



# Modeling wildfires via marked spatio-temporal Poisson processes

José J. Quinlan<sup>1</sup> · Carlos Díaz-Avalos<sup>2</sup> · Ramsés H. Mena<sup>2</sup>

Received: 4 August 2020 / Revised: 1 February 2021 / Accepted: 30 March 2021 / Published online: 7 May 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

From a statistical viewpoint, characteristics such as ignition time, location and duration are relevant components for wildfire modeling. The observed ignition sites and starting times constitute a space-time point pattern, and a natural framework to model this type of data is via point processes. In this work, we propose a marked Poisson process to model fire patterns in space-time, considering durations as marks. The collected data correspond to fires observed in the Valencian Community, Spain, between 2010 and 2015. The methodology relies on writing the intensity function of such a process, jointly for starting times, locations and durations, as a weighted Dirichlet process mixture model. A particular choice of the kernel that determines such mixture was made, compatible with data features. We conducted posterior inference on some characteristics of interest for understanding wildfire behavior, showing high flexibility to emulate data patterns.

**Keywords** Dirichlet process mixture model · Posterior inference · Wildfire duration

**Mathematics Subject Classification** 62F15 · 62M30

---

Handling Editor: Luiz Duczmal

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10651-021-00497-1>.

---

✉ José J. Quinlan  
jjquinla@uc.cl

Carlos Díaz-Avalos  
carlos@sigma.iimas.unam.mx

Ramsés H. Mena  
ramses@sigma.iimas.unam.mx

<sup>1</sup> Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>2</sup> Universidad Nacional Autónoma de México, Mexico, Mexico

## 1 Introduction

Fire is an ubiquitous ecological disturbance. Since the first plant communities were established on land, lightning and other ignition sources have started forest fires. In Mediterranean landscapes, fire is one of the main ecological disturbances, as it can affect forests severely enough to induce changes in forest dynamics and geohydrology. (Miguel et al. 2011; Shakesby 2011). Through time, living beings in most terrestrial ecosystems have adapted to the presence and effects of fire, turning it into a natural component in all forest ecosystems (Agee 1993). Despite this, fire represents a hazard to human life and property, encouraging society and policy makers to take measures that mitigate its effects. In European forests, wildfire incidence has increased in recent years, apparently associated to climate change and global warming (Aragó et al. 2016). Unlike other natural hazards, wildfire risk is predictable to some extent (Biro 2009), and thus models able to explain and predict the space and time variability of their incidence, as well as their variability in time to extinction, are of great importance (Faivre et al. 2018). Modeling forest fire dynamics helps to increase the structured knowledge about terrestrial ecosystems, to establish better management of them, and consequently, to set a sustainable use of forest resources (Ager et al. 2006; Lindner et al. 2010). Modeling the spatial variation of fires is important as their burning time is directly associated to the final size of the burned area and the amount of gasses and particulate matter emissions (van der Werf et al. 2009). Such association is also affected by local factors, like vegetation and lithology (Miranda et al. 1994).

Stochastic models for forest fire incidences have been in the literature since the pioneering modeling made by Dayananda (1977). With the advent of Geographic Information Systems (GIS), various attempts to model and screen out risk factors for forest fires have been developed (Amatulli et al. 2007). Though useful, some of these approaches lack necessary key components for adequate spatial modeling, e.g., the classical raster approach based on GIS assumes independence among pixels. In fact, most of the models based on GIS do not include a random component (Mandallaz and Ye 1997; Ager et al. 2006).

Ignition time and location of forest fires constitute a complex process, difficult to predict. However, historical data can be used to construct predictive maps for fire risk. In this sense, some authors have proposed models based on standard generalized linear models (Padilla and Garcia 2011; Vilar et al. 2010) and weighted generalized linear models (Marques et al. 2011). Other authors have proposed more elaborate models based on Markov random fields (Díaz-Avalos et al. 2001), maximum entropy models (Duane et al. 2015) (Duane et al., 2015), kernel density adaptive models (Amatulli et al. 2007), fire growth models (Bar Massada et al. 2009), quantile regression (Barros and Pereira 2014) and spatial point processes (e.g. Juan et al. 2012; Møller and Díaz-Avalos 2010; Turner 2009; Serra et al. 2014; Aragó et al. 2016; Díaz-Avalos et al. 2016), among others.

From an ecological viewpoint, characteristics such as ignition time, ignition location and fire duration, are relevant components for wildfire modeling. Indeed, these features are important as they are closely associated to fire dynamics, in the sense that they are amongst the main variables that control hazards and are also related to burning area and recurrence rate (McKenzie et al. 2011).

The observed sites and ignition times constitute a space-time point pattern and a natural framework to model this type of data is via marked point processes (see e.g. Daley and Vere-Jones 2008; Baddeley et al. 2015). These models have been used in fields such as biology, ecology, agronomy and physics, among others (Møller and Waagepetersen 2004; Illian et al. 2008; Diggle 2013). Within the forest fire modeling literature, various approaches have been proposed (see for example Møller and Díaz-Avalos 2010; Serra et al. 2014; Díaz-Avalos et al. 2016). While the theory and application of point processes is vast, in many cases their estimation remains a challenge. In this regard, the approach by Taddy and Kottas (2012), which models the intensity of a Poisson point process via weighted mixtures of Dirichlet processes, results in a flexible alternative to intensity modeling, as it borrows the practicality of Bayesian nonparametric inference techniques.

The complete randomness property inherent to Poisson point processes may be inadequate to model point patterns that exhibit regularity or clustering features. However, this effect is diminished when the intensity measure is modelled by another random measure, such as in the case of Cox processes. In fact, there are some intersections between non-homogeneous Poisson processes and renewal processes that fall within Cox processes theory (Yannaros 1994; Daley and Vere-Jones 2008).

In this work, we propose a marked Poisson point process for wildfire ignition times and locations as events, marked with durations. The central idea is to use a particular joint intensity function for events and marks, which adapts to the context at issue. Specifically, we apply the model to infer the intensity functions for ignition times and locations, the survival and hazard functions for durations, and the conditional survival function for durations given starting times and locations, for fires occurring in the Valencian Community, Spain, between January 1st, 2010, and December 31st, 2015. Wildfire ignitions may be intentional, accidental, or caused by natural forces such as lightning. In our study, we will only consider wildfires caused by lightning strikes.

## 2 Methodology

### 2.1 Study area

The Valencian Community is located in the east and southeast (SE) part of Spain, neighbouring the communities of Castilla - La Mancha, Cataluña, Aragón and Murcia (Fig. 1). The inland part of the territory is craggy, with the highest peaks in the Valencia and Castellón provinces forming part of the Iberian mountain range. The mountains in the Province of Alicante form part of the Subbaetic range. The coastal zone is mostly a fertile farming zone with some wetlands. According to the Agencia Estatal de Meteorología of Spain (2013), the climate in the study area is diverse, with Mediterranean climate in the coastal areas, continental Mediterranean climate in the continental zones and some zones with hot semi-arid climate.

The vegetation of the Valencian community is diverse as well due to the topography in the community. According to the Coordination for the Information on the Environment (CORINE), there are more than forty-four vegetation types in the Valencian Community.

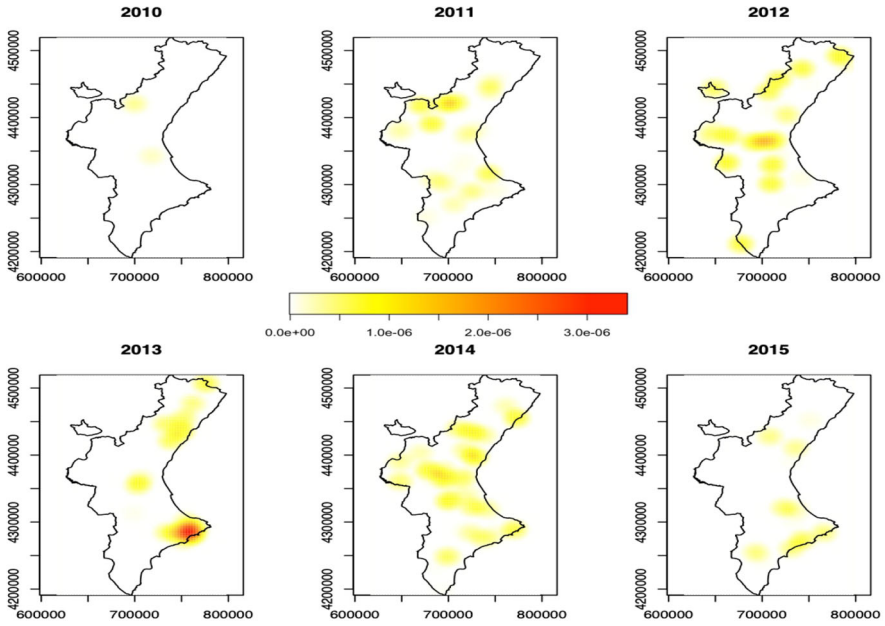


**Fig. 1** Political division map showing the location of the Valencian Community in the Iberic Peninsula

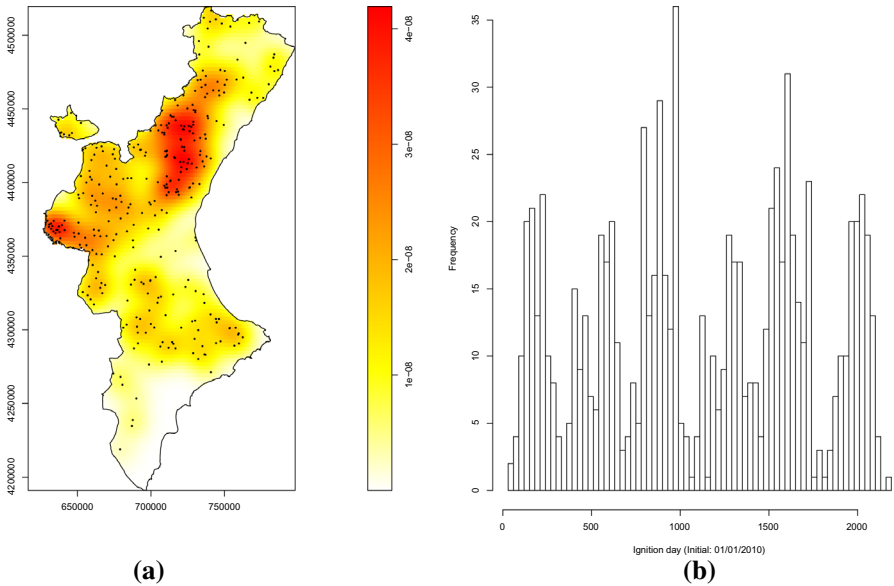
## 2.2 Data set

The data set used in our case study consists of  $m = 337$  wildfires reported for the Valencian Community during the years 2010 to 2015 by the “Comité de Lucha contra Incendios Forestales” in the Valencian Community. The time span is 2191 days, starting from January 1st, 2010, 00:00:01 GTM. Wildfires are originated by several causes, but here we will analyze only those originated by natural causes. The information used in our analysis considers starting time (days since January 1st, 2010, 00:00:01 GTM), duration (days) and geographic location of the ignition site in UTM. Figure 2 shows nonparametric kernel estimates (Diggle 2013) of the intensity function for the five years considered in this study, using the observed point pattern of wildfires in space and time. The incidence of wildfire ignitions shows variability among the six years, but the levels of the intensity estimates have similar maxima except for 2010, where the overall intensity was low, and 2013, which shows an intensity peak in the SE part of the study area.

To better highlight the areas where wildfires are more common, Fig. 3 presents the marginal spatial distribution of the wildfires in the Valencian Community, as well as the variation in 30 day bins of the number of wildfires between January 1st, 2010, and December 31st, 2015. The spatial incidence of the wildfires (Fig. 3a) shows only some areas where wildfires apparently form clusters. These areas are located in the western part of the study area, where forest zones and higher elevations are common. Wildfires at the north, south and coast zones of the Valencian Community were scarce during the study period. Figure 3b shows the seasonality of wildfire incidence in the study area, with high incidences during the warm and dry summer months, and lower incidence during winter. Due to the weather conditions, as well as to the closeness to firefight centers and accessibility, wildfire control times are random, analogous



**Fig. 2** Estimated spatial incidence of wildfires in the Valencian Community, Spain, for years 2010 to 2015, obtained by using a space-time kernel estimator with border correction



**Fig. 3** Spatial incidence of wildfires in the Valencian Community, Spain, between January 1st, 2010, and December 31st, 2015 **(a)**, and temporal incidence along time in 30 day bins **(b)**

to survival analysis (Daley and Vere-Jones 2003). Thus, the observed spatial point pattern of wildfire ignition sites, ignition times, and survival times can be considered as a realization of a space-time marked Poisson process. Hence, our goal here is to propose and fit a model for the joint intensity function of the point pattern of wildfires occurring in the Valencian Community during the time span mentioned above. Note that the areas where the nonparametric estimate of the intensity function in Figs. 2 and 3 is high suggest the presence of some degree of clustering. However, when the events are spread along the five years of the study, the observed groupings are a consequence of a static projection of all events during that period rather than a true clustering effect.

### 3 Modeling the spatio-temporal incidence of wildfires using Dirichlet process mixture models

We assume that the observed point pattern is a realization of a space-time marked point process and denote the point configuration of fire  $i$  by  $\mathbf{y}_i = (s_i, \mathbf{x}_i, l_i)$ , where  $s_i \in [0, \infty)$  and  $\mathbf{x}_i = (x_{1i}, x_{2i}) \in \mathbb{R}^2$  are the wildfire starting times and locations, respectively, each attached with its duration  $l_i \in [0, \infty)$ , say the marks. Specifically, we assume that  $\{\mathbf{y}_i : i = 1, \dots, m\}$  are  $m \in \mathbb{N}$  points generated from a marked space-time Poisson process (MPP) with a continuous intensity function  $\lambda : [0, \infty) \times \mathbb{R}^2 \times [0, \infty) \rightarrow [0, \infty)$ , such that

$$0 < \int_{[0, \infty)} \left\{ \int_{W_0} \lambda(s, \mathbf{x}, l) ds d\mathbf{x} \right\} dl < \infty \quad (1)$$

for all bounded subsets  $W_0$  of  $[0, \infty) \times \mathbb{R}^2$ . Hence, in some  $W := W_0 \times [0, \infty)$ , we observe  $m$  wildfire starting times  $\{s_1, \dots, s_m\}$ , their corresponding locations  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  and durations  $\{l_1, \dots, l_m\}$ . Using (1) we can estimate features of the MPP through its likelihood function, which is proportional to

$$\exp(-\rho)\rho^m \prod_{i=1}^m \pi(s_i, \mathbf{x}_i, l_i), \quad \text{with } \pi(\cdot) := \frac{\lambda(\cdot)}{\rho} \quad \text{and} \\ \rho := \int_W \lambda(s, \mathbf{x}, l) ds d\mathbf{x} dl. \quad (2)$$

In (2),  $\pi : W \rightarrow [0, \infty)$  is a continuous probability density function by construction. As  $\exp(-\rho)\rho^m$  depends on the data only through the sample size  $m$ ,  $\{\mathbf{y}_i : i = 1, \dots, m\}$  can be considered as a random sample coming from  $\pi$ , independent from  $\rho$ . Thus, modeling  $\lambda$  flexibly is equivalent to modeling  $\pi$ . This technique has been previously used by Taddy and Kottas (2012), who proposed modeling  $\pi$  as Dirichlet process mixture (DPM) model (Antoniak 1974; MacEachern 1994; Escobar and West 1995), and a Jeffreys prior for  $\rho$ , supported on  $(0, \infty)$ , with functional form  $\rho^{-1}$ . Under such framework, each  $\mathbf{y}_i$  is generated from a random probability density function

$\pi(\mathbf{y} \mid H) := \int_{\Theta} \mathcal{K}(\mathbf{y} \mid \boldsymbol{\theta}) dH(\boldsymbol{\theta})$ , which admits the hierarchical representation

$$\begin{aligned} \mathbf{y}_i \mid \boldsymbol{\theta}_i &\stackrel{\text{ind.}}{\sim} \mathcal{K}(\cdot \mid \boldsymbol{\theta}_i) : i = 1, \dots, m \\ \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m \mid H &\stackrel{\text{i.i.d.}}{\sim} H \\ H \mid \alpha, H_0 &\sim \text{DP}(\alpha, H_0). \end{aligned} \tag{3}$$

Here,  $\text{DP}(\alpha, H_0)$  represents a Dirichlet process (DP) with centering probability measure  $H_0$  and concentration parameter  $\alpha$  (Ferguson 1973; Blackwell and MacQueen 1973), and  $\mathcal{K}(\cdot \mid \boldsymbol{\theta})$  is a suitable probability kernel indexed by a latent parameter  $\boldsymbol{\theta} \in \Theta$ . Model (3) can be enriched by assigning prior distributions to  $\alpha$  and (possible) hyper-parameters associated with  $H_0$ .

Many theoretical properties regarding DPM models can be found in the specialized literature. Here, we highlight one that is important for model specification. Sethuraman (1994) shows that

$$H(\cdot) = V_1 \delta_{\boldsymbol{\theta}_1}(\cdot) + \sum_{r=2}^{\infty} \left\{ V_r \prod_{t=1}^{r-1} (1 - V_t) \right\} \delta_{\boldsymbol{\theta}_r}(\cdot), \tag{4}$$

where  $\delta_{\mathbf{a}}(\cdot)$  is the Dirac measure centred on  $\mathbf{a} \in \Theta$  and  $\{(\boldsymbol{\theta}_r, V_r) : r \in \mathbb{N}\}$  is a sequence of independent random vectors  $(\boldsymbol{\theta}_r, V_r)$  whose coordinates are mutually independent with common distributions  $\boldsymbol{\theta}_r \sim H_0$  and  $V_r \sim \text{Beta}(1, \alpha)$ . Equation (4) says that  $H$  is a discrete probability measure with probability one. Thus,  $\pi(\mathbf{y} \mid H)$  given by (3) is a countably infinite random mixture centered at the parametric model  $\pi(\mathbf{y} \mid H_0) := \int_{\Theta} \mathcal{K}(\mathbf{y} \mid \boldsymbol{\theta}) dH_0(\boldsymbol{\theta})$ . This feature can help in choosing  $\mathcal{K}$  by analyzing the marginal behaviour of  $s$ ,  $\mathbf{x}$  and  $l$ , or some transformations on them. Also, (4) implies that the induced random intensity function  $\lambda(\mathbf{y} \mid \rho, H) := \rho \pi(\mathbf{y} \mid H)$  is not separable in terms of  $s$ ,  $\mathbf{x}$  and  $l$ , even when  $\mathcal{K}$  and  $H_0$  are factorizable.

A full specification of model (3) requires selecting  $\mathcal{K}$  and  $H_0$ . Within a similar context, various choices have been studied, e.g., a Beta distribution for  $s$  (Kottas 2006a); a Sarmanov bivariate distribution with Beta marginals for  $\mathbf{x}$  (Kottas and Sansó 2007); a Weibull distribution (Kottas 2006b) or Gamma distribution (Fúquene 2015) for  $l$ . Implementing such choices for our context, we observed poor estimates for the corresponding parameters, which in turn affected every other posterior inference related to (3). After analysing our data, we found that working under transformations of the form  $z \mapsto \log\left(\frac{z-l_b}{u_b-z}\right) : z \in [l_b, u_b]$ , wildfire starting times and locations can be captured by mixtures of Normal distributions. The choice of  $l_b$  and  $u_b$  were selected according to the range of data. Specifically, we applied the following data transformation

$$\begin{aligned}
 f(s) &= \log\left(\frac{s}{2191-s}\right) \quad \text{and} \\
 \mathbf{G}(\mathbf{x}) &= (g_1(x_1), g_2(x_2)) \\
 &= \left(\log\left(\frac{x_1 - 591,685.7}{823,584.3 - x_1}\right), \log\left(\frac{x_2 - 4,157,261}{4,578,246 - x_2}\right)\right)
 \end{aligned} \tag{5}$$

for starting times and locations, respectively. For wildfire durations, we applied Cheng and Yuan (2013)'s approach. Hence,

$$\begin{aligned}
 \mathcal{K}(\mathbf{y} \mid \boldsymbol{\theta}) &= \{\text{Normal}(f(s) \mid \eta, \kappa) f'(s)\} \{\text{Normal}(\mathbf{G}(\mathbf{x}) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathbf{J}(\mathbf{x})\} \\
 &\quad \{\text{Normal}(\log(l) \mid \xi, \phi) l^{-1}\},
 \end{aligned} \tag{6}$$

where  $\boldsymbol{\theta} = (\eta, \kappa, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi, \phi)$  and  $\mathbf{J}(\mathbf{x}) = g'_1(x_1)g'_2(x_2)$ . Here,  $\text{Normal}(\cdot \mid \mathbf{m}, \mathbf{V})$  is the Normal probability density function with (scalar or vector) mean  $\mathbf{m}$  and (scalar or matrix) variance  $\mathbf{V}$ ,  $(\eta, \kappa) \in \mathbb{R} \times (0, \infty)$ ,  $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^2 \times \mathbb{S}^2$  and  $(\xi, \phi) \in \mathbb{R} \times (0, \infty)$ . The set  $\mathbb{S}^2$  is the space of positive-definite matrices of dimension  $2 \times 2$ . As for  $H_0$ , we use the following distribution

$$dH_0(\boldsymbol{\theta}) = dH_{01}(\eta, \kappa) dH_{02}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) dH_{03}(\xi, \phi), \tag{7}$$

where

$$\begin{aligned}
 dH_{01}(\eta, \kappa) &= \text{Normal}(\eta \mid m_s, c_s^{-1}\kappa) \text{Inv-Gamma}(\kappa \mid 0.5a_s, 0.5b_s) d\eta d\kappa \\
 dH_{02}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \text{Normal}(\boldsymbol{\mu} \mid \mathbf{m}_x, c_x^{-1}\boldsymbol{\Sigma}) \text{Inv-Wishart}(\boldsymbol{\Sigma} \mid a_x, \mathbf{B}_x) d\boldsymbol{\mu} d\boldsymbol{\Sigma} \\
 dH_{03}(\xi, \phi) &= \text{Normal}(\xi \mid m_l, c_l^{-1}\phi) \text{Inv-Gamma}(\phi \mid 0.5a_l, 0.5b_l) d\xi d\phi.
 \end{aligned}$$

Here,  $\text{Inv-Gamma}(\cdot \mid a, b)$  is the inverse-Gamma probability density function with shape  $a$  and scale  $b$ ,  $\text{Inv-Wishart}(\cdot \mid d, \mathbf{S})$  is the inverse-Wishart probability density function with degrees of freedom  $d$  and scale matrix  $\mathbf{S}$ ,  $(m_s, c_s, a_s, b_s) \in \mathbb{R} \times (0, \infty)^3$ ,  $(\mathbf{m}_x, c_x, a_x, \mathbf{B}_x) \in \mathbb{R}^2 \times (0, \infty)^2 \times \mathbb{S}^2$  and  $(m_l, c_l, a_l, b_l) \in \mathbb{R} \times (0, \infty)^3$ . Equation (7) is an appealing choice in that it is a conjugate distribution for (6), with hyperparameters that are easy to calibrate using the data in the transformed scale. With the idea of having posterior inference on weakly informative priors, we first set  $a_s = a_l = 3$  and  $a_x = 4$ , which are the least integer values such that  $\kappa$ ,  $\boldsymbol{\Sigma}$  and  $\phi$  have finite means and infinite variances. Next, we choose prior guesses for  $(m_s, c_s, b_s)$ ,  $(\mathbf{m}_x, c_x, \mathbf{B}_x)$  and  $(m_l, c_l, b_l)$  which determine the first two moments of  $H_{01}$ ,  $H_{02}$  and  $H_{03}$ , through the relations

$$\begin{aligned}
 E[\eta] &= m_s, \quad \text{Var}[\eta] = c_s^{-1}b_s, \quad E[\kappa] = b_s, \\
 E[\boldsymbol{\mu}] &= \mathbf{m}_x, \quad \text{Var}[\boldsymbol{\mu}] = c_x^{-1}\mathbf{B}_x, \quad E[\boldsymbol{\Sigma}] = \mathbf{B}_x, \\
 E[\xi] &= m_l, \quad \text{Var}[\xi] = c_l^{-1}b_l, \quad E[\phi] = b_l.
 \end{aligned}$$

These specifications ensure good coverage of the data along with a user-friendly specification.

Here is worth emphasizing that, besides the application at issue, the main difference between the approach by Kottas et al. (2005) and the one here proposed is the marked



feature of our spatio-temporal model plus the choices for kernel and prior guess at the shape, (6) and (7), respectively. These variations allow us to capture the duration and the data-based spatio-temporal features under the transformations (5).

### 4 Posterior Inference

We adapt Algorithm 8 (see Algorithm 1 below), proposed by Neal (2000), which relies on (4). This characteristic implies ties among the latent parameters  $\theta_1, \dots, \theta_m$  in (3), thus forming a set of  $k \leq m$  unique latent values  $\theta_1^*, \dots, \theta_k^*$ . A convenient way for updating  $\theta$ s in terms of  $\theta^*$ s, is by introducing a collection of auxiliary variables  $e_1, \dots, e_m \in \{1, \dots, k\}$ , such that  $e_i = j$  if  $\theta_i = \theta_j^*$ . As for the concentration parameter  $\alpha$ , we use a Gamma( $a_0, b_0$ ) prior distribution with shape  $a_0$  and rate  $b_0$  (see, e.g., Escobar and West 1995; Kottas et al. 2005), independent of all hyperparameters indexed in the centering measure  $H_0$ . Parameters  $a_0, b_0 \in (0, \infty)$  can be fixed by exploring the prior distribution of  $k$  given  $\alpha$ , whose mean and variance are equal to  $\sum_{r=1}^m \frac{\alpha}{\alpha+r-1}$  and  $\sum_{r=1}^m \frac{\alpha(r-1)}{(\alpha+r-1)^2}$ , respectively. Details about all full conditionals and steps in this pseudo-algorithm are given in Sections A–C of the online Supplementary Material.

---

#### Algorithm 1

---

- 1: **Input.**
    - a) Data:  $y_1, \dots, y_m$ .
    - b) Parameters:  $m_s, c_s, b_s, \mathbf{m}_x, c_x, \mathbf{B}_x, m_l, c_l, b_l, a_0, b_0$ .
    - c) Iterations:  $T$ .
  - 2: **Initial values.**
    - a) Auxiliary variables:  $(e_1(0), \dots, e_m(0))$ .
    - b) Unique parameters:  $\{(\eta_j^*(0), \kappa_j^*(0), \mu_j^*(0), \Sigma_j^*(0), \xi_j^*(0), \phi_j^*(0)) : j = 1, \dots, k(0)\}$ .
  - 3: **Updating.**

**for**  $t = 1$  to  $T$  **do**

    - a) Auxiliary variables:  $(e_1(t), \dots, e_m(t))$  following A.
    - b) Unique parameters:  $\{(\eta_j^*(t), \kappa_j^*(t), \mu_j^*(t), \Sigma_j^*(t), \xi_j^*(t), \phi_j^*(t)) : j = 1, \dots, k(t)\}$  following B.
    - c) Concentration parameter:  $\alpha(t)$  following C.
    - d) Scale factor:  $\rho(t)$  following a Gamma( $k(t), 1$ ) distribution.
  - 4: **end do**
  - 5: **Output.**
    - a) MCMC samples:  $\{(e_1(t), \dots, e_m(t)) : t = 1, \dots, T\}, \{(\eta_j^*(t), \kappa_j^*(t), \mu_j^*(t), \Sigma_j^*(t), \xi_j^*(t), \phi_j^*(t)) : j = 1, \dots, k(t) : t = 1, \dots, T\}, \{\alpha(t) : t = 1, \dots, T\}$  and  $\{\rho(t) : t = 1, \dots, T\}$ .
- 

Although there are numerical features that can be estimated directly from the MCMC samples, several quantities necessarily require posterior samples from the mixing measure  $H$  in (3), a topic that is not addressed by Neal (2000). To fix ideas, let  $\mathcal{F}(\rho, H) := \mathcal{F}(\lambda(\cdot | \rho, H))$  be a generic function of the the intensity function  $\lambda(\cdot | \rho, H) = \rho \pi(\cdot | H)$ . Recall that  $\pi(\cdot | H)$  is given by (3), and  $\rho$  is independent of  $H$ , following a Jeffreys prior supported on  $(0, \infty)$  with functional form  $\rho^{-1}$ . Given the observed data  $\{y_1, \dots, y_m\}$ , we use a method that allows reporting posterior point and interval estimates for  $\mathcal{F}(\rho, H)$ . This inferential problem requires to deal with

the posterior distribution of  $H$ . To address this issue, Ishwaran and Zarepour (2002) proposed a clever method which combines the MCMC posterior samples

$$\{(e_1(t), \dots, e_m(t)) : t = 1, \dots, T\}, \{\alpha(t) : t = 1, \dots, T\}, \{\rho(t) : t = 1, \dots, T\} \text{ and} \\ \{(\eta_j^*(t), \kappa_j^*(t), \mu_j^*(t), \Sigma_j^*(t), \xi_j^*(t), \phi_j^*(t)) : j = 1, \dots, k(t) : t = 1, \dots, T\}$$

with a “truncated” stick-breaking representation of the Dirichlet process (Sethuraman 1994). Algorithm 2 below provides a pseudo-code applying their technique in our context. Details can be found in Section D of the online Supplementary Material.

---

### Algorithm 2

---

1: **Input.**

- a) Truncation index:  $G$ .
- b) MCMC samples:  $\{(e_1(t), \dots, e_m(t)) : t = 1, \dots, T\}, \{(\eta_j^*(t), \kappa_j^*(t), \mu_j^*(t), \Sigma_j^*(t), \xi_j^*(t), \phi_j^*(t)) : j = 1, \dots, k(t) : t = 1, \dots, T\}, \{\alpha(t) : t = 1, \dots, T\}$  and  $\{\rho(t) : t = 1, \dots, T\}$ .

2: **Approximation.**

**for**  $t = 1$  to  $T$  **do**

- a) Truncated DP:  $H_G(t)$  following D.

3: **end do**

4: **Output.**

- a) Approximated posterior samples:  $\{\mathcal{F}(\rho(t), H_G(t)) : t = 1, \dots, T\}$ .
- 

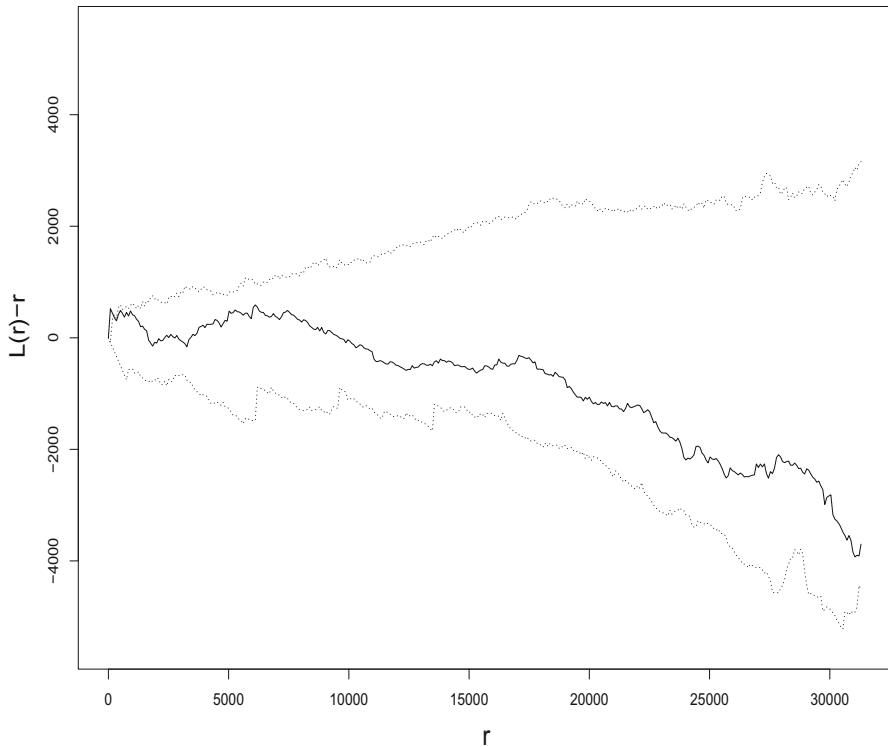
## 5 Results and discussion

Using the data set described in Section 2 and the algorithms described in the previous section, we fitted the model described in Section 3. Parameter elicitation suggested in Sections 3 and 4 is provided below.

- **Concentration parameter:**  $a_0 = 2.5$  and  $b_0 = 0.2$ . These values induce a priori  $E[k] \approx 41$  and  $\text{Var}[k] \approx 361$ .
- **Unique latent parameters:**  $m_s = 0, c_s = 0.004, b_s = 0.04, \mathbf{m}_x = (0, 0)^\top, c_x = 0.02, \mathbf{B}_x = \text{diag}(0.1, 0.1), m_l = -1, c_l = 0.1$  and  $b_l = 1$ . These values induce a priori  $E[\eta] = 0, \text{Var}[\eta] = 10, E[\kappa] = 0.04, E[\mu] = (0, 0)^\top, \text{Var}[\mu] = \text{diag}(5, 5), E[\Sigma] = \text{diag}(0.1, 0.1), E[\xi] = -1, \text{Var}[\xi] = 10$  and  $E[\phi] = 1$ .

We collected 5,000 MCMC iterations after discarding the first 100,000. The estimation of the posterior survival function for durations given starting times and locations was done using a truncation index  $G = 30$ .

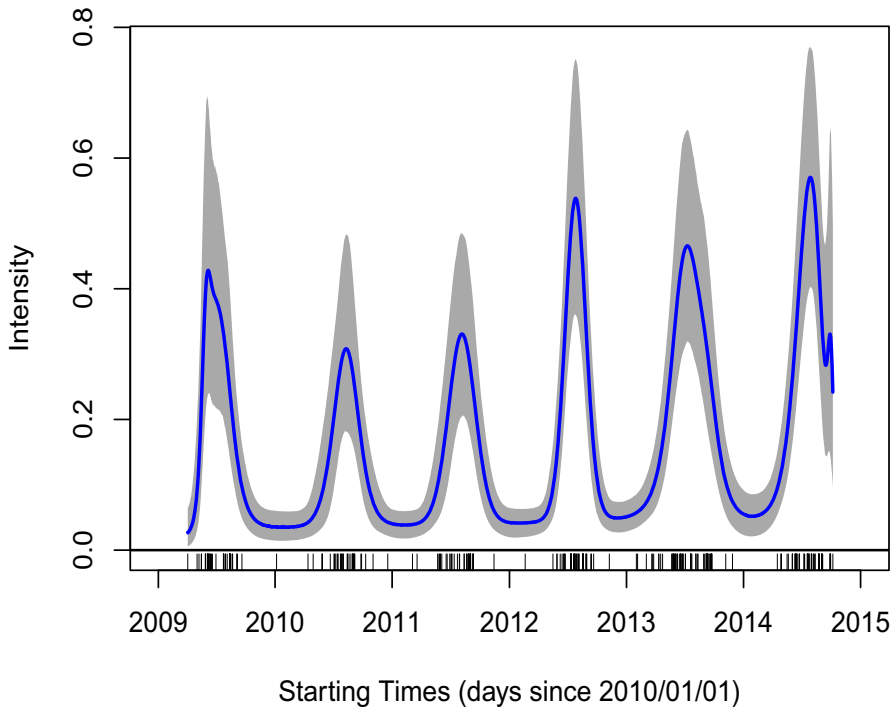
Given that the proposed model for the intensity function is based on a random distribution, under the assumption of an underlying Poisson point process, our model falls in the class of Cox point processes with random density (3). In order to check the validity of the Poisson assumption, we computed envelopes for the inhomogeneous  $L$ -function for a non-homogeneous Poisson process with density proportional to the intensity function shown in Fig. 3a (Møller and Waagepetersen 2004). Figure 4 shows the observed  $L$ -function computed from the observations and the 95% confidence



**Fig. 4** Confidence band for the  $(L(r) - r)$  function in terms of distance  $r$  (in meters), computed for the observed spatial marginal point pattern of wildfires in the Valencian Community between January 1st, 2010, and December 31st, 2015. Dotted lines represent the upper and lower band limits and the solid black line represents the observed non-homogeneous  $(L(r) - r)$  function

bands. The observed non-homogeneous  $L$ -function lies inside the confidence bands, showing that there is no strong evidence against the Poisson assumption. Although one would expect some level of interaction between wildfires as areas burned by one fire become less prone to have another ignition, one should also consider that only a small proportion of wildfires burn large areas, and that the random origin of the ignition source for naturally caused wildfires—mostly lightning strikes—is likely to start wildfires in other areas. Further, the spread of the small number of fires observed in this study explains the lack of evidence of interactions.

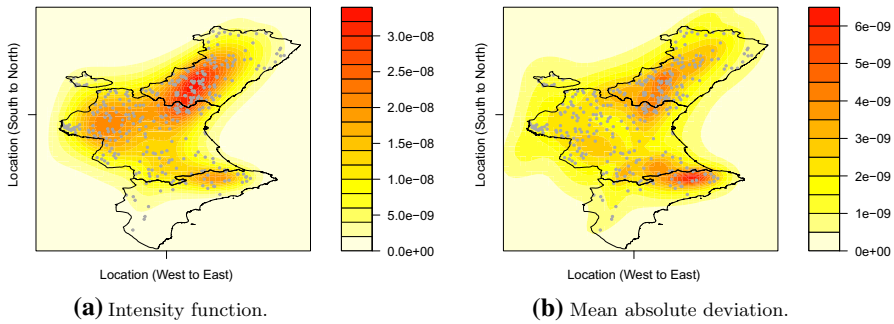
Figure 5 shows the estimated posterior intensity function for starting times. The flexibility of the model captures quite accurately the trend and local varying cycles across time. The figure also reveals that ignition times due to lightning strikes are highly concentrated in spring/summer, which is consistent with climatic factors in the region. In addition, the intensity function shows a positive trend over time in the number of outbreaking wildfires in the study area, a phenomenon that is probably associated to the change from agricultural to industrial society in Spain (Moreira et al. 2011). Land abandonment is a social factor related to wildfire increase in size, severity and risk in Europe (McLauchlan et al. 2020; Perpiña Castillo et al. 2021).



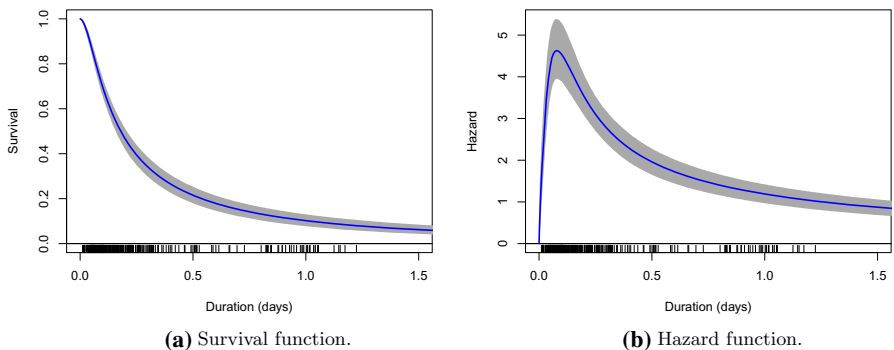
**Fig. 5** Estimated posterior intensity function (solid blue curve) for wildfire starting times. Shaded gray area represents the 95% credible interval. The symbols “|” below the  $x$ -axis are the observed starting times

Abandonment of agricultural lands results in shrub growth, increasing at a slow pace of highly flammable fuels (Jaber et al. 2001). The abandonment of farm lands induces an ecological succession process known as “old field succession”, which has been related to increased fire risk in Spain (Santana et al. 2010) as the abandoned farm lands represent an adequate habitat for grasses, weeds and other highly combustible pioneering vegetal species. The effect of the change from agricultural to industrial society on the incidence of forest fires in Spain has been discussed by Velez (1998), and for other Mediterranean countries by Rego (1992); Alexandrian and Esnault (1998); Pérez B. (2003). A detailed treatment of the subject is beyond the scope of this paper and the reader can find further details in the literature in ecology and ecological succession.

Figure 6a shows the estimated posterior intensity function for locations through a heat map. Notice that high-intensity zones are concentrated in the north-central portion of the Iberian System (Mira, Martés and Aledua mountain chains), where the type of flora that abounds corresponds to bushes, pastures, conifers and mountainous forests. The high risk of wildfires is possibly explained by the presence of these rapid combustion elements, together with the difficult access due to the altitude. Figure 6b displays the posterior mean absolute deviation as a measure of dispersion around the estimated intensity function.

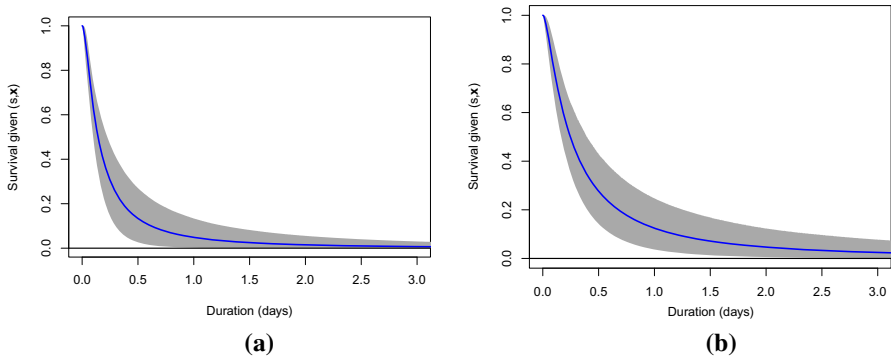


**Fig. 6** Estimated posterior intensity function and mean absolute deviation for wildfire locations. Solid gray dots represent the observed data



**Fig. 7** Estimated posterior survival and hazard functions for wildfire durations. Solid blue curves represent point-wise posterior means, and shaded gray areas, the corresponding 95% credible intervals. The symbols “|” below the  $x$ -axis indicate the observed durations

Wildfire duration data can be seen as lifetimes generated by a heavy-tailed survival distribution (Xi et al. 2019). In fact, the hazard function is also useful for understanding the severity and impact of wildfires on human communities and ecosystems. Indeed, for duration data, proportional hazard and frailty models (Kalbfleisch and Prentice 2002) are frequently used to assess how durations vary in space through a collection of fire compartments (Martell and Sun 2008). See also Morin (2014); Morin et al. (2015); Bayham (2015) for nice discussions and modeling details about fire durations. Here, we use the outcome of the Bayesian nonparametric estimation to estimate the posterior survival function. Figure 7a shows such a posterior estimate, revealing a right-asymmetric shape with heavy tail. For display purposes, we truncated the  $x$ -axis at 1.5 days. It should be noted that the estimated probability that a wildfire lasts more than 1.5 days lies between 0.044 and 0.087, with probability 0.95. Figure 7b shows the corresponding estimated posterior hazard function. There is little association between the pattern of lightning strikes and fire ignitions. Wildfires triggered by lightning strikes occur at random despite the existence of zones with high lightning-strike intensity (Agee 1993).



**Fig. 8** Estimated posterior survival function for wildfire durations given two events  $(s, \mathbf{x})$ . Solid blue curves represent point-wise posterior means and shaded gray areas, the corresponding 95% credible intervals

To illustrate the versatility of the proposed model in estimating functions of the duration, Figs. 8a and b show estimated posterior survival curves for durations given two different space-time observed events, selected at random. Notice the difference on the decay of the curves. In Fig. 8a, the selected event has an associated duration  $l = 0.0729$  days (1.75 h), whereas for Fig. 8b,  $l = 1.1458$  days (27.5 h). Notice that Fig. 8a corresponds to a low fire incidence area, while Fig. 8b to a high incidence area. Clearly, the corresponding survival function shows a shorter duration for low incidence areas, whereas for high fire incidence zones, the survival times are longer. This could be due to the different fuel availability (Agee 1993).

## 6 Conclusions

Wildfires are becoming a common event in the European Mediterranean zone. Their impact in landscape change,  $\text{CO}_2$  emissions to the atmosphere and the hazard they pose to human life and property make the study of the factors associated to their presence and the development of predicting models an important task. Here we have presented a model that considers the spatio-temporal variability and duration of wildfires, three of the most important properties for prevention, risk assessment and management. The case study is representative of a significant part of the European Mediterranean zone. Our proposal represents a good tradeoff between model generality and implementation feasibility.

Weighted DPM models are a rich class of stochastic processes that provide great flexibility in terms of modeling prior beliefs, avoiding the sometimes restrictive parametric assumptions about conditional intensity functions used in classical space-time modeling.

The approach undertaken here can potentially be explored in several directions. For example, other marks, such as the area burned by a wildfire, can be incorporated in the joint intensity function. Space-time covariates  $\mathbf{z}(s, \mathbf{x}) \in \mathbb{R}^p$ , such as weather conditions, wind speed, etc., could potentially be introduced in the joint intensity function, at the expense of adding more complexity.

**Acknowledgements** José J. Quinlan gratefully recognizes the support provided by the National Agency for Research and Development (ANID) through Fondecyt Grant 3190324 and CONACyT Grant 241195. Carlos Díaz-Avalos and Ramsés H. Mena wish to thank the support provided by PAPIIT-UNAM Project IG100221.

**Data availability** The data that support the findings of this study are available from the corresponding author upon request.

**Code availability** Computer programs used in the analysis are available from the corresponding author upon request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This work does not contain any studies with human participants and/or animals performed by any of the authors.

## References

- Agee JK (1993) Fire ecology of Pacific Northwest forests. Island Press, Washington, DC
- Ager A, Finney M, McMahan A (2006) A wildfire risk modeling system for evaluating landscape fuel treatment strategies. USDA Forest Service Proceedings RMRS-P-41
- Alexandrian D, Esnault F (1998) Policies Affecting Forest Fires in the Mediterranean Basin, FAO Meeting on Public Policies Affecting Forest Fires
- Amatulli G, Pérez-Cabello F, Riva J (2007) Mapping lightning/human-caused forest fires occurrence under ignition point location uncertainty. *Ecol Model* 200:321–333
- Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann Stat* 2:1152–1174
- Aragó P, Juan P, Díaz-Avalos C, Salvador P (2016) Spatial point process modeling applied to the assessment of risk factors associated with forest wildfires incidence in Castellón, Spain, *European Journal of Forest Research*, 451–464
- Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R, 1st edn. Chapman and Hall/CRC, Boca Raton
- Bar Massada A, Radeloff VC, Stewart SI, Hawbaker TJ (2009) Wildfire risk in the wildland-urban interface: a simulation study in northwestern Wisconsin. *For Ecol Manage* 258:1990–1999
- Barros A, Pereira J (2014) Wildfire selectivity for land cover type?: Does size matter? *PLoS ONE*. <https://doi.org/10.1371/journal.pone>
- Bayham J (2015) Characterizing Incentives: An Investigation of Wildfire Response and Environmental Entry Policy. PhD Dissertation, Washington State University. Outstanding Doctoral Dissertation Honorable Mention, *American Journal of Agricultural Economics* 97:656
- Birost Y (2009) Living with wildfires: What science can tell us—a contribution to the science-policy dialogue, EFI Discussion Paper, 15
- Blackwell D, MacQueen JB (1973) Ferguson Distributions Via Polya Urn Schemes, vol. 1
- Cheng N, Yuan T (2013) Nonparametric Bayesian lifetime data analysis using Dirichlet process Lognormal mixture model. *Naval Res Logist (NRL)* 60:208–221
- Daley DJ, Vere-Jones D (2003) An introduction to the theory of point processes, vol. i (elementary theory and methods) of probability and its applications. Springer-Verlag, New York
- Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes, vol. II (General theory and structure) of probability and its applications. Springer-Verlag, New York
- Dayananda P (1977) Stochastic models for forest fires. *Ecol Model* 3:309–313
- Díaz-Avalos C, Juan P, Serra-Saurina L (2016) Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes. *For Ecol Manage* 381:360–369
- Díaz-Avalos C, Peterson DL, Alvarado E, Ferguson SA, Besag JE (2001) Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canad J For Res* 31:1579–1593

- Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns. CRC Press, Chapman & Hall/CRC Monographs on Statistics & Applied Probability
- Duane A, Piqué M, Castellnou M, Brotons L (2015) Predictive modelling of fire occurrences from different fire spread patterns in Mediterranean landscapes. *Int J Wildl Fire* 24:407
- Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *J Am Stat Assoc* 90:577–588
- Faivre N, Rego FM, Rodríguez JM, Calzada VR, Xanthopoulos G (2018) Forest Fires—Sparkling firesmart policies in the EU
- Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Ann Stat* 1:209–230
- Fúquene J (2015) A semi-parametric Bayesian extreme value model using a Dirichlet process mixture of Gamma densities. *J Appl Stat* 42:267–280
- Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns, Statistics in Practice. John Wiley & Sons Ltd, Chichester
- Ishwaran H, Zarepour M (2002) Exact and approximate sum representations for the Dirichlet process. *Canad J Stat* 30:269–283
- Jaber A, Guarnieri F, Wybo I (2001) Intelligent software agents for forest fire prevention and fighting. *Saf Sci* 39:3–17
- Juan P, Mateu J, Saez M (2012) Pinpointing spatio-temporal interactions in wildfire patterns. *Stoch Environ Res Risk Assess* 26:1131–1150
- Kalbfleisch J, Prentice R (2002) The statistical analysis of failure time data. John Wiley and Sons Inc, New York
- Kottas A (2006a) Dirichlet process mixtures of Beta distributions, with applications to density and intensity estimation, in Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods
- Kottas A (2006b) Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *J Stat Plann Inference* 136:578–596
- Kottas A, Müller P, Quintana F (2005) Nonparametric Bayesian modeling for multivariate ordinal data. *J Comput Graph Stat* 14:610–625
- Kottas A, Sansó B (2007) Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *J Stat Plann Inference* 137:3151–3163
- Lindner M, Maroschek M, Netherer S, Kremer A, Barbati A, Garcia-Gonzalo J, Seidl R, Delzon S, Corona P, Kolström M, Lexer MJ, Marchetti M (2010) Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *For Ecol Manage* 259:698–709, adaptation of Forests and Forest Management to Changing Climate
- MacEachern SN (1994) Estimating Normal means with a conjugate style Dirichlet process prior. *Commun Stat Simul Comput* 23:727–741
- Mandallaz D, Ye R (1997) Prediction of forest fires with Poisson models. *Can J For Res* 27:1685–1694
- Marques S, Borges JG, Garcia-Gonzalo J, Moreira F, Carreiras JMB, Oliveira MM, Cantarinha A, Botequim B, Pereira JMC (2011) Characterization of wildfires in Portugal. *Eur J For Res* 130:775–784
- Martell DL, Sun H (2008) The impact of fire suppression, vegetation, and weather on the area burned by lightning-caused forest fires in Ontario. *Can J For Res* 38:1547–1563
- McKenzie D, Miller C, Falk DA (eds) (2011) The landscape ecology of fire. Springer, Dordrecht
- McLauchlan KK, Higuera PE, Miesel J, Rogers BM, Schweitzer J, Shuman JK, Tepley AJ, Varner JM, Veblen TT, Adalsteinsson SA, Balch JK, Baker P, Battlori E, Bigio E, Brando P, Cattau M, Chipman ML, Coen J, Crandall R, Daniels L, Enright N, Gross WS, Harvey BJ, Hatten JA, Hermann S, Hewitt RE, Kobziar LN, Landesmann JB, Lorant MM, Maezumi SY, Mearns L, Moritz M, Myers JA, Pausas JG, Pellegrini AFA, Platt WJ, Roozeboom J, Safford H, Santos F, Scheller RM, Sherriff RL, Smith KG, Smith MD, Watts AC (2020) Fire as a fundamental ecological process: research advances and frontiers. *J Ecol* 108:2047–2069
- Miguel C, Francisco M, Pedro C, Pedro V (2011) Land use and topography influences on wildfire occurrence in northern Portugal. *Landsc Urban Plann* 100:169–176
- Miranda A, Coutinho M, Borrego C (1994) Forest fire emissions in Portugal: a contribution to global warming? *Environ Pollut* 83:121–123
- Møller J, Díaz-Avalos C (2010) Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scand J Stat* 37:2–25
- Møller J, Waagepetersen RP (2004) Statistical inference and simulation for spatial point processes. Chapman & Hall/CRC, Boca Raton



- Moreira F, Viedma O, Arianoutsou M, Curt T, Koutsias N, Rigolot E, Barbatí A, Corona P, Vaz P, Xanthopoulos G, Mouillot F, Bilgili E (2011) Landscape—wildfire interactions in southern Europe: implications for landscape management. *J Environ Manage* 92:2389–2402
- Morin A (2014) A spatial analysis of forest fire survival and a marked cluster process for simulating fire load, Master's thesis, Electronic Thesis and Dissertation Repository, The University of Western Ontario, Canada <https://ir.lib.uwo.ca/etd/2192>
- Morin A, Albert-Green A, Woolford D, Martell D (2015) The use of survival analysis methods to model the control time of forest fires in Ontario, Canada. *Int J Wildl Fire* 24:964–973
- Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture models. *J Comput Graph Stat* 9:249–265
- Padilla M, García A (2011) On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. *Int J Wildl Fire* 20:46–58
- Pérez B, Cruz A, Fernández-González F, Moreno JM (2003) Effects of the recent land-use history on the postfire vegetation of uplands in Central Spain. *For Ecol Manage* 182:273–283
- Perpiñá Castillo C, Jacobs-Crisioni C, Diogo V, Lavalle C (2021) Modelling agricultural land abandonment in a fine spatial resolution multi-level land-use model: an application for the EU. *Environ Model Softw* 136:104946
- Rego F (1992) Land use changes and wildfires. Eisevier, London
- Santana V, Baeza M, Marrs R, Vallejo R (2010) Old-field secondary succession in SE Spain. *Plant Ecol* 211:337–349
- Serra L, Saez M, Mateu J, Varga D, Juan P, Díaz-Avalos C, Rue H (2014) Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008. *Environ Ecol Stat* 21:531–563
- Sethuraman J (1994) A constructive definition of Dirichlet priors. *Stat Sin* 4:639–650
- Shakesby R (2011) Post-wildfire soil erosion in the Mediterranean: review and future research directions. *Earth Sci Rev* 105:71–100
- Taddy MA, Kottas A (2012) Mixture modeling for marked Poisson processes. *Bayesian Anal* 7:335–362
- Turner R (2009) Point patterns of forest fire locations. *Environ Ecol Stat* 16:197–223
- van der Werf GR, Morton DC, DeFries RS, Olivier JGJ, Kasibhatla PS, Jackson RB, Collatz GJ, Randerson JT (2009) CO<sub>2</sub> emissions from forest loss. *Nat Geosci* 2:737–738
- Velez R (1998) Report of the Working Group on the Mediterranean Basin. *FAO Meeting on Public Policies Affecting Forest Fires* 1:59–60
- Vilar L, Woolford D, Martell D, Martín M (2010) A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *Int J Wildl Fire* 19:325–337
- Xi DD, Taylor SW, Woolford DG, Dean C (2019) Statistical models of key components of wildfire risk. *Annu Rev Stat Appl* 6:197–222
- Yannaros N (1994) Weibull renewal processes. *Ann Inst Stat Math* 46:641–648

**José J. Quinlan** is a post-doctoral researcher position at the Pontificia Universidad Católica de Chile (PUC), Chile. His current research interests include Bayesian analysis, point processes and change-point analysis.

**Carlos Díaz-Avalos** is Associate Professor at the Universidad Nacional Autónoma de México (UNAM), Mexico. His current research interests include spatial statistics, random fields, linear models and wildfire models.

**Ramsés H. Mena** is Professor at the Universidad Nacional Autónoma de México (UNAM), Mexico. His current research interests include Bayesian nonparametrics, time series analysis and computational statistics.