

# A test for directional-linear independence, with applications to wildfire orientation and size

Eduardo García-Portugués · Ana M. G. Barros ·  
Rosa M. Crujeiras · Wenceslao González-Manteiga ·  
José Pereira

© Springer-Verlag Berlin Heidelberg 2013

**Abstract** A nonparametric test for assessing the independence between a directional random variable (circular or spherical, as particular cases) and a linear one is proposed in this paper. The statistic is based on the squared distance between nonparametric kernel density estimates and its calibration is done by a permutation approach. The size and power characteristics of various variants of the test are investigated and compared with those for classical correlation-based tests of independence in an extensive simulation study. Finally, the best-performing variant of the new test is applied in the analysis of the relation between the orientation and size of Portuguese wildfires.

**Keywords** Directional data · Directional-linear density · Independence test · Nonparametric estimation · Permutations · Wildfires data

## 1 Introduction

Characterization of wildfire orientation patterns at landscape scale has important management implications (Moreira et al. 2001; Lloret et al. 2002; Moreira et al. 2011). It has been shown that landscape fuel reduction treatments will only be successful if strategically placed in

order to intersect fire spread in the heading direction (Finney 2001; Schmidt et al. 2008).

Barros et al. (2012) assessed the existence of preferential fire perimeter orientation at watershed level, to support the spatial layout of fuelbreak networks. Their analysis identified clusters of watersheds where fire perimeters were preferentially aligned along the NE/SW and the SE/NW axes. Those watersheds included fire perimeters that together account for roughly 65% of the overall burnt area in Portugal, over the period from 1975 to 2005, while in the remaining watersheds fire perimeters were randomly aligned. In Fig. 1, some descriptive maps of the data of interest are displayed. The left plot shows the total area burnt in each watershed, whereas the middle plot represents the mean slope of the fires in each region. Finally, the right plot indicates which watersheds exhibit a preferred fire orientation, versus a random orientation, according to Barros et al. (2012). The authors argued that spatial patterns of fire perimeter orientation found in the 31-year dataset could be explained by dominant weather during the Portuguese fire season (Pereira et al. 2005). However, given that fire perimeter orientation analysis is event-based (i.e., it is based on the orientation of each fire event) all perimeters are treated equally independently of their size. In this paper, a test for assessing independence between wildfire size and orientation is presented, complementing the work of Barros et al. (2012). Furthermore, orientation of the wildfire will be considered in two-dimensional and three-dimensional spaces.

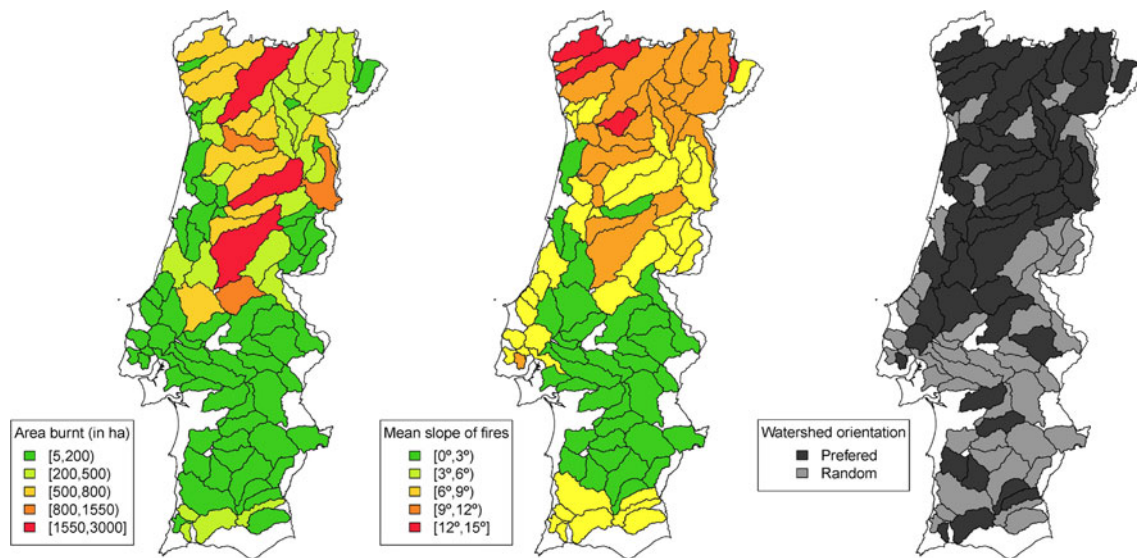
Spatial characterization of a wildfire, by means of its main orientation, and the associated burnt area, must be handled by non-standard statistical approaches, given the special nature of fire orientation. Specifically, it can be measured as an angle in the plane (two-dimensional orientation) or as a pair of angles identifying a direction in the

---

E. García-Portugués · R. M. Crujeiras ·  
W. González-Manteiga (✉)  
Department of Statistics and Operations Research, Faculty of  
Mathematics, University of Santiago de Compostela, Santiago,  
Spain  
e-mail: wenceslao.gonzalez@usc.es

E. García-Portugués  
e-mail: eduardo.garcia@usc.es

A. M. G. Barros · J. Pereira  
School of Agriculture, University of Lisbon, Lisbon, Portugal



**Fig. 1** Descriptive maps of wildfires in Portugal with the 102 watersheds delineated by Barros et al. (2012). The *left* map shows the number of hectares burnt from fire perimeters associated with each watershed. Each fire perimeter is associated with the watershed that contains its centroid. The *center* map represents the mean slope of the

fires of each watershed, where the slope is measured in degrees ( $0^\circ$  stands for plain slope and  $90^\circ$  for a vertical one). Finally, the *right* map shows watersheds where fires display preferential alignment according to Barros et al. (2012)

three-dimensional sphere, if the main slope of the wildfire is taken into account. Hence, appropriate methods for handling circular and, more generally, directional data must be considered, jointly with suitable combinations of directional and linear techniques.

The analysis of the relation between directional and linear variables has been classically approached through the construction of circular-linear correlation coefficients. The adaptation of the classical linear correlation coefficient to the circular-linear setting was introduced by Mardia (1976) and Johnson and Wehrly (1977) and further studied by Liddell and Ord (1978), who obtained its exact distribution under certain parametric assumptions. For the circular-linear case, a rank-based test of association was also proposed by Mardia (1976), who derived its asymptotic distribution. Later, Fisher and Lee (1981) adapted Kendall's  $\tau$  as a measure of circular-linear association based on the notion of concordance in the cylinder. To the best of the authors' knowledge, these three tests are the only ones available for testing the independence in directional-linear variables. As they are based on correlation coefficients, these tests are only powerful against deviations in the conditional expectation that can be measured by the corresponding coefficient. As a consequence, none of these tests is able to capture all possible types of dependence, neither for the conditional expectation nor for more complex types of dependence.

From a different perspective, circular and linear variables can also be jointly modeled by the construction of circular-linear distributions. Johnson and Wehrly (1978)

introduced a method for deriving circular-linear densities with specified marginals. A new family of circular-linear distributions based on nonnegative trigonometric sums, which proved to be more flexible in capturing the data structure, was proposed by Fernández-Durán (2007), adapting the method by Johnson and Wehrly (1978). More recently, García-Portugués et al. (2012) exploited the copula representation of the Johnson and Wehrly (1978) family, allowing for a completely nonparametric estimator, which was applied to analyze  $\text{SO}_2$  concentrations and wind direction. Nevertheless, the aforementioned methods are designed for the circular-linear case, whereas in our context, a more general tool for handling directional-linear relations is needed, provided that wildfire orientation may be reported in two or three dimensions.

In this paper, the assessment of the relation between a directional (circular or spherical, as particular cases) and a linear variable is approached through the construction of a formal test to check directional-linear independence. Inspired by the ideas of Rosenblatt (1975) and Rosenblatt and Wahlen (1992) for the linear setting (see also Ahmad and Li (1997)), the proposed test statistic is based on a nonparametric directional-linear kernel density estimator and an  $\mathcal{L}_2$  distance is taken as a discrepancy measure between the joint estimator and the one constructed under the independence hypothesis. The new test presents some interesting advantages: it is designed in a general fashion for directional variables of all dimensions and it is able to capture all kinds of deviations from independence by virtue the nonparametric density estimation. Besides, one gets a

kernel density estimate as a spin-off, which provides further information about the form of dependence when independence is rejected.

The remainder of the paper is organized as follows. In Sect. 2, some background to kernel density estimation, for linear, directional and directional-linear data is presented. Section 3 is devoted to the introduction of the test statistic, introducing a simplified version of the test and describing in detail its practical application. The finite sample performance of the test, in terms of size and power, is assessed through a simulation study for circular-linear and spherical-linear variables. Application to real data is provided in Sec. 4, including data description and results, focusing on the assessment of independence between wildfire orientation and burnt area size in Portugal. Some discussion and final comments are given in Sect. 5.

## 2 Background to kernel density estimation

In the linear setting, the basic building block for the independence test introduced by Rosenblatt (1975) is a kernel density estimator. Independence between two linear random variables is assessed through an  $\mathcal{L}_2$  distance between a bidimensional kernel density estimator and the product of the marginal kernel density estimators. In order to extend such a procedure to the directional-linear case, kernel density estimation for linear, directional and directional-linear variables is required. A brief background on kernel density estimators will be provided in this section.

### 2.1 Linear kernel density estimation

The well-known kernel density estimator for linear data was introduced by Rosenblatt (1956) and Parzen (1962). Given a random sample  $Z_1, \dots, Z_n$  from a linear random variable  $Z$  (i.e. with support  $Supp(Z) \subseteq \mathbb{R}$ ) with density  $f$ , the kernel density estimator at a point  $z \in \mathbb{R}$  is defined as

$$\hat{f}_g(z) = \frac{1}{ng} \sum_{i=1}^n K\left(\frac{z - Z_i}{g}\right),$$

where  $K$  is a kernel function, usually a symmetric density about the origin, and  $g > 0$  is the smoothing or bandwidth parameter, which controls the roughness of the estimator. Properties of this estimator have been deeply studied (see Silverman (1986) or Wand and Jones (1995) for comprehensive reviews). It is also well known that the choice of kernel (normal, Epanechnikov, etc.) has little effect on the overall shape of the kernel density estimate. However, the bandwidth is a key tuning parameter: large values produce oversmoothed estimates of  $f$ , whereas small values provide undersmoothed curves. Comprehensive reviews on

bandwidth selection are given in Cao et al. (1994), Chiu (1996) and Jones et al. (1996), among others.

### 2.2 Directional kernel density estimation

Denote by  $\mathbf{X}$  a directional random variable with density  $f$ . The support of such a variable is the  $q$ -dimensional sphere, namely  $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \dots + x_{q+1}^2 = 1\}$ , endowed with the Lebesgue measure in  $\Omega_q$ , that will be denoted by  $\omega_q$ . Therefore, a directional density is a nonnegative function that satisfies  $\int_{\Omega_q} f(\mathbf{x}) \mathbf{x} = 1$ .

The directional kernel density estimator was introduced by Hall et al. (1987) and Bai et al. (1988). Given a random sample  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , of a directional variable  $\mathbf{X}$  with  $Supp(\mathbf{X}) \subseteq \Omega_q$  and density  $f$ , at a point  $\mathbf{x} \in \Omega_q$  the estimator is given by

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^n L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right), \tag{1}$$

where  $L$  is the directional kernel,  $h > 0$  is the bandwidth parameter and  $c_{h,q}(L)$  is a normalizing constant depending on the kernel  $L$ , the bandwidth  $h$  and the sphere dimension  $q$ . The scalar product of two vectors,  $\mathbf{x}$  and  $\mathbf{y}$ , is denoted by  $\mathbf{x}'\mathbf{y}$ , where  $'$  denotes the transpose operator.

A common choice for the directional kernel is  $L(r) = e^{-r}$ ,  $r \geq 0$ , also known as the von Mises kernel due to its relation with the von Mises–Fisher distribution (Watson 1983). In a  $q$ -dimensional sphere, the von Mises density  $vM(\boldsymbol{\mu}, \kappa)$  is given by

$$f_{vM}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\{\kappa \mathbf{x}'\boldsymbol{\mu}\}, \tag{2}$$

$$C_q(\kappa) = \kappa^{\frac{q-1}{2}} \left[ (2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa) \right]^{-1},$$

where  $\boldsymbol{\mu} \in \Omega_q$  is the mean direction,  $\kappa \geq 0$  is the concentration parameter around the mean and  $\mathcal{I}_\nu$  is the modified Bessel function of order  $\nu$ ,

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2} \Gamma\left(\nu + \frac{1}{2}\right)} \int_{-1}^1 (1 - t^2)^{\nu - \frac{1}{2}} e^{zt} dt.$$

For the von Mises kernel, the value of  $c_{h,q}(L)$  is  $C_q(1/h^2) e^{1/h^2}$  and the directional estimator (1) can be interpreted as a mixture of  $q$ -von Mises–Fisher densities:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n f_{vM}(\mathbf{x}; \mathbf{X}_i, 1/h^2).$$

Note that large values of  $h$  provide a small concentration parameter, which results in a uniform model in the sphere, whereas small values of  $h$  give high concentrations around the sample observations, providing an undersmoothed curve. Cross-validation rules based on likelihood cross validation

(LCV) and least squares cross validation (LSCV) for bandwidth selection were discussed by Hall et al. (1987).

### 2.3 Directional-linear kernel density estimation

Consider a directional-linear random variable,  $(\mathbf{X}, Z)$  with support  $Supp(\mathbf{X}, Z) \subseteq \Omega_q \times \mathbb{R}$  and joint density  $f$ . For the simple case of circular data ( $q = 1$ ), the support of the variable is the cylinder and, in general, the support is a multidimensional cylinder. Following the ideas in the previous sections for the linear and directional cases, given a random sample  $(\mathbf{X}_1, Z_1), \dots, (\mathbf{X}_n, Z_n)$ , the directional-linear kernel density estimator at a point  $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$  can be defined as:

$$\hat{f}_{h,g}(\mathbf{x}, z) = \frac{c_{h,q}(L)}{ng} \sum_{i=1}^n LK\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}, \frac{z - Z_i}{g}\right), \tag{3}$$

where  $LK$  is a directional-linear kernel,  $g > 0$  is the linear bandwidth parameter,  $h > 0$  is the directional bandwidth and  $c_{h,q}(L)$  is the directional normalizing constant. The estimator (3) was introduced by García-Portugués et al. (2013a), who also studied its asymptotic properties in terms of bias and variance, and established its asymptotic normality.

A product kernel  $LK(\cdot, \cdot) = L(\cdot) \times K(\cdot)$ , specifically, the von Mises-normal kernel

$$LK(r, t) = e^{-r} \times \phi(t), \quad r \in [0, \infty), t \in \mathbb{R},$$

will be considered throughout this paper in order to simplify computations, where  $\phi$  denotes the standard normal density. Similarly to the linear and directional kernel density estimators, a smoothing parameter (bidimensional, in this case) is involved in the estimator construction. The cross-validation procedures introduced by Hall et al. (1987) can be adapted to the directional-linear setting, yielding the following bandwidth selectors:

$$(h, g)_{LCV} = \arg \max_{h,g > 0} \sum_{i=1}^n \log \hat{f}_{h,g}^{-i}(\mathbf{X}_i, Z_i),$$

$$(h, g)_{LSCV} = \arg \max_{h,g > 0} \left[ 2n^{-1} \sum_{i=1}^n \hat{f}_{h,g}^{-i}(\mathbf{X}_i, Z_i) - \int_{\Omega_q \times \mathbb{R}} \hat{f}_{h,g}(\mathbf{x}, z)^2 \mathbf{x}z \right],$$

where  $\hat{f}_{h,g}^{-i}$  represents the kernel density estimator computed without the  $i$ th datum.

### 3 A test for directional-linear independence

The new test statistic for assessing independence between a directional and a linear variable is described in this section.

#### 3.1 The test statistic

Consider the joint directional-linear density  $f_{(\mathbf{x},Z)}$  for the variable  $(\mathbf{X}, Z)$ .  $f_{\mathbf{X}}$  and  $f_Z$  denote the directional and linear marginal densities, respectively. The null hypothesis of independence between both components can be stated as

$$H_0 : f_{(\mathbf{x},Z)}(\mathbf{x}, z) = f_{\mathbf{X}}(\mathbf{x})f_Z(z), \quad \forall (\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$$

and the alternative hypothesis as

$$H_a : f_{(\mathbf{x},Z)}(\mathbf{x}, z) \neq f_{\mathbf{X}}(\mathbf{x})f_Z(z),$$

for any  $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ .

Following the idea of Rosenblatt (1975), a natural statistic to test  $H_0$  arises from considering the  $\mathcal{L}_2$  distance between the nonparametric estimation of the joint density  $f_{(\mathbf{x},Z)}$  by the directional-linear kernel estimator (3), denoted by  $\hat{f}_{(\mathbf{x},Z);h,g}$ , and the nonparametric estimation of  $f_{(\mathbf{x},Z)}$  under  $H_0$ , given by the product of the marginal directional and linear kernel estimators, denoted by  $\hat{f}_{\mathbf{X};h}$  and  $\hat{f}_{Z;g}$ , respectively. We therefore propose the following test statistic:

$$T_n = \Delta_2(\hat{f}_{(\mathbf{x},Z);h,g}, \hat{f}_{\mathbf{X};h}\hat{f}_{Z;g}), \tag{4}$$

where  $\Delta_2$  stands for the squared  $\mathcal{L}_2$  distance in  $\Omega_q \times \mathbb{R}$  between two functions  $f_1$  and  $f_2$ :

$$\Delta_2(f_1, f_2) = \int_{\Omega_q \times \mathbb{R}} (f_1(\mathbf{x}, z) - f_2(\mathbf{x}, z))^2 \omega_q(d\mathbf{x})dz.$$

The test statistic depends on a pair of bandwidths ( $h, g$ ), which is used for the directional-linear estimator, and whose components are also considered for the marginal directional and linear kernel density estimators. Under the null hypothesis of independence,  $H_0$ , it holds that  $\mathbb{E}[\hat{f}_{(\mathbf{x},Z);h,g}(\mathbf{x}, z)] = \mathbb{E}[\hat{f}_{\mathbf{X};h}(\mathbf{x})] \mathbb{E}[\hat{f}_{Z;g}(z)]$ .

Asymptotic properties of (4) have been studied by García-Portugués et al. (2013b), who proved its asymptotic normality under independence, but with a slow rate of convergence that does not encourage its use in practice. For that reason, a calibration mechanism will be needed for the practical application of the test.

In addition, the construction of  $T_n$  requires the calculation of an integral over  $\Omega_q \times \mathbb{R}$ , which may pose computational problems since it involves the calculation of several nested integrals. However, if the kernel estimators are obtained using von Mises and normal kernels, then an easy to compute expression for  $T_n$  can be obtained, as stated in the following lemma.

**Lemma 1** *If the kernel estimators involved in (4), obtained from a random sample  $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$  of  $(\mathbf{X}, Z)$ , are*

constructed with von Mises and normal kernels, the following expression for  $T_n$  holds:

$$T_n = \mathbf{1}_n \left( \frac{1}{n^2} \Psi(h) \circ \Omega(g) - \frac{2}{n^3} \Psi(h) \Omega(g) + \frac{1}{n^4} \Psi(h) \mathbf{1}_n \mathbf{1}'_n \Omega(g) \right) \mathbf{1}'_n, \tag{5}$$

where  $\circ$  denotes the Hadamard product and  $\Psi(h)$  and  $\Omega(g)$  are  $n \times n$  matrices given by

$$\Psi(h) = \left( \frac{C_q(1/h^2)^2}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)} \right)_{ij},$$

$$\Omega(g) = (\phi_{\sqrt{2}g}(Z_i - Z_j))_{ij},$$

where  $\mathbf{1}_n = (1, \dots, 1)$ , with length  $n$ ,  $C_q$  is the normalizing function (2) and  $\phi_{\sqrt{2}g}$  is the normal density with zero mean and standard deviation  $\sqrt{2}g$ .

The proof of this result can be seen in Appendix A. Note that expression (5) for  $T_n$  only requires matrix operations. This will be the expression used for computing the test statistic. It should also be noted that the effect of the dimension  $q$  appears only in the definition of  $C_q$  and in  $\|\mathbf{X}_i + \mathbf{X}_j\|$ , and both are easily scalable for large  $q$ . Thus, an important advantage of (5) is that computing requirements are similar for different dimensions  $q$ , something which is not the case if (4) is employed with numerical integration.

### 3.2 Calibration of the test

The null hypothesis of independence is stated in a non-parametric way, which determines the resampling methods used for calibration. However, as the null hypothesis is of a non-interaction kind, a permutation approach (which is not at all foreign to hypothesis testing) seems a reliable option. If  $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$  is a random sample from the directional-linear variable  $(\mathbf{X}, Z)$  and  $\sigma$  is a random permutation of  $n$  elements, then  $\{(\mathbf{X}_i, Z_{\sigma(i)})\}_{i=1}^n$ , represents the resulting  $\sigma$ -permuted sample.  $T_n^\sigma$  denotes the test statistic computed from the  $\sigma$ -permuted random sample. Under the assumption of independence between the directional and linear components, it is reasonable to expect that the distribution of  $T_n$  is similar to the distribution of  $T_n^\sigma$ , which can be easily approximated by Monte Carlo methods.

In addition to its simplicity, the main advantage of the use of permutations is its easy implementation using Lemma 1, as it is possible to reuse the computation of the matrices  $\Psi(h)$  and  $\Omega(g)$  needed for  $T_n$  to compute a  $\sigma$ -

permuted statistic  $T_n^\sigma$ . In virtue of expression (5) and the definition of  $T_n^\sigma$ , the  $\sigma$ -permuted test statistic is given by

$$T_n^\sigma = \mathbf{1}_n \left( \frac{1}{n^2} \Psi(h) \circ \Omega^\sigma(g) - \frac{2}{n^3} \Psi(h) \Omega^\sigma(g) + \frac{1}{n^4} \Psi(h) \mathbf{1}_n \mathbf{1}'_n \Omega^\sigma(g) \right) \mathbf{1}'_n,$$

where the  $ij$ th entry of the matrix  $\Omega^\sigma(g)$  is the  $\sigma(i)\sigma(j)$ -entry of  $\Omega(g)$ . For the computation of  $\Psi(h)$  and  $\Omega(g)$ , symmetry properties reduce the number of computations and can also be used to optimize the products  $\Psi(h) \circ \Omega^\sigma(g)$  and  $\Psi(h) \Omega^\sigma(g)$ . The last addend of  $T_n^\sigma$  is the same as that of  $T_n$  and there is no need to recompute it. The testing procedure can be summarized in the following algorithm.

**Algorithm 1** Let  $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$  be a random sample from a directional-linear variable  $(\mathbf{X}, Z)$ .

1. Obtain a suitable pair of bandwidths  $(h, g)$ .
2. Compute the observed value of  $T_n$  from (5), with kernel density estimators taking bandwidths  $(h, g)$ .
3. Permutation calibration. For  $b = 1, \dots, B \leq n!$  compute  $T_n^{\sigma_b}$  with bandwidths  $(h, g)$  for a random permutation  $\sigma_b$ .
4. Approximate the  $p$ -value by  $\#\{T_n \leq T_n^{\sigma_b}\} / B$ , where  $\#$  denotes the cardinal of the set.

In Steps 2 and 3, a pair of bandwidths must be chosen. For the directional-linear case, as commented in Section 2, cross-validation bandwidths, namely  $(h, g)_{\text{LCV}}$  and  $(h, g)_{\text{LSCV}}$ , can be considered. However, as usually happens with cross-validated bandwidths, these selectors tend to provide undersmoothed estimators, something which a priori is not desirable as introduces a substantial variability in the statistic  $T_n$ .

To mitigate this problem, a more sophisticated bandwidth selector will be introduced. Considering the von Mises–normal kernel, the bootstrap version for the mean integrated squared error (MISE) of the directional-linear kernel density estimator (3) was derived by García-Portugués et al. (2013a):

$$\begin{aligned} \text{MISE}_{h_p, g_p}^*(h, g) &= \left( C_q(1/h^2)^2 C_q(2/h^2)^{-1} 2\pi^{\frac{1}{2}} gn \right)^{-1} \\ &\quad + n^{-2} \mathbf{1}_n [(1 - n^{-1}) \Psi_2^*(h) \circ \Omega_2^*(g) \\ &\quad - 2\Psi_1^*(h) \circ \Omega_1^*(g) \\ &\quad + \Psi_0^* \circ \Omega_0^*] \mathbf{1}'_n, \end{aligned}$$

where matrices  $\Psi_a^*(h)$  and  $\Omega_a^*(g)$ ,  $a = 0, 1, 2$  are

$$\Psi_0^* = \left( \frac{C_q(1/h_p^2)^2}{C_q(\|\mathbf{X}_i + \mathbf{X}_j/h_p^2\|)} \right)_{ij},$$

$$\Psi_1^*(h) = \left( \frac{\int_{\Omega_q} C_q(1/h^2)C_q(1/h_p^2)^2 e^{\mathbf{x}^T \mathbf{X}_j/h_p^2} \omega_q(dx)}{C_q(\|\mathbf{x}/h^2 + \mathbf{X}_i/h_p^2\|)} \right)_{ij},$$

$$\Psi_2^*(h) = \left( \frac{\int_{\Omega_q} \frac{C_q(1/h^2)^2}{C_q(\|\mathbf{x}/h^2 + \mathbf{X}_i/h_p^2\|)} \cdot \frac{C_q(1/h_p^2)^2}{C_q(\|\mathbf{x}/h^2 + \mathbf{X}_j/h_p^2\|)} \omega_q(dx) \right)_{ij},$$

$$\Omega_0^* = \left( \phi_{\sqrt{2}g_p}(Z_i - Z_j) \right)_{ij},$$

$$\Omega_a^*(g) = \left( \phi_{\sigma_{a,g}}(Z_i - Z_j) \right)_{ij},$$

with  $\sigma_{a,g} = (ag^2 + 2g_p^2)^{\frac{1}{2}}$ ,  $a = 1, 2$ , and  $(h_p, g_p)$  a given pair of pilot bandwidths. Then, the estimation bandwidths are obtained as

$$(h, g)_{bo} = \arg \min_{h, g > 0} \text{MISE}_{h_p, g_p}^*(h, g).$$

The choice of  $(h_p, g_p)$  is needed in order to compute  $(h, g)_{bo}$ . This must be done by a joint criterion for two important reasons. Firstly, to avoid the predominance of smoothing in one component that may dominate the other (this could happen, for example, if the directional variable is uniform, as in that case the optimal bandwidth tends to infinity). Secondly, to obtain a test with more power against deviations from independence. Based on these comments, a new bandwidth selector, named Bootstrap LCV (BLCV), is introduced:

$$(h, g)_{\text{BLCV}} = \arg \min_{h, g > 0} \text{MISE}_{(h, g)_{\text{MLCV}}}^*(h, g),$$

where the pair of bandwidths  $(h, g)_{\text{MLCV}}$  are obtained by enlarging the order of  $(h, g)_{\text{LCV}}$  to be of the kind  $(\mathcal{O}(n^{-\frac{1}{6+q}}), \mathcal{O}(n^{-\frac{1}{4}}))$ , the order that one would expect for a pair of directional-linear pilot bandwidths. For the linear component, this can be seen in the paper by Cao (1993), where the pilot bandwidth is proved to be  $g_p = \mathcal{O}(n^{-\frac{1}{4}})$ , larger than the order of the optimal estimation bandwidth,  $n^{-\frac{1}{5}}$ . For the directional case there is no pilot bandwidth available, but considering that the order of the optimal estimation bandwidth is  $n^{-\frac{1}{4+q}}$  (García-Portugués et al. 2013a), then a plausible conjecture is  $h_p = \mathcal{O}(n^{-\frac{1}{6+q}})$ .

### 3.3 Simulation study

Six different directional-linear models were considered in the simulation study. The models are indexed by a  $\delta$

parameter that measures the degree of deviation from the independence, where  $\delta = 0$  represents independence and  $\delta > 0$  accounts for different degrees of dependence. The models show three kind of possible deviations from the independence: first order deviations, that is, deviations in the conditional expectation (models M1, M2 and M3); second order deviations or conditional variance deviations (models M4 and M5) and first and second order deviations (model M6). In order to clarify notation,  $\phi(\cdot; m, \sigma)$  and  $f_{LN}(\cdot; m, \sigma)$  represent the density of a normal and a log-normal, with mean/log-scale  $m$  and standard deviation/shape  $\sigma$ . Notation  $\mathbf{0}_q$  represents a vector of  $q$  zeros.

- M1.  $f_1(\mathbf{x}, z) = \phi(z; \delta(2 + \mathbf{x}^T \boldsymbol{\mu}), \sigma) \times f_{VM}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$ , with  $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$ ,  $\kappa = 1$  and  $\sigma = 1$ .
- M2.  $f_2(\mathbf{x}, z) = f_{LN}(z; \delta(1 + (\mathbf{x}^T \boldsymbol{\mu})^2), \sigma) \times f_{VM}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$ , with  $\boldsymbol{\mu} = (-1, \mathbf{0}_q)$ ,  $\kappa = 0$  and  $\sigma = \frac{1}{4}$ .
- M3.  $f_3(\mathbf{x}, z) = [rf_{LN}(z; \delta(1 + (\mathbf{x}^T \boldsymbol{\mu}_1)^3), \sigma_1) + (1 - r)\phi(z; m, \sigma_2)] \times [pf_{VM}(\mathbf{x}; \boldsymbol{\mu}_1, \kappa_1) + (1 - p)f_{VM}(\mathbf{x}; \boldsymbol{\mu}_2, \kappa_2)]$ , with  $\boldsymbol{\mu}_1 = (\mathbf{0}_q, 1)$ ,  $\boldsymbol{\mu}_2 = (\mathbf{0}_q, -1)$ ,  $\kappa_1 = 2$ ,  $\kappa_2 = 1$ ,  $p = \frac{3}{4}$ ,  $\sigma_1 = \sigma_2 = \frac{1}{4}$ ,  $m = 1$  and  $r = \frac{1}{4}$ .
- M4.  $f_4(\mathbf{x}, z) = \phi\left(z; m, \frac{1}{4} + \delta(1 - (\mathbf{x}^T \boldsymbol{\mu}_r)^3)\right) \times f_{VM}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$ , with  $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$ ,  $\kappa = 1$ ,  $\boldsymbol{\mu}_r = (-1, \mathbf{0}_q)$  and  $m = 0$ .
- M5.  $f_5(\mathbf{x}, z) = f_{LN}\left(z; m, (5 + \delta(3\mathbf{x}^T \boldsymbol{\mu}_2 - \mathbf{x}^T \boldsymbol{\mu}_1))^{-1}\right) \times [pf_{VM}(\mathbf{x}; \boldsymbol{\mu}_1, \kappa_1) + (1 - p)f_{VM}(\mathbf{x}; \boldsymbol{\mu}_2, \kappa_2)]$ , with  $\boldsymbol{\mu}_1 = (\mathbf{0}_q, 1)$ ,  $\boldsymbol{\mu}_2 = (\mathbf{0}_q, -1)$ ,  $\kappa_1 = \kappa_2 = 2$ ,  $p = \frac{1}{2}$  and  $m = 0$ .
- M6.  $f_6(\mathbf{x}, z) = [rf_{LN}(z; m, \sigma) + (1 - r)\phi(z; \delta(2 + \mathbf{x}^T \boldsymbol{\mu}_r), \frac{1}{4} + \delta(\mathbf{x}^T \boldsymbol{\mu}_r)^2)] \times f_{VM}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$ , with  $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$ ,  $\boldsymbol{\mu}_r = (-1, \mathbf{0}_q)$ ,  $\kappa = 1$ ,  $p = \frac{3}{4}$ ,  $\sigma = \frac{1}{2}$ ,  $m = 0$  and  $r = \frac{3}{4}$ .

The choice of the models was done in order to capture situations with heteroskedasticity, skewness in the linear component and different types of von Mises mixtures in the directional component.

For the proposed models, different deviations from independence have been considered, by setting  $\delta = 0, 0.25, 0.50$ . The proposed test statistic has been computed for all the models and sample sizes  $n = 50, 100, 200, 500, 1000$ . The new test based on the permutation resampling described in Algorithm 1, depending on the bandwidth choice, is denoted by  $T_n^{\text{LCV}}$  and  $T_n^{\text{BLCV}}$ . The number of permutations considered was  $B = 1000$  and the number of Monte Carlo replicates was  $M = 1000$ . Both the circular-linear and spherical-linear cases were explored. For the circular-linear case, the test was compared with the three tests available for circular-linear association, described as follows:

- Circular-linear correlation coefficient from Mardia (1976) and Johnson and Wehrly (1977), denoted by  $R_n^2$ .
- Rank circular-linear correlation coefficient from Mardia (1976), denoted by  $U_n$ .

**Table 1** Proportion of rejections for the  $R_n^2$ ,  $U_n$ ,  $\lambda_{4n}^*$ ,  $T_n^{LCV}$  and  $T_n^{BLCV}$  tests of independence as a function of sample size,  $n$ , and model, for a nominal significance level of 5 %

$n$	Model	Circular-linear					Spherical-linear	
		$R_n^2$	$U_n$	$\lambda_{4n}^*$	$T_n^{LCV}$	$T_n^{BLCV}$	$T_n^{LCV}$	$T_n^{BLCV}$
50	$H_{1,0.00}$	0.047	0.040	0.047	0.050	0.047	0.059	0.045
	$H_{2,0.00}$	0.051	0.044	0.042	0.055	0.052	0.057	0.052
	$H_{3,0.00}$	0.047	0.045	0.051	0.059	0.054	0.051	0.048
	$H_{4,0.00}$	0.047	0.040	0.047	0.050	0.046	0.059	0.046
	$H_{5,0.00}$	0.042	0.045	0.053	0.055	0.047	0.070	0.057
	$H_{6,0.00}$	0.058	0.065	0.062	0.057	0.055	0.065	0.054
	$H_{1,0.25}$	0.162	0.120	0.092	0.132	0.139	0.099	0.094
	$H_{2,0.25}$	0.055	0.074	0.088	0.143	0.071	0.072	0.050
	$H_{3,0.25}$	0.535	0.538	0.365	0.511	0.543	0.238	0.246
	$H_{4,0.25}$	0.051	0.044	0.067	0.239	0.234	0.103	0.097
	$H_{5,0.25}$	0.046	0.049	0.059	0.128	0.121	0.110	0.094
	$H_{6,0.25}$	0.354	0.332	0.239	0.436	0.432	0.284	0.275
	$H_{1,0.50}$	0.512	0.412	0.235	0.378	0.421	0.231	0.253
	$H_{2,0.50}$	0.054	0.124	0.261	0.633	0.291	0.219	0.078
	$H_{3,0.50}$	0.925	0.845	0.734	0.929	0.949	0.662	0.666
	$H_{4,0.50}$	0.058	0.050	0.081	0.424	0.420	0.149	0.139
	$H_{5,0.50}$	0.055	0.059	0.094	0.501	0.491	0.320	0.298
	$H_{6,0.50}$	0.782	0.706	0.536	0.754	0.756	0.556	0.540
100	$H_{1,0.00}$	0.052	0.054	0.063	0.068	0.061	0.072	0.068
	$H_{2,0.00}$	0.044	0.046	0.052	0.053	0.048	0.051	0.055
	$H_{3,0.00}$	0.047	0.050	0.046	0.061	0.054	0.064	0.049
	$H_{4,0.00}$	0.052	0.054	0.063	0.067	0.060	0.072	0.071
	$H_{5,0.00}$	0.056	0.050	0.057	0.073	0.063	0.074	0.063
	$H_{6,0.00}$	0.046	0.046	0.050	0.062	0.059	0.077	0.071
	$H_{1,0.25}$	0.291	0.227	0.102	0.211	0.213	0.155	0.163
	$H_{2,0.25}$	0.051	0.073	0.092	0.263	0.094	0.114	0.067
	$H_{3,0.25}$	0.889	0.851	0.407	0.805	0.849	0.487	0.500
	$H_{4,0.25}$	0.060	0.049	0.074	0.478	0.484	0.222	0.219
	$H_{5,0.25}$	0.063	0.050	0.067	0.260	0.251	0.171	0.171
	$H_{6,0.25}$	0.547	0.574	0.283	0.720	0.718	0.492	0.479
	$H_{1,0.50}$	0.847	0.721	0.290	0.669	0.718	0.416	0.460
	$H_{2,0.50}$	0.053	0.122	0.279	0.940	0.660	0.530	0.123
	$H_{3,0.50}$	1.000	0.997	0.872	0.999	0.999	0.942	0.957
	$H_{4,0.50}$	0.058	0.053	0.103	0.784	0.803	0.341	0.355
	$H_{5,0.50}$	0.083	0.056	0.107	0.836	0.860	0.602	0.630
	$H_{6,0.50}$	0.965	0.951	0.642	0.968	0.967	0.864	0.845
200	$H_{1,0.00}$	0.049	0.056	0.064	0.057	0.054	0.065	0.060
	$H_{2,0.00}$	0.055	0.063	0.055	0.053	0.054	0.046	0.041
	$H_{3,0.00}$	0.051	0.054	0.058	0.053	0.050	0.045	0.042
	$H_{4,0.00}$	0.049	0.056	0.064	0.057	0.053	0.065	0.057
	$H_{5,0.00}$	0.049	0.056	0.043	0.066	0.063	0.060	0.061
	$H_{6,0.00}$	0.048	0.059	0.052	0.049	0.046	0.054	0.048
	$H_{1,0.25}$	0.529	0.444	0.099	0.349	0.373	0.192	0.208
	$H_{2,0.25}$	0.058	0.081	0.106	0.551	0.154	0.178	0.052
	$H_{3,0.25}$	0.996	0.995	0.431	0.980	0.987	0.795	0.818
	$H_{4,0.25}$	0.054	0.057	0.085	0.839	0.862	0.337	0.348

**Table 1** continued

n	Model	Circular-linear					Spherical-linear	
		$R_n^2$	$U_n$	$\lambda_{4n}^*$	$T_n^{LCV}$	$T_n^{BLCV}$	$T_n^{LCV}$	$T_n^{BLCV}$
	$H_{5,0.25}$	0.056	0.052	0.058	0.459	0.487	0.303	0.343
	$H_{6,0.25}$	0.830	0.896	0.277	0.974	0.971	0.842	0.830
	$H_{1,0.50}$	0.982	0.957	0.299	0.924	0.940	0.721	0.750
	$H_{2,0.50}$	0.061	0.145	0.325	0.999	0.967	0.899	0.249
	$H_{3,0.50}$	1.000	1.000	0.913	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.053	0.058	0.122	0.981	0.984	0.595	0.618
	$H_{5,0.50}$	0.124	0.051	0.105	0.991	0.995	0.921	0.950
	$H_{6,0.50}$	1.000	1.000	0.691	1.000	1.000	0.994	0.993

For the six different models the values of the deviation from independence parameter are  $\delta = 0$  (independence), 0.25 and 0.50. Each proportion was calculated using  $B = 1000$  permutations for each of  $M = 1000$  random samples of size  $n$  simulated from the specified model

- $\lambda_{4n}$  measure of cylindrical association of Fisher and Lee (1981), implemented with its incomplete version  $\lambda_{4n}^*$  considering  $m = 5000$  random 4-tuples.

Although there exists an exact distribution for  $R_n^2$  under certain normality assumptions on the linear response and asymptotic distributions for  $U_n$  and  $\lambda_{4n}^*$ , for a fair comparison, the calibration of these tests has also been done by permutations ( $B = 1000$ ). The exact and asymptotic distributions for  $R_n^2$  and  $U_n$  were also tried instead of the permutation approach, providing empirical levels and powers quite similar to the ones based on permutations.

The proportion of rejections under  $H_{k,\delta}$  (for model number  $k$  with  $\delta$  deviation) is reported in Tables 1 and 2, for the circular-linear and spherical-linear cases, with different sample sizes. In the circular-linear case, the empirical size is close to the nominal level for all the competing tests. The  $T_n^{LCV}$  test for this case shows in general a satisfactory behavior under the null hypothesis, except for some cases in models M1, M4 and M5, where the test tends to reject the null hypothesis more than expected. This is mostly corrected by  $T_n^{BLCV}$ , with a decrease of power with respect to  $T_n^{LCV}$  in model M2. For the spherical-linear case, the improvement in size approximation  $T_n^{BLCV}$  is notable, specially for small sample sizes. If the tests maintain the nominal significance level of 5 %, it is expected that approximately 95% of the observed proportions of rejections under the null hypothesis (i.e. when  $\delta = 0$ ) to lie within the interval (0.036,0.064) to three decimal places.

Regarding power, the test for  $R_n^2$  is the most powerful one for models M1 and M3, although the performance of  $T_n^{LCV}$  and  $T_n^{BLCV}$ , specially for model M3, is quite similar. This was to be expected, as the circular-linear association tests should present more power against deviations of the first order. However, for models M2, M4 and M5, all these tests are not able to distinguish the alternatives and the

rejection ratios are close to the nominal level, resulting in  $\lambda_{4n}^*$  being the test with better behavior among them. In contrast,  $T_n^{LCV}$  and  $T_n^{BLCV}$  correctly detect the deviations from the null. In Model M6,  $R_n^2$  is only the most competitive for the situation with  $n = 50$ , with  $T_n^{LCV}$  and  $T_n^{BLCV}$  the most competitive for the remainder of situations.  $U_n$  shows a similar performance to  $R_n^2$ , but with more power in M2 and less in M5.  $\lambda_{4n}^*$  is less affected than  $R_n^2$  and  $U_n$  by the change of models, but also has lower power than them for models M1, M3 and M6. The results for the spherical-linear case are quite similar to the previous ones for the empirical size, but with lower power in comparison with the circular-linear scenario, something expected as a consequence of the difference in dimensionality.

Some final comments on the simulation results follow. For the different sample sizes and dimensions, the running times for  $T_n^{LCV}$  and  $T_n^{BLCV}$  are collected in Table 3. Computation times for  $T_n^{LCV}$  are very similar for different dimensions  $q$ , whereas  $T_n^{BLCV}$  is affected by  $q$  due to the choice of the bandwidths  $(h, g)_{BLCV}$ . The choice of the kernels was corroborated to be non-important for testing, as similar results were obtained for the test  $T_n^{LCV}$  using the directional-linear kernel  $LK(r, t) = (1 - r)1_{[0,1]}(r) \times \frac{3}{4}(1 - t^2)1_{[-1,1]}(t)$ . Cross-validatory bandwidths LSCV and BLSCV were also tried in the simulation study, providing worse results (this is also what usually happens with directional data, as it can be seen in García-Portugués (2013)). Finally, it is worth mentioning that bootstrap calibration was also tried as an alternative to the permutation approach, using a pair of bandwidths for estimation and another pair for the smooth resampling. The results in terms of size, power and computing times were substantially worse than the ones obtained for permutations.

In conclusion, both  $T_n^{LCV}$  and  $T_n^{BLCV}$  tests show a competitive behavior in all the simulation models, sample sizes

**Table 2** Proportion of rejections for the  $R_n^2$ ,  $U_n$ ,  $\lambda_{4n}^*$ ,  $T_n^{LCV}$  and  $T_n^{BLCV}$  tests of independence as a function of sample size,  $n$ , and model, for a nominal significance level of 5 %

$n$	Model	Circular-linear					Spherical-linear	
		$R_n^2$	$U_n$	$\lambda_{4n}^*$	$T_n^{LCV}$	$T_n^{BLCV}$	$T_n^{LCV}$	$T_n^{BLCV}$
500	$H_{1,0.00}$	0.053	0.060	0.054	0.069	0.064	0.055	0.045
	$H_{2,0.00}$	0.060	0.050	0.053	0.062	0.055	0.046	0.048
	$H_{3,0.00}$	0.054	0.059	0.064	0.050	0.044	0.045	0.042
	$H_{4,0.00}$	0.053	0.060	0.054	0.069	0.062	0.056	0.050
	$H_{5,0.00}$	0.042	0.038	0.058	0.050	0.047	0.051	0.059
	$H_{6,0.00}$	0.052	0.050	0.053	0.059	0.062	0.059	0.055
	$H_{1,0.25}$	0.916	0.842	0.088	0.698	0.727	0.422	0.447
	$H_{2,0.25}$	0.050	0.073	0.095	0.973	0.511	0.557	0.073
	$H_{3,0.25}$	1.000	1.000	0.443	1.000	1.000	0.995	0.997
	$H_{4,0.25}$	0.057	0.060	0.077	0.999	0.999	0.764	0.786
	$H_{5,0.25}$	0.080	0.038	0.068	0.850	0.865	0.679	0.750
	$H_{6,0.25}$	0.998	1.000	0.263	1.000	1.000	1.000	0.998
	$H_{1,0.50}$	1.000	1.000	0.264	0.998	0.999	0.982	0.986
	$H_{2,0.50}$	0.053	0.125	0.322	1.000	1.000	1.000	0.910
	$H_{3,0.50}$	1.000	1.000	0.942	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.064	0.060	0.090	1.000	1.000	0.982	0.987
	$H_{5,0.50}$	0.258	0.043	0.108	1.000	1.000	1.000	1.000
	$H_{6,0.50}$	1.000	1.000	0.709	1.000	1.000	1.000	1.000
1,000	$H_{1,0.00}$	0.059	0.053	0.060	0.056	0.057	0.061	0.060
	$H_{2,0.00}$	0.043	0.042	0.070	0.045	0.046	0.058	0.051
	$H_{3,0.00}$	0.063	0.054	0.062	0.057	0.054	0.038	0.037
	$H_{4,0.00}$	0.059	0.053	0.060	0.056	0.054	0.061	0.054
	$H_{5,0.00}$	0.055	0.060	0.047	0.053	0.051	0.078	0.074
	$H_{6,0.00}$	0.045	0.047	0.054	0.057	0.058	0.052	0.048
	$H_{1,0.25}$	0.997	0.992	0.084	0.938	0.947	0.730	0.747
	$H_{2,0.25}$	0.046	0.067	0.109	1.000	0.910	0.947	0.123
	$H_{3,0.25}$	1.000	1.000	0.459	1.000	1.000	1.000	1.000
	$H_{4,0.25}$	0.061	0.052	0.074	1.000	1.000	0.989	0.991
	$H_{5,0.25}$	0.129	0.059	0.059	0.993	0.995	0.936	0.971
	$H_{6,0.25}$	1.000	1.000	0.257	1.000	1.000	1.000	1.000
	$H_{1,0.50}$	1.000	1.000	0.281	1.000	1.000	1.000	1.000
	$H_{2,0.50}$	0.049	0.125	0.305	1.000	1.000	1.000	1.000
	$H_{3,0.50}$	1.000	1.000	0.954	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.058	0.057	0.099	1.000	1.000	1.000	1.000
	$H_{5,0.50}$	0.486	0.058	0.106	1.000	1.000	1.000	1.000
	$H_{6,0.50}$	1.000	1.000	0.751	1.000	1.000	1.000	1.000

For the six different models the values of the deviation from independence parameter are  $\delta = 0$  (independence), 0.25 and 0.50. Each proportion was calculated using  $M = 1000$  permutations of each of  $B = 1000$  random samples of size  $n$  simulated from the specified model

and dimensions considered, only being outperformed by  $R_n^2$  in models M1 and M3. Nevertheless, for those models, the rejection rates of both tests are in general close to the ones of  $R_n^2$ . The test  $T_n^{BLCV}$  corrects the over rejection of  $T_n^{LCV}$  in certain simulation models, without a significant loss in power but at the expense of a high computational cost.

Finally, the classical tests  $R_n^2$ ,  $U_n$  and  $\lambda_{4n}^*$  presented critical problems on detecting second order and some first order deviations from the independence. For all those reasons, the final recommendation is to preferably use the test  $T_n^{BLCV}$  for inference on directional-linear independence and  $T_n^{LCV}$  for a less computing intensive exploratory analysis.

**Table 3** Computing times (in s) for  $T_n^{\text{LCV}}$  and  $T_n^{\text{BLCV}}$  as a function of sample size and dimension  $q$ , with  $q = 1$  for the circular-linear case and  $q = 2$  for the spherical-linear case

Test	$q$	Sample size				
		50	100	200	500	1,000
$T_n^{\text{LCV}}$	1	0.17	0.25	0.93	6.15	28.41
	2	0.09	0.27	0.93	6.14	28.83
$T_n^{\text{BLCV}}$	1	0.66	1.11	2.05	9.12	33.79
	2	4.27	5.18	9.98	26.27	71.89

The tests were run with  $B = 1000$  permutations and the times were measured in a 3.5 GHz core

## 4 Real data analysis

### 4.1 Data description

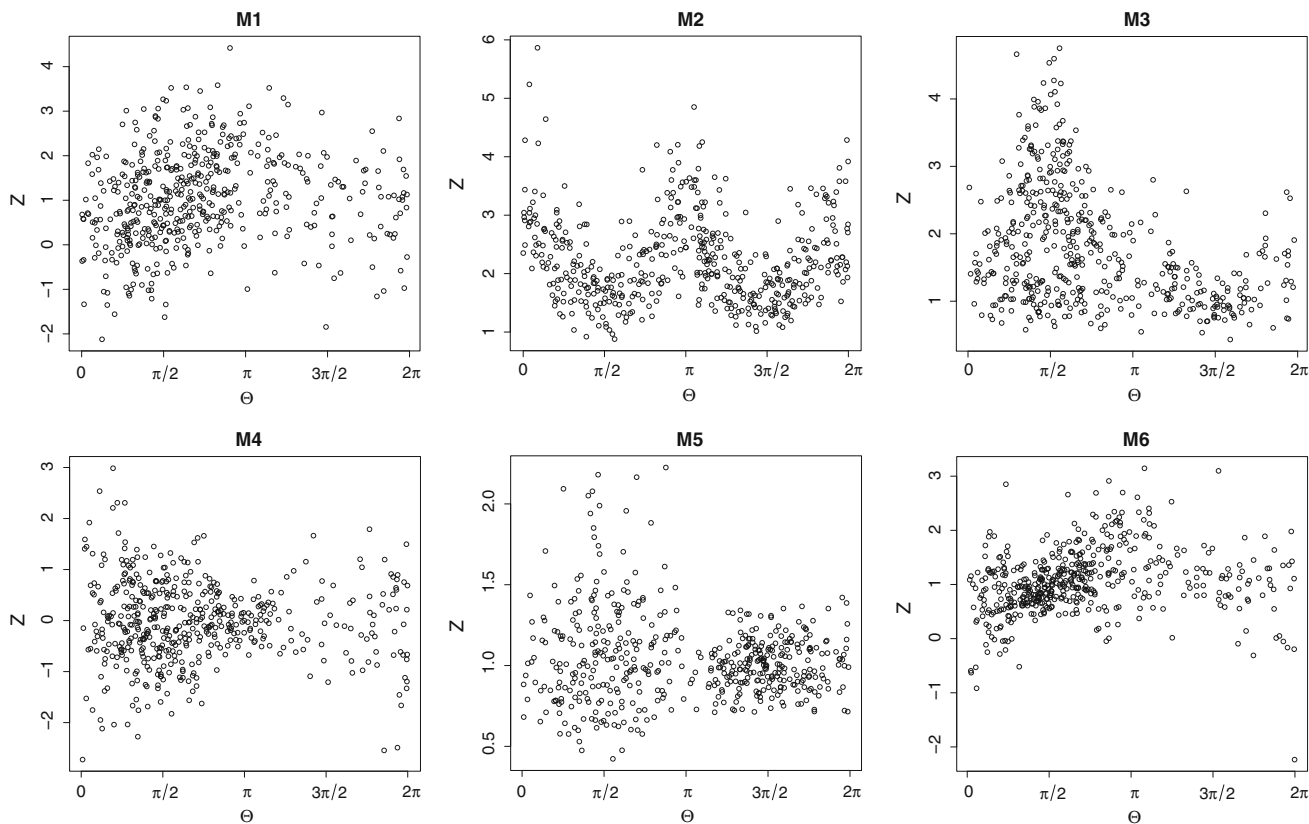
The original Portuguese fire atlas, covering the period from 1975 to 2005, is the longest annual and country-wide cartographic fire database in Europe (Pereira and Santos 2003). Annual wildfire maps were derived from Landsat data, which represents the world's longest and continuously acquired collection of moderate resolution land remote sensing data, providing a unique resource for those who work in forestry, mapping and global change research. For each year in the dataset, Landsat imagery covering Portugal's mainland was acquired after the end of the fire season, thus providing a snapshot of the fires that occurred during the season. Annual fire perimeters were derived through a semi-automatic procedure that starts with supervised image classification, followed by manual editing of classification results. Minimum Mapping Unit (MMU), i.e., the size of the smallest fire mapped, changed according to available data. Between 1975 and 1983 (the MultiSpectral Scanner era), spatial resolution of satellite images is 80 m and MMU of 35 ha. From 1984 onwards with data availability at spatial resolution of 30 m (Thematic Mapper and Enhanced Thematic Mapper era) MMU is 5 ha, allowing to map a larger number of smaller fires than in the 1975–1983 era. Below an MMU of  $\sim 5$  ha the burnt area classification errors increase substantially, and given the very skewed nature of fire size distribution, the 5 ha threshold ensures that over 90% of total area actually burned is mapped. For consistency, and due to discrepancies in minimum mapping unit between 1975–1983 and 1985–2005, in this study only fire perimeters mapped in the latter period were considered, which results in 26,870 fire perimeters.

This application is based on the watershed delineation proposed by Barros et al. (2012). In their work, watersheds were derived from the shuttle radar topography mission (SRTM) digital terrain model (Farr et al. 2007) using the ArcGIS hydrology toolbox (ESRI 2009). Minimum

watershed size was interactively increased so that each watershed contained a minimum of 25 fire observations (see the cited work for more details). Fire perimeters straddling watershed boundaries were allocated to the watershed that contained its centroid.

The orientation of fire perimeters and watersheds was determined by principal component analysis, following the approach proposed by Luo (1998, pp. 131–136). Specifically, principal component analysis was applied to the points that constitute the object's boundary (fire or watershed), with orientation given by the first principal component (PC1). Boundary points can be represented either in bidimensional space defined by each vertex's latitude and longitude coordinates, or in tridimensional space, taking also into account the altitude. Then, the PC1 corresponds to an axis that passes through the center of mass of the object and maximizes the variance of the projected vertices, represented in  $\mathbb{R}^2$  or in  $\mathbb{R}^3$ . The fact of computing the PC1 also in  $\mathbb{R}^3$  aims to take into account the variability of fires according to their slope, which, as the center plot of Fig. 1 shows, presents marked differences between regions. Then, the orientation of the object is taken as the direction given by its PC1.

It is important to notice that an orientation is an axial observation, and that some conversion is needed for applying the directional-linear independence test. In the two-dimensional case, the orientations can be encoded by an angular variable  $\Theta \in [0, \pi)$ , with period  $\pi$ , so  $2\Theta$  is a circular variable. Then, with this codification, the angles  $0, \pi/2, \pi, 3\pi/2$  represent the E/W, NE/SW, N/S and NW/SE orientations, respectively. In the three-dimensional space, the orientation is coded by a pair of angles  $(\Theta, \Phi)$  using spherical coordinates, where  $\Theta \in [0, \pi)$  plays the same role as the previous setting and  $\Phi \in [0, \pi/2]$  measures the inclination ( $\Phi = \frac{\pi}{2}$  for flat slope and  $\Phi = 0$  for vertical; only positive angles are considered as the slope of a certain angle  $\omega$  equals the slope of  $-\omega$ ). Therefore, points with spherical coordinates  $(2\Theta, \Phi)$ , which lie on the upper semisphere, can be regarded as a realization of a spherical variable.



**Fig. 2** Random samples of  $n = 500$  points for the simulation models in the circular-linear case, with  $\delta = 0.50$  (situation with dependence). From up to down and left to right, models M1 to M6. Models M1, M2 and M3 present a deviation from the independence in terms of the

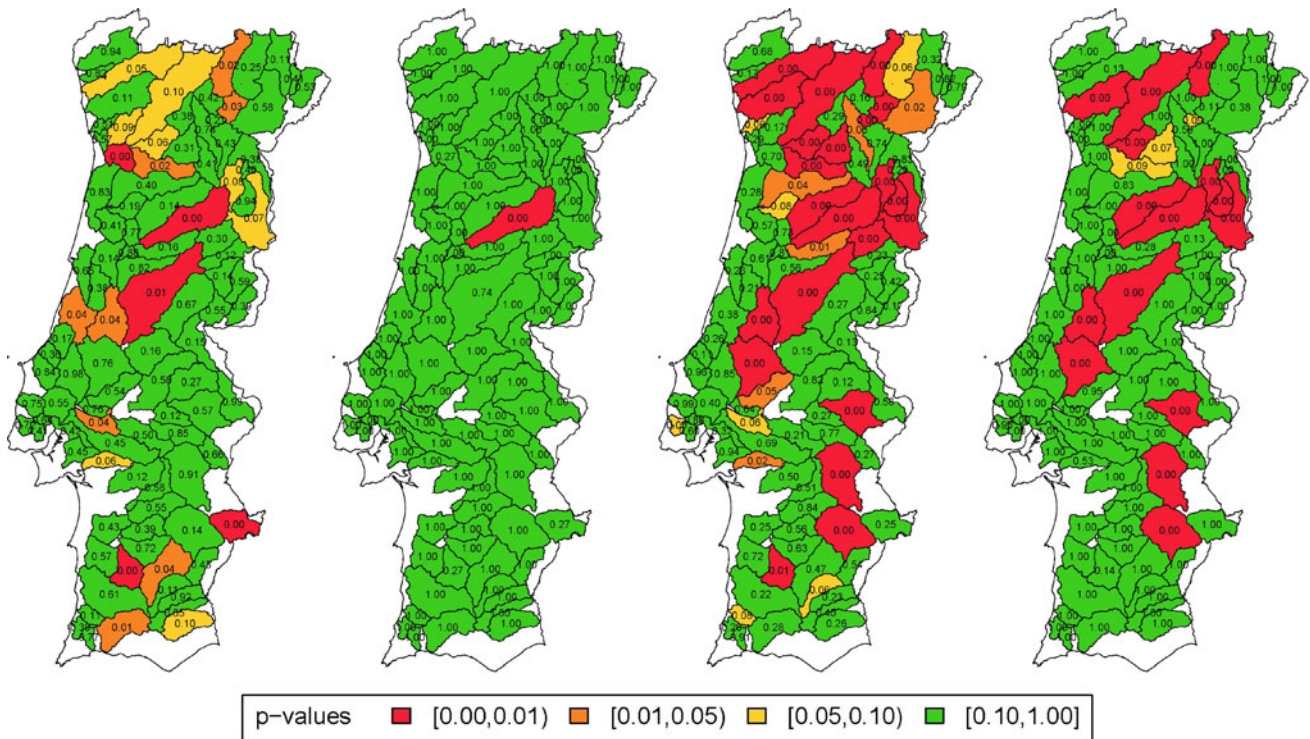
conditional expectation; models M4 and M5 account for a deviation in terms of the conditional variance and model M6 includes deviations both in conditional expectation and variance

### 4.2 Results

The null hypothesis of independence between wildfire orientation and its burnt area (in log scale) is rejected, either using orientations in  $\mathbb{R}^2$  or in  $\mathbb{R}^3$ , with a common  $p$ -value 0.000. The test is carried out using the bandwidth selector BLCV (considered from now on) and all the 26,870 observations for years 1985–2005, ignoring stratification by watershed, and with  $B = 1000$  permutations. The  $p$ -values for the null hypothesis of independence between the orientation of a watershed and the total burnt area of fires within the region are 0.008 and 0.000 for orientations in  $\mathbb{R}^2$  and in  $\mathbb{R}^3$ , respectively. Therefore, the null hypothesis is emphatically rejected.

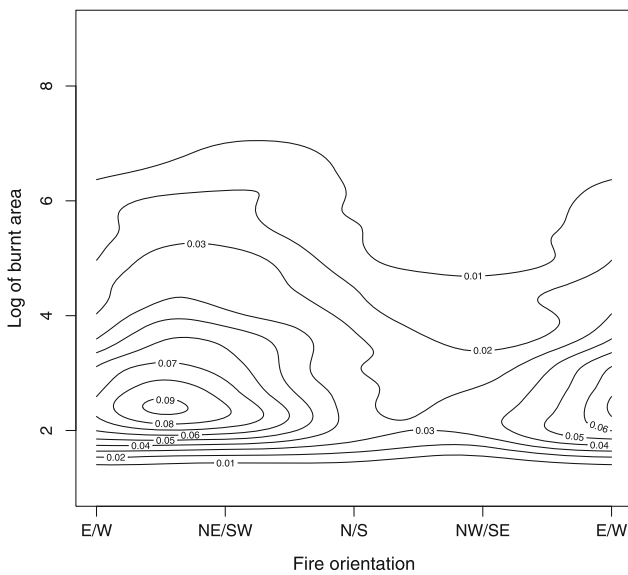
After identifying the presence of dependence between wildfire orientation and size, it is possible to carry out a watershed-based spatial analysis by applying the test to each watershed, in order to detect if the presence of dependence is homogeneous, or if it is only related to some particular areas. Figures 2 and 3 represents maps of  $p$ -values of the test applied to the observations of each watershed, using PC1 in  $\mathbb{R}^2$  and in  $\mathbb{R}^3$  (from left to right,

first and third plots of Fig. 3, respectively). The maps reveal the presence of 13 and 27 watersheds where the null hypothesis of independence is rejected with significance level  $\alpha = 0.05$ , for the circular-linear and the spherical-linear cases, respectively. This shows that the presence of dependence between fire orientation and size is not homogeneous and it is located in specific watersheds (see Fig. 4). It is also interesting to note that the inclusion of the altitude coordinate in the computation of the PC1 leads to a richer detection of dependence between the wildfire orientation and size at the watershed level. This is due to the negative relation between the fire slope and size (see Fig. 4), as large fires tend to have a flatter PC1 in  $\mathbb{R}^3$  because they occur over highly variable terrain. Finally, the resulting  $p$ -values from the watershed analysis can also be adjusted using the false discovery rate (FDR) procedure of Benjamini and Yekutieli (2001) (from left to right, second and fourth plots of Fig. 3). It is also possible to combine the  $p$ -values of the unadjusted maps with the FDR to test for independence between the wildfire orientation and the log-burnt area. The resulting  $p$ -values are 0.000 for the circular-linear and spherical-linear cases.

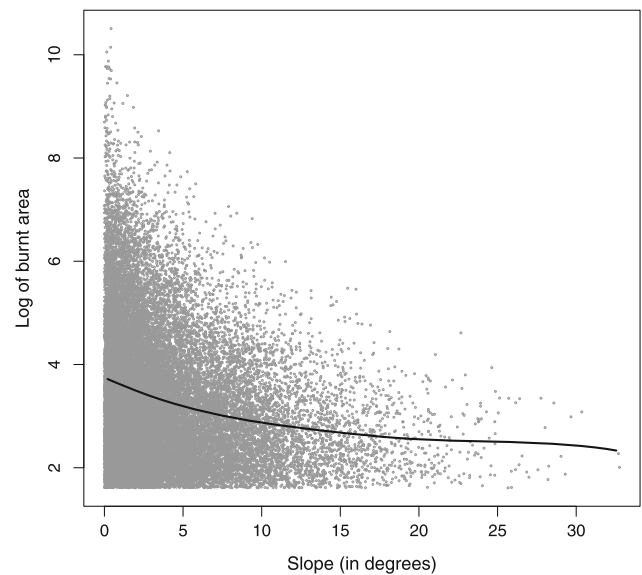


**Fig. 3**  $p$ -values from the independence test for the first principal component PC1 of the fire perimeter and the burnt area (on a log scale), by watersheds. From left to right, the first and second maps represent the circular-linear  $p$ -values (PC1 in  $\mathbb{R}^2$ ) and their corrected

versions using the FDR, respectively. The third and fourth maps represent the spherical-linear situation (PC1 in  $\mathbb{R}^3$ ), with uncorrected and corrected  $p$ -values by FDR, respectively



**Fig. 4** Left: density contour plot for fires in watershed number 31, the watershed in the second plot on the left of Fig. 3 with  $p$ -value = 0.000. The number of fires in the watershed is  $n = 1543$ . The contour plot shows that the size of the area burnt is related with



the orientation of the fires in the watershed. Right scatter plot of the fires slope and the burnt area for the whole dataset, with a nonparametric kernel regression curve. The plot evidences the negative correlation between fire slope and size

### 5 Discussion

A nonparametric test for assessing independence between a directional and a linear component has been proposed, and its finite sample performance has been investigated in a simulation study. Simulation results support a satisfactory behavior of the permutation test implemented with LCV and BLCV bandwidths, in most cases outperforming the available circular-linear testing proposals, and being competitive in other cases. The proposed BLCV bandwidths presents better results in terms of empirical size, although further study is required in bandwidth selection. In addition, when the null hypothesis of independence is rejected, the kernel density estimate can be used to explore the form of dependence, at least for the circular-linear and spherical-linear cases.

The application of the test to the entire wildfire orientation and size dataset makes possible the detection of dependence between these two variables, for both two-dimensional or three-dimensional orientation. The same conclusion holds for watershed orientation and total area burnt. A detailed study of each watershed allows for a more specific insight into the problem. The evidence of independence between fire size and fire orientation in some watersheds, suggests that an event-based analysis (such as the work of Barros et al. (2012)) should yield results similar to those that would be expected from an area-based analysis. On the other hand, detection of dependence between fire size and orientation in watersheds with uniform orientation (Barros et al. 2012) highlights cases where there may be a mixture of orientations. In such cases, an analysis taking fire size into account might find evidence of preferential orientation in fire perimeters. In watersheds where fire events show preferential orientation (non-uniform distribution) and there is dependence between size and orientation, fire orientation distributions are structured in relation to fire size, especially considering the typically asymmetric nature of fire size distributions, dominated by a small number of very large events (Strauss et al. 1989). In these cases, an area-weighted analysis of fire perimeter orientation might lead to different results than those found by Barros et al. (2012). When altitude is included in calculation of the PC1 in  $\mathbb{R}^3$ , it highlights the negative relation between fire slope and size, which is mostly due to the fact that larger fires present flatter PC1. Slope has a skewed distribution, with low mean value and a relatively long right tail. Thus, while small fires usually occur on high slopes, large fires on consistently steep areas are unlikely.

Finally, it can be argued that the data are probably not independent and identically distributed over space and time. Unfortunately, given the data gathering procedure

(detailed at the beginning of Section 4) dependence patterns cannot be clearly identified. Accounting for temporal or spatial dependence directly in the directional-linear kernel estimator and in the testing procedure is an open problem.

**Acknowledgements** The authors acknowledge the support of Project MTM2008-03010, from the Spanish Ministry of Science and Innovation, Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia and IAP network StUDyS, from Belgian Science Policy. Work of E. García-Portugués has been supported by FPU grant AP2010-0957 from the Spanish Ministry of Education and work of A. M. G. Barros, by Ph.D. Grant SFRH/BD/40398/2007 from the Fundação para a Ciência e Tecnologia. J. M. C. Pereira participated in this research under the framework of research projects “Forest fire under climate, social and economic changes in Europe, the Mediterranean and other fire-affected areas of the world (FUME)”, EC FP7 Grant Agreement No. 243888 and “Fire-Land-Atmosphere Inter-Relationships: understanding processes to predict wildfire regimes in Portugal” (FLAIR), PTDC/AAC/AMB/104702/2008. Authors gratefully acknowledge an anonymous referee for the suggestion of employing permutations for the test calibration and the careful revision of the paper. Authors also acknowledge the suggestions raised by another referee.

### Appendix

#### A Proof of Lemma 1

*Proof* The closed expression (just involving matrix computations) for  $T_n$  is obtained by splitting the calculation into three addends:

$$\begin{aligned}
 T_n &= \int_{\Omega_q \times \mathbb{R}} \left( \hat{f}_{(\mathbf{x}, Z); h, g}(\mathbf{x}, z) - \hat{f}_{\mathbf{x}; h}(\mathbf{x}) \hat{f}_{Z; g}(z) \right)^2 \omega_q(d\mathbf{x}) dz \\
 &= \int_{\Omega_q \times \mathbb{R}} \left( \frac{c_{h, q}(L)}{ng} \sum_{i=1}^n L \left( \frac{1 - \mathbf{x}' \mathbf{X}_i}{h^2} \right) K \left( \frac{z - Z_i}{g} \right) \right. \\
 &\quad \left. - \hat{f}_{\mathbf{x}; h}(\mathbf{x}) \hat{f}_{Z; g}(z) \right)^2 \omega_q(d\mathbf{x}) dz \\
 &= \sum_{i=1}^n \sum_{j=1}^n \int_{\Omega_q \times \mathbb{R}} \frac{c_{h, q}(L)^2}{n^2 g^2} L \left( \frac{1 - \mathbf{x}' \mathbf{X}_i}{h^2} \right) K \left( \frac{z - Z_i}{g} \right) \\
 &\quad \cdot L \left( \frac{1 - \mathbf{x}' \mathbf{X}_j}{h^2} \right) K \left( \frac{z - Z_j}{g} \right) \omega_q(d\mathbf{x}) dz \\
 &\quad - 2 \sum_{i=1}^n \int_{\Omega_q \times \mathbb{R}} \frac{c_{h, q}(L)}{ng} L \left( \frac{1 - \mathbf{x}' \mathbf{X}_i}{h^2} \right) K \left( \frac{z - Z_i}{g} \right) \\
 &\quad \cdot \hat{f}_{\mathbf{x}; h}(\mathbf{x}) \hat{f}_{Z; g}(z) \omega_q(d\mathbf{x}) dz \\
 &\quad + \int_{\Omega_q \times \mathbb{R}} \hat{f}_{\mathbf{x}; h}(\mathbf{x})^2 \hat{f}_{Z; g}(z)^2 \omega_q(d\mathbf{x}) dz \\
 &= (6) - (7) + (8).
 \end{aligned}$$

The first addend is:

$$\begin{aligned}
 (6) &= \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\Omega_q \times \mathbb{R}} L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right) K\left(\frac{z - Z_i}{g}\right) \\
 &\quad \cdot L\left(\frac{1 - \mathbf{x}'\mathbf{X}_j}{h^2}\right) K\left(\frac{z - Z_j}{g}\right) \omega_q(d\mathbf{x}) dz \\
 &= \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\Omega_q} e^{-2/h^2} e^{\mathbf{x}'(\mathbf{X}_i + \mathbf{X}_j)/h^2} \omega_q(d\mathbf{x}) \\
 &\quad \cdot \int_{\mathbb{R}} K\left(\frac{z - Z_i}{g}\right) K\left(\frac{z - Z_j}{g}\right) dz \\
 &= \frac{c_{h,q}(L)^2}{n^2} e^{-2/h^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\phi_{\sqrt{2g}}(Z_i - Z_j)}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)} \\
 &= \frac{C_q(1/h^2)^2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\phi_{\sqrt{2g}}(Z_i - Z_j)}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)}.
 \end{aligned}$$

For the second addend,

$$\begin{aligned}
 (7) &= 2 \sum_{i=1}^n \int_{\Omega_q \times \mathbb{R}} \frac{c_{h,q}(L)}{ng} L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right) K\left(\frac{z - Z_i}{g}\right) \\
 &\quad \cdot \hat{f}_{\mathbf{X};h}(\mathbf{x}) \hat{f}_{Z;g}(z) \omega_q(d\mathbf{x}) dz \\
 &= 2 \frac{c_{h,q}(L)}{ng} \sum_{i=1}^n \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right) \hat{f}_{\mathbf{X};h}(\mathbf{x}) \omega_q(d\mathbf{x}) \\
 &\quad \cdot \int_{\mathbb{R}} K\left(\frac{z - Z_i}{g}\right) \hat{f}_{Z;g}(z) dz \\
 &= 2 \frac{c_{h,q}(L)}{ng} \sum_{i=1}^n \left\{ \left[ \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right) \right. \right. \\
 &\quad \cdot \left. \frac{c_{h,q}(L)}{n} \sum_{j=1}^n L\left(\frac{1 - \mathbf{x}'\mathbf{X}_j}{h^2}\right) \omega_q(d\mathbf{x}) \right] \\
 &\quad \cdot \left. \left[ \int_{\mathbb{R}} K\left(\frac{z - Z_i}{g}\right) \frac{1}{ng} \sum_{k=1}^n K\left(\frac{z - Z_k}{g}\right) dz \right] \right\} \\
 &= \frac{2}{n^3} \sum_{i=1}^n \left\{ \left[ \int_{\Omega_q} c_{h,q}(L)^2 e^{-2/h^2} e^{\mathbf{x}'(\mathbf{X}_i + \mathbf{X}_j)/h^2} \omega_q(d\mathbf{x}) \right. \right. \\
 &\quad \cdot \left. \left. \left( \sum_{k=1}^n \phi_{\sqrt{2g}}(Z_i - Z_k) \right) \right\} \\
 &= \frac{2}{n^3} \sum_{i=1}^n \left\{ \left[ \sum_{j=1}^n \frac{C_q(1/h^2)^2}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)} \right] \right. \\
 &\quad \cdot \left. \left( \sum_{k=1}^n \phi_{\sqrt{2g}}(Z_i - Z_k) \right) \right\}.
 \end{aligned}$$

Finally, the third addend is obtained as

$$\begin{aligned}
 (8) &= \int_{\Omega_q \times \mathbb{R}} \hat{f}_{\mathbf{X};h}(\mathbf{x})^2 \hat{f}_{Z;g}(z)^2 \omega_q(d\mathbf{x}) dz \\
 &= \int_{\Omega_q} \hat{f}_{\mathbf{X};h}(\mathbf{x})^2 \mathbf{x} \int_{\mathbb{R}} \hat{f}_{Z;g}(z)^2 dz \\
 &= \left[ \sum_{i=1}^n \sum_{j=1}^n \frac{c_{h,q}(L)^2}{n^2} \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}'\mathbf{X}_i}{h^2}\right) \right. \\
 &\quad \cdot L\left(\frac{1 - \mathbf{x}'\mathbf{X}_j}{h^2}\right) \omega_q(d\mathbf{x}) \left. \right] \\
 &\quad \cdot \left[ \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^2 g^2} \int_{\mathbb{R}} K\left(\frac{z - Z_i}{g}\right) K\left(\frac{z - Z_j}{g}\right) dz \right] \\
 &= \left[ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{C_q(1/h^2)^2}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)} \right] \\
 &\quad \cdot \left[ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \phi_{\sqrt{2g}}(Z_i - Z_j) \right].
 \end{aligned}$$

From the previous results and after applying some matrix algebra, it turns out that

$$\begin{aligned}
 T_n &= \mathbf{1}_n \left( \frac{1}{n^2} \Psi(h) \circ \Omega(g) - \frac{2}{n^3} \Psi(h) \Omega(g) \right. \\
 &\quad \left. + \frac{1}{n^4} \Psi(h) \mathbf{1}_n \mathbf{1}_n' \Omega(g) \right) \mathbf{1}_n',
 \end{aligned}$$

where

$$\begin{aligned}
 \Psi(h) &= \left( \frac{C_q(1/h^2)^2}{C_q(\|\mathbf{X}_i + \mathbf{X}_j\|/h^2)} \right)_{ij}, \\
 \Omega(g) &= \left( \phi_{\sqrt{2g}}(Z_i - Z_j) \right)_{ij}.
 \end{aligned}$$

### References

Ahmad IA, Li Q (1997) Testing independence by nonparametric kernel method. *Stat Probab Lett* 34(2):201–210

Bai ZD, Rao CR, Zhao LC (1988) Kernel estimators of density function of directional data. *J Multivar Anal* 27(1):24–39

Barros AMG, Pereira JMC, Lund UJ (2012) Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *For. Ecol Manage* 264:98–107

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188

Cao R (1993) Bootstrapping the mean integrated squared error. *J Multivar Anal* 45(1):137–160

Cao R, Cuevas A, Gonzalez Manteiga W (1994) A comparative study of several smoothing methods in density estimation. *Comput Stat Data Anal* 17(2):153–176

Chiu ST (1996) A comparative review of bandwidth selection for kernel density estimation. *Stat Sin* 6(1):129–145

- Costa L, Thonicke K, Poulter B, Badeck FW (2011) Sensitivity of Portuguese forest fires to climatic, human, and landscape variables: subnational differences between fire drivers in extreme fire years and decadal averages. *Reg Environ Change* 11(3):543–551
- ESRI (2009) ArcMap 9.21. Environmental Systems Research Institute, Redlands
- Farr T, Rosen P, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Douglas A (2007) The shuttle radar topography mission. *Rev Geophys* 45(2). doi:10.1029/2005RG000183
- Fernández-Durán JJ (2007) Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics* 63(2):579–585
- Finney M (2001) Design of regular landscape fuel treatment patterns for modifying fire growth and behavior. *For Sci* 47(2):219–228
- Fisher NI, Lee AJ (1981) Nonparametric measures of angular-linear association. *Biometrika* 68(3):629–636
- García-Portugués E (2013) Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron J Stat* 7:1655–1685
- García-Portugués E, Crujeiras R, González-Manteiga W (2012) Exploring wind direction and SO<sub>2</sub> concentration by circular-linear density estimation. *Stoch Environ Res Risk Assess* 27(5):1055–1067.
- García-Portugués E, Crujeiras R, González-Manteiga W (2013a) Kernel density estimation for directional-linear data. *J Multivar Anal* 121:152–175
- García-Portugués E, Crujeiras R, González-Manteiga W (2013b) Central limit theorems for directional-linear data with applications (preprint)
- Hall P, Watson GS, Cabrera J (1987) Kernel density estimation with spherical data. *Biometrika* 74(4):751–762
- Johnson RA, Wehrly T (1977) Measures and models for angular correlation and angular-linear correlation. *J R Stat Soc B* 39(2):222–229
- Johnson RA, Wehrly TE (1978) Some angular-linear distributions and related regression models. *J Am Stat Assoc* 73(363):602–606
- Jones C, Marron J, Sheather S (1996) Progress in data-based bandwidth selection for kernel density estimation. *Comput Stat* 11:337–381
- Liddell IG, Ord JK (1978) Linear-circular correlation coefficients: some further results. *Biometrika* 65(2):448–450
- Lloret F, Calvo E, Pons X, Diaz-Delgado R (2002) Wildfires and landscape patterns in the eastern Iberian peninsula. *Landsc Ecol* 17:745–759
- Luo D (1998) Pattern recognition and image processing. Horwood, Chichester
- Mardia K (1976) Linear-circular correlation coefficients and rhythmometry. *Biometrika* 63(2):403–405
- Moreira F, Rego F, Ferreira P (2001) Temporal (1985–1995) pattern of change in a cultural landscape of northwestern Portugal: implications for fire occurrence. *Landsc Ecol* 16:557–567
- Moreira F, Viedma O, Arianoutsou M, Curt T, Koutsias N, Rigolot E, Barbati A, Corona P, Vaz P, Xanthopoulos G, Mouillot F, Bilgili E (2011) Landscape-wildfire interactions in southern Europe: implications for landscape management. *J Environ Manage* 92(10):2389–2402
- Parzen E (1962) On estimation of a probability density function and mode. *Ann Math Stat* 33:1065–1076
- Pereira J, Santos T (2003) Fire risk and burned area mapping in Portugal. Technical report, Direção-Geral das Florestas, Lisboa
- Pereira M, Trigo R, da Camara C, Pereira J, Leite S (2005) Synoptic patterns associated with large summer forest fires in Portugal. *Agric For Meteorol* 129(1-2):11–25
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *Ann Math Stat* 27:832–837
- Rosenblatt M (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann Stat* 3:1–14
- Rosenblatt M, Wahlen BE (1992) A nonparametric measure of independence under a hypothesis of independent components. *Stat Probab Lett* 15(3):245–252
- Schmidt DA, Taylor AH, Skinner CN (2008) The influence of fuels treatment and landscape arrangement on simulated fire behavior, Southern Cascade Range, California. *For Ecol Manag* 255(8–9):3170–3184
- Silva JS, Vaz P, Moreira F, Catry F, Rego FC (2011) Wildfires as a major driver of landscape dynamics in three fire-prone areas of Portugal. *Landsc Urban Plan* 101(4):349–358
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, London
- Strauss D, Bednar L, Mees R (1989) Do one percent of forest fires cause ninety-nine percent of the damage? *For Sci* 35(2):319–328
- Wand MP, Jones MC (1995) Kernel smoothing, Chapman & Hall, London
- Watson GS (1983) Statistics on spheres. University of Arkansas lecture notes in the mathematical sciences, vol 6. Wiley, New York