

## FIRE SEASONALITY IDENTIFICATION WITH MULTIMODALITY TESTS

BY JOSE AMEIJERAS-ALONSO<sup>\*,1,2</sup>, AKLI BENALI<sup>†,3</sup>,  
ROSA M. CRUJEIRAS<sup>‡,1</sup>, ALBERTO RODRÍGUEZ-CASAL<sup>‡,1</sup> AND  
JOSÉ M. C. PEREIRA<sup>†,3</sup>

*KU Leuven<sup>\*</sup>, Universidade de Lisboa<sup>†</sup> and  
Universidade de Santiago de Compostela<sup>‡</sup>*

Understanding the role of vegetation fires in the Earth system is an important environmental problem. Although fire occurrence is influenced by natural factors, human activity related to land use and management has altered the temporal patterns of fire in several regions of the world. Hence, for a better insight into fires regimes it is of special interest to analyze where human activity has altered fire seasonality. For doing so, multimodality tests are a useful tool for determining the number of annual fire peaks. The periodicity of fires and their complex distributional features motivate the use of nonparametric circular statistics. The unsatisfactory performance of previous circular nonparametric proposals for testing multimodality justifies the introduction of a new approach, considering an adapted version of the excess mass statistic, jointly with a bootstrap calibration algorithm. A systematic application of the test on the Russia–Kazakhstan area is presented in order to determine how many fire peaks can be identified in this region. A False Discovery Rate correction, accounting for the spatial dependence of the data, is also required.

**1. Introduction and motivation.** Vegetation fires are caused by several factors, and their occurrence is strongly influenced by natural factors such as fuel availability, temperature, precipitation, wind, humidity and the location of lightning strikes (Westerling et al. (2003)). In general, in different areas of the world climatic conditions favor the occurrence of fires around one specific annual period. For instance, in most of the temperate regions north of the Tropic of Cancer, fires occur from May to September, when dry conditions prevail (Le Page et al. (2010)). Apart from climate, human activity also influences fire regimes, and in some cases it can significantly alter the fire seasonality. Fire is used for many purposes related

---

Received October 2018; revised May 2019.

<sup>1</sup>Supported by the Project MTM2016-76969-P from the Spanish State Research Agency (AEI) co-funded by the European Regional Development Fund (ERDF), the Competitive Reference Groups 2017–2020 (ED431C 2017/38) from the Xunta de Galicia through the ERDF.

<sup>2</sup>Supported by the FWO research project G.0826.15N (Flemish Science Foundation), GOA/12/014 project (Research Fund KU Leuven) and was partially supported by the grant BES-2014-071006 from the Spanish Ministry of Science, Innovation and Universities.

<sup>3</sup>Supported by Centro de Estudos Florestais (CEF) Strategic Project (UID/AGR/00239/2013).

*Key words and phrases.* Circular data, multimodality, multiple testing, wildfires.

to land use practices with different timings throughout the year. For example, humans use fire as a tool for hunting, pasture management, clearing fields for agriculture, eliminating crop and forest harvest residues and managing fuels for wildfire risk reduction. Analysis of fire seasonality and its relationship with the temporal patterns of fire weather conditions provides useful information concerning the extent to which fire regimes are anthropogenic (Benali et al. (2017), Le Page et al. (2010), Magi et al. (2012)).

In this paper, a new tool for analyzing how many fire peaks can be identified is presented. The issue of determining the number of fire peaks can be translated into the statistical problem of testing the number of modes, defined as local maxima of the density. When studying this random variable, fire periodicity (jointly with the possibility of having a period with fires in December and January) must be accounted for, motivating the use of circular statistics to analyze this kind of data (Benali et al. (2017), Xu and Schoenberg (2011)). From a parametric approach, Benali et al. (2017) tackled this problem using a mixture of two von Mises (circular) distributions. However, temporal patterns of fire occurrence can be very complex (Keeley et al. (2009)), presenting, for example, prominent asymmetries. In such a context, simple parametric models may not capture the data characteristics appropriately. This fact motivates the consideration of nonparametric techniques for determining the number of modes with a testing approach. For scalar (real-valued) data, different alternatives have been presented in the statistical literature; some of them are based on the idea of the *critical bandwidth*, defined by Silverman (1981) and others, using as a test statistic the *excess mass* introduced by Müller and Sawitzki (1991). In the circular case, just Fisher and Marron (2001) provided an approach for testing multimodality, using the  $U^2$  of Watson as a test statistic, but computational results show a poor calibration in practice even for “large” values of sample size (see Section 3). The proposal for solving the multimodality testing problem presented in this paper considers an adapted version of the excess mass statistic for circular data. A correct calibration is guaranteed using a bootstrap procedure where resampling is based on a nonparametric estimator (a modified kernel density estimator) of the circular density function.

The study area used herein, straddling the border between Russia and Kazakhstan, is one of the main agricultural regions of the world. Fire is widely used here (Le Page et al. (2010)), both before planting and after harvesting, resulting in a multimodal fire season pattern (Hall et al. (2016), Figure 2, Benali et al. (2017)). The method proposed in this paper will be applied to each grid cell in the study area (Figure 2). The analysis of the number of fire peaks in each cell is helpful to assess human pyrogenic influence. Previous studies (see, e.g., Korontzi et al. (2006)) have shown that in the study region, the dry season, when fire weather severity is higher, lasts from June to September. The summer fire season coincides with this period, but it seems that there is also another annual fire period that occurs earlier, in early spring (March and April), when natural conditions are not particularly favorable for wildfires. It is this temporal mismatch between an observed fire

season and the natural conditions most suitable for fire that has been considered indicative of anthropogenic vegetation burning (Benali et al. (2017), Le Page et al. (2010)).

Another issue that needs to be taken into account for the practical application of the statistical method is related with the spatial area division. When fire data is distributed in space along grid cells, the nonparametric test must be applied systematically to each cell. In this context, a *False Discovered Rate* (FDR) procedure is required in order to control the incorrect rejections of the null hypothesis, that is, the identification of unimodal fire regimes as multimodal. Note also that the temporal pattern of fires can be spatially correlated with the neighboring cells (Nichols et al. (2011)). Land cover information may be useful to identify spatially adjacent sets of cells (*patches*) where the temporal patterns of fire occurrence are expected to be similar (see, e.g., Benali et al. (2017)). Then, an adaptation of the Benjamini and Heller (2007) proposal is applied to correct the FDR as well as accounting for the spatial dependence of the data.

In summary, a nonparametric testing procedure for determining the number of modes in a circular density is presented. This procedure is designed with the goal of determining the number of fire peaks in the Russia–Kazakhstan area and treated in a lattice division in such a way that the test is applied systematically to each cell, requiring therefore an FDR correction.

The organization of the paper is the following: Section 2 details the circular excess mass approach for testing the null hypothesis that the data underlying distribution has  $k$  modes. The method is validated in Section 3, presenting a complete simulation study and comparing the new proposal with the one by Fisher and Maron (2001), regarding empirical size and power. The FDR correction, accounting for the spatial dependence of the data, and the analysis of the number of fire peaks in the study area is presented in Section 4. Some final comments and discussion are given in Section 5. Details on the models employed in the simulation study, a complete description of the calibration function used to generate the resamples in the bootstrap procedure with some theoretical background, some further simulation results showing rejection rates for different scenarios and the construction of the land cover patches cells, where a similar fire behavior is expected, are provided as Supplementary Material (SM, Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)).

**2. Statistical tools: A nonparametric test for circular multimodality.** As mentioned in the Introduction, the first goal is to provide a statistical tool for determining the number of fire peaks in a specific region. This practical problem can be formulated as a testing problem on the number of modes in a circular density.

Directional data, observations on directions, arise quite frequently in many natural sciences and in particular in wildfires modeling (several examples are provided in Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2018)). The need for circular statistics appears when the periodicity must be taken into account, and the

sample can be represented on the circumference. As mentioned before, this is the case of wildfires that commonly have a strong seasonal pattern.

Given a circular random variable  $\Theta$  with probability density function  $f$ , the goal is to test if the number of modes of  $f$  (fire peaks), namely  $j$ , is equal to a given value  $k \in \mathbb{Z}^+$  (fire weather seasons), against if it is greater than  $k$ . In general, rejecting  $H_0$  can suggest a strong human influence on fire seasonality in the region. The statistical testing problem can be formulated as assessing

$$(2.1) \quad H_0 : j = k \quad \text{vs} \quad H_a : j > k.$$

Several methods were proposed for testing  $H_0$  in the linear setting following a nonparametric approach (see Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019a)). These procedures are not suitable for our problem since the circular nature of the data is not taken into account. To the best of the authors' knowledge, only the proposal of Fisher and Marron (2001) is suitable for this purpose in the circular literature, but, as it is shown (see Section 3), the performance in practice is quite unsatisfactory. Hence, a well-calibrated test is required in the circular setting for testing  $H_0$  with a general  $k$ . Different options for constructing a test statistic have been inspired by testing methods for scalar data. For instance, an adapted version of the critical bandwidth, called critical concentration (defined below), could be employed. However, if a critical bandwidth approach is used, then at least two turning points (mode and antimode) appear in a circular density; Hall and York (2001) showed that a bootstrap test based on the critical bandwidth cannot be directly calibrated with more than one turning point. Therefore, the use of an excess mass statistic (introduced by Müller and Sawitzki (1991)) is advised. Actually, for testing the number of modes, the excess mass statistic admits a natural reformulation in the circular setting. In the linear case, the asymptotic calibration results of the excess mass statistic are provided by Cheng and Hall (1998). In what follows, the objective is to provide the specific expression of the excess mass in the circular setting, together with a way of calibrating its distribution in practice.

2.1. *The excess mass test statistic.* The proposed test statistic requires, first, the construction of the following empirical excess mass function for  $k$  modes which, given a sample  $\Theta = (\Theta_1, \dots, \Theta_n)$  from  $\Theta$ , is defined as

$$E_{n,k}(\mathbb{P}_n, \lambda) = \sup_{C_1(\lambda), \dots, C_k(\lambda)} \left\{ \sum_{m=1}^k \mathbb{P}_n(C_m(\lambda)) - \lambda \|C_m(\lambda)\| \right\},$$

where the supremum is taken over all families  $\{C_m(\lambda) : m = 1, \dots, k\}$  of disjoint connected sets (closed arcs),  $\|C_m(\lambda)\|$  denotes the set measure (the arcs length),  $\mathbb{P}_n(C_m(\lambda)) = (1/n) \sum_{i=1}^n \mathcal{I}(\Theta_i \in C_m(\lambda))$  and  $\mathcal{I}$  is the indicator function. An example of the theoretical excess mass is provided in Figure 1 (left) for illustration purposes. A way of determining the plausibility of the null hypothesis ( $f$  has  $k$  modes) is by observing if the difference  $D_{n,k+1}(\lambda) = E_{n,k+1}(\mathbb{P}_n, \lambda) - E_{n,k}(\mathbb{P}_n, \lambda)$

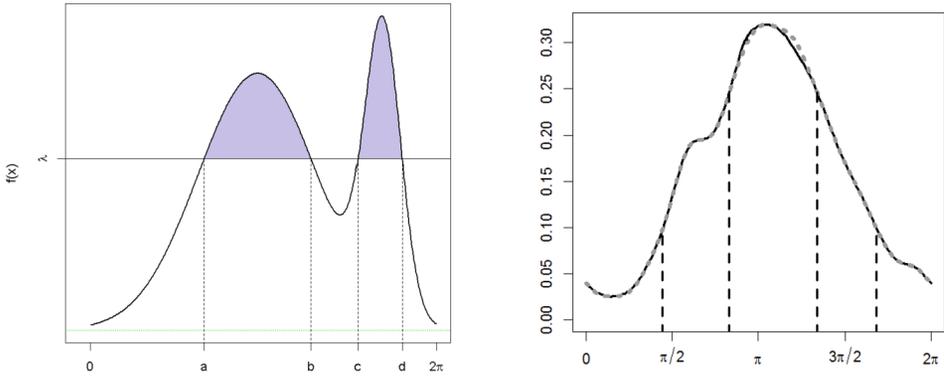


FIG. 1. *Left: theoretical excess mass function for two modes (in gray), that is, largest probability of mass exceeding the level  $\lambda$  (horizontal line) when taking two arcs. Right: kernel density estimation (2.4) with critical concentration parameter for one mode (2.5),  $\hat{f}_{v_1}$  (dotted gray line) and calibration function,  $\hat{g}$  ((2.7), solid line). The sample ( $n = 200$ ) was obtained from the model M9 (described in SM1). Dashed lines: neighborhoods where the J functions (SM2.3 in Section SM2) are defined for  $\hat{\theta}_1$  and  $\hat{\theta}_2$ .*

is “large.” Using these differences for different thresholds ( $\lambda$  values), the test statistic for (2.1) is described below and  $H_0$  is rejected when its value is large:

$$(2.2) \quad \Delta_{n,k+1} = \max_{\lambda} \{D_{n,k+1}(\lambda)\}.$$

2.2. *Results for scalar random variables.* As previously mentioned, the test statistic in (2.2) was originally proposed for scalar random variables. For the linear case, the asymptotic behavior of the excess mass for  $k = 1$  was provided by Cheng and Hall (1998) who claimed that their results could also be extended for a general value of  $k$ . From now on, a sub or superindex  $l$  is used for the linear counterparts. Under some assumptions, which include  $f_l$  being continuously differentiable and the existence of a finite number of stationary points that are the modes and anti-modes (denoted as  $x_1, \dots, x_{2k-1}$ ), jointly with some regularity conditions on  $f_l$  in a neighborhood of these points, the distribution of  $\Delta_{n,k+1}^l$  only depends on the following values:

$$(2.3) \quad d_i^l = \frac{|f_l''(x_i)|}{f_l(x_i)^3}, \quad \text{with } i = 1, \dots, 2k - 1.$$

If  $f_l$  has  $k$  modes, Cheng and Hall (1998) also indicated that the distribution of  $\Delta_{n,k+1}^l$  can be approximated by  $\Delta_{n,k+1}^{l*}$  calculated from bootstrap resamples generated from a ‘calibration distribution’ with  $k$  modes. For guaranteeing a correct asymptotic behavior, this calibration function, namely  $\hat{g}_l$ , must satisfy that its associated values of  $\hat{d}_i^l = |\hat{g}_l''(\hat{x}_i)|/\hat{g}_l(\hat{x}_i)^3$  in its modes and antimodes, denoted as  $\hat{x}_i$ , converge in probability to the value of  $d_i^l$  in (2.3), as  $n \rightarrow \infty$ , for  $i = 1, \dots, 2k - 1$

(see Section SM2, Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)). These ideas can be adapted to the circular case with the complexity of defining an adequate calibration function to generate bootstrap samples. In what follows, the construction of this calibration function is illustrated.

2.3. *The calibration function.* The construction of an adequate calibration function  $\hat{g}$  must be done in the following way: (i) preserving the structure of the data under the assumption that  $f$  has  $k$  modes and antimodes; (ii) verifying that  $\hat{d}_i = |\hat{g}''(\hat{\theta}_i)|/\hat{g}(\hat{\theta}_i)^3$  converges in probability to  $d_i = |f''(\theta_i)|/f(\theta_i)^3$ , for  $i = 1, \dots, 2k$ , as  $n \rightarrow \infty$ , where  $\theta_i$  and  $\hat{\theta}_i$  are, respectively, the modes and antimodes of  $f$  and  $\hat{g}$ ; (iii) satisfying some regularity conditions (see C.1, C.2 and C.3). Assuming that  $f$  has  $k$  modes, the required regularity conditions are:

C.1  $f$  is bounded, and it has a continuous derivative;

C.2 when  $\theta \in [0, 2\pi)$ ,  $f'(\theta) = 0$  holds only for  $f(\theta) = 0$  or in the modes and antimodes, namely,  $\theta_i$  with  $i = 1, \dots, 2k$ ;

C.3  $f''$  exists and is Hölder continuous within a neighborhood of  $\theta_i$ .

REMARK 1. Condition (i) is not strictly necessary and, following Cheng and Hall (1998), a parametric approach could be employed for obtaining the calibration density. The main issue is that while for the linear test of unimodality only an estimator of  $d_1$  is needed, for the circular case an estimator of  $d_2$  (for the antimode) is also required. So, providing the parametric test in the circular case is almost as “complicated” as the linear test for bimodality. Note also that if such a parametric distribution is found, the second-order limit properties of the test will depend on the form of the density function.

The calibration function  $\hat{g}$  is obtained as follows. First, for estimating the unknown circular density  $f$  in (i), given the random sample of angles  $\Theta$ , the kernel density estimator is employed. This estimator is defined as

$$\hat{f}_\nu(\theta) = \frac{1}{n} \sum_{i=1}^n K(\theta; \Theta_i, \nu), \quad \text{with } \theta \in [0, 2\pi),$$

where  $K(\cdot; \Theta_i, \nu)$  is a kernel function, centered in  $\Theta_i$  and concentration parameter  $\nu$  (see, e.g., Oliveira, Crujeiras and Rodríguez-Casal (2012)). The chosen kernel function is the wrapped normal density with mean direction  $\Theta_i$  and concentration parameter  $\nu \in (0, 1)$ . This specific kernel leads to the following representation for the kernel density estimator:

$$(2.4) \quad \hat{f}_\nu(\theta) = \frac{1}{2\pi n} \sum_{i=1}^n \left( 1 + 2 \sum_{p=1}^{\infty} \nu^{p^2} \cos(p(\theta - \Theta_i)) \right) \quad \text{with } \theta \in [0, 2\pi).$$

The kernel density estimator preserves the structure of the sample, depending the number of modes on the concentration parameter  $\nu$ . With this particular kernel,

the number of modes of  $\hat{f}_\nu$  is always a nondecreasing function of  $\nu$  (Huckemann et al. (2016)). Hence, with the aim of preserving the structure of the data under the  $k$ -modality hypothesis, that is, objective (i), an analog of the critical bandwidth of Silverman (1981), namely the critical concentration  $\nu_k$ , can be employed as the concentration parameter for (2.4),

$$(2.5) \quad \nu_k = \max\{\nu : \hat{f}_\nu \text{ has at most } k \text{ modes}\}.$$

A representation of the kernel density estimation, employing  $\nu_1$  as the concentration parameter, can be observed in Figure 1 (right, dotted gray line). A unimodal estimation can be observed, and a second mode will appear between  $\pi/2$  and  $\pi$  if a higher value of  $\nu$  is taken. With this concentration parameter,  $\hat{f}_{\nu_k}$  should provide a correct estimation of the density function and also of the modes and antimodes locations. Nonetheless, a reasonable estimator of  $f''(\theta_i)$  is also needed in order to ensure (ii), that is,  $\hat{d}_i$  converges in probability to  $d_i$ , as  $n \rightarrow \infty$ . In that case, for correctly estimating  $f$  and  $f''$  different concentration parameters are required.  $f''$  can be properly estimated taking the value of  $\nu$ , which minimizes the asymptotic mean integrated squared error expression of  $\hat{f}_\nu''$ , replacing  $f$  by a mixture of  $M$  von Mises (a similar procedure for estimating  $f$  was proposed by Oliveira, Crujeiras and Rodríguez-Casal (2012)). If  $\nu_{PI}$  denotes this parameter, then  $d_i$  can be estimated from the sample with

$$(2.6) \quad \hat{d}_i = \frac{|f''_{\nu_{PI}}(\hat{\theta}_i)|}{\hat{f}_{\nu_k}(\hat{\theta}_i)^3}, \quad \text{with } i = 1, \dots, 2k.$$

To construct the calibration function  $\hat{g}$  from  $\hat{f}_\nu$ , satisfying (ii) and (iii), two modifications are needed. First, it is necessary to remove the  $t$  saddle points denoted as  $\zeta_p$ , for  $p \in \{1, \dots, t\}$ ; this is done below with the  $L$  function. Second, the density estimator is modified with the function  $J$  in a neighborhood of the estimated turning points in order to guarantee that  $|\hat{g}''(\hat{\theta}_i)|/(\hat{g}(\hat{\theta}_i))^3$  is equal to the value of  $\hat{d}_i$  in (2.6), for  $i \in \{1, \dots, 2k\}$ . Then, the calibration function is obtained by modifying the kernel density estimator with the critical concentration in the stationary points, using a similar procedure as in Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019a). In particular, the employed calibration function is

$$(2.7) \quad \hat{g}(\theta; \nu_k, \nu_{PI}, \boldsymbol{\varsigma}) = \begin{cases} J(\theta; \hat{\theta}_i, \nu_k, \nu_{PI}, \varsigma_i) & \text{if } \theta \text{ is in a neighborhood of } \hat{\theta}_i, \\ L(\theta; z_{(2p-1)}, z_{(2p)}) & \text{if } \theta \text{ is in a neighborhood of } \hat{\zeta}_p, \\ \hat{f}_{\nu_k}(\theta) & \text{in other case.} \end{cases}$$

Functions  $J$  and  $L$  are applied in each  $i \in \{1, \dots, 2k\}$  and  $p \in \{1, \dots, t\}$ , respectively. The complete characterization of  $\hat{g}$  is provided in Section SM2 (Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)) and an example of its representation is given in Figure 1 (right, continuous line) where the effect of the  $J$  function can be observed.

Using the calibration function defined in (2.7), the proposal in this paper for testing the null hypothesis (2.1) is to consider a bootstrap procedure in order to calibrate the excess mass statistic defined in (2.2). Given the sample  $\Theta$ ,  $B$  resamples  $\Theta^{*b}$  ( $b = 1, \dots, B$ ) of size  $n$  are generated from  $\hat{g}(\cdot; \nu_k, \nu_{PI}, \varsigma)$ . If  $\Delta_{n,k+1}^*$  is the excess mass statistic obtained from the generated resamples, for a significance level  $\alpha$ , the null hypothesis is rejected if  $\mathbb{P}(\Delta_{n,k+1}^* \leq \Delta_{n,k+1} | \Theta) \geq 1 - \alpha$ .

**3. Simulation study.** The aim of the following simulation study is to analyze the performance of the testing method, using the bootstrap procedure proposed in Section 2. The empirical size and power of the new method are also compared with the proposal introduced by Fisher and Marron (2001) for testing multimodality for circular data. For testing  $k$ -modality, they suggest using the  $U^2$  statistic of Watson (1961) as a test statistic, that is,

$$U^2 = n \int_0^{2\pi} \left[ F_n(x) - F_0(x) - \int_0^{2\pi} (F_n(y) - F_0(y)) dF_0(y) \right]^2 dF_0(x),$$

estimating  $F_0$  (circular distribution function) employing a kernel distribution estimation with  $k$  modes. In this simulation study, the distribution function associated with  $\hat{f}_{\nu_k}$  is used to estimate  $F_0$ , and its associated distribution is used for generating the bootstrap resamples to calibrate the test statistic.

Samples of size  $n = 50$ ,  $n = 200$  and  $n = 1000$  ( $n = 100$  instead of  $n = 1000$  in power studies) were drawn from 25 different distributions, 10 of them unimodal (MU1–MU10), 10 bimodal (MB1–MB10) and five trimodal (MT1–MT5), including unimodal (reflective) symmetric models, mixtures of them and asymmetric models. These distributions models are described in Section SM1 (Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)). For each model (MU1–MU10, MB1–MB10 and MT1–MT5) and sample size, 500 sample realizations were generated. Conditionally, on each sample and for each test, 500 resamples of size  $n$  were drawn using the calibration function of each test ( $\hat{g}$  for the new proposal and  $\hat{f}_{\nu_k}$  for the  $U^2$  statistic). Results are reported for significance levels  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ .

Table 1 includes the representative models (MU1–MU6, MB1–MB6 and MT1–MT4) where the main behavior can be seen (see below). Further tables of rejection with some extra models (MU7–MU10, MB7–MB10 and MT5) and the power results of the Fisher and Marron (2001) proposal are provided in Section SM3 (Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)). Results are organized as follows: Tables 1 and 2 shows empirical sizes (1(a) and 2(a)) and power (1(c) and 2(b)) for testing  $H_0 : j = 1$ . Tables 1 and 3 show the same results, empirical sizes (1(b) and 3(a)) and power (1(d) and 3(b)), for  $H_0 : j = 2$ .

From Tables 1(a), (b), 2(a) and 3(a), the poor calibration of the Fisher and Marron (2001) proposal can be observed. Even for sample size equal to 1000 the percentage of rejections is sometime under the significance level, as in the distributions where unimodality is tested: MU1, MU5, MU6, MU7, MU8 or MU10; or the

TABLE 1

Percentage of rejections for testing  $H_0 : j = 1$  vs.  $H_a : j > 1$  ((a) and (c)) and  $H_0 : j = 2$  vs.  $H_a : j > 2$  ((b) and (d)), with 500 simulations (1.96 times their estimated standard deviation in parenthesis) and  $B = 500$  bootstrap samples. For models under the null (a): MU1–MU6 and (b): MB1–MB6; and under the alternative hypothesis (c): MB1–MB4 and (d): MT1–MT4

(a) $H_0 : j = 1$		Unimodal	$\alpha$	0.01	0.05	0.10	0.01	0.05	0.10
		MU1	$U^2$ of Watson			Excess mass			
		$n = 50$	0(0)	0.022(0.013)	0.068(0.022)	0.012(0.010)	0.054(0.020)	0.094(0.026)	
		$n = 200$	0(0)	0.012(0.010)	0.058(0.020)	0.008(0.008)	0.038(0.017)	0.092(0.025)	
		$n = 1,000$	0.004(0.006)	0.008(0.008)	0.032(0.015)	0.006(0.007)	0.040(0.017)	0.086(0.025)	
		MU2	$U^2$ of Watson			Excess mass			
		$n = 50$	0.654(0.042)	0.828(0.033)	0.886(0.028)	0(0)	0.018(0.012)	0.038(0.017)	
		$n = 200$	0.708(0.040)	0.822(0.034)	0.884(0.028)	0.002(0.004)	0.012(0.010)	0.024(0.013)	
		$n = 1,000$	0.574(0.043)	0.718(0.039)	0.770(0.037)	0(0)	0.014(0.010)	0.022(0.013)	
		MU3	$U^2$ of Watson			Excess mass			
		$n = 50$	0.370(0.042)	0.520(0.044)	0.598(0.043)	0.008(0.008)	0.044(0.018)	0.104(0.027)	
		$n = 200$	0.648(0.042)	0.776(0.037)	0.848(0.031)	0.014(0.010)	0.036(0.016)	0.076(0.026)	
		$n = 1,000$	0.498(0.044)	0.648(0.042)	0.728(0.039)	0.012(0.010)	0.040(0.017)	0.074(0.023)	
		MU4	$U^2$ of Watson			Excess mass			
		$n = 50$	0.006(0.007)	0.040(0.017)	0.064(0.021)	0.002(0.004)	0.046(0.018)	0.094(0.026)	
		$n = 200$	0.004(0.006)	0.044(0.018)	0.108(0.027)	0.020(0.012)	0.058(0.020)	0.118(0.028)	
		$n = 1,000$	0.016(0.011)	0.070(0.022)	0.132(0.030)	0.006(0.007)	0.048(0.019)	0.098(0.026)	
		MU5	$U^2$ of Watson			Excess mass			
		$n = 50$	0.004(0.006)	0.030(0.015)	0.062(0.021)	0(0)	0.024(0.013)	0.044(0.018)	
		$n = 200$	0.004(0.006)	0.032(0.015)	0.052(0.019)	0.006(0.007)	0.044(0.018)	0.100(0.026)	
		$n = 1,000$	0.002(0.004)	0.018(0.012)	0.048(0.019)	0.010(0.009)	0.048(0.019)	0.100(0.026)	
		MU6	$U^2$ of Watson			Excess mass			
		$n = 50$	0.014(0.010)	0.056(0.020)	0.106(0.027)	0(0)	0.030(0.015)	0.070(0.022)	
		$n = 200$	0.012(0.010)	0.064(0.021)	0.122(0.029)	0.012(0.010)	0.038(0.017)	0.082(0.024)	
		$n = 1,000$	0.004(0.006)	0.026(0.014)	0.064(0.021)	0.002(0.004)	0.038(0.017)	0.086(0.024)	
(b) $H_0 : j = 2$	Bimodal	MB1	$U^2$ of Watson			Excess mass			
		$n = 50$	0.006(0.007)	0.024(0.013)	0.066(0.022)	0.010(0.009)	0.034(0.016)	0.068(0.022)	
		$n = 200$	0.024(0.013)	0.064(0.021)	0.090(0.025)	0.002(0.004)	0.024(0.013)	0.058(0.020)	
		$n = 1,000$	0.038(0.017)	0.094(0.026)	0.132(0.030)	0(0)	0.038(0.016)	0.074(0.023)	
		MB2	$U^2$ of Watson			Excess mass			
		$n = 50$	0.012(0.010)	0.048(0.019)	0.096(0.026)	0.002(0.004)	0.028(0.014)	0.060(0.021)	
		$n = 200$	0.010(0.009)	0.032(0.015)	0.066(0.022)	0.006(0.007)	0.030(0.015)	0.082(0.024)	
		$n = 1,000$	0(0)	0.004(0.006)	0.022(0.013)	0.004(0.006)	0.040(0.017)	0.088(0.025)	
		MB3	$U^2$ of Watson			Excess mass			
		$n = 50$	0.006(0.007)	0.026(0.014)	0.058(0.020)	0.004(0.006)	0.042(0.018)	0.100(0.026)	
		$n = 200$	0.002(0.004)	0.020(0.012)	0.048(0.019)	0.008(0.008)	0.046(0.018)	0.104(0.027)	
		$n = 1,000$	0.014(0.010)	0.044(0.018)	0.098(0.026)	0.006(0.007)	0.056(0.020)	0.112(0.028)	
		MB4	$U^2$ of Watson			Excess mass			
		$n = 50$	0.014(0.010)	0.060(0.021)	0.100(0.026)	0.004(0.006)	0.034(0.016)	0.054(0.020)	
		$n = 200$	0.002(0.004)	0.020(0.012)	0.046(0.018)	0.004(0.006)	0.036(0.016)	0.086(0.025)	
		$n = 1,000$	0(0)	0.012(0.010)	0.030(0.015)	0.002(0.004)	0.034(0.016)	0.082(0.024)	
		MB5	$U^2$ of Watson			Excess mass			
		$n = 50$	0.030(0.015)	0.084(0.024)	0.156(0.032)	0.006(0.007)	0.024(0.013)	0.060(0.021)	
		$n = 200$	0.026(0.014)	0.068(0.022)	0.118(0.028)	0.002(0.004)	0.040(0.017)	0.078(0.024)	
		$n = 1,000$	0.002(0.004)	0.010(0.009)	0.034(0.016)	0.010(0.009)	0.048(0.019)	0.100(0.026)	
		MB6	$U^2$ of Watson			Excess mass			
		$n = 50$	0.032(0.015)	0.144(0.031)	0.214(0.036)	0.018(0.012)	0.078(0.024)	0.140(0.030)	
		$n = 200$	0.054(0.020)	0.144(0.031)	0.230(0.037)	0.016(0.011)	0.064(0.021)	0.126(0.029)	
		$n = 1,000$	0.024(0.013)	0.102(0.027)	0.192(0.035)	0.006(0.007)	0.040(0.017)	0.092(0.025)	
(c) $H_0 : j = 1$	Bimodal	$\alpha$	0.01	0.05	0.10	$\alpha$	0.01	0.05	0.10
		Excess mass (MB1)			Excess mass (MB2)				
		$n = 50$	0.534(0.044)	0.758(0.038)	0.854(0.031)		0.002(0.004)	0.040(0.017)	0.076(0.023)
		$n = 100$	0.914(0.025)	0.968(0.015)	0.984(0.011)		0.010(0.009)	0.058(0.020)	0.112(0.028)
$n = 200$	0.996(0.006)	1(0)	1(0)		0.040(0.017)	0.116(0.028)	0.238(0.037)		
Bimodal		Excess mass (MB3)			Excess mass (MB4)				
		$n = 50$	0.022(0.013)	0.072(0.023)	0.140(0.030)		0.044(0.018)	0.164(0.032)	0.284(0.040)
		$n = 100$	0.032(0.015)	0.084(0.024)	0.156(0.032)		0.208(0.036)	0.438(0.043)	0.554(0.044)
		$n = 200$	0.044(0.018)	0.102(0.027)	0.204(0.035)		0.578(0.043)	0.796(0.035)	0.870(0.029)
(d) $H_0 : j = 2$	Trimodal	$\alpha$	0.01	0.05	0.10	$\alpha$	0.01	0.05	0.10
		Excess mass (MT1)			Excess mass (MT2)				
		$n = 50$	0.062(0.021)	0.204(0.035)	0.282(0.039)		0.008(0.008)	0.054(0.020)	0.102(0.027)
		$n = 100$	0.184(0.034)	0.380(0.043)	0.506(0.044)		0.018(0.012)	0.066(0.022)	0.124(0.029)
$n = 200$	0.436(0.043)	0.692(0.040)	0.780(0.036)		0.036(0.016)	0.088(0.025)	0.178(0.034)		
Trimodal		Excess mass (MT3)			Excess mass (MT4)				
		$n = 50$	0.002(0.004)	0.058(0.020)	0.108(0.027)		0.030(0.015)	0.160(0.032)	0.244(0.038)
		$n = 100$	0.022(0.013)	0.078(0.024)	0.154(0.032)		0.154(0.032)	0.340(0.042)	0.448(0.044)
		$n = 200$	0.036(0.016)	0.126(0.029)	0.200(0.035)		0.456(0.044)	0.660(0.042)	0.746(0.038)

models where bimodality is assessed: MB2, MB4, MB5 or MB8. For other scenarios, as in models MU2 and MU3 (unimodality) and MB1 and MB6 (bimodality), the percentage of rejections is above  $\alpha$ .

For the new proposal, as shown in Tables 1(a), (b), 2(a) and 3(a), a reasonable level accuracy is obtained in general, except for model MU2. Even for small sample sizes ( $n = 50$ ), when the null hypothesis of unimodality is tested, the percentage of rejections is close to the significance level  $\alpha$ . Exceptions to this general behavior include MU2 and also the models MU5 ( $n = 50$ ), MU6 ( $n = 50$ ), MU7 ( $n = 50$ ), MU8 ( $n = 200$ ) and MU10 ( $n = 200$ ), where the percentage of rejections is slightly below the significance level. To test bimodality, when the sample size is equal to or larger than  $n = 200$ , our proposal seems to calibrate correctly, except for model MB1 where the percentage of rejections is slightly below  $\alpha$ . When the sample size is not large enough, our new proposal presents a conservative performance in the leptokurtic models, such as model MU2. In this last model, this behavior is corrected when considering a larger sample size ( $n = 2000$ ).

As mentioned before, the test size of Fisher and Marron (2001) can exceed the significance level, as shown for models MU2 and MU3, where the percentage of rejections under the null for  $\alpha = 0.05$  (when  $n = 1000$ ) is greater than 0.6. For that reason, their proposal should not be employed in practice. For the reader interested in analyzing the behavior of the Fisher and Marron (2001) method under the alternative, Tables 2(b) and 3(b), in Section SM3 (Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)), include the associated power results. Power results in Tables 1(c), (d), 2(b) and 3(b) show that the new proposal, which seems to be the only one which is well calibrated, appears to have also good power in terms that the percentage of rejections increases with the sample size. The method rejects the null hypothesis on the bimodal model MB1 and the trimodal models MT1 and MT5. This new proposal also detects the small blips, for example, on models MB4, MB7 (bimodal) and MT4 (trimodal). However, in other distributions it has some difficulties when the small blips represent a low percentage of the data and the sample size is small (e.g., model MB2). In the difficult cases with almost overlapping peaks, when the sample size is small, the new method presents some limitations to detect the rejection of unimodality (with  $n = 50$  in MB3) and the rejection of bimodality (with  $n = 100$  in MT2 and with  $n = 50$  in MT3). However, as expected, the percentage of rejections increases with  $n$ .

**4. Data analysis: Detection of fire season multimodality.** As explained in Section 1, occurrence of a larger number of fire activity peaks than of fire weather severity peaks has been considered an indicator of anthropogenic vegetation burning. Although a general  $k$  is allowed in the testing problem, reflecting the possible occurrence of more than one peak of weather-driven fires, it should be noted that only one fire season under favorable meteorological conditions is expected in the Russia–Kazakhstan border, as shown in the Central Eurasia panel of Figure 3 in Benali et al. (2017). Thus, using the data described in Section 4.1, our goal is to

assess if there are one or more fire activity peaks in the study area. As shown in Section 3, the simulation study supports, in the finite-sample case, that the proposal introduced in Section 2 presents a correct behavior in terms of calibration and power. Thus, if just one cell is considered, this problem can be tackled employing the new procedure for testing  $H_0 : j = 1$ . However, since the goal is to analyze the number of fire activity peaks in the entire Russia–Kazakhstan border region and this area is divided in a cell grid, the proposed procedure can be applied systematically in each cell. As mentioned, an FDR procedure is required to control incorrect rejections of the null hypothesis. To perform such correction, the spatial correlation between the test p-values computed at different cells must be considered. These two issues are solved in Section 4.2, and the results obtained are provided in Section 4.3.

4.1. *Fire data.* The dataset analysed was obtained from the *MODerate resolution Imaging Spectroradiometer* (MODIS), launched by the *National Aeronautics and Space Administration* (NASA) on board the Terra (*EOS AM*) and the Aqua (*EOS PM*) satellites at a  $1 \text{ km}^2$  resolution at nadir. The MODIS algorithm (see Giglio et al. (2003) for further details) identifies the location of fires burning at the time of satellite overpass, based on the contrasting responses of the middle-infrared and longwave infrared bands in areas containing hot targets. Cloud and water pixels are previously excluded from analysis using multiple numerical thresholds on visible and near-infrared reflectance and thermal infrared temperature values. The size of the smallest flaming fire having at least a 50% chance of being detected by the MODIS algorithm, under both ideal daytime and nighttime conditions, is approximately  $100 \text{ m}^2$ . The combination of Terra and Aqua overpasses provide, on average, four daily overpasses depending on the location in the globe. Integrating the information from all available MODIS overpasses, from 10 July 2002 to 9 July 2012, the day of the year and the location of each active fire (thermal anomaly) was recorded. Note that a single wildfire may create more than one active fire in different days. The main issue is that distinguishing which different active fires correspond to the same wildfire is not an easy task. Some models were simulated, trying to replicate the real-data scenario (data may be repeated, but not so frequently, with the corresponding day's delay). In general, under the hypothesis of unimodality similar results to those reported in this paper are obtained.

Since MODIS data are recorded in a discretized time scale, in order to apply the testing procedure, it is necessary to recover the continuous underlying structure. For that purpose, denote by  $(X_1, \dots, X_n)$  the days of the year when the  $n$  active fires were detected with  $X_i \in \{1, \dots, 366\}$ . The dataset used for the study of the number of fire peaks is the following:

$$\Theta_i = 2\pi(X_i + \mathcal{E}_i)/366; \quad \text{with } i = 1, \dots, n,$$

$\mathcal{E}_i$  being generated from the uniform distribution  $U(-1, 0)$ . This means that it is assumed that fires occurred at any time of the day. Provided that there are not

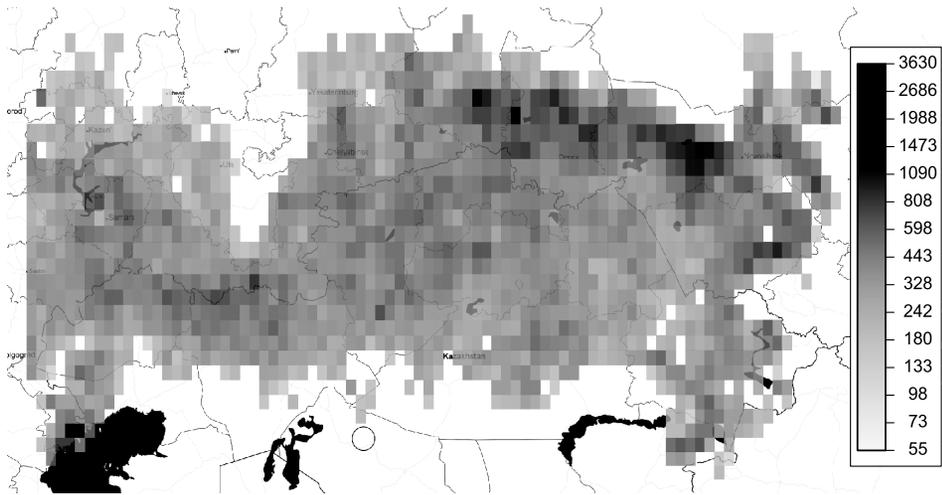


FIG. 2. Number of fires detected by the MODIS in the different cells of the studied region from from 10 July 2002 to 9 July 2012 (in translucent gray scale, as reflected in the right legend). In white, cells with low fire incidence or outside the study area. In black, surface water.

repeated observations, as this can considerably alter the test statistic, other ways of modifying the data can be considered, but, in general, this perturbation does not show relevant impacts in the results.

Once this modification is done, the area analyzed is divided into grid cells of size  $0.5^\circ$  (areas of approximately 55.5 km latitude and 32 km longitude at  $55^\circ$  North). Then, from the resulting cells those with low fire incidence, that is, cells having fewer than 10 active fires in more than seven out of 10 years, were excluded from the study. This leaves 1500 grid cells in the area, each one having between 55 and 3630 fires. Figure 2 shows a map including the study area and a summary of the total fire counts in the study period.

**4.2. Spatial false discovered rate.** The proposed testing procedure (see Section 2) will be applied systematically over these cells. Thus, it is clear that an FDR correction must be applied. Also, the subdivision of the study area into grid cells is not necessarily designed for producing “independent” areas, in the sense that occurrence of fires in a cell may not be independent of occurrence in neighboring cells. So, the spatial dependence must be taken into account in the FDR correction procedure. This was done following the ideas by [Benjamini and Heller \(2007\)](#), Procedure 3, with some modifications. Their method allows controlling the FDR accounting for the spatial dependence of the data and employing prior information about the aggregation of different locations (where the size and shape of the different groups do not need to be equal). This is the case in our study, as it is expected that the temporal patterns of fires will be similar in regions that are

close and with the same type of land use (Benali et al. (2017)). Once this aggregation is done, Benjamini and Heller (2007) propose testing first on those large units (patches) while allowing a single p-value for each patch. Then, if the null hypothesis is rejected in the patch, it controls the dependence of the aggregate and the cells' p-values in order to correct them in the multiple testing problem and to detect properly the rejected locations.

The method employed here can be outlined as follows: in an initial step (Step 1), the testing procedure (introduced in Section 2) is applied locally, to each cell, obtaining the corresponding p-value. Second (Step 2), *land cover patches* are defined in order to create groups of cells that are spatially adjacent and have the same land cover type. Finally (Step 3), a hierarchical testing procedure (similar to that one of Benjamini and Heller (2007)) is applied. This final step consists of, first, deciding in which of the previous patches the null hypothesis is rejected (patch testing; Step 3a) and, second, within each rejected patch, determining in which of its cells  $H_0$  is rejected (trimming procedure; Step 3b). Further details in the specific problem and solution are provided below.

*Step 1. Local application of the test.* In each of the 1500 grid cells, the method proposed in Section 2 was applied (with  $B = 5000$  bootstrap replicates) to obtain the corresponding p-values when it is tested if there is one or more fire activity peaks.

*Step 2. Land cover patches.* The land cover patches were created to identify areas expected to display “similar” fire season modality patterns. From the  $0.5^\circ$ -cells, the patches were constructed considering the contiguous cells and using the information of the land cover data provided by the European Space Agency Climate Change Initiative project (*Land Cover version 1.6.1*, data from 2008 to 2012, available at <http://www.esa-landcover-cci.org>). Land cover is defined as 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. Patch construction is detailed in Section SM4 (Ameijeiras-Alonso, Crujeiras and Rodríguez-Casal (2019b)), where the specific land cover for each cell is also given (see Figure 6, left). The 80 different patches obtained in the study area are represented in Figure 3.

*Step 3. Hierarchical testing procedure.* Some notation is required for this part. Let  $j = 1, \dots, J$  be the different patches created in Step 2,  $l = 1, \dots, L_j$  be the cells within the patch  $j$  and  $\tilde{p}_{lj}$  the p-value obtained in Step 1 for the cell  $l$  within the patch  $j$ . Then,  $z_{lj} = \Phi^{-1}(1 - \tilde{p}_{lj})$  is the corresponding z-score for the cell  $l$  in patch  $j$ , where  $\Phi$  is the cumulative distribution of a standard normal distribution. Note that since a bootstrap procedure is used to compute the p-values  $\tilde{p}_{lj}$ , they can be equal to zero or one and, in that case, the z-score is nonfinite. In order to correct that, if  $\tilde{p}_{lj}$  is equal to zero, following the ideas of Jeffreys (1946) prior, if  $B$  is the number of bootstrap resamples, the p-value is replaced by a random value from the distribution  $\text{Beta}(1/2, B + 1/2)$  and, if it is equal to one, a random value from  $\text{Beta}(B + 1/2, 1/2)$  is taken. Once the z-scores for each cell in the different patches

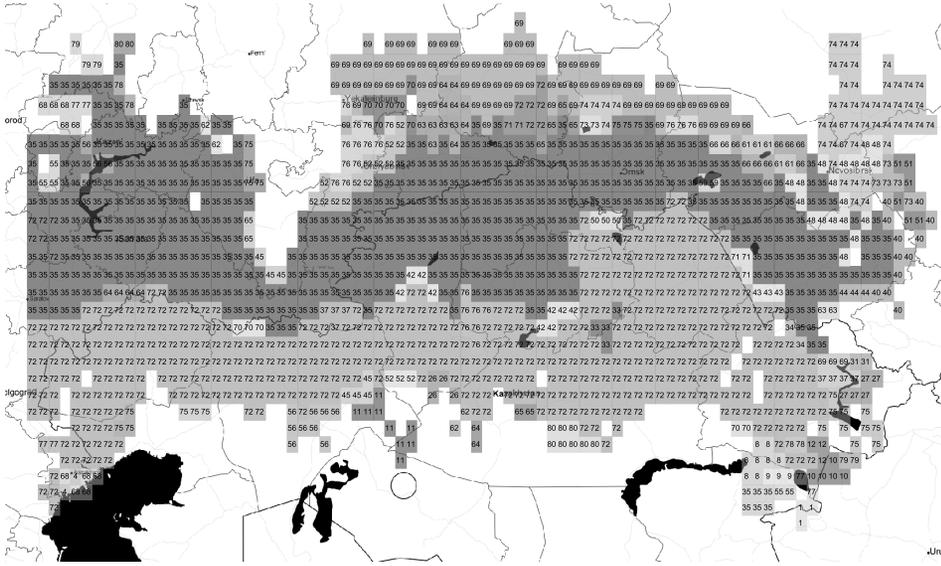


FIG. 3. In a gray (translucent) scale, each different shade of gray represents a different land cover patch (distinguished by a superimposed number). In white, cells with low fire incidence or outside the study area. In black, surface water.

are calculated, a hierarchical method is used. The testing procedure is divided into two steps: first it tests, at significance level  $\alpha_c$ , in which land cover patches the null hypothesis can be rejected (Step 3a) and, second, tests  $H_0$ , at level  $\alpha_r$ , for the cells within the rejected patches (Step 3b).

*Step 3a. Patch testing.* In this stage, land cover patches where the null hypothesis is rejected are identified. This step consists in computing a global p-value for each patch and then, since each patch has a different number of cells, the weighted FDR procedure of Benjamini and Hochberg (1997), at level  $\alpha_c$ , is applied in order to correct for multiple testing. The global p-value of the patch  $j$  is calculated as  $\check{p}_j = \Phi(\bar{Z}_j/\hat{\sigma}_{\bar{Z}_j})$ , that is, the right tail probability of the standard normal distribution calculated in the standardized z-score average of the cells in a land cover patch. More precisely:

1. In each patch  $j$ , with  $j \in \{1, \dots, J\}$  the z-score average is calculated as  $\bar{Z}_j = (1/L_j) \sum_{l=1}^{L_j} z_{lj}$ , where  $z_{lj}$  are the z-scores defined above.

2. The standard error is computed:  $\hat{\sigma}_{\bar{Z}_j} = (\hat{\sigma}_j/L_j) \sqrt{L_j + 2 \sum_{l=1}^{L_j} \sum_{m=1}^{L_j} \hat{\rho}_{l,m}^j}$ , where  $\hat{\rho}_{l,m}^j = 1 - \hat{\gamma}(s_{lj} - s_{mj})/\hat{\sigma}_j$  is the estimated correlation between cells  $l$  and  $m$  within the patch  $j$ . The value  $\hat{\gamma}(s_{lj} - s_{mj})$  is an estimation of the semivariogram evaluated at the distance (of the centroids in the map) between cells  $l$  and  $m$  and  $\hat{\sigma}_j$  the estimated variance of the cells in patch  $j$ . Differently from Benjamini and Heller (2007), the semivariogram estimator is obtained, by (weighted) least

squares on an exponential family, in order to ensure that such an estimator is indeed a valid semivariogram (something that may be not satisfied by nonparametric estimators) and to control the parameters of interest. Specifically, in an exponential family two parameters drive the behavior of the spatial covariance, the point variance and the range. Hence, the estimated variance is obtained from this parametric procedure. It should be noted that least squares procedures for variogram estimation require the use of a nonparametric pilot estimator. In this case, the robust version of the empirical variogram was used (see [Cressie \(1993\)](#), Ch. 2).

3. The weighted FDR procedure, at level  $\alpha_c$ , is applied on the p-values  $\check{p}_j$ , being the weight proportional to the patches size. Given the ordered p-values  $\check{p}_{(1)} \leq \dots \leq \check{p}_{(J)}$ , unimodality is rejected in the  $k$  patches with the smallest p-values, being  $k = \max\{v : \check{p}_{(v)} \leq (\sum_{j=1}^v L_{(j)} / \sum_{j=1}^J L_{(j)})\alpha_c\}$  and  $L_{(j)}$  the number of cells in the land cover patch associated with  $\check{p}_{(j)}$ .

*Step 3b. Trimming procedure.* Once a decision about which patches are candidates for rejecting the null hypothesis is made (and hence exhibiting a multimodal fire pattern), specific cells where this rejection holds are identified. It should be noted that the cell test statistic is correlated with the test statistic at the patch level. This means that an FDR correction cannot be directly applied over all the cells belonging to the same patch. A correction is proposed by [Benjamini and Hochberg \(1997\)](#). First, the conditional p-value of a cell, within a patch that was rejected, is calculated,  $\hat{p}_{lj}$ . Then, over these p-values the two-stage procedure, introduced by [Benjamini, Krieger and Yekutieli \(2006\)](#), at level  $\alpha_r$  is applied to enhance the power. This last method, in its first stage, consists in estimating the sum of weights of null cells, using for that purpose the classical FDR procedure at level  $\alpha_r$ . In a second stage, this quantity is used to determine the number of rejected cells within the patch. To be more precise and to summarize Step 3b, the following steps are detailed:

4. The conditional p-value of each cell  $l$  is calculated within the patch that was rejected  $j$ , denoted as  $\hat{p}_{lj}$ . These p-values are computed as follows:

$$\begin{aligned} \hat{p}_{lj} = & \int_{z_{lj}}^{\infty} \left( \frac{\hat{J}_0}{J} \tilde{\Phi} \left( \frac{\tilde{\Phi}^{-1}(u_1) - \hat{\rho}_{lj}u}{\sqrt{1 - \hat{\rho}_{lj}^2}} \right) \right. \\ & \left. + \left( 1 - \frac{\hat{J}_0}{J} \right) \tilde{\Phi} \left( \frac{\tilde{\Phi}^{-1}(u_1) - \hat{\rho}_{lj}u - \hat{\mu}_j}{\sqrt{1 - \hat{\rho}_{lj}^2}} \right) \right) \phi(u) du \\ & \times \left( \frac{\hat{J}_0}{J} u_1 + \left( 1 - \frac{\hat{J}_0}{J} \right) \tilde{\Phi}(\tilde{\Phi}^{-1}(u_1) - \hat{\mu}_j) \right)^{-1}, \end{aligned}$$

being  $\phi$  a standard normal density and noting that

(a)  $u_1 = (\sum_{j=1}^k L_{(j)} / \sum_{j=1}^J L_{(j)})\alpha_c$  is the cutoff point of the largest p-value rejected in Step 3a.

- (b)  $\hat{J}_0 = (J - \mathfrak{k}) / (1 - \alpha_c)$  is the estimated sum of weights of null patches.
- (c)  $\hat{\mu}_j = ((\sum_{j=1}^J \sum_{l=1}^{L_j} z_{lj}) / (\sum_{j=1}^J L_j)) / \hat{\sigma}_{\bar{z}_j}$  is the estimation of the standardized expectation of the patch test statistic under the alternative.
- (d)  $\hat{\rho}_{lj} = (1 + \sum_{m=1, m \neq l}^{L_j} \hat{\rho}_{l,m}^j) \hat{\sigma}_j / (L_j \hat{\sigma}_{\bar{z}_j})$  is the estimated correlation between the z-score in a given cell and the average z-score of the patch.
5. Given these  $L_j$  p-values in the patch  $j$ , a two-stage procedure is applied at level  $\alpha_r$ :
- (a) The classic FDR procedure is applied at level  $\alpha'_r = \alpha_r / (1 + \alpha_r)$ . Given the ordered p-values  $\hat{p}_{(1)j} \leq \dots \leq \hat{p}_{(L_j)j}$ , let  $\mathfrak{k}_{1j} = \max\{v : \hat{p}_{(v)j} \leq (v/L_j)\alpha'_r\}$ .
- (b) The classic FDR procedure is applied again at level  $\alpha'_r$ , being in this case the sum of weights of null cells,  $\hat{J}_{0j} = L_j - \mathfrak{k}_{1j}$ . Reject the unimodality in the  $\mathfrak{k}_{2j}$  cells with the smallest p-values, being  $\mathfrak{k}_{2j} = \max\{v : \hat{p}_{(v)j} \leq (v/\hat{J}_{0j})\alpha'_r\}$ .

4.3. *Results.* In what follows, the application of the new testing proposal, jointly with the FDR correction, to the active fire dataset is presented. As a first step, the p-values, applying the new procedure provided in Section 2 (with  $B = 5000$  bootstrap replicates), were computed in all the cells of the study area. In a second step, the different land cover patches were created using the land cover database and the spatial distribution of the cells. Finally, the hierarchical testing procedure was applied. First, to determine in which of the previously created patches the null hypothesis is rejected at significance level  $\alpha_c = 0.01$ . Second, within the rejected patches, it was determined which cells can be definitely rejected at the trimming significance level  $\alpha_r = 0.01$ .  $H_0$  was rejected in all the patches. The rejected cells are shown in light gray (marked with an X) in Figure 4.

Analyzing the fire modality map (Figure 4), results suggest that a multimodal pattern prevails in most of the study area. A large area where unimodality is not rejected stands out in the eastern part, mostly dominated by grassland cover. These findings are in agreement with previous results that have associated a multimodal pattern with the occurrence of croplands (Benali et al. (2017), Korontzi et al. (2006), Le Page et al. (2010), Magi et al. (2012)) and steppe covers (Loboda et al. (2012)). In croplands, the multimodal fire seasonality is a consequence of April and May burning in preparation for spring planting of grain crops, and postharvest crop residue burning, typically in August–September. In steppes, burning occurs in the summer and fall, probability due to a combination of human and climate factors (Loboda et al. (2012)). Recognition that multimodal fire regimes are more widespread than previously thought, and typically associated with anthropogenic burning, is important for various reasons. It will contribute to improve the parametrization of dynamic global vegetation models used to predict environmental impacts of changes in land use and climate, to refine estimates of the seasonality of greenhouse gas emissions from vegetation fires and to support fire management activities that aim at reducing the exposure of human populations and the losses of valuable resources and assets.

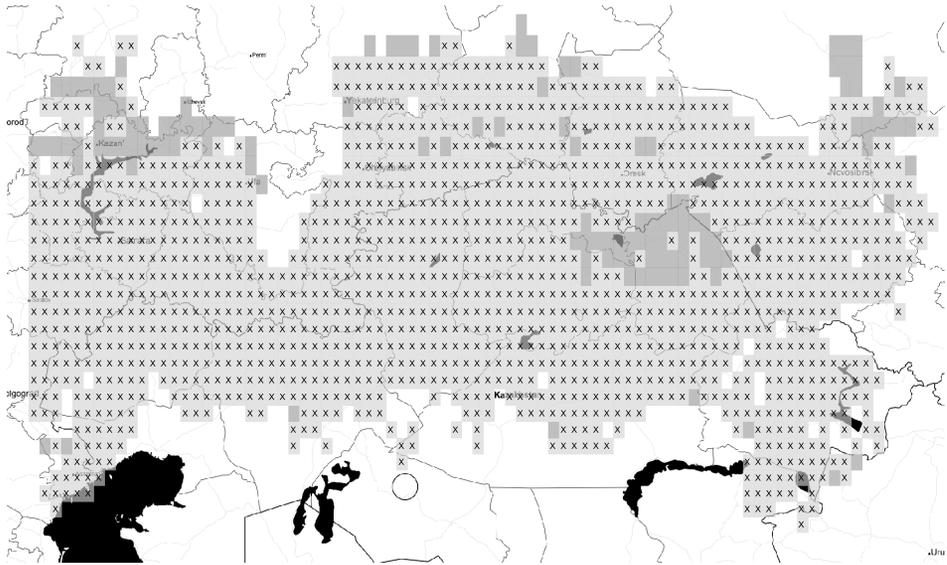


FIG. 4. Results after applying the procedure described in Section 4.2, with  $\alpha_c = 0.01$  and  $\alpha_r = 0.01$ , in the study area. In light gray (marked with an X), cells where  $H_0$  is rejected (multimodal). In dark gray, cells where there is no evidence to reject  $H_0$  (unimodal). In white, cells with low fire incidence or outside the study area. In black, surface water.

**5. Conclusions.** A new and effective nonparametric method for testing circular multimodality is presented with the objective of assessing the number of fire activity peaks and their mismatch with fire weather seasonality. Uncoupling of temporal patterns of fire activity from favorable climatic conditions, and coupling with land use dynamics, in the form of a substantial time lag between the peaks of the fire season and the fire weather season (Le Page et al. (2010)), or as the occurrence of a bimodal fire season in climates with a single dry season (Benali et al. (2017)) is considered indicative of an anthropogenic fire regime.

For a better understanding of vegetation fires in the Earth system, future research may include the use of multimodality test as a preliminary tool in different regions of the world for exploring when the fire season peaks occur and their associated mass. This would allow reviewing different studies in the environmental science literature with nonparametric techniques. For instance, one could determine when the principal fire season peaks are produced in each  $0.5^\circ$  cell (Le Page et al. (2010)), the delay of the anthropogenic fire season peaks with respect to the peak of dry season (Magi et al. (2012)) or the mass associated to each peak for better understanding the importance of the different human pyrogenic activities (Korontzi et al. (2006)).

Related with the proposed multimodality test and due to the growing interest in the last few years in more flexible models in circular data (see Ley and Verdebout (2017), Ch. 2), the new proposal can be used as a preliminary tool for determining

if a multimodal model is needed. Also, this test could be used for determining the minimum number of components in a mixture of parametric unimodal distributions when the objective is modeling fire data.

When the FDR, accounting for the spatial dependence of the data, needs to be considered, the proposed method provides a useful algorithm for any context where prior information about the neighboring locations is known.

## SUPPLEMENTARY MATERIAL

**Supplementary material of fire seasonality identification with multimodality test** (DOI: [10.1214/19-AOAS1273SUPP](https://doi.org/10.1214/19-AOAS1273SUPP); .pdf). This Supplementary Material provides details on the models employed in the simulation study; a complete description of the calibration function used to generate the resamples in the bootstrap procedure with some theoretical background; some further simulation results showing rejection rates for different scenarios; and the construction of the land cover patches cells where a similar fire behavior is expected.

## REFERENCES

- AMEIJEIRAS-ALONSO, J., CRUJEIRAS, R. M. and RODRÍGUEZ-CASAL, A. (2018). Applied directional statistics: Modern methods and case studies. In *Chap. Directional Statistics for Wildfires* (C. Ley and T. Verdebout, eds.) Chapman and Hall/CRC Press, Boca Raton, FL.
- AMEIJEIRAS-ALONSO, J., CRUJEIRAS, R. M. and RODRÍGUEZ-CASAL, A. (2019a). Mode testing, critical bandwidth and excess mass. *Test*. To appear. DOI:<https://www.doi.org/10.1007/s11749-018-0611-5>.
- AMEIJEIRAS-ALONSO, J., CRUJEIRAS, R. M. and RODRÍGUEZ-CASAL, A. (2019b). Supplement to “Fire seasonality identification with multimodality tests.” DOI:[10.1214/19-AOAS1273SUPP](https://doi.org/10.1214/19-AOAS1273SUPP).
- BENALI, A., MOTA, B., CARVALHAIS, N., OOM, D., MILLER, L. M., CAMPAGNOLO, M. L. and PEREIRA, J. M. C. (2017). Bimodal fire regimes unveil a global scale anthropogenic fingerprint. *Glob. Ecol. Biogeogr.* **26** 799–811.
- BENJAMINI, Y. and HELLER, R. (2007). False discovery rates for spatial signals. *J. Amer. Statist. Assoc.* **102** 1272–1281. MR2412549
- BENJAMINI, Y. and HOCHBERG, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.* **24** 407–418. MR1481424
- BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438
- CHENG, M.-Y. and HALL, P. (1998). Calibrating the excess mass and dip tests of modality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 579–589. MR1625938
- CRESSIE, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1239641
- FISHER, N. I. and MARRON, J. S. (2001). Mode testing via the excess mass estimate. *Biometrika* **88** 499–517. MR1844848
- GIGLIO, L., DESCLOITRES, J., JUSTICE, C. O. and KAUFMAN, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment* **87** 273–282.
- HALL, P. and YORK, M. (2001). On the calibration of Silverman’s test for multimodality. *Statist. Sinica* **11** 515–536. MR1844538
- HALL, J. V., LOBODA, T. V., GIGLIO, L. and MCCARTY, G. W. (2016). A MODIS-based burned area assessment for Russian croplands: Mapping requirements and challenges. *Remote Sensing of Environment* **184** 506–521.

- HUCKEMANN, S., KIM, K.-R., MUNK, A., REHFELDT, F., SOMMERFELD, M., WEICKERT, J. and WOLLNIK, C. (2016). The circular SiZer, inferred persistence of shape parameters and application to early stem cell differentiation. *Bernoulli* **22** 2113–2142. [MR3498025](#)
- JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **186** 453–461. [MR0017504](#)
- KEELEY, J. E., SAFFORD, H., FOTHERINGHAM, C. J., FRANKLIN, J. and MORITZ, M. (2009). The 2007 southern California wildfires: Lessons in complexity. *Journal of Forestry* **107** 287–296.
- KORONTZI, S., MCCARTY, J., LOBODA, T., KUMAR, S. and JUSTICE, C. O. (2006). Global distribution of agricultural fires in croplands from 3 years of Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Global Biogeochemical Cycles* **20** 202101–202115.
- LE PAGE, Y., OOM, D., SILVA, J., JÖNSSON, P. and PEREIRA, J. M. C. (2010). Seasonality of vegetation fires as modified by human action: Observing the deviation from eco-climatic fire regimes. *Glob. Ecol. Biogeogr.* **19** 575–588.
- LEY, C. and VERDEBOUT, T. (2017). *Modern Directional Statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. [MR3752655](#)
- LOBODA, T. V., GIGLIO, L., BOSCHETTI, L. and JUSTICE, C. O. (2012). Regional fire monitoring and characterization using global NASA MODIS fire products in dry lands of Central Asia. *Front. Earth Sci.* **6** 196–205.
- MAGI, B. I., RABIN, S., SHEVLIAKOVA, E. and PACALA, S. (2012). Separating agricultural and non-agricultural fire seasonality at regional scales. *Biogeosciences* **9** 3003–3012.
- MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. [MR1147099](#)
- NICHOLS, K., SCHOENBERG, F. P., KEELEY, J. E., BRAY, A. and DIEZ, D. (2011). The application of prototype point processes for the summary and description of California wildfires. *J. Time Series Anal.* **32** 420–429. [MR2857337](#)
- OLIVEIRA, M., CRUJEIRAS, R. M. and RODRÍGUEZ-CASAL, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.* **56** 3898–3908. [MR2957840](#)
- SILVERMAN, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* **43** 97–99. [MR0610384](#)
- WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114. [MR0131930](#)
- WESTERLING, A. L., GERSHUNOV, A., BROWN, T. J., CAYAN, D. R. and DETTINGER, M. D. (2003). Climate and wildfire in the western United States. *Bull. Am. Meteorol. Soc.* **84** 595–604.
- XU, H. and SCHOENBERG, F. P. (2011). Point process modeling of wildfire hazard in Los Angeles County, California. *Ann. Appl. Stat.* **5** 684–704. [MR2840171](#)

J. AMEJEIRAS-ALONSO  
 STATISTICS SECTION  
 DEPARTMENT OF MATHEMATICS  
 KU LEUVEN  
 CELESTIJNENLAAN 200B  
 LEUVEN 3001  
 BELGIUM  
 E-MAIL: [jose.amejeirasalonso@kuleuven.be](mailto:jose.amejeirasalonso@kuleuven.be)

A. BENALI  
 J. M. C. PEREIRA  
 CENTRO DE ESTUDOS FLORESTAIS  
 INSTITUTO SUPERIOR DE AGRONOMIA  
 UNIVERSIDADE DE LISBOA  
 TAPADA DA AJUDA  
 LISBOA 1349-017  
 PORTUGAL  
 E-MAIL: [aklibenali@gmail.com](mailto:aklibenali@gmail.com)  
[jmcpereira@gmail.com](mailto:jmcpereira@gmail.com)

R. M. CRUJEIRAS  
A. RODRÍGUEZ-CASAL  
DEPARTMENT OF STATISTICS,  
MATHEMATICAL ANALYSIS AND OPTIMIZATION  
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA  
LOPE GÓMEZ DE MARZOA, S/N  
SANTIAGO DE COMPOSTELA, A CORUÑA 15782  
SPAIN  
E-MAIL: [rosa.crujeiras@usc.es](mailto:rosa.crujeiras@usc.es)  
[alberto.rodriquez.casal@usc.es](mailto:alberto.rodriquez.casal@usc.es)