Received: 22 June 2020 | Revised: 1 January 2021 | Accepted: 8 January 2021

DOI: 10.1111/biom.13431

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Density estimation for circular data observed with errors

Marco Di Marzio[1] | **Stefania Fensore**[1] | **Agnese Panzera**[2] | **Charles C. Taylor**[3]

[1] DMQTE, Università di Chieti-Pescara, Viale Pindaro 42, Pescara, Italy

[2] DiSIA, Università di Firenze, Viale Morgagni 59, Firenze, Italy

[3] Department of Statistics, University of Leeds, Leeds, UK

Correspondence
Charles C. Taylor, Department of Statistics, University of Leeds, Leeds LS2 9JT, UK.
Email: charles@maths.leeds.ac.uk

Abstract

Until now the problem of estimating circular densities when data are observed with errors has been mainly treated by Fourier series methods. We propose kernel-based estimators exhibiting simple construction and easy implementation. Specifically, we consider three different approaches: the first one is based on the equivalence between kernel estimators using data corrupted with different *levels* of error. This proposal appears to be totally unexplored, despite its potential for application also in the Euclidean setting. The second approach relies on estimators whose weight functions are circular deconvolution kernels. Due to the periodicity of the involved densities, it requires ad hoc mathematical tools. Finally, the third one is based on the idea of *correcting* extra bias of kernel estimators which use contaminated data and is essentially an adaptation of the standard theory to the circular case. For all the proposed estimators, we derive asymptotic properties, provide some simulation results, and also discuss some possible generalizations and extensions. Real data case studies are also included.

KEYWORDS

circular kernels, deconvolution, equivalence, Fourier coefficients, measurement errors, movements of ants, smoothing, surface wind directions

1 | INTRODUCTION

Circular data arise when the sample space is described by a unit circle. If compared to a linear scale, the main features of circular observations are that the beginning and the end of the measurement scale coincide, and their common location, which determines the origin (or zero direction), is arbitrarily chosen. Once the origin and the direction of rotation have been fixed, any circular observation can be measured by an angle ranging, in radians, from 0 to 2π. Circular data often arise in biology, meteorology, and geology; other examples include phenomena that are periodic in time. For comprehensive accounts of *circular statistics* see, for example, Fisher (1993) and Jammalamadaka and SenGupta (2001), and for collections of recent advances see Ley and Verdebout (2017, 2018).

We discuss the problem of nonparametrically estimating a circular density when, instead of observing a random sample from that density, a version contaminated by measurement errors is available. This is the classical errors-in-variables problem. Differently from the Euclidean setting, where kernel-type estimators have been widely employed for this problem (see, e.g., Delaigle (2014) and the references therein), in the circular setting only trigonometric series estimators have been developed. In particular, Efromovich (1997) proposed an estimator constructed by approximating the target density as a truncated series where the theoretical coefficients of the trigonometric basis are replaced by the empirical ones. Then Comte and Taupin (2003), using a model selection procedure, derived an adaptive penalized contrast estimator, and Johannes and Schwarz (2013) proposed

an orthogonal series estimator which is optimal in the minimax sense.

In this paper, we introduce estimators which have the advantage of being defined in terms of simple averages, and thus favoring intuition, flexibility, and saving computational time. Specifically, we pursue three routes.

The first one originates from the quite general idea that the two following links have the same nature: (a) the link between an estimate based on unavailable data and that which is based on current sample and (b) the link between this latter and an estimate based on the sample artificially corrupted by adding noise drawn from the error distribution. This idea provides the basis to formulate equations where the uncorrupted estimate is the unknown, leading to estimators which are corrected by means of a difference or a ratio. Importantly, due to its generality, we note that this equivalence idea applies *in principle* to all estimation methods, both for regression and density estimation, regardless of whether data are directional or not. In fact, the only constraint seems to be knowledge of the error distribution.

The second one relies on estimators which share the structure of classical density deconvolution estimators, whose weight functions are defined as Fourier series whose coefficients are represented by the ratio of the Fourier coefficients of a circular kernel and those of the error density. The fact that an infinite summation is involved in the Fourier expansion also poses the challenge of selecting the number of terms to obtain a truncated version of it.

The third approach is based on the idea of *correcting* the extra bias due to the measurement error of the naive kernel estimator which uses contaminated data. In particular, using the idea of low-order approximations of Carroll and Hall (2004), starting from a Taylor-like series expansion, the estimator is obtained as the difference between the naive kernel density estimator and a consistent estimator of the excess bias due to the measurement errors. A possible generalization is to consider different smoothing degrees for the two terms appearing in its formulation.

To motivate our research, we note that determining the distribution of wind or marine current directions constitutes a very relevant field of application for our proposed methods because direction data are typically affected by various sources of noise. In particular, *surface* wind direction data are the object of different fields of study. The main features of such data are the instantaneous nature and an inherent, strong variability even in very small periods of time. Surely, there is a widespread interest in establishing *prevailing winds*, defined as the dominant wind directions in an area. Classical problems involving prevailing winds analysis are forecasting wildfire directions, determining seasonal wind direction variations, or optimizing wind turbine locations. Also, in aviation, a *crosswind landing* is a typical maneuver in which a significant component of the wind is perpendicular to the runway axis. Surface winds can be obviously conceived as prevailing winds perturbed by random noise which may be due to changes of wind speed and other meteorological conditions. Additionally, land-based surface wind measurements without exposure problems hardly exist. The requirement of open, level terrain is difficult to meet, and most wind stations over land are perturbed by topographic effects or surface cover, or by both. Finally, instruments are typically prone to measurement error, including deterioration and miscalibration.

A practical way used to obtain the prevailing wind direction lies in averaging the observations belonging to more or less prolonged time intervals, and then to depict the distribution of these averages by a rose diagram. An alternative to this somewhat arbitrary approach is to deconvolve wind data after appropriately modeling the error distribution.

Typical targets are the average direction or the most common ones. More robust indicators, based on the cumulative distribution function, like probability of intervals centered on the mode, are often required.

The paper is organized as follows. Section 2 collects some preliminaries about Fourier series representation of circular densities, and Section 3 recalls some theory about kernel estimation of circular densities in the errors-free case. In Section 4, we discuss the errors-in-variables problem, and we study the proposed approaches for kernel estimation of a circular density when data are observed with error, providing some asymptotic properties. Then, in Section 5 we present some simulation results, and in Section 6 we report two illustrative examples using real datasets, one on ant directions, and the other on wind directions. In Section 7, we end with some conclusions.

2 | SOME PRELIMINARIES

Denote as Q and f_Q a circular random variable and its probability density function, respectively. Due to the circular domain, f_Q is 2π-periodic, that is $f_Q(\theta) = f_Q(\theta + 2m\pi)$ for any integer m; then its characteristic function $\varphi_Q(\ell) = \mathsf{E}[e^{i\ell Q}]$ is just defined for integer ℓ and satisfies $\varphi_Q(\ell) = \varphi_{Q+2\pi}(\ell), \ell \in \mathbb{Z}$, with $|\varphi_Q(\ell)| \leq 1$, and $\varphi_Q(0) = 1$. Notice that the complex numbers $\{\varphi_Q(\ell), \ell \in \mathbb{Z}\}$ are the coefficients in the Fourier series representation of f_Q and correspond to the *trigonometric moments* of Q about the mean direction, that is

$$\varphi_Q(\ell) = \alpha_\ell + i\beta_\ell, \quad \alpha_\ell = \mathsf{E}[\cos(\ell Q)], \quad \beta_\ell = \mathsf{E}[\sin(\ell Q)].$$

Clearly, for any $\ell \in \mathbb{Z}, \alpha_{-\ell} = \alpha_\ell, \beta_{-\ell} = -\beta_\ell, |\alpha_\ell| \leq 1$, and $|\beta_\ell| \leq 1$.

Assuming that f_Q is square integrable on $[0, 2\pi)$, analogously to the inversion formula for characteristic functions of real-valued random variable, one can represent $f_Q(q)$, $q \in [0, 2\pi)$, through the Fourier series

$$f_Q(q) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \varphi_Q(\ell) \exp(-i\ell q) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{\ell=1}^{\infty} (\alpha_\ell \cos(\ell q) + \beta_\ell \sin(\ell q)) \right\}. \quad (1)$$

When $Q = X(\mathrm{mod} 2\pi)$, where X is a real valued random variable with probability density function f_X, then Q has probability density function

$$f_Q(q) = \sum_{k=-\infty}^{\infty} f_X(q + 2\pi k)$$

and its distribution is said to be the *wrapped* version of the distribution of X. The trigonometric moment of order ℓ of the resulting wrapped distribution is equal to the value of the characteristic function of X, say φ_X, at (integer) ℓ, that is $\varphi_Q(\ell) = \varphi_X(\ell)$.

The smoothness of f_Q, which is usually measured by the number of continuous derivatives it has over some domain, can be defined according to the rate of decay to zero of the coefficients in its Fourier representation. Formally, following Efromovich (1997), f_Q is said to be *supersmooth* if $\varphi_Q(\ell)$, $\ell \in \mathbb{Z}$, has exponential decay, that is

$$c_0(|\ell| + 1)^{a_0} e^{-b|\ell|^a} \leq |\varphi_Q(\ell)| \leq c_1(|\ell| + 1)^{a_1} e^{-b|\ell|^a},$$

while it is *ordinary smooth* if $\varphi_Q(\ell)$ exhibits polynomial decay, that is

$$c_0(|\ell| + 1)^{-a_0} \leq |\varphi_Q(\ell)| \leq c_1(|\ell| + 1)^{-a_1},$$

where a, b, c_0, c_1 are constants in $\mathbb{R}^+$ and a_0, a_1 are both in $\mathbb{R}$.

Examples of supersmooth densities include the densities of wrapped Normal, wrapped Cauchy, and von Mises distributions; conversely, the wrapped Laplace and the wrapped Gamma densities are examples of ordinary smooth ones.

3 | CIRCULAR DENSITY ESTIMATION IN THE ERRORS-FREE CASE

Given a random sample of angles $\Theta_1, \ldots, \Theta_n$ from an unknown circular density f_Θ, the kernel estimator of f_Θ at $\theta \in [0, 2\pi)$ is given by

$$\hat{f}_\Theta(\theta; \kappa) = \frac{1}{n} \sum_{i=1}^{n} K_\kappa(\Theta_i - \theta), \quad (2)$$

where K_κ is a circular kernel, that is a periodic, unimodal, symmetric density function with the concentration parameter $\kappa > 0$, which admits the convergent Fourier series representation

$$K_\kappa(\theta) = \frac{1 + 2\sum_{\ell=1}^{\infty} \gamma_\ell(\kappa) \cos(\ell\theta)}{2\pi}.$$

Notice that, with respect to the general Fourier series representation as formulated in (1), due to the symmetry, the Fourier coefficients of K_κ satisfy $\beta_\ell = 0$ and $\alpha_\ell = \gamma_\ell(\kappa)$ for any ℓ. As it happens in the linear setting, the role of the kernel function is to emphasize, in the estimation process, the contribution of the observations which are in a neighborhood of the estimation point. Here, the concentration parameter κ controls the width of that neighborhood playing the inverse role of the bandwidth in the linear case, in the sense that smaller values of κ give wider neighborhoods.

In the following sections, estimator (2) will be denoted as kernel density estimator (KDE).

Letting $\eta_j(K_\kappa) = \int_{-\pi}^{\pi} K_\kappa(u) \sin^j(u) du$, we say that K_κ is a rth sin-order kernel if $\eta_j(K_\kappa) = 0$ for $0 < j < r$ and $\eta_r(K_\kappa) \neq 0$. Classical examples of second sin-order kernels include the von Mises density with $\gamma_\ell(k) = \mathcal{I}_\ell(k)/\mathcal{I}_0(k)$, where $\mathcal{I}_\ell(k)$ is the modified Bessel function of the first kind and order ℓ; the wrapped Normal and wrapped Cauchy densities with $\gamma_\ell(k) = k^{\ell^2}$ and $\gamma_\ell(k) = k^\ell$, respectively.

The asymptotic properties of $\hat{f}_\Theta(\theta; \kappa)$, as obtained by Di Marzio *et al.* (2009), are collected in the following.

Result 1. Given the random sample $\Theta_1, \ldots, \Theta_n$ from f_Θ, consider estimator $\hat{f}_\Theta(\theta; \kappa)$, $\theta \in [0, 2\pi)$, with a second sin-order kernel K_κ. If

(i) f_Θ is twice continuously differentiable in a neighborhood of θ,

(ii) κ increases with n in such a way that, for $\ell \in \mathbb{Z}^+$,

$$\lim_{\kappa \to \infty} \frac{1 - \gamma_\ell(\kappa)}{1 - \gamma_2(\kappa)} = \frac{\ell^2}{4},$$

(iii) κ increases with n in such a way that, for $\ell \in \mathbb{Z}^+$,

$$\lim_{n \to \infty} \gamma_\ell(\kappa) = 1 \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^{\infty} \gamma_\ell^2(\kappa) = 0,$$

then

$$\mathrm{E}[\hat{f}_\Theta(\theta; \kappa)] - f_\Theta(\theta) = \frac{(1 - \gamma_2(\kappa))}{4} f_\Theta^{(2)}(\theta) + o(1 - \gamma_2(\kappa)),$$

and

$$\mathrm{Var}[\hat{f}_\Theta(\theta; \kappa)] = \frac{\left(1 + 2\sum_{\ell=1}^{\infty} \gamma_\ell^2(\kappa)\right)}{2\pi n} f_\Theta(\theta) + o\left(\frac{\sum_{\ell=1}^{\infty} \gamma_\ell^2(\kappa)}{n}\right).$$

4 | KERNEL DENSITY ESTIMATION IN THE ERRORS-IN-VARIABLES CASE

Now, we consider the problem of estimating the density of a circular random variable Θ, say f_Θ, when data are *contaminated* by measurement errors, that is we have n realizations $\Phi_1, \dots, \Phi_n$ of the random variable

$$\Phi = (\Theta + \varepsilon) \mathrm{mod}(2\pi), \tag{3}$$

where ε is a random angle independent of Θ, with density f_ε assumed to be known and symmetric around zero. We also assume that f_Θ, f_ε and the density f_Φ of Φ are square integrable densities on $[0, 2\pi)$ such that all of them admit an absolutely convergent Fourier series representation.

In the Euclidean setting, some variations of the above model have been studied. The case where ε is not independent of Θ, named Berkson errors case, has been considered, for example, in Delaigle (2007). A further model with classical measurement errors having heteroscedastic nature has been studied, for example, by Delaigle and Meister (2008). The case of unknown error density has been considered, among others, by Delaigle *et al.* (2008) and Delaigle and Meister (2008).

In the sequel, we discuss three different approaches. The first one relies on the equivalence between estimators with different levels of errors, the second one exploits the fact that f_Φ is a convolution between f_Θ and f_ε, and the third one is based on the estimation of the increase in bias due to the measurement error.

4.1 | Equivalence-based approach

We hypothesize that the link between the estimate based on the Θ_is and the estimate based on the corrupted data Φ_is is the same as the link between this latter and the estimate based on sample data corrupted by an additional (simulated) level of error, that is

$$\hat{f}_\Theta(\theta;\kappa) : \hat{f}_\Phi(\theta;\kappa) = \hat{f}_\Phi(\theta;\kappa) : \hat{f}_\Psi(\theta;\kappa),$$

where

$$\Psi_i = (\Phi_i + \varepsilon_i^*) \bmod (2\pi),$$

with the ε_i^*s being drawn from the error density.

Considering the symbol ":" either as a difference or a ratio, one can, respectively, define estimators like the following ones:

$$\mathrm{EQD}_\kappa(\theta) = 2\hat{f}_\Phi(\theta;\kappa) - \hat{f}_\Psi(\theta;\kappa) \tag{4}$$

$$\mathrm{EQR}_\kappa(\theta) = \frac{(\hat{f}_\Phi(\theta;\kappa))^2}{\hat{f}_\Psi(\theta;\kappa)}. \tag{5}$$

We observe that this method cannot be considered a resampling one because we draw the ε_i^*s from the known f_ε, rather than from a sample of a smoothed version of the data. As in resampling schemes, particularly for small datasets, it will be better to generate $B > 1$ artificial samples and use an average of the estimates $\hat{f}_{\Psi,j}(\theta;\kappa)$, $j = 1, \dots, B$ in order to reduce the effect of random fluctuations.

Concerning the asymptotic properties, we get the following.

Result 2. Given random samples $\Phi_1, \dots, \Phi_n$ and $\Psi_1, \dots, \Psi_n$, under assumptions (i)–(iii) of Result 1, and assuming that the derivatives of f_Θ are continuous up to order 4, and that ε has finite second-order moment and concentrates around 0, for estimator $\mathrm{EQD}_\kappa(\theta)$ we get

$$\mathrm{E}[\mathrm{EQD}_\kappa(\theta)] - f_\Theta(\theta) = \frac{1}{4}\left\{ f_\Theta^{(2)}(\theta)(1-\gamma_2(\kappa)) - \frac{(1-\lambda_2(\kappa_\varepsilon))^2}{4} f_\Theta^{(4)}(\theta) \right\} + o(1-\gamma_2(\kappa)) + o\left(\{1-\lambda_2(\kappa_\varepsilon)\}^2\right).$$

and

$$\mathrm{Var}[\mathrm{EQD}_\kappa(\theta)] = \frac{\left(1 + 2\sum_{\ell=1}^{\infty} \gamma_\ell^2(\kappa)\right)}{2\pi n} f_\Theta(\theta) + o\left(\frac{\sum_{\ell=1}^{\infty} \gamma_\ell^2(\kappa)}{n}\right),$$

where $\gamma_\ell(\kappa)$ and $\lambda_\ell(\kappa_\varepsilon)$ are the ℓth coefficients of the cosine terms in the Fourier series representation of K_κ and f_ε, respectively.

Proof. See Appendix. □

Remark 1. It seems clear that the more the measurement error is concentrated, the more accurate is the estimator, due to a bigger value of $\lambda_2(\kappa_\varepsilon)$. Specifically, $\lim_{\kappa_\varepsilon \to \infty} \lambda_2(\kappa_\varepsilon) = 1$ gives the same properties as the error-free case. This is true for all the estimators presented in the sequel.

Remark 2. Using the following *linearization arguments:*

$$\begin{aligned}
&\mathrm{EQR}_\kappa(\theta) - f_\Theta(\theta) \\
&= \frac{(\hat{f}_\Phi(\theta;\kappa))^2 - f_\Theta(\theta)\hat{f}_\Psi(\theta;\kappa)}{\hat{f}_\Psi(\theta;\kappa)} \\
&= \frac{(\hat{f}_\Phi(\theta;\kappa))^2 - f_\Theta(\theta)\hat{f}_\Psi(\theta;\kappa)}{f_\Theta(\theta)} \left[1 - \frac{\{\hat{f}_\Psi(\theta;\kappa) - f_\Theta(\theta)\}}{\hat{f}_\Psi(\theta;\kappa)}\right],
\end{aligned}$$

observe that, if ε concentrates around 0, by a first-order Taylor-series expansion of $\hat{f}_\Psi$ for Ψ_i around Θ_i, the

second term in squared brackets, being $O_p(1)$, can be dropped. So, using the assumptions in Result 2, starting from the approximation

$$\mathrm{EQR}_\kappa(\theta) - f_\Theta(\theta) \approx \frac{(\hat{f}_\Phi(\theta;\kappa))^2 - f_\Theta(\theta)\hat{f}_\Psi(\theta;\kappa)}{f_\Theta(\theta)},$$

we have that the estimator $\mathrm{EQR}_\kappa(\theta)$ shares the asymptotic properties of the estimator $\mathrm{EQD}_\kappa(\theta)$.

Remark 3. If ε and ε^* are error terms having different distributions, such that $\varepsilon \perp \varepsilon^*$, and $\varepsilon^* \perp \Theta$, using the assumption of Result 2 and assuming that both ε and ε^* have finite second sin-order moments and concentrate around 0, it can be shown that the asymptotic bias of estimator (4) depends on both the levels of error via their second sin-order moments. In particular, denoting as $\delta_2(\kappa_{\varepsilon^*})$ the second Fourier coefficient of the density of ε^*, and reasoning as in the proof of Result 2, with the caveat that the second sin-order moments of ε and ε^*—which are, respectively, given by $\{1-\gamma_2(\kappa_\varepsilon)\}/2$ and $\{1-\delta_2(\kappa_{\varepsilon^*})\}/2$—do not cancel, it can be shown that the asymptotic bias is

$$\mathsf{E}[\mathrm{EQD}_\kappa(\theta)] - f_\Theta(\theta) \sim \frac{f_\Theta^{(2)}(\theta)}{4}\{(1-\gamma_2(\kappa)) + (1-\lambda_2(\kappa_\varepsilon)) - (1-\delta_2(\kappa_{\varepsilon^*}))\}.$$

More general versions of the above estimators can be also defined by using two distinct smoothing parameters for the two terms in the difference and the ratio, respectively, appearing in (4) and (5). For example, for the first case, one can define

$$\mathrm{EQD}_{\kappa_1,\kappa_2}(\theta) = 2\hat{f}_\Phi(\theta;\kappa_1) - \hat{f}_\Psi(\theta;\kappa_2).$$

Remark 4. Curiously, despite its simplicity, we notice that the proposed equivalence approach appears to be unexplored in the Euclidean setting. When linear variables are observed with error, assuming that the error density is known, the same scheme can be used by simply replacing circular kernels by linear ones. It should come as no surprise that the asymptotic properties of such defined Euclidean estimators have identical rates of convergence as those of Result 2.

4.2 | Deconvolution approach

Considering that f_Φ is the *circular convolution* of f_Θ and f_ε, that is, for $\theta \in [0, 2\pi)$,

$$f_\Phi(\theta) = \int_0^{2\pi} f_\Theta(\omega) f_\varepsilon(\theta - \omega) d\omega, \tag{6}$$

the estimation of f_Θ reduces to a circular density deconvolution problem. Due to (6), for $\ell \in \mathbb{Z}$, we have

$$\varphi_\Phi(\ell) = \varphi_\Theta(\ell)\varphi_\varepsilon(\ell),$$

then, if $\varphi_\varepsilon(\ell) \neq 0$ for any $\ell \in \mathbb{Z}$, a possible estimator of $f_\Theta(\theta)$ is

$$\frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty} \frac{\hat{\varphi}_\Phi(\ell)}{\varphi_\varepsilon(\ell)} e^{-i\ell\theta},$$

where $\hat{\varphi}_\Phi(\ell) = \frac{1}{n}\sum_{j=1}^n e^{i\ell\Phi_j}$ is the empirical version of $\varphi_\Phi(\ell)$. Now, the decay of $\varphi_\varepsilon(\ell)$ requires some *regularization* technique, which can be produced by using the characteristic function of a circular kernel K_κ, say $\varphi_{K_\kappa}(\ell)$, as a *tapering factor*. According to this approach, a kernel-type estimator for f_Θ can be defined as

$$\begin{aligned}\mathrm{D}_\kappa(\theta) &= \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty} \frac{\hat{\varphi}_\Phi(\ell)}{\varphi_\varepsilon(\ell)}\varphi_{K_\kappa}(\ell)e^{-i\ell\theta} \\ &= \frac{1}{2\pi n}\sum_{j=1}^n \left(1 + 2\sum_{\ell=1}^{\infty}\frac{\gamma_\ell(\kappa)}{\lambda_\ell(\kappa_\varepsilon)}\cos(\ell(\theta-\Phi_j))\right). \end{aligned} \tag{7}$$

Note that $\mathrm{D}_\kappa(\cdot)$ has the form of a classical kernel density estimator whose weight function is

$$\tilde{K}_\kappa(\theta) = \frac{1}{2\pi}\left\{1 + 2\sum_{\ell=1}^{\infty}\frac{\gamma_\ell(\kappa)}{\lambda_\ell(\kappa_\varepsilon)}\cos(\ell\theta)\right\}.$$

In order to guarantee its definiteness, we also assume that (a) the error density is an infinitely divisible distribution, that is it has nonvanishing Fourier coefficients $\lambda_\ell(\kappa_\varepsilon)$ for any integer ℓ, and (b) both K_κ and $\tilde{K}_\kappa$ are square integrable functions, that is, using Parseval's identity,

$$\frac{1}{2\pi}\left(1 + 2\sum_{\ell=1}^{\infty}\gamma_\ell^2(\kappa)\right) < \infty \quad \text{and} \quad \frac{1}{2\pi}\left(1 + 2\sum_{\ell=1}^{\infty}\frac{\gamma_\ell^2(\kappa)}{\lambda_\ell^2(\kappa_\varepsilon)}\right) < \infty.$$

Alternatively, estimator (7) can be derived by reference to the so-called *unbiased score method*, which has been introduced in Stefanski and Carroll (1990) for the Euclidean setting. It requires that the conditional expectation of the *unknown* kernel L_κ evaluated at $\theta - \Phi_j$ is equal to a given kernel K_κ evaluated at $\theta - \Theta_j$

$$\mathsf{E}[L_\kappa(\theta - \Phi_j)|\Theta_j] = K_\kappa(\theta - \Theta_j). \tag{8}$$

Then, by working in the Fourier domain, one has

$$\int_0^{2\pi} e^{i\ell\theta}\mathsf{E}[L_\kappa(\theta-\Phi_j)|\Theta_j]d\theta = \int_0^{2\pi} e^{i\ell\theta}K_\kappa(\theta-\Theta_j)d\theta.$$

Hence, assuming that we can interchange integral and expectation and using a change of variable, this leads to

$$\mathsf{E}\left[e^{i\ell\varepsilon_j}\int_0^{2\pi} e^{i\ell v}L_\kappa(v)dv|\Theta_j\right] = \int_0^{2\pi} e^{i\ell v}K_\kappa(v)dv,$$

which finally yields $\varphi_\varepsilon(\ell)\varphi_{L_\kappa}(\ell) = \varphi_{K_\kappa}(\ell)$. Hence, we obtain

$$\begin{aligned} L_\kappa(\theta) =& \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty}\varphi_{L_\kappa}(\ell)\exp(-i\ell\theta) \\ =& \frac{1}{2\pi}\left\{1+2\sum_{\ell=1}^{\infty}\frac{\gamma_\ell(\kappa)}{\lambda_\ell(\kappa_\varepsilon)}\cos(\ell\theta)\right\}, \end{aligned}$$

and so $L_\kappa(\theta) = \tilde{K}_\kappa(\theta)$.

The asymptotic properties of estimator (7) are collected in the following

Result 3. Given a random sample $\Phi_1,\dots,\Phi_n$ from f_Φ, assume (3). Then, for estimator $\mathrm{D}_\kappa(\theta)$ with K_κ being a second sin-order kernel, under assumptions (i)–(iii) of Result 1, one has

$$\mathsf{E}[\mathrm{D}_\kappa(\theta)] - f_\Theta(\theta) = \frac{(1-\gamma_2(\kappa))}{4}f_\Theta^{(2)}(\theta) + o(1-\gamma_2(\kappa)),$$

and

$$\begin{aligned} \mathsf{Var}[\mathrm{D}_\kappa(\theta)] =& \frac{\left(1+2\sum_{\ell=1}^{\infty}\gamma_\ell^2(\kappa)/\lambda_\ell^2(\kappa_\varepsilon)\right)}{2\pi n}f_\Theta(\theta) \\ &+ o\left(\frac{\sum_{\ell=1}^{\infty}\gamma_\ell^2(\kappa)/\lambda_\ell^2(\kappa_\varepsilon)}{n}\right). \end{aligned}$$

Proof. See Appendix. □

Note that, as expected after considering Equation (8), only the variance of $\mathrm{D}_\kappa(\theta)$ is affected by the measurement errors. Thus, differently from the errors-free case, the convergence rate of the estimator is driven by the rate of decay of the coefficients in the Fourier series representation of f_ε, as well as by the smoothness of f_Θ.

The practical implementation of estimator (7) always requires a truncation of the infinite summation appearing in its formulation, by using a sufficiently large number of terms. However, according to the nature of the error density, the coefficients $\lambda_\ell(\kappa_\varepsilon)$ can go to zero too fast yielding instability problems; therefore, we could select the number of coefficients, say p, with the specific aim of reducing this instability. This leads to a further estimator, which depends on two tuning parameters, κ and p. Specifically, when we select also the number of coefficients we obviously have no longer a deconvolution estimator, but a kind of *trigonometric series* estimator as follows:

$$\mathrm{D}_{\kappa,\mathrm{p}}(\theta) = \frac{1}{2\pi n}\sum_{j=1}^{n}\left(1+2\sum_{\ell=1}^{p}\frac{\gamma_\ell(\kappa)}{\lambda_\ell(\kappa_\varepsilon)}\cos(\ell(\theta-\Phi_j))\right).$$

Concerning the asymptotic properties, Result 3 holds for the bias, while in the variance the infinite summation reduces to a p-term one.

4.3 | Removing an estimate of the excess of bias

Removing an estimate of the bias due to measurement errors is an alternate route. Consider the *naive* kernel estimator of $f_\Theta(\theta)$ as defined in Equation (2), but based on $\Phi_1,\dots,\Phi_n$. By expanding $K_\kappa(\Phi_i-\theta)$ for Φ_i around Θ_i, one has

$$\begin{aligned} \hat{f}_\Phi(\theta;\kappa) \approx& \hat{f}_\Theta(\theta;\kappa) + \frac{1}{n}\sum_{i=1}^{n}\sin(\Phi_i-\Theta_i)K_\kappa^{(1)}(\Theta_i-\theta) \\ &+ \frac{1}{2n}\sum_{i=1}^{n}\sin^2(\Phi_i-\Theta_i)K_\kappa^{(2)}(\Theta_i-\theta). \end{aligned}$$

Then, taking the expectation, and observing that $\mathsf{E}[\sin^2(\varepsilon_i)] = (1-\lambda_2(\kappa_\varepsilon))/2$, leads to the asymptotic expression of the excess of bias conditional on the Θ_is

$$\frac{(1-\lambda_2(\kappa_\varepsilon))}{4n}\sum_{i=1}^{n}K_\kappa^{(2)}(\Theta_i-\theta), \tag{9}$$

which can be estimated on the basis of the corrupted sample. This enables us to define a bias-corrected estimator as

$$\mathrm{B}_\kappa(\theta) = \hat{f}_\Phi(\theta;\kappa) - \frac{(1-\lambda_2(\kappa_\varepsilon))}{4n}\sum_{i=1}^{n}K_\kappa^{(2)}(\Phi_i-\theta). \tag{10}$$

Surely, the second term of the RHS of (10) is an estimate of (9) because of the use of Φ_is. Then, the precision of estimator (10) heavily relies on the variance of the measurement errors. Moreover, differently from estimator (7), it requires only the knowledge of $\lambda_2(\kappa_\varepsilon)$ and is not limited to the cases where $\lambda_\ell(\kappa_\varepsilon)\neq 0$, $\ell\in\mathbb{Z}^+$. Further, the above estimator shares the structure of a classical kernel density estimator with weight function $K_\kappa(\theta) - (1-\lambda_2(\kappa_\varepsilon))/4K_\kappa^{(2)}(\theta)$. The

Euclidean counterpart of this estimator has been studied by Stefanski (1985), Carroll and Hall (2004), and Delaigle (2008).

The asymptotic properties are collected in the following.

Result 4. Given the random sample $\Phi_1, \dots, \Phi_n$ from f_Φ, consider estimator $B_\kappa(\theta), \theta \in [0, 2\pi)$, where K_κ is a second sin-order kernel satisfying assumptions (ii)–(iii) of Result 1. If f_Θ and f_Φ have continuous derivatives up to order 2 and 4, respectively, and f_ε has finite second sin-moment, then we have

$$\mathrm{E}[B_\kappa(\theta)] - f_\Theta(\theta) = \frac{1-\gamma_2(\kappa)(2-\lambda_2(\kappa_\varepsilon))}{4} f_\Phi^{(2)}(\theta) + o(1-\gamma_2(\kappa)),$$

and

$$\mathrm{Var}[B_\kappa(\theta)] = \frac{f_\Phi(\theta)}{2\pi n}\left\{1 + 2\sum_{\ell=1}^{\infty}\gamma_\ell^2(\kappa)\left[1 + \frac{\ell^2(1-\lambda_2(\kappa_\varepsilon))}{4}\right]^2\right\} + o\left(\frac{\sum_{\ell=1}^{\infty}\gamma_\ell^2(\kappa)}{n}\right).$$

Proof. See Appendix. □

Notice that, as for Result 2, Result 4 uses a double asymptotic approach. For some considerations about the double asymptotic approach for the Euclidean counterpart, see Delaigle (2008).

We note that the asymptotic bias has the same order as in the deconvolution approach.

Since the second term of Equation (10) is an estimate of the extra bias, it could reasonably have a separate smoothing parameter with respect to the naive estimator $\hat{f}_\Phi(\theta;\kappa)$, leading to a slight modification

$$B_{\kappa_1,\kappa_2}(\theta) = \hat{f}_\Phi(\theta;\kappa_1) - \frac{(1-\lambda_2(\kappa_\varepsilon))}{4n}\sum_{i=1}^{n} K_{\kappa_2}^{(2)}(\Phi_i - \theta).$$

5 | SIMULATIONS

In order to explore the potential of each method, we firstly propose a simulation study where the best possible smoothing degrees are selected, then we consider the case where the smoothing degree is data-driven. Notice that the best smoothing degree analysis allows us to establish which is the best estimator regardless of the smoothing selection rule behavior. Also, consider that for the circular setting such a rule does not still exist when the data are affected by measurement errors.

Since the proposed methods produce estimates which, although integrating to one, can be negative, in the following we consider their normalized versions by replacing the negative values by zero and then rescaling.

5.1 | Simulation models

Our simulation setting considers a number of models where the target population f_Θ is the von Mises density (vM), while for the error densities f_ε we specify a wrapped Normal (wN) error model for the supersmooth case and a wrapped Laplace (wL) model for the ordinary smooth one. The noise-to-signal ratio (NSR), which is defined as the ratio between the *circular variance* of ε and that of Θ, is taken as 25%, 33%, and 45%. For each of these cases, we consider both a supersmooth and ordinary smooth error density with zero mean direction and different values of the concentration parameter chosen in order to obtain the values of NSR as described in the following scenarios:

- Scenario 1: NSR = 25%
  - (a) target density: vM(π, 2), supersmooth error density: wN(0, .92)
  - (b) target density: vM(π, 1), ordinary smooth error density: wL(0, .40)
- Scenario 2: NSR = 33%
  - (a) target density: vM(π, 2), supersmooth error density: wN(0, .90)
  - (b) target density: vM(π, 1), ordinary smooth error density: wL(0, .47)
- Scenario 3: NSR = 45%
  - (a) target density: vM(π, 8), supersmooth error density: wN(0, .97)
  - (b) target density: vM(π, 1.3), ordinary smooth error density: wL(0, .50).

The simulation models are depicted in Figure 1, where, for illustrative purposes, we set the mean of Θ equal to the error mean. Notice that the concentration parameter takes nonnegative real values for both wL and vM but for wL lower values of the concentration parameter give higher concentration, while for vM the opposite holds. As for wN, the concentration parameter ranges from 0 to 1 with the concentration increasing with the value of the parameter. Let $\nu_\ell(\kappa_\Theta)$ and κ_Θ be, respectively, the ℓth Fourier coefficient, $\ell \in \mathbb{Z}^+$, and the smoothing parameter of f_Θ. We have $\nu_\ell(\kappa_\Theta) = \mathcal{I}_\ell(\kappa_\Theta)/\mathcal{I}_0(\kappa_\Theta)$, and $\lambda_\ell(\kappa_\varepsilon)$ equals $\kappa_\varepsilon^{\ell^2}$ and $\kappa_\varepsilon^{-2}/(\ell^2 + \kappa_\varepsilon^{-2})$, respectively, for the wN and wL error distributions.

5.2 | Best possible smoothing degree

In this simulation study, we compare the performance of some of the proposed estimators using the *best possible smoothing degree*. Specifically, we use 200 samples drawn according to the previous scenarios where, for each estimator, we select the smoothing degrees as the minimizers,

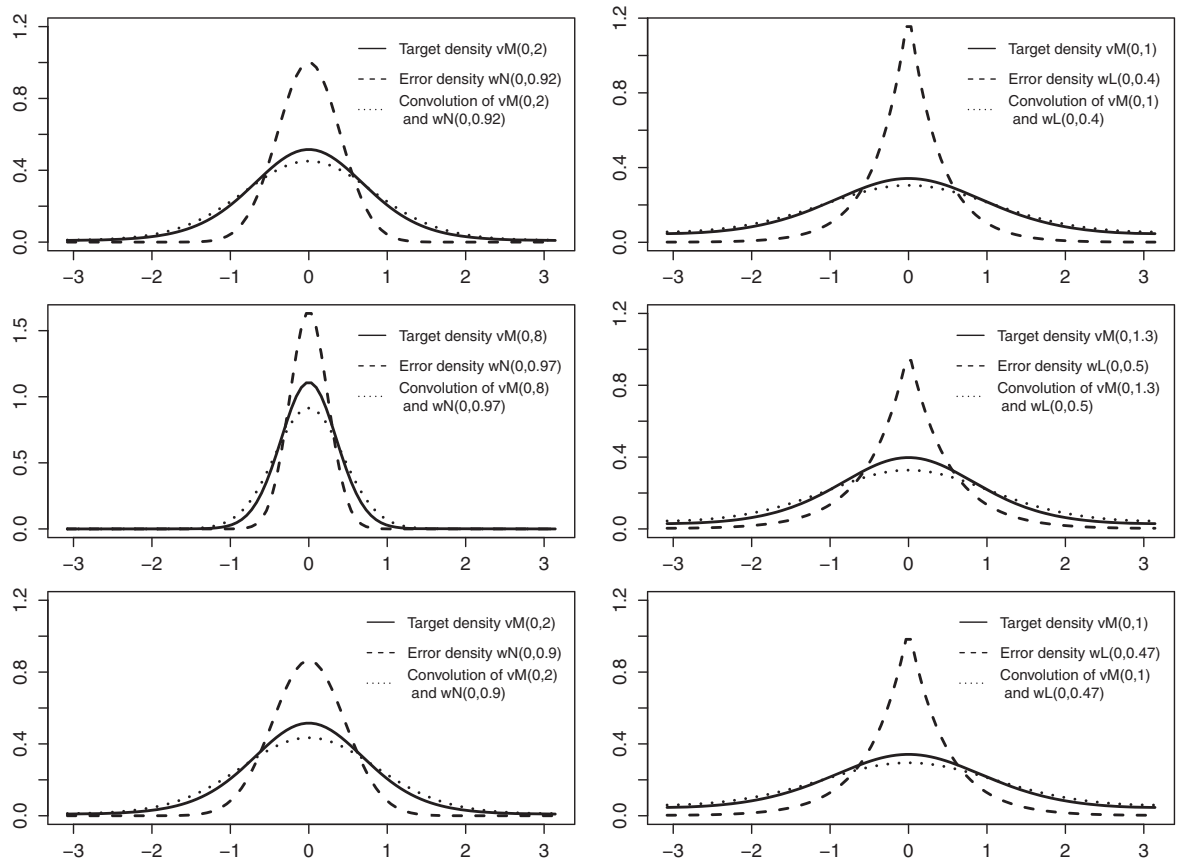

FIGURE 1 Scenarios 1–3 simulation models

over a grid of values, of the averaged integrated squared error (AISE).

We compare the proposed estimators with the naive kernel density estimator and evaluate performance in terms of AISE. Data are drawn from the simulation models described in the previous section, and a von Mises density is employed as the kernel. For each simulation, we generate 200 samples of size $n = 200, 500$, and 1000. The results are collected in Table 1.

In general, we can see that, for a given combination of target and error density, every estimator deteriorates when the noise to signal ratio increases and the error density is supersmooth.

In further comparisons, we see that the naive kernel density estimator KDE shows the highest values of AISE and the lowest convergence rates. The deconvolution-based estimator D_κ, where the infinite sum is approximated by the sum of 20 ratio coefficients, in agreement with our theoretical results, performs reasonably when the error density is ordinary smooth, otherwise the result is very poor. Notice that our sequence of ratio coefficients is not necessarily decreasing as the order increases. However, if we apply to this estimator a simple regularization strategy, consisting in using only the decreasing part of the ratio series, we greatly improve the performance obtaining an estimator, which we call *regularized* D_κ and denote by $r\mathrm{D}_\kappa$, which is generally superior to the naive one and nonregularized one.

The p-term deconvolution-based estimator $\mathrm{D}_{\kappa,\mathrm{p}}$, where both κ and p are *smoothing* parameters, seems to present the best results for every sample size and simulation setting. In particular, we select κ and p by minimizing the AISE over a two-dimensional grid.

The bias-correction estimator B_κ has generally a good performance although affected by the type of error density. We notice that, when the error density is ordinary smooth, it does not have the same efficiency as the deconvolution ones because the bias correction refers only to the leading term. The bias-correction estimator with two different smoothing parameters $\mathrm{B}_{\kappa_1,\kappa_2}$ shows a certain improvement compared to B_κ.

Finally, the results of the equivalence-based estimator $\mathrm{EQD}_{\kappa_1,\kappa_2}$ seem to be very similar to the best ones. This estimator improves the correction of the bias due to the measurement error. Simulations for the equivalence-based estimator $\mathrm{EQR}_{\kappa_1,\kappa_2}$ lead to very similar results, which have not been presented here. Indeed, such similarity was expected on the basis of Remark 2 which shows that estimators EQR_κ and EQD_κ are asymptotically equivalent in the case of one smoothing parameter. Surely, for small

TABLE 1 AISE (× 1000) over 200 samples of size 200, 500 and 1000 drawn from the target population contaminated by noise from different error populations

NSR	f_Θ	f_ε	n	KDE	D_κ	rD_κ	$D_{\kappa,p}$	B_κ	B_{κ_1,κ_2}	EQD_{κ_1,κ_2}
25%	vM(π, 2)	wN(0, .92)	200	1.758	2.390	1.576	**0.965**	1.564	1.392	1.031
			500	1.282	2.089	1.121	**0.577**	1.018	0.867	0.613
			1000	1.052	1.931	0.958	**0.333**	0.749	0.614	0.416
	vM(π, 1)	wL(0, .40)	200	1.030	0.932	0.887	0.741	0.985	0.888	**0.727**
			500	0.698	0.555	0.566	**0.355**	0.621	0.539	0.376
			1000	0.545	0.374	0.430	**0.191**	0.443	0.370	0.238
33%	vM(π, 2)	wN(0, .90)	200	2.221	5.221	2.342	**1.076**	1.914	1.707	1.184
			500	1.702	4.900	1.908	**0.721**	1.296	1.105	0.746
			1000	1.455	4.716	1.749	**0.453**	0.987	0.817	0.538
	vM(π, 1)	wL(0, .47)	200	1.265	1.079	0.925	0.869	1.182	1.062	**0.766**
			500	0.919	0.661	0.736	**0.431**	0.793	0.686	0.445
			1000	0.762	0.456	0.650	**0.240**	0.600	0.502	0.280
45%	vM(π, 8)	wN(0, .97)	200	5.779	6.149	4.738	2.675	4.529	4.004	**2.492**
			500	4.541	4.877	3.770	1.547	2.957	2.492	**1.395**
			1000	4.093	4.451	3.603	1.028	2.333	1.912	**0.991**
	vM(π, 1.3)	wL(0, .50)	200	1.915	1.476	1.552	1.096	1.707	1.533	**1.019**
			500	1.450	0.877	1.160	**0.510**	1.159	0.989	0.603
			1000	1.292	0.636	1.079	**0.328**	0.934	0.782	0.443

KDE assumes no error, while codes D, B, and EQD, respectively, refer to the deconvolution, bias-correction, and equivalence method, all addressing observation error. Bold font indicates the best performance.

samples, slight differences in the estimate are also due to the fact that EQD_κ requires both clipping and normalizing, while EQR_κ only rescaling.

5.3 | Data-driven smoothing degree

In this section, we provide some evidence about the performance of the estimators when the smoothing degrees are data-driven. The simulation models remain the same as before. Our smoothing degree selection method implements the *plug-in* principle, where the unknown quantities in the asymptotic mean integrated squared error formulations are calculated on the basis of a parametric assumption of the population of the error free data. A simple plug-in selector can be obtained by replacing the unknown density appearing in the asymptotic mean integrated squared error formula by a reference density, say g. In the special case where g and f_ε are assumed to be circular densities sharing the same wrapped stable distribution, with respective concentration parameters ρ and κ_ε, then their convolution is still the same wrapped stable density with concentration parameter being the product between ρ and κ_ε. Then, assuming that κ_ε is known, ρ can be directly estimated from the data by the ratio of the estimated concentration parameter of the convolution and κ_ε. Beyond this special case, a naive estimate of ρ can be obtained using corrupted data. In our simulation study, we assume a von Mises population for g whose concentration parameter is estimated from corrupted data using classical maximum likelihood.

Clearly, the use of a data-driven smoothing degree leads to an increase of the AISE. The average deterioration observed for the estimators KDE, D_κ, rD_κ, B_κ, and EQD_{κ_1,κ_2} are, respectively, 13.8%, 28.8%, 73.4%, 34.8%, and 38%. The results are depicted in Table 2. We see that relative merits remain similar to the previous study with the equivalence-based estimator being clearly superior. Notice that the smoothing selection task for this latter estimator is much less problematic than the usual in errors-in-variable problems because each estimator of the ratio (difference) is estimated using the *appropriated* sample, avoiding the classical situation where we have a sample drawn from a density different from the target one. On the other hand, the deconvolution-based estimator clearly suffers from a badly selected smoothing degree in supersmooth cases.

6 | REAL DATA EXAMPLES

6.1 | Ant data

As an illustrative example, we apply our estimators to a dataset previously used by Efromovich (1997) for circular density estimation with errors-in-variables. The dataset has been firstly described by Fisher (1993, Appendix B.7)

TABLE 2 AISE (× 1000) obtained using a plug-in approach over 200 samples of size 200, 500, and 1000 drawn from the target population contaminated by noise from different error populations. Others settings as in Table 1

NSR	f_Θ	f_ε	n	KDE	D_κ	rD_κ	B_κ	EQD_{κ_1,κ_2}
25%	vM(π, 2)	wN(0, .92)	200	1.908	2.408	2.367	1.772	**1.435**
			500	1.410	2.099	2.046	1.245	**0.875**
			1000	1.162	1.938	1.884	0.994	**0.605**
	vM(π, 1)	wL(0, .40)	200	1.282	1.651	1.651	1.342	**0.764**
			500	0.795	0.646	0.646	0.756	**0.490**
			1000	0.620	0.437	0.424	0.557	**0.315**
33%	vM(π, 2)	wN(0, .90)	200	2.422	5.275	5.216	2.229	**1.681**
			500	1.876	4.977	4.913	1.647	**1.055**
			1000	1.607	4.765	4.713	1.375	**0.755**
	vM(π, 1)	wL(0, .47)	200	1.664	2.803	2.803	1.756	**0.769**
			500	1.063	0.866	0.865	1.016	**0.560**
			1000	0.874	0.628	0.597	0.786	**0.379**
45%	vM(π, 8)	wN(0, .97)	200	6.314	6.175	6.161	5.829	**4.084**
			500	4.948	4.884	4.793	4.517	**2.503**
			1000	4.487	4.455	4.252	4.143	**1.882**
	vM(π, 1.3)	wL(0, .50)	200	2.262	2.350	2.350	2.255	**1.198**
			500	1.663	1.208	1.179	1.519	**0.759**
			1000	1.466	1.126	1.748	1.285	**0.564**

and concerns the directions chosen by 100 ants in response to an evenly illuminated black target placed at π. To estimate the density of the chosen directions, Fisher (1993) showed that classical parametric models, like von Mises, are not suited. However, he argued that the population is unimodal since the ants move toward the target with some variation. The rationale behind considering this density estimation problem as an errors-in-variables one is that, due to the typical jerky movement of the insect, the point where the ant intersects the circle can be treated as indirect observation of the direction chosen by the ant.

Efromovich (1997) used a nonparametric approach based on orthogonal trigonometric series and obtained a remarkable result. In fact, his estimate revealed the presence of three modes, in contrast with unimodality detected by previous studies. However, from Figure 2 we conclude that his estimate appears to be artificially symmetric and also shows the pitfall of detecting the modes in partial contrast with data location (see the mode estimation in the right tail).

As an error model for our estimators, we use a wrapped Normal error with zero mean and concentration equal to 0.88, which is very similar to the scenario proposed by Efromovich (1997). We compare the p-term deconvolution-based estimator $D_{\kappa,p}$, the bias-correction estimator B_κ, and the equivalence-based one EQD_{κ_1,κ_2}, suitably normalized, with both the orthogonal series estimator of Efromovich (1997), here denoted by orthogonal series (OS), and the standard circular KDE. A von Mises kernel is used throughout. As for the smoothing degree selection, classical least square cross-validation has been employed. According to this criterion, given a random sample $X_1, \ldots, X_n$ from a density f, for a generic kernel-type estimator of f with smoothing parameter κ, $\hat{f}(\cdot;\kappa)$, the optimal value of κ is the minimizer of

$$\int \hat{f}(x;\kappa)^2 dx - 2n^{-1} \sum_i \hat{f}_{-i}(X_i;\kappa),$$

where $\hat{f}_{-i}$ is the leave-one-out version of $\hat{f}$, obtained after removing X_i from the sample. As can be seen in Figure 2, our estimators confirm multimodality, differently from the standard circular kernel density estimator. However, our modes are differently located from those ones highlighted by the trigonometric series method. We are also able to endorse the asymmetry of the sample.

6.2 | Surface wind data

In this application, we estimate prevailing winds as described in the Introduction. We use wind data from NOAA database. Specifically, we consider Station 42059, which lies in the Eastern Caribbean Sea, 180 nautical miles SSW of Ponce, Puerto Rico. We focus on instantaneous wind directions observed at 06.00 a.m. during Summer 2009. Only odd calendar days have been considered in order to satisfy a stochastic independence assumption.

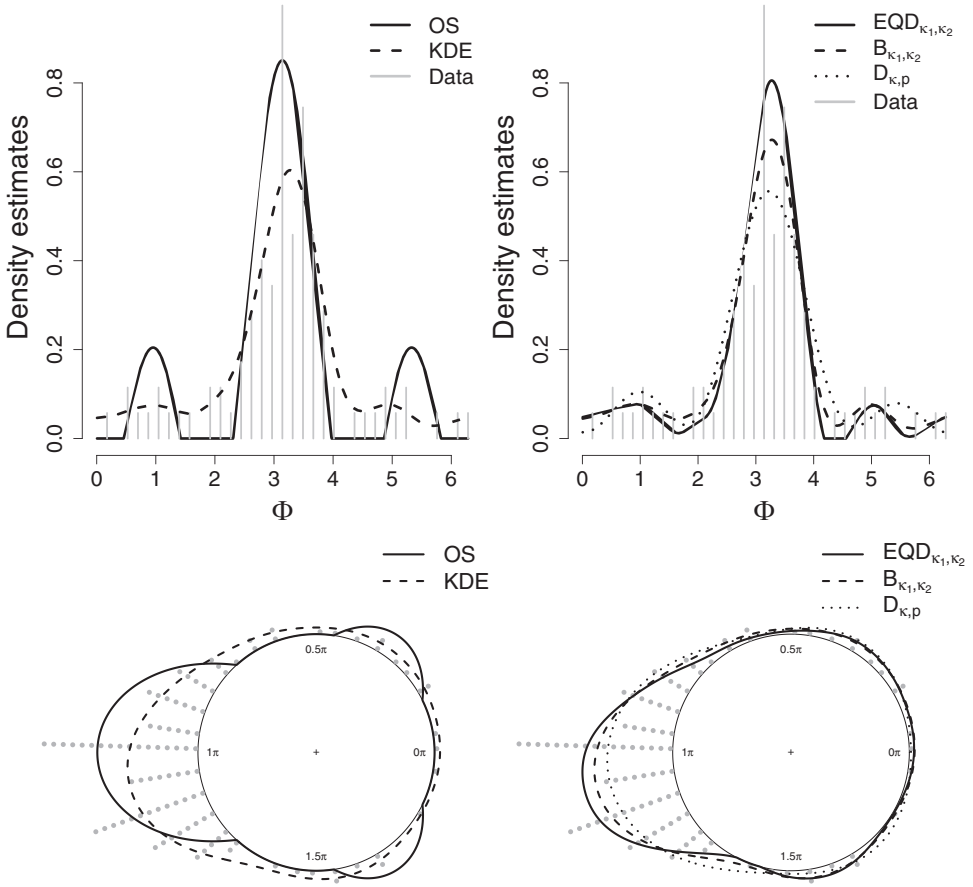

FIGURE 2 Ants data and density estimates of the directions. KDE is based on direct data, while the other estimates assume a wrapped Normal error with zero mean and concentration equal to 0.88

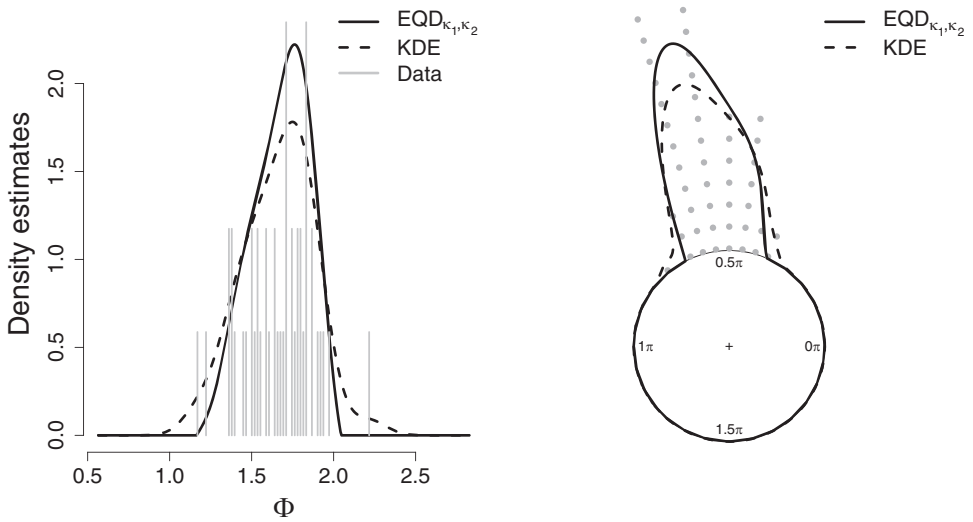

FIGURE 3 Surface wind data and density estimates of the directions. KDE is based on direct data, while EQD$_{\kappa_1,\kappa_2}$ estimate assumes a wrapped Normal error with zero mean and concentration equal to 0.975

Concerning the error distribution, based on observed ranges of moment-to-moment fluctuations over 10 min, we conclude that the measurement error can be approximated by a wrapped Normal error with zero mean and concentration equal to 0.975. The results, using a von Mises kernel, are shown in Figure 3. Due to the shape of the data, we use plug-in rule, where the reference curve is von Mises, and the population concentration is estimated by maximum likelihood. In Table 2, we have seen that, when the plug in selector is used, the most successful method is the equivalence one. This was seen to hold true also for these data, and so only this estimate is shown.

Although mode is confirmed, we can observe a clear effect of deconvolution in generating a more concentrated and regular shape due to the reduction of the effect of noise in the data.

7 | CONCLUSIONS

In this paper, we have explored the errors-in-variables density estimation problem for circular data. We have pursued the kernel approach, as an alternative to the trigonometric series estimators. Intuition, flexibility, and ease of implementation are features of our approach.

However, we notice that research on kernel density estimation for circular data with errors-in-variables requires more attention. Surely, the selection of the smoothing degree is a challenging, nearly unexplored field. Also, consider the case of errors which depend on unobserved data values. Practical applications for such a model are, for example, time recording data, where some clock positions, like integers, half or quarter of hours are more frequently recorded due to the natural attitude of the observer to *rounding*. Regression, that is when predictor variables are observed with errors, and multivariate settings, that is hyperspherical and toroidal data, remain, at the moment, unexplored as well.

ACKNOWLEDGEMENTS

The authors thank a co-editor, an associate editor, and two anonymous referees for very useful comments that improved the presentation of the paper.

DATA AVAILABILITY STATEMENT

The Ants data are provided by Fisher (1993, Appendix B.7); the wind data are freely available from the NOAA Institutional Repository, see references for the URL.

ORCID

Charles C. Taylor https://orcid.org/0000-0003-0181-1094

REFERENCES

Carroll, R.J. and Hall, P. (2004) Low order approximations in deconvolution andregression with errors in variables. *Journal of the Royal Statistical Society, Series B*, 66, 31–46.

Comte, F. and Taupin, M.L. (2003) Adaptive density deconvolution for circular data. *Prépublication MAP5 2003-10 report*. Paris, France: Université Paris Descartes.

Delaigle, A. (2007) Nonparametric density estimation from data with a mixture of Berkson and classical errors. *The Canadian Journal of Statistics*, 35, 89–104.

Delaigle, A. (2008) An alternative view of the deconvolution problem. *Statistica Sinica*, 18, 1025–1045.

Delaigle, A. (2014) Nonparametric kernel methods with errors-in-variables: constructing estimators, computing them, and avoiding common mistakes. *Australian & New Zealand Journal of Statistics*, 56, 105–124.

Delaigle, A., Hall, P. and Meister, A. (2008) On deconvolution with repeated measurements. *Annals of Statistics*, 36, 665–685.

Delaigle, A. and Meister, A. (2008) Density estimation with heteroscedastic error. *Bernoulli*, 14, 562–579.

Di Marzio, M., Panzera, A. and Taylor, C.C. (2009) Local polynomial regression for circular predictors. *Statistics & Probability Letters*, 79, 2066–2075.

Efromovich, S. (1997) Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92, 526–535.

Fisher, N.I. (1993) *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University Press.

Jammalamadaka, S.R. and SenGupta, A. (2001) *Topics in Circular Statistics*. Singapore: World Scientific.

Johannes, J. and Schwarz, M. (2013) Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19, 1576–1611.

Ley, C. and Verdebout, T. (2017) *Modern Directional Statistics*. London, UK: Chapman and Hall/CRC.

Ley, C. and Verdebout, T. (2018) *Applied Directional Statistics. Modern Methods and Case Studies*. London, UK: Chapman and Hall/CRC.

NOAA. (2020) National Data Buoy Center. https://www.ndbc.noaa.gov/station_page.php?station=42059 (accessed August 18, 2020).

Stefanski, L.A. (1985) The effect of measurement error on parameter estimation. *Biometrika*, 72, 583–592.

Stefanski, L.A. and Carroll, R.J. (1990) Deconvolving kernel density estimators. *Statistics*, 21, 169–184.

SUPPORTING INFORMATION

R scripts for implementing the two real examples proposed in Section 6 are available at the Biometrics website on Wiley Online Library, along with code for replicating the simulation results reported in Section 5.

How to cite this article: Di Marzio M, Fensore S, Panzera A, Taylor CC. Density estimation for circular data observed with errors. *Biometrics*. 2021;1–13. https://doi.org/10.1111/biom.13431

APPENDIX A

Proof of Result 2. For the bias, we start by observing that

$$\mathsf{E}[\mathrm{EQD}_\kappa(\theta)] = 2\int_0^{2\pi} K_\kappa(\phi-\theta)f_\Phi(\phi)d\phi - \int_0^{2\pi} K_\kappa(\psi-\theta)f_\Psi(\psi)d\psi.$$

Now, for a circular convolution, say f_Ω, of a circular density f_Q and a circular density f_U, if U concentrates around 0, one can consider the following pth order Taylor series representation:

$$f_\Omega(q) = f_Q(q) + \sum_{j=1}^{p} \frac{f_Q^{(j)}(q)}{j!}\int_0^{2\pi} \sin^j(u)f_U(u)du + o\left(\int_0^{2\pi} \sin^{p+1}(u)f_U(u)du\right).$$

Then recalling that f_Φ and f_Ψ, respectively, are the circular convolutions of f_Θ and f_ε, and of f_Φ and f_ε, the fact that ε concentrates around 0 enables the use of the above expansion for both f_Φ and f_Ψ. Then, the same expansion applies also for the jth term in the expansion of f_Ψ, which is the circular convolution of $f_\Phi^{(j)}$ and f_ε. In particular, by considering all the expansions up to the second order, and using $\int_0^{2\pi} \sin^2(u)f_\varepsilon(u)du = (1-\lambda_2(\kappa_\varepsilon))/2$, one has

$$\begin{aligned}
\mathsf{E}[\mathrm{EQD}_\kappa(\theta)] =& 2\int_0^{2\pi} K_\kappa(\omega-\theta)\left\{ f_\Theta(\omega) + \frac{(1-\lambda_2(\kappa_\varepsilon))}{4} f_\Theta^{(2)}(\omega) + o(1-\lambda_2(\kappa_\varepsilon))\right\}d\omega \\
&- \int_0^{2\pi} K_\kappa(\omega-\theta)\left\{ f_\Theta(\omega) + \frac{(1-\lambda_2(\kappa_\varepsilon))}{2} f_\Theta^{(2)}(\omega) + \frac{(1-\lambda_2(\kappa_\varepsilon))^2}{1} 6 f_\Theta^{(4)}(\omega)\right. \\
&\left. + o([1-\lambda_2(\kappa_\varepsilon)]^2)\right\}d\omega \\
=& \int_0^{2\pi} K_\kappa(\omega-\theta)f_\Theta(\omega)d\omega - \frac{(1-\lambda_2(\kappa_\varepsilon))^2}{1} 6\int_0^{2\pi} K_\kappa(\omega-\theta)f_\Theta^{(4)}(\omega)d\omega \\
&+ \; o([1-\lambda_2(\kappa_\varepsilon)]^2).
\end{aligned}$$

Now, note that the first term in the leading term of the above expectation corresponds to the expectation of a standard kernel estimator of f_Θ. Then using the fact K_κ is a circular kernel satisfying the assumptions in Result 1, standard asymptotic arguments for this quantity along with a first-order approximation of the second term lead to the bias result. For the asymptotic variance, by using the first-order version of the above expansion of convolution for both f_Φ and f_Ψ, we can finally use

$$\mathrm{Var}[\mathrm{EQD}_\kappa(\theta)] \approx \frac{1}{n}\int_0^{2\pi} K_\kappa^2(\omega-\theta)f_\Theta(\theta)d\omega,$$

which, using classical circular kernel density estimation theory, leads to the result.

Proof of Result 3. The asymptotic bias directly follows by considering identity (8) and using Result 1. The asymptotic variance directly follows by using Parseval's identity.

Proof of Result 4. We start by observing that, for $\theta \in \mathbb{R}$, letting

$$W_\kappa(\theta) = K_\kappa(\theta) - \frac{(1-\lambda_2(\kappa_\varepsilon))}{4}K_\kappa^{(2)}(\theta),$$

the estimator can be rewritten as a standard kernel estimator with kernel W_κ. Now, we have that the ℓth coefficient in the Fourier series representation of W_κ, say $w_\ell(\kappa)$, satisfies

$$w_\ell(\kappa) = \gamma_\ell(\kappa)\left(1 + \ell^2\frac{(1-\lambda_2(\kappa_\varepsilon))}{4}\right).$$

Hence, W_κ is a second sin-order kernel, such that, as ε is concentrated around 0, $\eta_j(W_\kappa) = O(\eta_j(K_\kappa))$. Then, using Result 1, with $w_\ell(\kappa)$ as the Fourier coefficients, leads to both the asymptotic bias and the asymptotic variance. □