096840-B-I00 and PID2022-136878NB-I00, and by the Valencian Government, through project PROMETEO-2021-063. It also benefited from a study grant from the Manuel Ventura Figueroa Foundation. Last but not least, we would like to thank the Galician Government for their financial support through project ED431C-2021/25.

Author affiliations

[A]Center of Operations Research, Miguel Hernández University of Elche, Elche, Alicante, Spain.

[B]Higher Polytechnic Engineering School, University of Santiago de Compostela, Santiago de Compostela, A Coruña, Spain.