WILEY

# A circular nonhomogeneous hidden Markov field for the spatial segmentation of wildfire occurrences

Jose Ameijeiras-Alonso[1]  |  Francesco Lagona[2]  |  Monia Ranalli[2]  |  Rosa M. Crujeiras[1]

[1]Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

[2]Department of Political Sciences, University of Roma Tre, Rome, Italy

**Correspondence**
Francesco Lagona, Deptartment of Political Sciences, University of Roma Tre, via G. Chiabrera 199, 00145 Rome, Italy.
Email: francesco.lagona@uniroma3.it

**Abstract**

Motivated by studies of wildfire seasonality, we propose a nonhomogeneous hidden Markov random field to model the spatial distribution of georeferenced fire occurrences during the year, by representing occurrence times as circular data. The model is based on a mixture of Kato–Jones circular densities, whose parameters vary across space according to a latent nonhomogeneous Potts model, modulated by georeferenced covariates. It allows us to segment fire occurrences according to a finite number of latent classes that represent the conditional distributions of the data under specific periods of the year, simultaneously accounting for unobserved heterogeneity and spatial autocorrelation. Further, it parsimoniously accommodates specific features of wildfire occurrence data such as multimodality, skewness, and kurtosis. Due to the numerical intractability of the likelihood function, estimation of the parameters is based on composite likelihood methods. It reduces to a computationally efficient expectation–maximization algorithm that iteratively alternates the maximization of a weighted composite likelihood function with weights updating. The proposal is illustrated in a study of wildfire occurrences in the Iberian Peninsula during a decade.

**KEYWORDS**

composite likelihood, fires, Kato–Jones density, land use, Markov random field, spatial circular data

## 1 | INTRODUCTION

Wildfires represent a major environmental and economic issue in Southern Europe. In Spain and Portugal, for example, some areas have been recently devastated by wildfires. In mid-June 2017, more than 45,000 hectares were devastated in the district of Leiria (central Portugal), with around 150 active fires in a week. Some years before, in 2006, the damaged region was Galicia (NW Spain), with almost 2,000 active fires during the first fortnight of August, which destroyed an area larger than the one affected by wildfires in the five previous years.

Studies on wildfires typically focus on specific aspects of wildfires, such as fire frequency, fire intensity, and fire extension. This paper focuses on the spatial distribution of wildfire seasonality. As a part of fire regimes characterization, the spatial analysis of fire seasonality is crucial for understanding the behavior of fires at global and local spatial scales (Benali et al., 2017). In addition, knowledge of seasonal patterns of fires across space is useful for designing appropriate precautionary and intervention measures against wildfires.

Basic information available for our study includes occurrence of georeferenced wildfire events (day of the year) across a decade (2002–2012). On a yearly scale, event times can be placed on a unit circle corresponding to a period of 365 days

which, without loss of generality, can be rescaled to the support $[0, 2\pi)$. The idea of studying seasonality by viewing event times as circular data is not new and has been, for example, recently used by Shirota and Gelfand (2017) in a study of crime events. By taking this approach, georeferenced event times can be represented as spatial circular data.

While methods for handling independent circular data are well established (Ley & Verdebout, 2017), the analysis of spatially dependent circular data is a new emerging area of research. Some proposals in this context rely on the extension of geostatistical models to the circular setting. Spatial wrapped Gaussian processes (Jona-Lasinio, Gelfand, & Jona-Lasinio, 2012) and spatial projected-Gaussian processes (Wang & Gelfand, 2014) have been proposed for modeling sea motion. Consistent with the geostatistical paradigm, these models consider spatial circular processes that vary continuously over space. Alternative approaches are based on circular processes that vary discretely over space. Some proposals in this context make use of adaptations of spatial autoregressive processes to the circular setting. For example, Modlin, Fuentes, and Reich (2012) introduced a circular autoregressive model, based on the wrapped normal distribution for studying hurricane wind directions. Circular autoregressive models, based on the multivariate von Mises distribution, have been proposed for characterizing sea wave directions (Lagona, 2016; Lagona, Picone, Maruotti, & Cosoli, 2015). Other proposals consider hidden Markov random fields (MRFs) for spatial circular data (Lagona & Picone, 2016; Ranalli, Lagona, Picone, & Zambianchi, 2018). In this setting, spatial circular data are conditionally independent given a latent MRF, that is, a multinomial process in discrete space that fulfills a spatial Markov property (Gaetan & Guyon, 2010) and that segments the study area according to a finite number of latent classes. This approach is useful when the interest is focused on segmenting the area under study according to a small number of classes, each one associated with a specific distribution of the data.

Hidden MRFs are popular models in spatial statistics, since the seminal paper by Besag (1975). They can be seen as a spatial extension of the hidden Markov models that are exploited in time series analysis. Circular extensions of hidden Markov models have been widely used for the analysis of time series with circular components (Holzmann, Munk, Suster, & Zucchini, 2006; Lagona & Picone, 2013; Lagona, Picone, & Maruotti, 2015; Maruotti, Punzo, Mastrantonio, & Lagona, 2016; Mastrantonio, Maruotti, & Jona-Lasinio, 2015). On the contrary, the widespread use of circular hidden MRFs has been limited by the intractability of the likelihood function of these models. Following Alfó, Nieddu, and Vicari (2008), Lagona and Picone (2016) adapted a mean-field approximation for Gaussian hidden MRFs to the circular setting and developed a computationally intensive expectation–maximization (EM) algorithm. Unfortunately, the method is numerically unstable, and little is known about the distributional properties of the estimators. More recently, Ranalli et al. (2018) suggested composite likelihood (CL) methods to estimate a (homogeneous) circular hidden MRF, showing that this method provides a good solution to balance statistical and computational efficiency, through an extensive simulation study.

In studies of wildfire seasonality, circular hidden Markov fields provide a natural approach to segment an area of interest according to regions that are associated with specific seasonal patterns of fire events. In this paper, specifically, we assume that the distribution of the occurrence of fires (recorded as days on a yearly base and viewed as circular data) is well approximated by a mixture of Kato–Jones densities, whose parameters vary across space according to a nonhomogeneous Potts model. The Kato–Jones density is a four-parameter unimodal density that flexibly accommodates skewness and kurtosis on the circle (Kato & Jones, 2015). The proposed nonhomogeneous Potts model is an MRF, whose parameters depend on georeferenced covariates. The model is estimated by extending the proposal by Ranalli et al. (2018) to the case of a nonhomogeneous hidden MRF that is modulated by georeferenced covariates.

The rest of the paper is organized as follows. Section 2 briefly describes the wildfire data that motivated this study. Section 3 presents the structure of the proposed nonhomogeneous hidden MRF, and Section 4 illustrates the CL methods that we suggest for estimation. Section 5 is devoted to the results that have been obtained by the proposed methods on the real data. Section 6 finally summarizes relevant points of discussion.

## 2 | WILDFIRE OCCURRENCES IN THE IBERIAN PENINSULA

The identification of fire peaks in certain spatial areas should serve to warn the authorities in order to organize appropriate interventions or campaigns, especially when a critical period is highly probable (a critical day for fire risk is identified by the 30-30-30 conditions: more than 30°, winds of more than 30 km/h, and less than 30% of humidity).

In the Iberian Peninsula, most fires occur during summer with a peak of activity in August–September. In terms of climatological and weather conditions, summer is the most suitable season for fires in latitudes over the Tropic of Cancer, because dry weather conditions are predominant. Authorities are certainly concerned with wildfire problems during

summer season, and rangers and firefighters groups are reinforced in this period. Nevertheless, wildfires do not only occur in summer. A recent example was October 2017, when northern Portugal counted 600 fires in 5 days, and in Galicia, more than 100 fires occurred in a day. In this case, authorities were not prepared for fighting wildfires, because October is not considered as a *peak season*.

In addition to summer weather conditions, land management practices also influence fire seasonality, showing in this case preferential timings. Although some human fires are produced intentionally or unintentionally during dry months, there are other activities that cause fires in periods that are outside the principal peak of fires. These activities include, for example, agriculture burning for preparing fields (for harvest work) or for clearing the crop residues (after harvesting or in order to avoid future burnings during the climatological season of fires). The distribution of fire activity peaks during the year and across the study area helps explain where and how human activity changed fire seasonality using fires as a land management tool (Ameijeiras-Alonso, Crujeiras, & Rodríguez-Casal, 2018).

For studying the distribution of the times at which wildfires occur, we collected data of the fires in the Iberian Peninsula from July 10, 2002 to July 9, 2012, detected by the MODerate resolution Imaging Spectroradiometer (MODIS), launched into the Earth orbit by NASA on board of the Terra (*EOS AM*) and the Aqua (*EOS PM*) satellites. MODIS identifies locations where fires are actively burning at the time of satellite overpass, using for this purpose an algorithm that summarizes a number of measures such as brightness, temperatures, cloud and water masks, and the sun glint (Giglio, Descloitres, Justice, & Kaufman, 2003; Oom & Pereira, 2013).

During the study period, a total of 63,127 fires were detected by MODIS in the inland territory of Spain and Portugal. By wrapping the day of occurrence of these fires around a circle of circumference 365 days, we obtain the rose diagram in Figure 1 (left side), where the area of each circular sector is proportional to the class frequency. The overlapped von Mises kernel nonparametric density estimate (Taylor, 2008) shows the multimodality of these data. The data distribution is, however, better visualized by unwrapping the kernel estimate and overlapping a histogram of the data distribution with support $[0, 2\pi]$ (Figure 1, middle panel). As expected, most of the fires are concentrated around the first fortnight of August (climatological peak of fires). However, two additional peaks appear during the second fortnight of March and at the beginning of October. This alteration of fire seasonality is well documented in the forestry literature (see, e.g., Benali et al., 2017; Korontzi, McCarty, Loboda, Kumar, & Justice, 2006; Le Page, Oom, Silva, Jönsson, & Pereira, 2010; Magi, Rabin, Shevliakova, & Pacala, 2012), and the regularity of these multimodal patterns was recently studied by Ameijeiras-Alonso (2017) in a worldwide setting.

Ignoring the spatial dimension of the data, Figure 1 suggests that the distribution of fire occurrence over the year is a mixture of either two or three components that pertain to specific seasonal patterns of fire occurrences. We are interested in modeling the spatial distribution of these components through a hidden MRF. The model accommodates the spatial aspect of the data by a neighborhood structure between wildfire sites.

Such a spatial neighborhood structure can be specified in different ways (Bivand, Pebesma, & Gómez-Rubio, 2008, chapter 9). Let $S = \{1, \dots, n\}$ be the set of the wildfire locations in the Iberian Peninsula. In order to define a neighborhood structure on $S$, the study area is partitioned by a grid of three-hectare square cells. Under this setting, a cell may include 0, 1, or more fire locations. Let $\mathbb{A}$ be the set of the grid cells where at least one fire event occurred during the
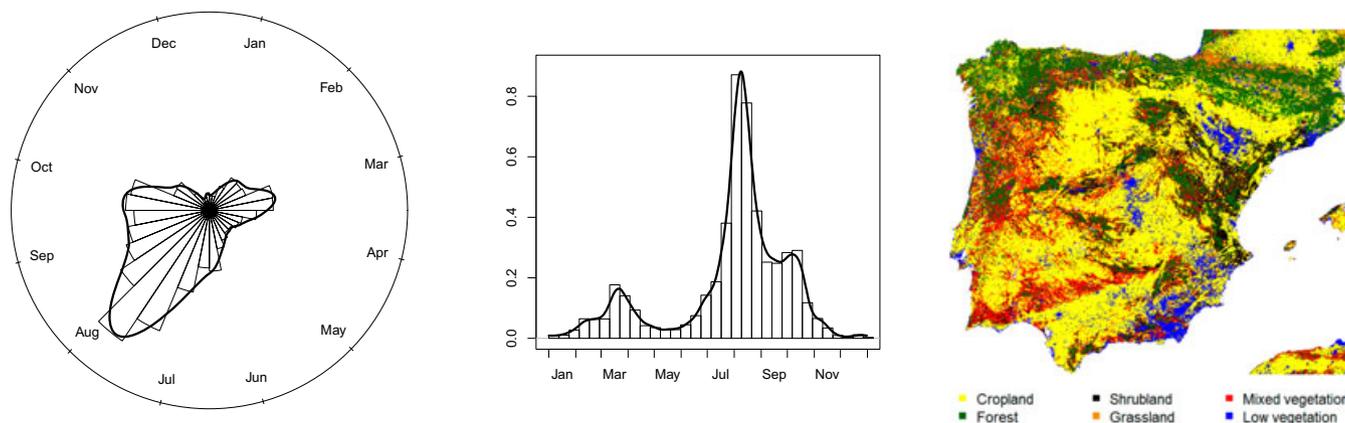


**FIGURE 1** Distribution of fire occurrence times and land cover in the Iberian Peninsula during the period 2002–2012. Left: rose diagram of the observed fire occurrence times with a circular kernel density estimate (continuous line). Middle: the unwrapped circular kernel density estimate, overlapped on the histogram of fire occurrence times (the ticks represent the beginning or the end of the month that is indicated between them). Right: most frequent land cover of the Iberian Peninsula during the study period
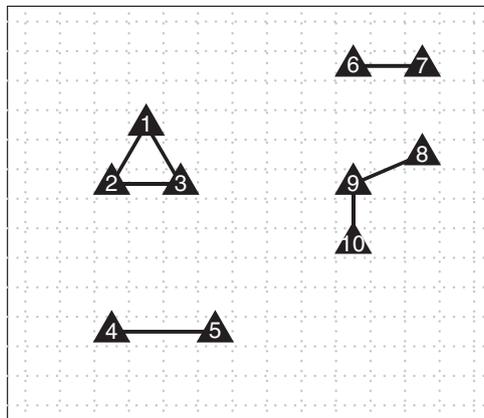
**FIGURE 2** A stylized picture of the neighborhood structure exploited in this study. The study area is partitioned according to a regular square grid where each square cell is an area of three hectares. Triangles indicate the occurrence of at least one fire in the grid cell. Neighboring cells are indicated by segments

observation period. Precisely, for each grid cell $a \in \mathbb{A}$, the set $N(a)$ of the neighboring cells was defined by including all the nearest cells, according to the Euclidean distance between the cell centroids. Figure 2 displays a stylized example with 10 grid cells: Triangles indicates cells where at least one fire occurred, and segments indicate their nearest neighbors. For each fire location $i \in S$, let $a(i)$ be the cell that includes location $i$, and let $N(a(i))$ be the set of those cells that are neighbors of cell $a(i)$. For each fire location $i$, the neighborhood of location $i$ was defined by including all the other fire locations that belong to $a(i)$ (if $a(i)$ includes further fire locations besides $i$) and all the fire locations in the cells that are neighbors of $a(i)$, formally

$$N(i) = \{j \in a(i) : j \neq i\} \cup \{j \in a : a \in N(a(i))\}. \tag{1}$$

Referring to Figure 2, for example, the neighboring fire locations of a fire location that belongs to cell 9 include all the other fire locations in cell 9 and, additionally, all the fire locations in cells 8 and 10.

Figure 1 displays the distribution of time of fire occurrence (left and middle panels) and also shows a summary of land cover characterization on the study period (right panel). This covariate is essential in our analysis given its impact on fire occurrence timings. We obtained annual land cover at a three-hectare resolution from the European Space Agency Climate Change Initiative project (Land Cover version 2.0.7; available at http://www.esa-landcover-cci.org), which describes the physical material at the surface of the earth, including various types of vegetation, bare rock and soil, water, snow and ice, and artificial surfaces. In our study, each fire occurrence is associated with the land cover of the year when the fire was detected. Annual land cover is included as a factor with six levels that respectively indicate *cropland* (rainfed; irrigated or postflooding), *forest* (tree cover; broadleaved, needleleaved, or mixed leaf type; evergreen or deciduous; closed or open), *shrubland*, *grassland* (herbaceous cover, grassland), *mixed vegetation*, and *low vegetation* (sparse vegetation; tree cover, flooded; urban areas; bare areas; water bodies; permanent snow and ice). In the Iberian Peninsula, a total of 16,145 fires were produced in cropland, 25,963 in forest, 9,868 in shrubland, 1,621 in grassland, 3,096 in mixed vegetation, and 6,434 in low vegetation areas. Figure 1 (right) displays the modal land cover during the 10-year study period.

## 3 | A CIRCULAR HIDDEN MRF

The data that motivated this work are in the form of a spatial series of circular observations, say, $\mathbf{y} = (y_i, i = 1, \dots, n)$, $y_i \in [0, 2\pi)$. Each observation is associated with the $i$th row vector $\mathbf{x}_i^\top$ of a design matrix $\mathbf{X}$. In our case study, $y_i$ indicates the day of the year of the $i$th fire event, while $\mathbf{X}$ is a six-column coding matrix that includes land cover (of the year when the fire was detected) as a factor with six levels. In this section, we describe a circular hidden MRF model that spatially segments these data according to $K$ latent classes. The model can be seen as a mixture of $K$ unimodal circular densities, whose parameters vary across space according to a latent multinomial process. We specify the mixture components as Kato–Jones densities (Kato & Jones, 2015). The latent multinomial process is instead specified by a nonhomogeneous Potts model (Strauss, 1977), whose parameters are modulated by the design matrix $\mathbf{X}$. In Section 3.1, we briefly describe the Kato–Jones density, and Section 3.2 is devoted to the Potts model. Finally, Section 3.3 describes the proposed HMRF model, by combining the Kato–Jones distribution with the latent Potts model.
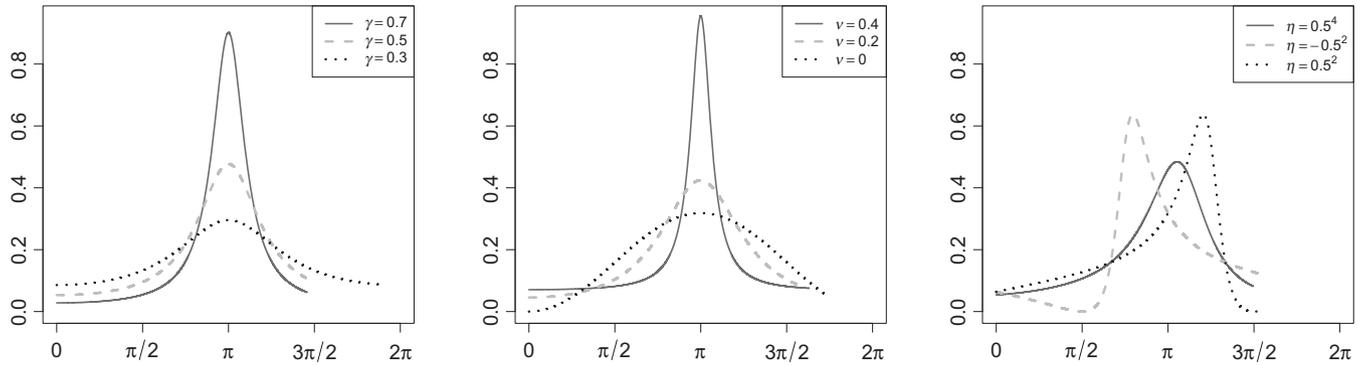
**FIGURE 3** The Kato–Jones density. Left: the shape of the density with $\mu = \pi, \nu = \gamma^2, \eta = 0$ and varying concentration parameter $\gamma$. Middle: shape with $\mu = \pi, \gamma = 0.5, \eta = 0$ and varying kurtosis parameter $\nu$. Right: shape with $\mu = \pi, \gamma = 0.5, \nu = 0.25$ and varying skewness parameter $\eta$

## 3.1 | The Kato–Jones density

The choice of the component density in a mixture model should be always driven by the purpose of the analysis. If the purpose is model-based classification, then the component should be unimodal and appropriate in light of the data (McNicholas, 2016, chapter 9, p. 157). A recent proposal that suits our needs is the four-parameter Kato–Jones density, namely,

$$f(y; \mu, \gamma, \nu, \eta) = \frac{1}{2\pi} \left( 1 + 2\gamma^2 \frac{\gamma \cos(y - \mu) - \nu}{\gamma^2 + \nu^2 + \eta^2 - 2\gamma(\nu \cos(y - \mu) + \eta \sin(y - \mu))} \right), \quad y \in [0, 2\pi), \tag{2}$$

where $(\nu, \eta) \neq (\gamma, 0)$ and

$$0 \leq \mu < 2\pi$$
$$0 \leq \gamma < 1 \tag{3}$$
$$(\nu - \gamma^2)^2 + \eta^2 \leq \gamma^2(1 - \gamma)^2.$$

Figure 3 displays the density shape for this circular distribution model by varying the values of some parameters. In our approach, such a model will be used for characterizing the distribution of fire occurrence times, $y_i$. The properties of this density are extensively described by Kato and Jones (2015). We summarize here the most practical advantages of this density. First, the meaning of the parameters is intuitively appealing. Precisely, $\mu$ indicates the circular mean, $\gamma$ is the concentration parameter (if $\gamma = 0$, then the density reduces to a uniform density on the circle), $\nu$ measures kurtosis (the greater the value of $\nu$, the sharper the peakedness of the density), and $\eta$ indicates skewness (the density is symmetric if and only if $\eta = 0$). Second, samples from the Kato–Jones distribution can be easily generated by drawing samples from the wrapped Cauchy distribution and then by using the acceptance/rejection algorithm proposed by Kato and Jones (2015). Third, maximum likelihood is numerically tractable by using a maximization algorithm that allows for constraints such as that provided by the `nlminb` function in R. This function requires that the range of each parameter should not depend on other parameters, and a suitable reparametrization is necessary, such as that suggested by Kato and Jones (2015).

## 3.2 | A nonhomogeneous Potts model

The Potts model is a multinomial process in discrete space (lattice) with $K$ classes. Given a lattice that divides an area of interest according to $n$ observation sites $i = 1, \dots, n$, a sample that is drawn from a spatial multinomial process is a segmentation of this area, obtained by associating each site with a segmentation label $k = 1, \dots, K$. Formally, each observation site $i$ is associated with a multinomial random variable $\mathbf{U}_i = (U_{i1}, \dots, U_{iK})$ with one trial and $K$ classes, where $U_{ik}$ is a Bernoulli random variable that is equal to 1 if $i$ is labeled by $k$ and 0 otherwise. A specific segmentation of the area can be accordingly represented as a sample drawn from the multinomial process $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_n)$. The Potts model is a spatial multinomial process that accounts for a neighborhood structure $N(i), i = 1, \dots, n$ among the observation sites, which associates each site with a set $N(i)$ of neighbors. The neighborhood structure that was exploited for our case study is formally defined by Equation 1 (see Figure 2 for a graphical display). By taking the class $K$ as a reference, each

segmentation $\mathbf{u}$ is associated with $K$ sufficient statistics: the number of subsets of two neighboring sites that share the same class $k \neq K$, as follows:

$$n(\mathbf{u}) = \sum_{i=1}^{n} \sum_{j>i : j \in N(i)} \sum_{k=1}^{K-1} u_{ik}u_{jk},$$

and $K - 1$ sufficient statistics, as follows:

$$n_k(\mathbf{u}) = \sum_{i=1}^{n} u_{ik} \quad k = 1, \dots, K-1,$$

which indicate the number of neighboring sites that are associated with latent class $k$. Under the Potts model, the probability of a specific segmentation $\mathbf{u}$ is known up to $K$ parameters $\alpha_1, \dots, \alpha_{K-1}, \rho$, and it is given by

$$p(\mathbf{u}; \boldsymbol{\alpha}, \rho) = \frac{\exp\left(\sum_{k=1}^{K-1} n_k(\mathbf{u})\alpha_k + n(\mathbf{u})\rho\right)}{W(\boldsymbol{\alpha}, \rho)}, \quad (4)$$

where $W(\boldsymbol{\alpha}, \rho)$ is the normalizing constant. The parameter $\rho$ is an autocorrelation parameter: If it is positive (negative), then it penalizes segmentations with a few concordant (discordant) neighbors. This parameter is often referred to as a regularization parameter, given that large values of $\rho$ are associated with segmentations where areas with the same label are geometrically regular. Figure 4 shows three samples that have been drawn from a three-color Potts model with $\rho = 0, 0.5, 1$, defined on a regular lattice of $32 \times 32$ sites, by assuming that all the $\alpha_k$ parameters are equal to zero. The R library potts was used to generate these images. Each parameter $\alpha_k$ penalizes segmentations with a few sites that belong to class $k$. When $\rho = 0$, then (4) reduces to a multinomial distribution where the parameters $\alpha$ are class-specific log odds, as follows:

$$\alpha_k = \log \frac{P(u_{ik} = 1)}{P(u_{iK} = 1)}.$$

Model (4) is a homogeneous Potts model, because the parameters $\alpha_k$ do not vary across the observation sites. We extend this setting by assuming that these parameters depend on site-specific covariates, through a multinomial logistic model, as follows:

$$\alpha_{ik} = \mathbf{x}_i^\top \boldsymbol{\beta}_k \quad i = 1, \dots, n \quad k = 1, \dots, K-1, \quad (5)$$

where the regression coefficients $\boldsymbol{\beta}_k$ indicate the influence of the available covariates on the frequency of label $k$ in the segmentation. As a result, we obtain the nonhomogeneous Potts model

$$p(\mathbf{u}; \boldsymbol{\beta}, \rho) = \frac{\exp\left(\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\beta} + n(\mathbf{u})\rho\right)}{W(\boldsymbol{\beta}, \rho)}, \quad (6)$$

where $W(\boldsymbol{\beta}, \rho)$ is the normalizing constant. For each site $i$ and each label $k$, let

$$n_k\left(\mathbf{u}_{\bar{N}(i)}\right) = u_{ik} \sum_{j \in N(i)} u_{jk}$$

be the number of sites in the neighborhood of $i$ that are labeled by $k$, where $\bar{N}(i) = N(i) \cup \{i\}$ indicates the neighborhood of $i$, completed by $i$. Under model (4), the conditional distribution of each site depends only on the labels taken by the neighboring sites, namely,

$$p(u_{ik} = 1 \mid \mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \dots, \mathbf{u}_n) = \frac{\exp\left(\alpha_k + \rho n_k\left(\mathbf{u}_{\bar{N}(i)}\right)\right)}{1 + \sum_{k=1}^{K-1} \exp\left(\alpha_k + \rho n_k\left(\mathbf{u}_{\bar{N}(i)}\right)\right)}, \quad (7)$$
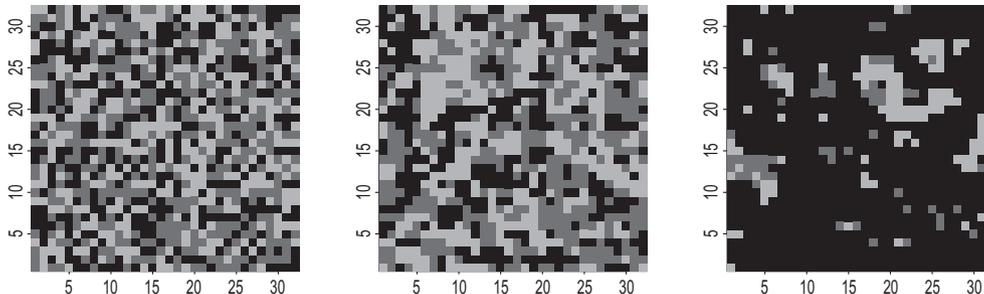


**FIGURE 4** Three samples respectively drawn from 3-color Potts models with $\rho = 0$ (left), $\rho = 0.5$ (middle) and $\rho = 1$ (right)

Accordingly, the Potts model is an MRF with respect to the neighborhood structure. Under the nonhomogeneous random field (6), these conditional probabilities reduce to

$$p(u_{ik} = 1 \mid \mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n) = \frac{\exp\left(\mathbf{x}^\top \boldsymbol{\beta}_k + \rho n_k\left(\mathbf{u}_{\bar{N}(i)}\right)\right)}{1 + \sum_{k=1}^{K-1} \exp\left(\mathbf{x}^\top \boldsymbol{\beta}_k + \rho n_k\left(\mathbf{u}_{\bar{N}(i)}\right)\right)}, \tag{8}$$

and the autocorrelation coefficient $\rho$ can be viewed as an autoregression coefficient that is associated with the spatially lagged outcome $n_k(\mathbf{u}_{\bar{N}(i)})$.

## 3.3 | A circular hidden MRF

The proposed hidden MRF is specified by assuming that the observed data are conditionally independent, given a segmentation generated by the nonhomogeneous Potts model (6). Precisely, we assume that the conditional distribution of the observed process, given the latent segmentation $\mathbf{u}$, takes the form of a product density, say,

$$f(\mathbf{y} \mid \mathbf{u}; \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} f(y_i; \boldsymbol{\theta}_k)^{u_{ik}}, \tag{9}$$

where the vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ includes $K$ label-specific parameters, and $f(y; \boldsymbol{\theta}_k)$, $k = 1, \ldots, K$, are $K$ Kato–Jones circular densities (2). The joint density of the observed data and the unobserved class memberships is therefore given by

$$f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}, \boldsymbol{\beta}, \rho) = f(\mathbf{y} \mid \mathbf{u}; \boldsymbol{\theta}) p(\mathbf{u}; \rho, \boldsymbol{\beta}). \tag{10}$$

By integrating this distribution with respect to the segmentation $\mathbf{u}$, we obtain the likelihood function of the unknown parameters

$$L(\boldsymbol{\theta}, \rho, \boldsymbol{\beta}) = \sum_{\mathbf{u}} f(\mathbf{y} \mid \mathbf{u}; \boldsymbol{\theta}) p(\mathbf{u}; \rho, \boldsymbol{\beta}). \tag{11}$$

When $\rho = 0$, then the model reduces to a latent class model with concomitant covariates. Otherwise, when $\rho \neq 0$, the model is a mixture of circular densities, whose parameters vary across space according to a Markovian process.

# 4 | PARAMETER ESTIMATION THROUGH CL METHODS

## 4.1 | An EM algorithm

Direct maximization of the likelihood function (11) is unfeasible. As a result, we propose to estimate the parameters by maximizing a surrogate function, namely, a CL function (Lindsay, 1988). Following Ranalli et al. (2018), our proposal relies on covering the set $S = \{1 \ldots n\}$ of the observation sites by all the pairs $S_2 = \{i, j\}$ of neighboring sites, that is, such that $i \in N(j)$ and $j \in N(i)$. The neighborhood structure that was exploited for our case study is formally defined by Equation 1 (see Figure 2 for a graphical display). For each subset $S_2$, we define

$$L_{S_2}(\boldsymbol{\theta}, \rho, \boldsymbol{\beta}) = \sum_{\mathbf{u}_{S_2}} p\left(\mathbf{u}_{S_2}; \rho, \boldsymbol{\beta}\right) \prod_{i \in S_2} \prod_{k=1}^{K} f(y_i; \boldsymbol{\theta}_k)^{u_{ik}}$$

as the contribution of the data in $S_2$ to the CL function, where $\mathbf{u}_{S_2} = \{u_i : i \in S_2\}$ and

$$p\left(\mathbf{u}_{S_2}; \rho, \boldsymbol{\beta}\right) = \frac{\exp\left(\sum_{i \in S_2} \mathbf{x}_i^\top \boldsymbol{\beta} + n\left(\mathbf{u}_{S_2}\right) \rho\right)}{W_2(\boldsymbol{\beta}, \rho)},$$

with $W_2(\boldsymbol{\beta}, \rho)$ being the normalizing constant and a two-site nonhomogeneous Potts model.

We propose to estimate the parameters by maximizing the following composite log-likelihood function:

$$c\ell(\boldsymbol{\theta}, \rho, \boldsymbol{\beta}) = \sum_{S_2} \log L_{S_2}(\boldsymbol{\theta}, \rho, \boldsymbol{\beta}). \tag{12}$$

To this aim, we use an EM algorithm (the R code is included as supplementary material) that iteratively generates a sequence of parameter values by alternating an E step and an M step, until convergence. During the E step, it

computes the expected value of the complete-data composite log-likelihood with respect to the predictive distribution of the segmentation. This step reduces to the computation of the following expected value for each pair of sites:

$$\mathbb{E} \log L_{S_2}(\boldsymbol{\theta}, \rho, \boldsymbol{\beta}) = \sum_{i \in S_2} \sum_{k=1}^{K} \hat{u}_{ik} \log f(y_i; \boldsymbol{\theta}_k) + \sum_{\mathbf{u}_{S_2}} \hat{\mathbf{u}}_{S_2} \log p(\mathbf{u}_{S_2}; \rho, \boldsymbol{\beta}),$$

where the $K \times K$ predictive probabilities $\hat{\mathbf{u}}_{S_2}$ are given by

$$\hat{\mathbf{u}}_{S_2} = p\left(\mathbf{u}_{S_2} \mid \mathbf{y}_{S_2}, \hat{\rho}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}\right) = \frac{p\left(\mathbf{u}_{S_2}; \hat{\rho}, \hat{\boldsymbol{\beta}}\right) f\left(\mathbf{y}_{S_2}; \hat{\boldsymbol{\theta}}\right)}{\sum_{\mathbf{u}_{S_2}} p\left(\mathbf{u}_{S_2}; \hat{\rho}, \hat{\boldsymbol{\beta}}\right) f\left(\mathbf{y}_{S_2}; \hat{\boldsymbol{\theta}}\right)}, \tag{13}$$

whereas $\hat{\rho}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\beta}}$ are the parameter values that were available from the previous step of the algorithm. The normalizing constant of these probabilities is numerically tractable, as it involves a summation over two sites. Suitable marginalization of (13) provides the univariate probabilities $\hat{\mathbf{u}}_i = p(\mathbf{u}_i \mid \mathbf{y}_{S_2}, \hat{\rho}, \hat{\theta}, \hat{\boldsymbol{\beta}})$, with $i = 1, \ldots, n$.

During the M step, the algorithm maximizes the expected complete-data composite log-likelihood with respect to the unknown parameters. Because this function is the sum of two components that depend on different sets of parameters, the M step reduces to the separate maximization of two functions, namely,

$$Q(\boldsymbol{\theta}) = \sum_{S_2} \sum_{i \in S_2} \sum_{k=1}^{K} \hat{u}_{ik} \log f(y_i; \boldsymbol{\theta}_k) \tag{14}$$

$$Q(\rho, \boldsymbol{\beta}) = \sum_{S_2} \sum_{\mathbf{u}_{S_2}} \hat{\mathbf{u}}_{S_2} \log p\left(\mathbf{u}_{S_2}; \rho, \boldsymbol{\beta}\right). \tag{15}$$

Maximization of both $Q(\boldsymbol{\theta})$ and $Q(\rho, \boldsymbol{\beta})$ can be carried out by a standard optimization routine that allows for parameter constraints, such as that provided by the `nlminb` function in `R`.

## 4.2 | Computational details

The proposed CL has been defined by covering the study area with pairs of neighboring sites. Covering the area by a larger subset might have been an option. However, the numerical tractability of the EM algorithm dramatically decreases with the cardinality of the largest subset of the cover. A cover that includes subsets with two elements is therefore a natural strategy. When the cover includes all the subsets of two elements, Equation (12) reduces to the pairwise likelihood function (Varin, Reid, & Firth, 2011). In a spatial setting, a pairwise likelihood can be further simplified by discarding all the pairs $\{i, j\}$ that do not include neighboring sites. Ranalli et al. (2018) provided an extensive simulation study that shows that this choice provides a computationally efficient EM algorithm, without a relevant loss in statistical efficiency.

It is well known that the EM algorithm suffers from two drawbacks: It is sensitive to the choice of starting points, and it may converge to local maxima. These two aspects are strictly linked to each other. To avoid local maxima, we follow a short-runs strategy, by running the EM algorithm from 50 random initializations and by stopping the algorithm without waiting for full convergence, that is, when the relative increase in two consecutive composite log-likelihoods is less than $10^{-2}$. The best solution is taken as starting point to run the EM algorithm until full convergence, that is, when the difference in two consecutive composite log-likelihoods is less than $10^{-5}$.

Standard errors could in principle be obtained by numerically approximating the observed Godambe matrix (Godambe, 1960), which is, however, known to present numerical instability. This computation requires both the numerical approximation of variability and sensitivity matrices, and the inversion of the variability matrix (i.e., the covariance of the CL score), which is usually a matrix of large size. A feasible alternative can be found in parametric bootstrap methods, to obtain quantiles of the distribution of the estimates. In this paper, we refitted the model to $R = 500$ bootstrap samples, which were simulated from the estimated model parameters. We then computed the 2.5% and the 97.5% quantiles of the empirical distribution of each bootstrap estimate. Simulation of the circular hidden MRF is straightforward, by taking advantage of standard simulation routines available for the Potts model and the Kato–Jones distribution. Specifically, we exploited the Gibbs sampler algorithm (Feng, Tierney, & Magnotta, 2012) that is available in the R package `PottsUtils` to simulate a configuration of segmentation labels. Given a configuration of segmentation labels, a circular observation $y_i$ is drawn at each lattice site $i$, according to the appropriate Kato–Jones distribution, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_k$, where $k$ is the

segmentation label that has been generated by the Potts model at location $i$. Kato and Jones (2015) suggested a simple acceptance/rejection algorithm to simulate from their distribution, and we follow their proposal.

## 5 | SPATIAL SEGMENTATION OF FIRE OCCURRENCE TIMES

A collection of hidden MRF models has been estimated in order to obtain the spatial segmentation of the wildfire data described in Section 2, by varying the number of states $K$ from two to six. The number $K$ of latent classes was chosen by selecting the model minimizing the composite integrated classification likelihood (Ranalli & Rocci, 2016), which extends the integrated classification likelihood to the CL framework. According to Table 1, a model with three components attains the minimum composite integrated classification likelihood value.

Table 2 (top panel) includes both the estimates and the bootstrap confidence intervals of the parameters of the three class-specific Kato–Jones densities that have been obtained by estimating a hidden MRF model with $K = 3$ components. These values should be interpreted by recalling the specific support of each parameter, displayed also in Table 2. In particular, we remark that the supports of the kurtosis and the skewness parameters depend on the value taken by the concentration parameter. By rescaling the mean parameters to the scale of a circle of circumference 365 days, the values $\mu_1 = 1.335$, $\mu_2 = 3.798$, and $\mu_3 = 4.800$ are respectively associated with the dates March 18, August 8, and October 6. The three latent classes can therefore be interpreted as three different periods of the year that respectively occur in spring, summer, and autumn, in the corresponding latitudes. The data are rather concentrated within these classes, as shown by the large values attained by the concentration parameters ($\gamma_1 = 0.861$, $\gamma_2 = 0.858$, and $\gamma_3 = 0.816$, in a unit scale). The kurtosis parameter indicates that the summer latent class is more peaked than the other two classes. This is an expected result, as summer is the season associated to most wildfires ($\nu_1 = 0.699$, $\nu_2 = 0.736$, and $\nu_3 = 0.666$). More interestingly, while the Spring component is left skewed, the shape of the other two components is quite symmetric and well

**TABLE 1** Composite integrated classification likelihood values, estimated by fitting a hidden Markov random field model with components $K = 2, \ldots, 6$

| | Number of components | | | | |
| --- | --- | --- | --- | --- | --- |
| | 2 | 3 | 4 | 5 | 6 |
| composite integrated classification likelihood | 2,128,333 | **1,974,793** | 2,280,843 | 2,033,279 | 2,346,882 |

**TABLE 2** Parameter estimates of a hidden Markov random field model with $K = 3$ states and related bootstrap 95% confidence intervals (within brackets) and parameter supports (italics)

| | Latent classes | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | 1 (Spring) | | 2 (Summer) | | 3 (Autumn) | |
| Parameter | Est. | Conf. Int. | Est. | Conf. Int. | Est. | Conf. Int. |
| $\mu$ (mean) | 1.335 | (1.320,3.781) | 3.798 | (3.795,3.837) | 4.800 | (4.540,4.822) |
| | | $[0, 2\pi)$ | | $[0, 2\pi)$ | | $[0, 2\pi)$ |
| $\gamma$ (concentration) | 0.861 | (0.741,0.873) | 0.858 | (0.827,0.869) | 0.816 | (0.374,0.818) |
| | | $[0, 1)$ | | $[0, 1)$ | | $[0, 1)$ |
| $\nu$ (kurtosis) | 0.699 | (0.628,0.762) | 0.736 | (0.718,0.755) | 0.666 | (0.297,0.672) |
| | | $[0.622, 0.861]$ | | $[0.614, 0.858]$ | | $[0.516, 0.816]$ |
| $\eta$ (skewness) | 0.047 | (−0.003,0.060) | 0.000 | (−0.010,0.016) | 0.000 | (−0.018,0.105) |
| | | $[−0.120, 0.120]$ | | $[−0.122, 0.122]$ | | $[−0.150, 0.150]$ |
| | Latent classes (reference = latent class 3) | | | | | |
| Land cover | 1 (Spring) | | 2 (Summer) | | | |
| Cropland | −0.636 | (−0.777,0.080) | 0.523 | (−0.249,0.663) | | |
| Forest | −0.366 | (−0.469,0.553) | 0.956 | (0.297,1.259) | | |
| Shrubland | 0.066 | (−0.075,0.740) | 1.331 | (0.433,1.604) | | |
| Grassland | 0.390 | (0.388,1.159) | 1.474 | (0.559,2.058) | | |
| Mixed vegetation | −0.377 | (−0.453,0.336) | 0.874 | (−0.080,1.036) | | |
| Low vegetation | −0.713 | (−0.789,0.883) | 1.142 | (0.498,1.620) | | |
| $\rho$ (spatial) | 0.952 | (0.751,1.005) | | | | |

approximated by a wrapped Cauchy density (when $\eta = 0$ and $\nu = \gamma^2$, the Kato–Jones distribution reduces to the two-parameter wrapped Cauchy density).

Table 2 (bottom panel) shows the estimated influence of land cover on the marginal distribution of the latent classes, by taking the class 3 (Autumn) as reference. As expected, latent class 2 is associated with positive estimates, indicating that the proportion of fires during summer is larger than during autumn, regardless of land cover. More interestingly, the proportion of fires in spring is larger than in autumn only when fires occur in shrublands and grasslands. In all the other cases, the proportion of fires in spring is smaller than in autumn.

The estimated model segments the Iberian Peninsula according to three latent classes that are associated with three well distinct seasonal patterns of fire occurrences. Each fire event can be allocated to a given class by comparing the posterior probabilities $\hat{\mathbf{u}}_i$ and associating each event with the most likely class. By counting the events that the model allocated to each class across land cover, we obtain the estimated marginal distribution of the latent classes, say, $\hat{p}_k, k = 1, 2, 3$ (Table 3, first bold row) and the conditional distribution of latent classes within each land type $h$, say $\hat{p}_{hk}, k = 1, 2, 3, h = 1, \ldots, 6$ (the six rows in Table 3). Table 3 further includes the proportions $p_h$ of the study area that are covered by each land type $h$ (first bold column). By integrating these probabilities with the three Kato–Jones distributions $f(y; \hat{\boldsymbol{\theta}}_k), k = 1, 2, 3$, displayed by Table 2, we obtain six land-specific curves, as follows:

$$\hat{f}_h(y) = p_h \sum_{k=1}^{3} \hat{p}_{hk} f(y; \hat{\boldsymbol{\theta}}_k),$$

which indicate the estimated conditional distribution of the times of fire occurrence given the land cover, scaled by the proportion $p_h$. Figure 5 overlaps these curves on the observed land-specific distributions of fire occurrences, scaled by the proportions $p_h$. Notice that the area under each curve is simply the proportion $p_h$. These pictures clearly indicate that wildfire seasonality varies across lands of different types. They further indicate that the model adequately predicts the distribution of fire occurrence times within each land type. Obviously, goodness of fit can be improved by estimating a model with more than three latent classes, if desired. Figure 5 also includes the estimated marginal distribution of fire occurrences, obtained by summing the scaled distributions, say,

$$\hat{f}(y) = \sum_{h=1}^{6} \hat{f}_h(y),$$

and overlapped on the marginal distribution of the data. This picture indicates that the model correctly decomposes the marginal distribution of the data into land-specific components, which are associated with a mixture of Kato–Jones densities with land-specific weights.

Figure 6 displays the final segmentation of the wildfire occurrence data across the study area, obtained by associating each georeferenced event with the posterior probability of being in each class. In particular, dark colors indicate

**TABLE 3** Conditional latent class distribution of wildfire events given the land use, estimated by a three-state hidden Markov random field

| Land cover | | Latent classes | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | | 0.113 | 0.700 | 0.187 |
| Cropland | **0.256** | 0.122 | 0.584 | 0.294 |
| Forest | **0.411** | 0.110 | 0.728 | 0.162 |
| Shrubland | **0.156** | 0.126 | 0.748 | 0.126 |
| Grassland | **0.026** | 0.151 | 0.687 | 0.162 |
| Mixed vegetation | **0.049** | 0.136 | 0.684 | 0.180 |
| Low vegetation | **0.102** | 0.060 | 0.813 | 0.127 |

*Note.* Bold row: estimated marginal latent class distribution. Bold column: observed proportions of the study area covered by each type of land.
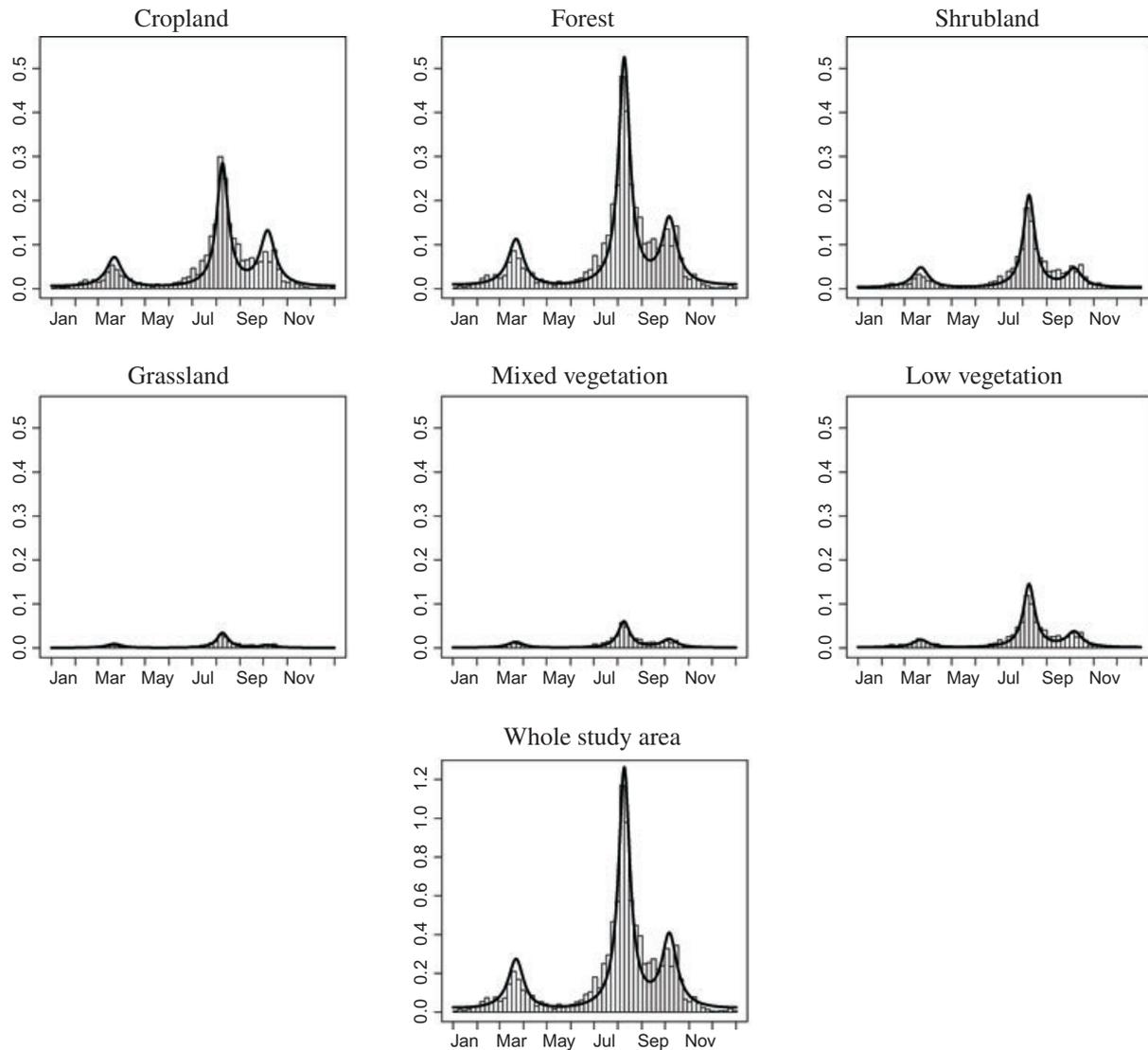
**FIGURE 5** Top: the conditional six distributions of fire occurrence times given the land cover, scaled by the proportion of the study area covered by each type of land. Continuous line: estimated scaled distribution. Histogram: observed scaled distribution. Bottom: estimated (continuous line) and observed (histogram) marginal distribution of fire occurrence times, respectively, obtained by summing the six curves and histograms above

a higher posterior probability of latent class membership. It can be clearly seen from top and bottom-left panels that the wildfire distribution across space is quite different for the three latent classes. Although we do not aim at drawing thorough conclusions about fire dynamics, this figure depicts some relevant association between fire occurrence and land use. Patterns of wildfires can be clustered within three regions: the northeast district of Bragança, affected by spring fires; the south-central region of Alentejo, which suffers from autumn fires; and the rest of the country, with a high incidence of summer fires. Actually, these summer fires are spatially concentrated in the western part of the peninsula, both in Portugal and Galicia. With a closer look at the Spanish territory, the autumn component seems to be related with regions where winter cereal crops are frequent. This is the case of the central part of Castile and León and the south-central and southeastern part of the Iberian Peninsula (south-central region of Alentejo, south of Extremadura, around Sevile and Córdoba in Andalusia, and some regions in Castile-La Mancha), which are also the predominant areas where barley, wheat, and rye are harvested (the area harvested for each crop in the Iberian Peninsula can be obtained at http://www.earthstat.org). In addition, maize and sunflower crops play an important role in these areas, because also maize (and potato) is harvested in the western part of the Iberian Peninsula that is characterized by summer fires.
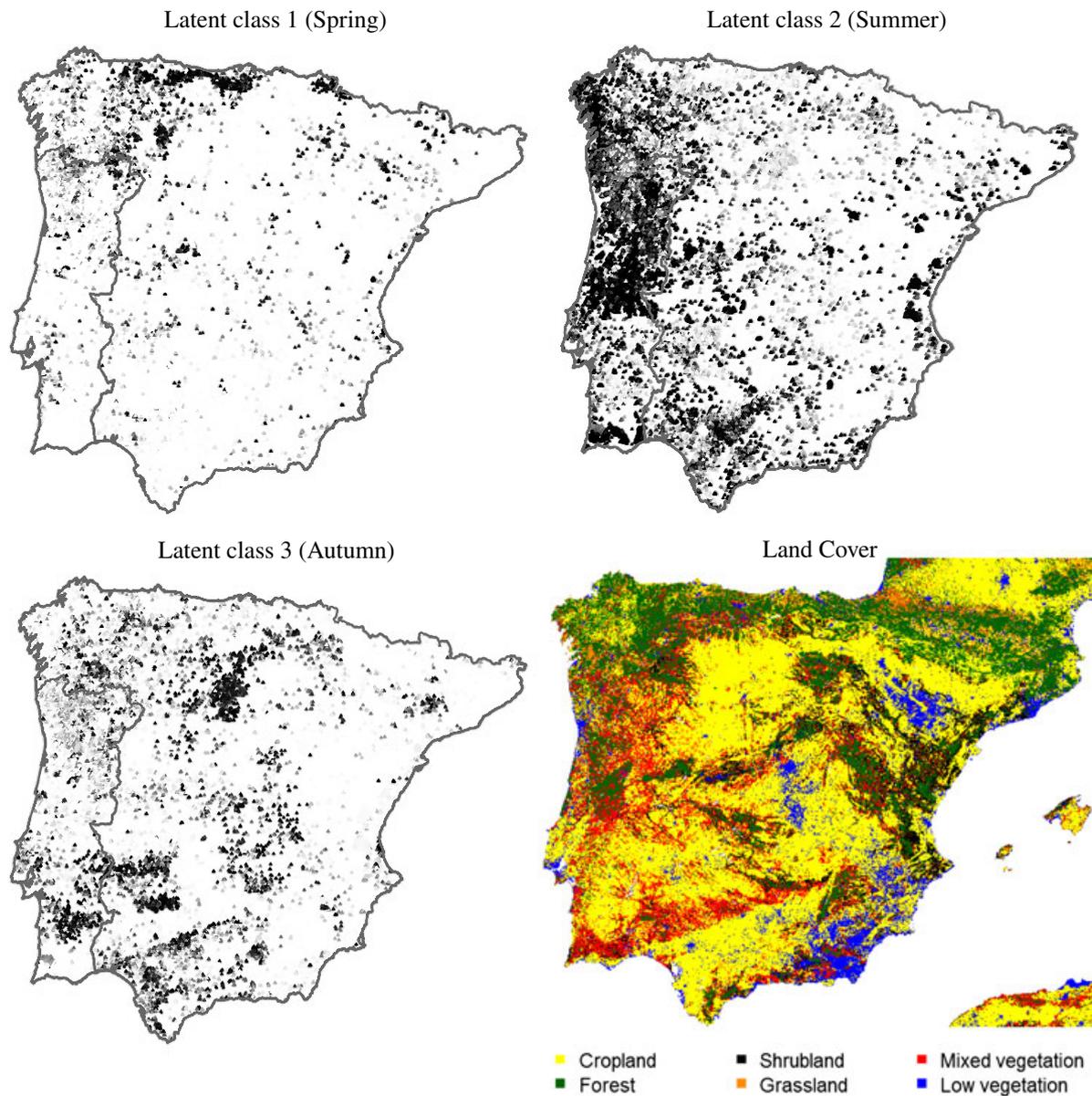
**FIGURE 6** Top and bottom left: locations of wildfires in the Iberian Peninsula in the period 2002–2012. Each point is associated to a gray level that is proportional to the posterior probability of latent class membership (darker levels indicate higher probabilities). Bottom right: most frequent land cover of the Iberian Peninsula during the study period

## 6 | CONCLUDING REMARKS

Scaling georeferenced event times to a unit circle allows us to view fire occurrences as a spatial series of circular data. Circular spatial series, however, require special methods that address the circular nature of the data in a spatial setting. Our proposal is based on a hidden MRF for circular data. It segments the study area according to latent classes that represent specific seasons of fire occurrence. Simultaneously, it decomposes the marginal distribution of the data into land-specific components, which are associated with a mixture of Kato–Jones densities with land-specific weights. By taking this approach, we were able to indicate the most likely places where fires could occur in specific periods of the year and to capture the association between fire occurrences and land use within each season of the year.

From a technical viewpoint, this approach offers a number of advantages. First, it flexibly accommodates multimodality, skewness, and kurtosis, simultaneously accounting for spatial autocorrelation. Second, it provides a parsimonious representation of the data distribution by means of a small number of latent classes that offer an intuitively appealing interpretation of fire regimes. Third, despite of the huge sample size, model estimation is computationally feasible.

The model provides a platform for a number of possible extensions. For example, it could be exploited to investigate at what extent covariates of land morphology and weather conditions might influence the seasonality of fires. Furthermore, it could be extended to a cylindrical setting, by simultaneously modeling the time of fire occurrence and the fire intensity. Although further research is needed to explore these possibilities, in its present form, the model is already capable to capture relevant aspects of the seasonality of fires and provide crucial information for policy planning.

## ACKNOWLEDGEMENTS

## REFERENCES

Alfó, M., Nieddu, L., & Vicari, D. (2008). A finite mixture model for image segmentation. *Statistics and Computing*, *18*(2), 137–150.

Ameijeiras-Alonso, J. (2017). *Assessing simplifying hypotheses in density estimation* (PhD dissertation). Universidade de Santiago de Compostela, Spain. Retrieved from https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1573365

Ameijeiras-Alonso, J., Crujeiras, R. M., & Rodríguez Casal, A. (2018). Directional statistics for wildfires. In C. Ley, & T. Verdebout (Eds.), *Applied directional statistics: Modern methods and case studies*. Boca Raton, FL: Chapman and Hall/CRC Press.

Benali, A., Mota, B., Carvalhais, N., Oom, D., Miller, L. M., Campagnolo, M. L., & Pereira, J. M. C. (2017). Bimodal fire regimes unveil a global-scale anthropogenic fingerprint. *Global Ecology and Biogeography*, *26*(7), 799–811.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, *24*(3), 179–195.

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York, NY: Springer.

Feng, D., Tierney, L., & Magnotta, V. (2012). MRI tissue classification using high-resolution Bayesian hidden Markov normal mixture models. *Journal of the American Statistical Association*, *107*(497), 102–119.

Gaetan, C., & Guyon, X. (2010). *Spatial statistics and modeling*. New York, NY: Springer.

Giglio, L., Descloitres, J., Justice, C., & Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, *87*(2–3), 273–282.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, *31*(4), 1208–1211.

Holzmann, H., Munk, A., Suster, M., & Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, *13*(3), 325–347.

Jona-Lasinio, G., Gelfand, A., & Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics*, *6*(4), 1478–1498.

Kato, S., & Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, *102*(1), 181–190.

Korontzi, S., McCarty, J., Loboda, T., Kumar, S., & Justice, C. (2006). Global distribution of agricultural fires in croplands from 3 years of Moderate Resolution Imaging Spectroradiometer (modis)data. *Global Biogeochemical Cycles*, *20*(2), GB2021.

Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von Mises distribution. *Environmental and Ecological Statistics*, *23*(1), 89–113.

Lagona, F., & Picone, M. (2013). Maximum likelihood estimation of bivariate circular hidden Markov models from incomplete data. *Journal of Statistical Computation and Simulation*, *83*(7), 1223–1237.

Lagona, F., & Picone, M. (2016). Model-based segmentation of spatial cylindrical data. *Journal of Statistical Computation and Simulation*, *86*(13), 2598–2610.

Lagona, F., Picone, M., & Maruotti, A. (2015). A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, *26*(8), 534–544.

Lagona, F., Picone, M., Maruotti, A., & Cosoli, S. (2015). A hidden Markov approach to the analysis of space-time environmental data with linear and circular components. *Stochastic Environmental Research and Risk Assessment*, *29*(2), 397–409.

Le Page, Y., Oom, D., Silva, J. M. N., Jönsson, P., & Pereira, J. M. C. (2010). Seasonality of vegetation fires as modified by human action: Observing the deviation from eco-climatic fire regimes. *Global Ecology and Biogeography*, *19*(4), 575–588.

Ley, C., & Verdebout, T. (2017). *Modern directional statistics*. Boca Raton, FL: Chapman and Hall/CRC Press.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*(1), 221–239.

Magi, B. I., Rabin, S., Shevliakova, E., & Pacala, S. (2012). Separating agricultural and non-agricultural fire seasonality at regional scales. *Biogeosciences*, *9*, 3003–3012.

Maruotti, A., Punzo, A., Mastrantonio, G., & Lagona, F. (2016). A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. *Stochastic Environmental Research and Risk Assessment*, *30*(6), 1725–1740.

Mastrantonio, G., Maruotti, A., & Jona-Lasinio, G. (2015). Bayesian hidden Markov modelling using circular-linear general projected normal distribution. *Environmetrics*, *26*(2), 145–158.

McNicholas, P. D. (2016). *Mixture model-based classification*. Boca Raton, FL: Chapman and Hall/CRC Press.

Modlin, D., Fuentes, M., & Reich, B. (2012). Circular conditional autoregressive modeling of vector fields. *Environmetrics*, *23*(1), 46–53.

Oom, D., & Pereira, J. M. C. (2013). Exploratory spatial data analysis of global MODIS active fire data. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 326–340.

Ranalli, M., Lagona, F., Picone, M., & Zambianchi, E. (2018). Segmentation of sea current fields by cylindrical hidden Markov models: A composite likelihood approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *67*(3), 575–598.

Ranalli, M., & Rocci, R. (2016). Standard and novel model selection criteria in the pairwise likelihood estimation of a mixture model for ordinal data. In A. F. Wilhelm, & H. A. Kestler (Eds.), *Analysis of large and complex data. Studies in classification, data analysis, and knowledge organization* (pp. 53–68). Cham, Switzerland: Springer.

Shirota, S., & Gelfand, A. E. (2017). Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, *11*(2), 481–503.

Strauss, D. J. (1977). Clustering on coloured lattices. *Journal of Applied Probability*, *14*(1), 135–143.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis*, *52*(7), 3493–3500.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*(1), 5–42.

Wang, F., & Gelfand, A. E. (2014). Modeling space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association*, *109*(508), 1565–1580.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.