**ORIGINAL PAPER**

# An adaptive method for bandwidth selection in circular kernel density estimation

Stanislav Zámečník[1] · Ivana Horová[1] · Stanislav Katina[1,2] · Kamila Hasilová[3]

## Abstract

Kernel density estimations of circular data are an effective type of nonparametric estimation. The performance of these estimations depends significantly on a smoothing parameter referred to as bandwidth. Selecting suitable bandwidths for these types of estimation pose fundamental challenges, therefore fixed bandwidth selectors are often the initial choice. The study investigates common bandwidth selection methods and proposes novel methods which adopt the idea from the linear case. The attention is also paid to variable bandwidth selection. Using simulations which incorporate a range of circular distributions that exhibit multimodality, peakedness and skewness, the proposed methods were evaluated and then compared with other bandwidth selectors to determine their potential advantages. Two real datasets, one containing animal movements and the other wind direction data, were applied to illustrate the utility of the proposed methods.

**Keywords** Circular density · Bandwidth selector · Adaptive kernel estimator · Von Mises density · Smoothed cross validation

## 1 Introduction

Circular observations are data considered as points on a circle, measured in degrees or radians. The term 'circular data' is used to distinguish them from the data with the real line as their support, which are then referred to as 'linear data'. Circular data are often encountered in field such as biology, bioinformatics, meteorology, geology and environmetrics. Examples of studies in these fields include the analysis of patterns in animal navigation, wind direction and circadian rhythms. Circular

✉ Stanislav Zámečník
zamecnik@math.muni.cz

1    Department of Mathematics and Statistics, Masaryk University, Brno, Czech Republic

2    Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic

3    Department of Quantitative Methods, University of Defence, Brno, Czech Republic

🖄 Springer

data are essentially different from linear data. Not only is there no true zero, but any designation of low or high values is arbitrary. Furthermore, the periodic nature of the circular data complicates their analysis, since standard methods for observations in Euclidean space are inappropriate for descriptive and inferential analysis of such data (instead of differences between observations, the angle between two vectors has to be considered when using data on the circle).

Statistical methods for describing and analysing circular data are relatively new and are still undergoing development. Authors such as Batschelet (1981), Fisher (1995), Mardia and Jupp (2009) have systematically analysed the theory and methodology of circular statistics. Monographs by Jammalamadaka and SenGupta (2001), Ley and Verdebout (2017), Ley and Verdebout (2018) and reviews by Pewsey (2000) and Mardia (2021) also contain valuable bibliographies.

Kernel density estimation is a type of nonparametric estimation which is frequently applied and depends on a kernel as a weight function and on bandwidth as a smoothing parameter. Although bandwidth has an important task in kernel smoothing, selection of the optimal bandwidth is not straightforward. The literature on this subject for linear data is quite extensive and includes monographs authored by Wand and Jones (1994), Härdle et al. (2012), Silverman (1986), Scott (1992), Simonoff (2012), Horová et al. (2012) and other studies by Marron and Ruppert (1994), Park and Marron (1990), Scott and Terrell (1987), Jones and Kappenman (1992), Cao et al. (1994).

Considerable attention has been paid to methods of kernel estimation on the circle. Similarly to the linear case, the choice of the bandwidth plays a key role. Some useful concepts behind bandwidth selection methods for linear kernel density estimation can be applied to circular densities, but a straightforward conversion from a linear to a circular case is often not possible because of the inherent characteristics of circular data. Practical rules for selecting the degree of smoothing were investigated. These were based on a cross-validation method, bootstrap and plug-in rule ideas. Hall et al. (1987) described data driven bandwidth selectors based on cross-validation methods. They developed least square cross-validation and likelihood cross-validation bandwidth selectors. Taylor (2008) proposed a rule of thumb for bandwidth selection. Oliveira et al. (2012) devised a novel plug-in rule. Oliveira et al. (2013) demonstrated that robustifying the Taylor estimation of the concentration parameter improved the estimate of the smoothing parameter. The error criteria for the bandwidth selection are often based on minimising the asymptotic mean integrated squared error (AMISE); however, García-Portugués et al. (2013) determined that the AMISE and mean integrated squared error (MISE) can differ significantly for moderate and even large sample sizes. García-Portugués (2013) proposed new bandwidth selectors for kernel density estimation with directional data. Tsuruta and Sagae (2017) also tackled bandwidth selection and employed higher order kernels. Tenreiro (2022) proposed a direct plug-in approach based on an alternative Fourier series to select the smoothing parameter. Variable bandwidth selection methods for linear data have also been modified and applied to circular data. Pham Ngoc (2019) proposed a "spherical penalized comparison to overfitting" procedure and analysed its performance in a numerical study with spherical data (i.e., three-dimensional data).

The aim of this paper is to introduce new bandwidth selection methods for selecting the smoothing parameter in circular kernel density estimation. We propose a novel bandwidth selection method based on an alternative to the cross-validation method. We follow the concept applied by Hall et al. (1992) for linear densities. We also suggest modified adaptive bandwidth selection methods based on the concepts applied by Breiman et al. (1977) and Silverman (1986) and the procedure described by Demir (2018).

The paper is organized as follows. Section 2 introduces the subject of circular kernel density estimation and reviews the respective literature and methods; Section 3 details current bandwidth selection methods and presents a novel smoothed cross-validation (SCV) method adapted from the linear case. Section 4 explores adaptive bandwidth selection methods that have not been studied extensively in circular statistics and proposes new adaptive methods. Section 5 compares the performance of these methods through a simulation; Sect. 6 shows the application of these methods to real world data. Section 7 discuss findings and concludes the paper.

## 2 Preliminaries

Let $\theta_1, \ldots, \theta_n \in [0, 2\pi)$ be a sample of angles coming from an unknown circular density $f(\theta)$. Estimating the function $f$ is a natural part of exploratory data analysis. The circular kernel density estimator $\hat{f}$ of $f$ is defined as

$$\hat{f}(\theta, v) = \frac{1}{n} \sum_{i=1}^{n} K_v(\theta - \theta_i), \qquad \theta \in [0, 2\pi). \tag{2.1}$$

where $K_v(\cdot)$ is a circular kernel function, $v$ is a parameter such that $v \to \infty$, $\sqrt{v}n^{-1} \to 0$ for $n \to \infty$ (Di Marzio et al. 2011). This parameter takes the role of the smoothing parameter, called the bandwidth. As the kernel function, we consider the von Mises density (*VM*)

$$g(\theta, \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \tag{2.2}$$

where $\mu \in [0, 2\pi)$ is the mean direction, $\kappa$ is a concentration parameter and $I_r(\cdot)$ is the modified Bessel function of order $r$ (see Appendix 8). In this case, the density estimator of $f$ is given by

$$\hat{f}(\theta, v) = \frac{1}{2\pi n I_0(v)} \sum_{i=1}^{n} e^{v \cos(\theta - \theta_i)}, \qquad \theta \in [0, 2\pi). \tag{2.3}$$

Although the selected circular kernel function is important, the estimate quality crucially depends on the choice of bandwidth $v$, which determines the smoothness of the estimate. From an exploratory point of view, every bandwidth choice delivers a useful density estimate. Small values illustrate the global structure in un unknown density whereas large bandwidth values reveal local structure which may or may not be present in the actual density.

Bandwidth selection methods are generally based on an appropriate error criterion, a suitable candidate being the MISE, where

$$\mathrm{MISE}(v) = \mathrm{E} \int_0^{2\pi} (\hat{f}(\theta, v) - f(\theta))^2 \mathrm{d}\theta.$$

As this expression is not mathematically tractable, the asymptotic expression for the MISE (AMISE) derived by Di Marzio et al. (2019) is given by

$$\mathrm{MISE}(v) = \mathrm{AMISE}(v) + o(1),$$

where, using (2.3) and under the conditions $v \to \infty$, $\sqrt{v}n^{-1} \to 0$ for $n \to \infty$ and assuming that the density $f$ is such that its second derivative $f''$ is continuous and square integrable, the AMISE can be written as

$$\mathrm{AMISE}(v) = \frac{1}{16}\left(1 - \frac{I_2(v)}{I_0(v)}\right) \int_0^{2\pi} \left(f''(\theta)\right)^2 \mathrm{d}\theta + \frac{I_0(2v)}{2n\pi(I_0(v))^2}. \tag{2.4}$$

The optimal bandwidth value which minimizes (2.4) is obtained from

$$v_{opt} = \left(2\pi^{1/2}n \int_0^{2\pi} \left(f''(\theta)\right)^2 \mathrm{d}\theta\right)^{2/5}.$$

However, a significant problem exists since $v_{opt}$ depends on an unknown density $f$.

To obtain an approximate computation of $v_{opt}$, several common methods can be applied, for example rule of thumb (RT), plug-in (PI), least square cross-validation (LSCV), likelihood cross-validation (LCV) and bootstrap (BT). However, no universally applicable method exists for kernel estimation of circular densities. The following section discusses this in greater detail.

## 3 Choice of bandwidth – fixed bandwidth methods

### 3.1 Known methods

Taylor (2008) proposed an RT method based on an approach described by Silverman (1986). RT uses the von Mises density (2.2) as a reference density $f$ to overcome the problem with unknown density in (2.4). The AMISE consequently takes the form

$$\mathrm{AMISE}(v) = \frac{3\kappa I_2(2\kappa)}{32\pi v^2(I_0(\kappa))^2} + \frac{v^{1/2}}{2n\pi^{1/2}}. \tag{3.1}$$

The bandwidth $v_{\mathrm{RT}} = v_{opt}$, which minimizes AMISE (3.1), is estimated as

$$\hat{v}_{\mathrm{RT}} = \left(\frac{3n\hat{\kappa}^2 I_2(2\hat{\kappa})}{4\pi^{1/2}(I_0(\hat{\kappa}))^2}\right)^{2/5}, \tag{3.2}$$

where $\hat{\kappa}$ is the estimate of the concentration parameter $\kappa$ obtained using the maximum likelihood method.

Oliveira et al. (2012) suggested an alternative method by plugging in a more flexible distribution family as a reference density in the AMISE formula (2.4). To obtain the bandwidth selector $\hat{v}_{PI}$, the three-step procedure described below begins with applying a finite mixture of $M$ von Mises densities, i.e., $f(\theta) = \sum_{i=1}^{M} w_i g_i(\theta, \mu_i, \kappa_i)$, $\sum_{i=1}^{M} w_i = 1$ and $g_i$ is defined by (2.2):

1. Select the number of mixture components $M$ for the reference distribution.
2. Estimate the AMISE (2.4) as follows:

   2.1. Estimate the parameters in the von Mises mixture of $M$ densities.
   2.2. Calculate the integral $\int_0^{2\pi} \left( f''(\theta) \right)^2 d\theta$.
   2.3. Plug-in the quantity above in (2.4) to obtain $\widehat{\text{AMISE}}(v)$.

3. Minimize $\widehat{\text{AMISE}}$ with respect to $v$ and obtain selector $\hat{v}_{PI}$.

The integral in step 2.2 is computed numerically using the composite trapezoidal rule.

The earliest, fully automated fixed bandwidth selection methods for linear data are those based on cross-validation. The LSCV for circular data proposed by Taylor (2008) targets the MISE and employs an objective function

$$\text{LSCV}(v) = \int_0^{2\pi} \hat{f}^2(\theta, v) d\theta - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(\theta_i, v), \tag{3.3}$$

where $\hat{f}_{-i}$ denotes the circular kernel density estimator (2.3) without the $i$-th observation

$$\hat{f}_{-i}(\theta_i, v) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} K_v(\theta_i - \theta_j), \qquad i = 1, \ldots, n.$$

Minimizing objective function (3.3), we obtain the bandwidth $\hat{v}_{LSCV}$.

Another popular approach cross-validation method is LCV (e.g., Di Marzio et al. (2011), Oliveira et al. (2012)). The optimal bandwidth $\hat{v}_{LCV}$ is obtained by maximizing the likelihood function

$$\text{LCV}(v) = \prod_{i=1}^{n} \hat{f}_{-i}(\theta_i, v), \tag{3.4}$$

where $\hat{f}_{-i}$ represents the same as above.

Di Marzio et al. (2011) introduced a BT method for bandwidth selection procedure inspired by the method applied to linear data by Taylor (2008). Using the von Mises kernel, the MISE of the BT has a closed expression, and $\hat{v}_{BT}$ provides the value for minimizing the expression

$$\mathrm{BT}(v) = \int_0^{2\pi} \mathrm{E_{BT}} \big[\hat{f}^*(\theta, v) - \hat{f}(\theta, v)\big]^2 d\theta, \tag{3.5}$$

where $\mathrm{E_{BT}}$ denotes the bootstrap expectation with respect to random samples $\theta_1, \ldots, \theta_n$ generated from density $f(\theta, v)$.

## 3.2 Proposed method

An alternative to the methods described above is the SCV method originally introduced by Hall et al. (1992) for linear densities. Here, we propose an adaptation of this method for circular densities. It resembles RT selection by applying a kernel estimator to estimate an integrated squared bias component (ISB) of the MISE

$$\mathrm{MISE}(\hat{f}(\theta, v)) = \frac{1}{n} \underbrace{\int_0^{2\pi} (K_v^2 * f)(\theta) - (K_v * f)^2(\theta) d\theta}_{\mathrm{IV}}$$
$$+ \underbrace{\int_0^{2\pi} ((K_v * f)(\theta) - f(\theta))^2 d\theta}_{\mathrm{ISB}} \tag{3.6}$$

where IV is an integrated variance and $*$ is a convolution operator. The SCV method's difference is that it estimates an exact formula of ISB rather than its asymptotic approximation, as in the case of RT. An appealing feature of this method is less dependency on asymptotic approximation.

The SCV's objective function is obtained by replacing the unknown function $f$ in (3.6) with its estimate $\hat{f}$ (2.3) and selecting the von Mises density (2.2) as the kernel function:

$$\mathrm{SCV}(v) = \widehat{\mathrm{ISB}}(v) + \widehat{\mathrm{IV}}(v), \tag{3.7}$$

where

$$\widehat{\mathrm{ISB}}(v) = \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \Bigg\{ \frac{1}{4\pi^2 I_0^4(v)} \int_0^{2\pi} I_0\big(v\sqrt{2(1 + \cos(\theta - \theta_i))}\big)$$
$$\times I_0\big(v\sqrt{2(1 + \cos(\theta - \theta_j))}\big) d\theta - \frac{1}{2\pi^2 I_0^3(v)} \int_0^{2\pi} I_0\big(v\sqrt{2(1 + \cos(\theta - \theta_i))}\big)$$
$$\times e^{v\cos(\theta - \theta_j)} d\theta + \frac{1}{2\pi I_0^2(v)} I_0\big(v\sqrt{2(1 + \cos(\theta_i - \theta_j))}\big) \Bigg\} \tag{3.8}$$

and

$$\widehat{\text{IV}}(v) \simeq \frac{v^{1/2}}{2n\pi^{1/2}},$$

which is a good estimate of the integrated variance based on its asymptotic counter-partner. Minimization of the above function results in $\hat{v}_{\text{SCV}}$.

## 4 Choice of bandwidth – adaptive bandwidth methods

Generally, in the linear case, the problem with fixed bandwidth kernel density estimators arises when a true distribution has long-tails or multiple modes. Such a bandwidth performs well near the peak of the distribution, but poorly at its tails. Breiman et al. (1977) proposed an efficient solution that varies the bandwidth at each data point:

$$\hat{f}(x) = \sum_{i=1}^{n} \frac{1}{nh(X_i)} K\left(\frac{x - X_i}{h(X_i)}\right),$$

where $h(X_i)$ is the variable bandwidth at data point $X_i$. Silverman (1986) followed up on this work, suggesting that $h(X_i)$ has to be proportional to $g/\hat{f}(X_i)^{1/2}$, where $g$ is a geometric mean, and proposed a three-step procedure: Step 1 – obtain a pilot estimate, Step 2 – define a local bandwidth factor, Step 3 − evaluate the adaptive estimator.

A similar problem occurs with directional data: this method can be easily adapted by applying the von Mises density as the kernel function $K(\cdot)$ to define a variable bandwidth estimator

$$\hat{f}(\theta, v) = \frac{1}{2n\pi} \sum_{i=1}^{n} \frac{1}{I_0(v(\theta_i))} e^{v(\theta_i)\cos(\theta - \theta_i)} \tag{4.1}$$

and the corresponding three-step procedure:

1. Obtain a pilot estimate $\hat{f}(\theta_i)$ using the fixed bandwidth method (see Sect. 3).
2. Define the local bandwidth factors $\lambda_i$ ($i = 1, \ldots, n$) according to the equation

   $$\lambda_i = \left\{g/\hat{f}(\theta_i)\right\}^{\alpha},$$

   where $\alpha \in [0, 1]$ is a sensitivity parameter, typically preferred with an optimal value of 1/2.
3. Evaluate the new adaptive estimator, where $v(\theta_i) = \lambda_i v$ in (4.1), and thus obtain

   $$\hat{f}(\theta) = \frac{1}{2n\pi} \sum_{i=1}^{n} \frac{1}{I_0(\lambda_i v)} e^{\lambda_i v \cos(\theta - \theta_i)}. \tag{4.2}$$

To calculate $\lambda_i$, factors other than the geometric mean $g$ may be considered, for example the arithmetic mean (denoted $a$). Applying a procedure similar to Demir (2018), multiple adaptive estimators can be defined:

**Table 1** Bandwidth selection methods compared in the simulations

| | | |
|---|---|---|
| $\hat{v}_{RT}$ | Rule of thumb | Eq. (3.2) |
| $\hat{v}_{LSCV}$ | Least square cross-validation | Eq. (3.3) |
| $\hat{v}_{LCV}$ | Likelihood cross-validation | Eq. (3.4) |
| $\hat{v}_{PI}$ | Plug-in | procedure on p. 5 |
| $\hat{v}_{BT}$ | Bootstrap | Eq. (3.5) |
| $\hat{v}_{SCV}$ | Smoothed cross-validation | Eq. (3.7) |

- For estimators with a factor $\lambda$ based on the geometric mean, the optimal bandwidth is denoted $v_B^g$;
- For estimators with a factor $\lambda$ based on the arithmetic mean, the optimal bandwidth is denoted $v_B^a$;

where B stands for the bandwidth selection method (i.e., RT, LSCV, LCV, PI, BT or SCV).

## 5 Simulation study

Using Monte Carlo simulations, this section of the current study compares the efficiency of the proposed SCV method and proposed adaptive methods with the existing methods presented above, namely RT, LSCV, LCV, PI and BT methods.

Various circular distributions – von Mises, cardioid, wrapped normal, wrapped Cauchy and their mixtures with different skewness, kurtosis and multimodality – are used (see Appendix 8.2 for details).

The simulation comprised two parts

1. A performance comparison of the proposed SCV method (see (3.6)) with common methods on *simple models* which include typical circular distributions, i.e., von Mises (M1), wrapped normal (M2), wrapped Cauchy (M3) and cardioid (M4).
2. A performance comparison of the proposed adaptive methods (see (4.2)) on *mixture models*: mixture of two von Mises distributions (M5–M12). It should be noted that mixture models consider a diverse range of von Mises distributions with different concentration parameters, for example antipodal, symmetric and asymmetric distributions.

For each distribution, 100 random samples of size $n = 100$ and $n = 250$ were generated. Tables 2, 3 and 4 compare the integrated squared errors (ISE), where ISE $= \int (f - \hat{f})^2$, of the circular kernel density estimator (2.3) for various bandwidth selectors (using 500 points for numerical calculation of the integrals).

The performance of these commonly used methods is compared to the bandwidth selector obtained with the proposed SCV method; the notations for the respective methods are listed in Table 1.

**Table 2** Average ISE for models M1–M4

| Estimator | M1 | | M2 | | M3 | | M4 | |
|---|---|---|---|---|---|---|---|---|
| | 100 | 250 | 100 | 250 | 100 | 250 | 100 | 250 |
| $\hat{v}_{RT}$ | 0.7568 | 0.3746 | 0.5393 | 0.2938 | 1.9390 | 1.1269 | 0.4736 | 0.2496 |
| $\hat{v}_{LSCV}$ | 0.9100 | 0.4292 | 0.6763 | 0.3365 | 1.4215 | 0.6717 | 0.5698 | 0.2996 |
| $\hat{v}_{LCV}$ | 0.8755 | 0.4245 | 0.6135 | 0.3154 | 1.6238 | 0.7558 | 0.5888 | 0.3067 |
| $\hat{v}_{PI}$ | 0.9103 | 0.3998 | 0.7497 | 0.3143 | 1.3573 | 0.6606 | 0.8476 | 0.3398 |
| $\hat{v}_{BT}$ | 0.8871 | 0.4053 | 0.6604 | 0.2948 | 3.4518 | 0.8587 | 0.7276 | 0.2495 |
| $\hat{v}_{SCV}$ | 1.2078 | 0.4331 | 0.6929 | 0.2996 | 8.1770 | 1.0710 | 0.6305 | 0.2614 |

**Table 3** Average ISE for models M5–M8

| Estimator | M5 | | M6 | | M7 | | M8 | |
|---|---|---|---|---|---|---|---|---|
| | 100 | 250 | 100 | 250 | 100 | 250 | 100 | 250 |
| $\hat{v}_{RT}$ | 0.4520 | 0.4571 | 2.8653 | 2.8875 | 0.9247 | 0.8051 | 0.6519 | 0.3607 |
| $\hat{v}_{RT}^{a}$ | 0.4539 | 0.4709 | 2.8603 | 2.8873 | 0.9431 | 0.8474 | 0.7280 | 0.4571 |
| $\hat{v}_{RT}^{g}$ | 0.4539 | 0.4707 | 2.8602 | 2.8873 | 0.9426 | 0.8463 | 0.7257 | 0.4520 |
| $\hat{v}_{LSCV}$ | 0.6051 | 0.3133 | 0.9639 | 0.4081 | 0.6663 | 0.3382 | 0.5285 | 0.2592 |
| $\hat{v}_{LSCV}^{a}$ | 0.5189 | 0.2853 | 1.0003 | 0.4233 | 0.6287 | 0.3398 | 0.5232 | 0.2588 |
| $\hat{v}_{LSCV}^{g}$ | 0.5272 | 0.2873 | 1.0024 | 0.4245 | 0.6357 | 0.3410 | 0.5274 | 0.2595 |
| $\hat{v}_{LCV}$ | 0.5821 | 0.2930 | 0.9537 | 0.3953 | 0.6351 | 0.3258 | 0.4942 | 0.2444 |
| $\hat{v}_{LCV}^{a}$ | 0.5058 | 0.2694 | 1.0025 | 0.4115 | 0.6119 | 0.3283 | 0.5029 | 0.2441 |
| $\hat{v}_{LCV}^{g}$ | 0.5131 | 0.2711 | 1.0030 | 0.4127 | 0.6167 | 0.3294 | 0.5055 | 0.2449 |
| $\hat{v}_{PI}$ | 0.8316 | 0.3050 | 0.9981 | 0.3774 | 0.7390 | 0.3393 | 0.9016 | 0.3361 |
| $\hat{v}_{PI}^{a}$ | 0.7296 | 0.2829 | 0.9885 | 0.3813 | 0.6740 | 0.3284 | 0.8553 | 0.3409 |
| $\hat{v}_{PI}^{g}$ | 0.7416 | 0.2845 | 0.9968 | 0.3833 | 0.6839 | 0.3304 | 0.8641 | 0.3416 |
| $\hat{v}_{BT}$ | 0.6543 | 0.6572 | 2.9163 | 1.7077 | 1.3028 | 1.3034 | 1.8030 | 0.8697 |
| $\hat{v}_{BT}^{a}$ | 0.6544 | 0.6574 | 2.9163 | 1.7290 | 1.3030 | 1.3036 | 1.7953 | 0.8804 |
| $\hat{v}_{BT}^{g}$ | 0.6544 | 0.6574 | 2.9163 | 1.7266 | 1.3030 | 1.3036 | 1.7964 | 0.8787 |
| $\hat{v}_{SCV}$ | 0.6396 | 0.6379 | 2.8760 | 0.8510 | 1.2500 | 1.2362 | 1.6320 | 0.3039 |
| $\hat{v}_{SCV}^{a}$ | 0.6403 | 0.6381 | 2.8760 | 0.9112 | 1.2523 | 1.2384 | 1.6282 | 0.3360 |
| $\hat{v}_{SCV}^{g}$ | 0.6403 | 0.6381 | 2.8760 | 0.9057 | 1.2523 | 1.2384 | 1.6292 | 0.3332 |

Performance is also compared to the selectors (based on the geometric and arithmetic means) obtained from the adaptive bandwidth methods. Figures 1, 2 and 3 graph the behaviour of the average of ISE values. The simulations were performed in statistical software R version 4.2.2 (R Core Team 2022), using our code for the proposed SCV kernel density estimator and existing codes from R packages *NPCirc* and *circular*.

**Table 4** Average ISE for models M9–M12

| Estimator | M9 | | M10 | | M11 | | M12 | |
|---|---|---|---|---|---|---|---|---|
| | 100 | 250 | 100 | 250 | 100 | 250 | 100 | 250 |
| $\hat{v}_{RT}$ | 0.3816 | 0.3641 | 0.9669 | 0.7185 | 0.8591 | 0.8651 | 0.9613 | 0.5494 |
| $\hat{v}_{RT}^{a}$ | 0.3743 | 0.3636 | 1.0115 | 0.8016 | 0.8610 | 0.8721 | 1.0425 | 0.6381 |
| $\hat{v}_{RT}^{g}$ | 0.3744 | 0.3636 | 1.0099 | 0.7987 | 0.8609 | 0.8720 | 1.0268 | 0.6223 |
| $\hat{v}_{LSCV}$ | 0.6122 | 0.2702 | 0.6784 | 0.3328 | 0.7351 | 0.3470 | 0.7786 | 0.3938 |
| $\hat{v}_{LSCV}^{a}$ | 0.5200 | 0.2538 | 0.6553 | 0.3320 | 0.6774 | 0.3502 | 0.7501 | 0.4019 |
| $\hat{v}_{LSCV}^{g}$ | 0.5265 | 0.2547 | 0.6608 | 0.3334 | 0.6852 | 0.3512 | 0.7534 | 0.4016 |
| $\hat{v}_{LCV}$ | 0.6080 | 0.2660 | 0.6087 | 0.3246 | 0.7395 | 0.3381 | 0.7307 | 0.3659 |
| $\hat{v}_{LCV}^{a}$ | 0.5197 | 0.2508 | 0.6051 | 0.3271 | 0.6883 | 0.3474 | 0.7146 | 0.3783 |
| $\hat{v}_{LCV}^{g}$ | 0.5259 | 0.2517 | 0.6072 | 0.3284 | 0.6956 | 0.3480 | 0.7174 | 0.3778 |
| $\hat{v}_{PI}$ | 0.8389 | 0.2678 | 0.8025 | 0.3347 | 0.9218 | 0.3274 | 0.9361 | 0.3593 |
| $\hat{v}_{PI}^{a}$ | 0.7232 | 0.2464 | 0.7422 | 0.3281 | 0.8383 | 0.3269 | 0.8665 | 0.3604 |
| $\hat{v}_{PI}^{g}$ | 0.7334 | 0.2474 | 0.7516 | 0.3298 | 0.8503 | 0.3280 | 0.8822 | 0.3617 |
| $\hat{v}_{BT}$ | 0.3653 | 0.3652 | 1.6663 | 1.5989 | 0.9539 | 0.9547 | 3.0033 | 0.6187 |
| $\hat{v}_{BT}^{a}$ | 0.3653 | 0.3652 | 1.6667 | 1.5994 | 0.9540 | 0.9548 | 3.0067 | 0.6943 |
| $\hat{v}_{BT}^{g}$ | 0.3653 | 0.3652 | 1.6667 | 1.5992 | 0.9540 | 0.9548 | 3.0067 | 0.6805 |
| $\hat{v}_{SCV}$ | 0.3632 | 0.3549 | 1.5497 | 1.2418 | 0.9437 | 0.9185 | 2.9350 | 0.6298 |
| $\hat{v}_{SCV}^{a}$ | 0.3629 | 0.3545 | 1.5546 | 1.2558 | 0.9440 | 0.9189 | 2.9419 | 0.7067 |
| $\hat{v}_{SCV}^{g}$ | 0.3629 | 0.3545 | 1.5546 | 1.2550 | 0.9440 | 0.9190 | 2.9347 | 0.6922 |

## 5.1 Simple models

An advantage of the proposed novel method is its direct derivation from the exact formula of the ISB. The method does not require the selection of hyperparameters, and thus it is a simple and efficient process. For a sample size $n = 100$, the SCV selector is competitive with other selectors for models M2 and M4. However, it does not perform as satisfactorily for models M1 and M3. By contrast, for sample sizes $n = 250$, the SCV selector provides the best results for all the models (M1–M4). It is important to note that even though the SCV selector is based on an exact formula, it still exhibits a certain level of dependence on the sample size. The graphs for models M1 and M3 clearly indicate that the number of observations in the sample have an effect on the SCV selector. In summary, the SCV selector does not demonstrate any consistent tendency to outperform other methods, but it achieves very respectable results in models M2 and M4 and is comparable with the common selectors for all of the models at $n = 250$.

## 5.2 Mixture models

One of the steps in the adaptive method is selection of the sensitivity parameter $\alpha$ and function $\lambda(\cdot)$ for variability in the kernel window. Note that when $\alpha = 0$, the
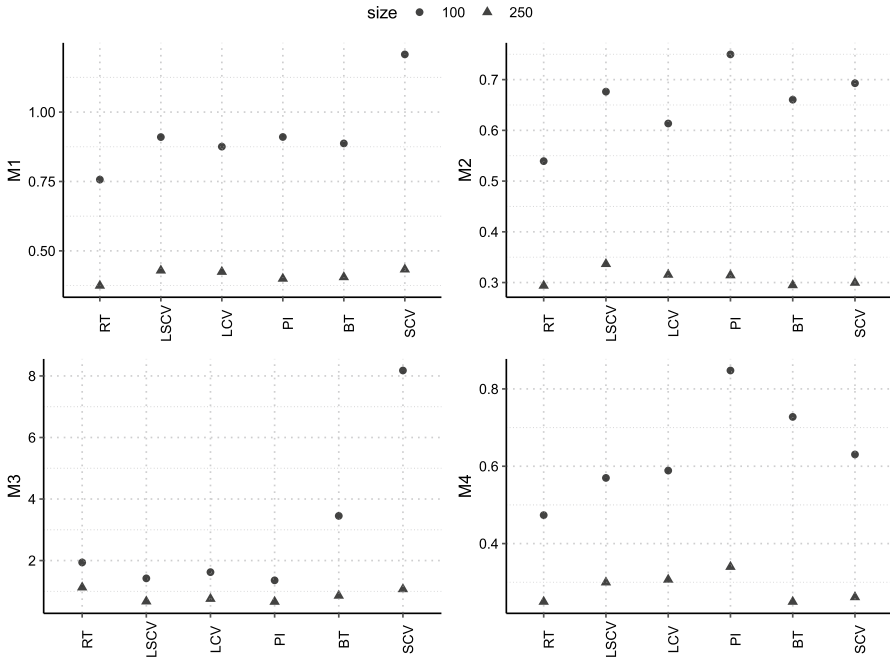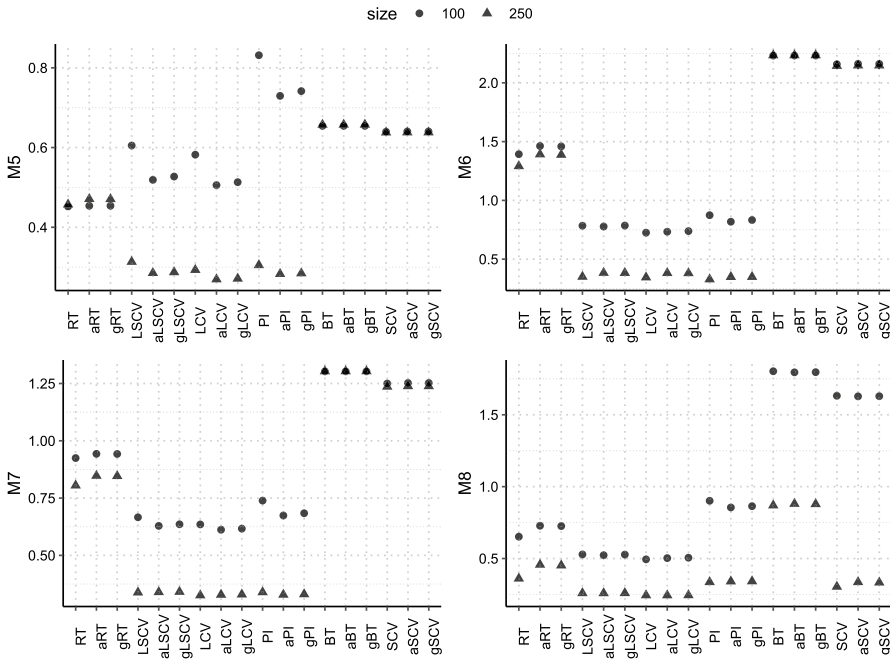
**Fig. 1** Average ISE for models M1–M4 in the simulation



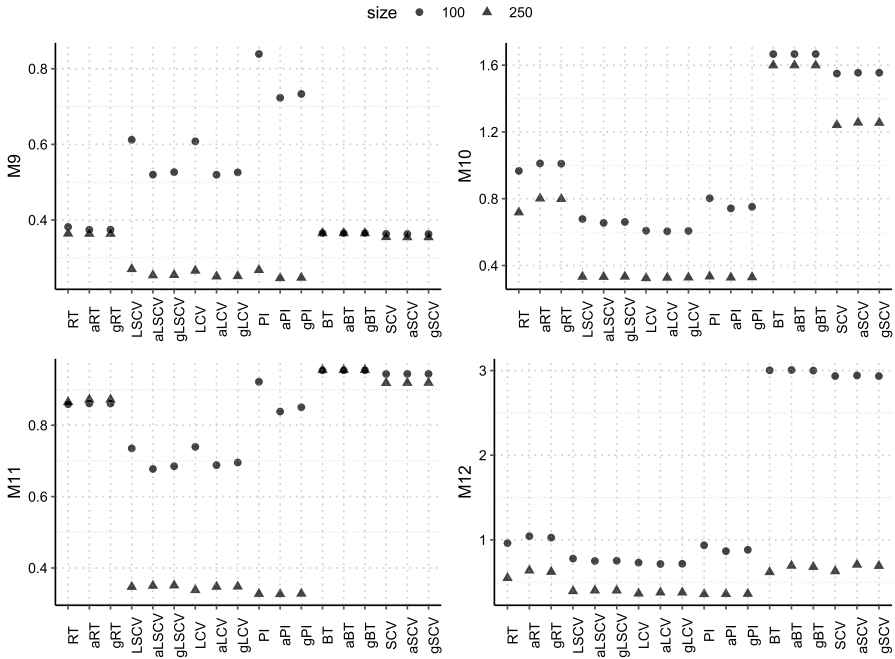**Fig. 2** Average ISE for models M5–M8 in the simulation

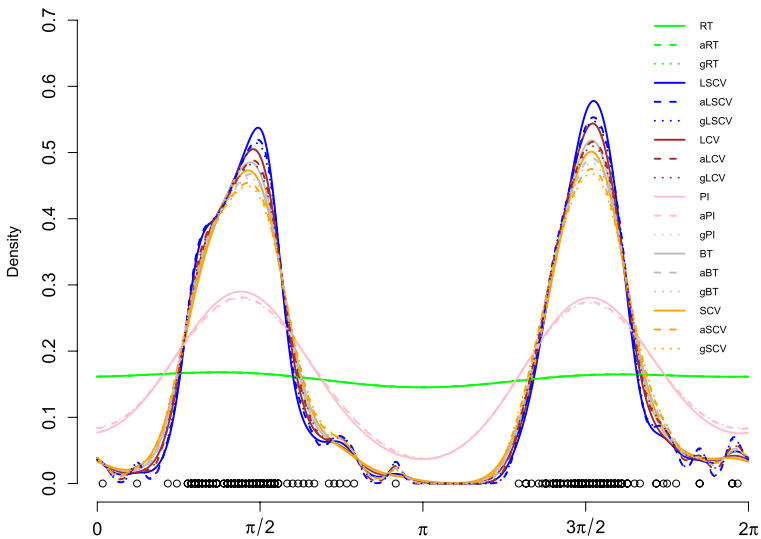**Fig. 3** Average ISE for models M9–M12 in the simulation



**Fig. 4** Comparison of classic and adaptive methods using the dragonfly orientations dataset
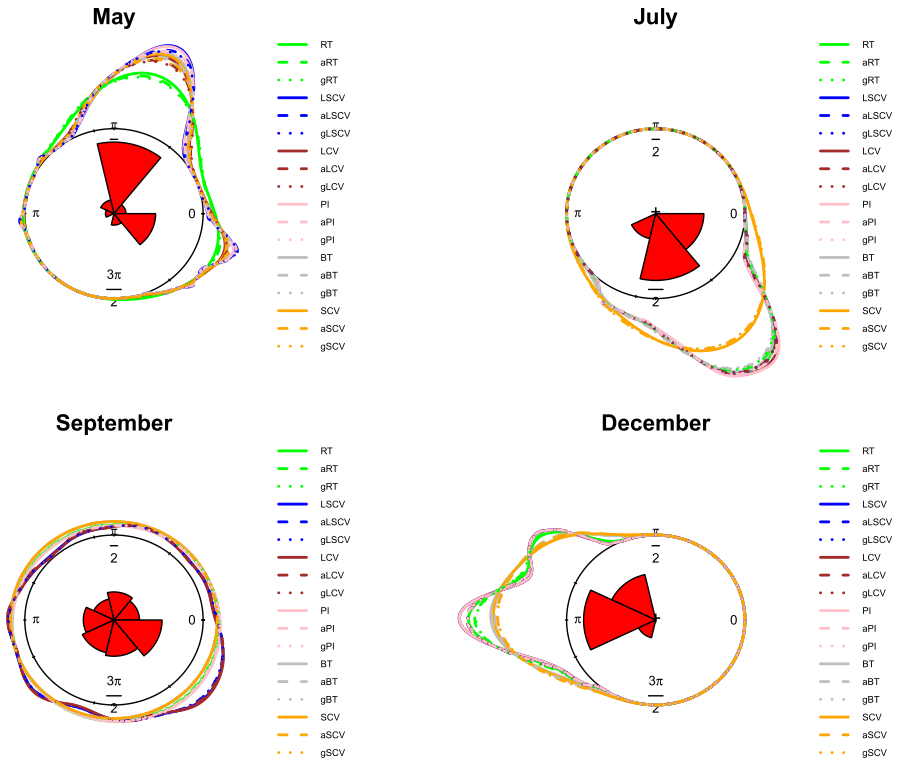
**Fig. 5** Comparison of classic and adaptive methods using the wind direction dataset

adaptive method is reduced to a fixed method, while higher values result in greater sensitivity in the pilot estimate. Using either the geometric or arithmetic mean, the simulation tests different cases for $\lambda(\cdot)$. Different values for $\alpha$ are also used, but the findings are the same as those observed by Wand and Jones (1994): no biased difference could be seen with the selected weight functions applied in every case. This finding is highlighted in Tables 3 and 4 and in Figs. 2 and 3, which compare the average ISEs.

For a sample size $n = 100$, both $v_B^g$ and $v_B^a$ are very competitive. The LSCV, LCV and PI methods especially produce significant differences in performance; adaptive types tend to outperform their fixed bandwidth counterparts and are therefore a preferable choice. Improvements are obtained in mixture models with varying location and distances between modes (see average ISEs for M5–M7 and M9–M11). It is not surprising that if the pilot estimate is poor, in most cases the adaptive method is not able to significantly improve the performance. This behaviour is observed with the BT and SCV methods in most models, indicating a consequence of these methods performing poorly with multimodal data. For a sample size $n = 250$, the results show improvement, but not as significantly as in the previous case. Notably, the best improvement is observed in models with similar concentration parameters and
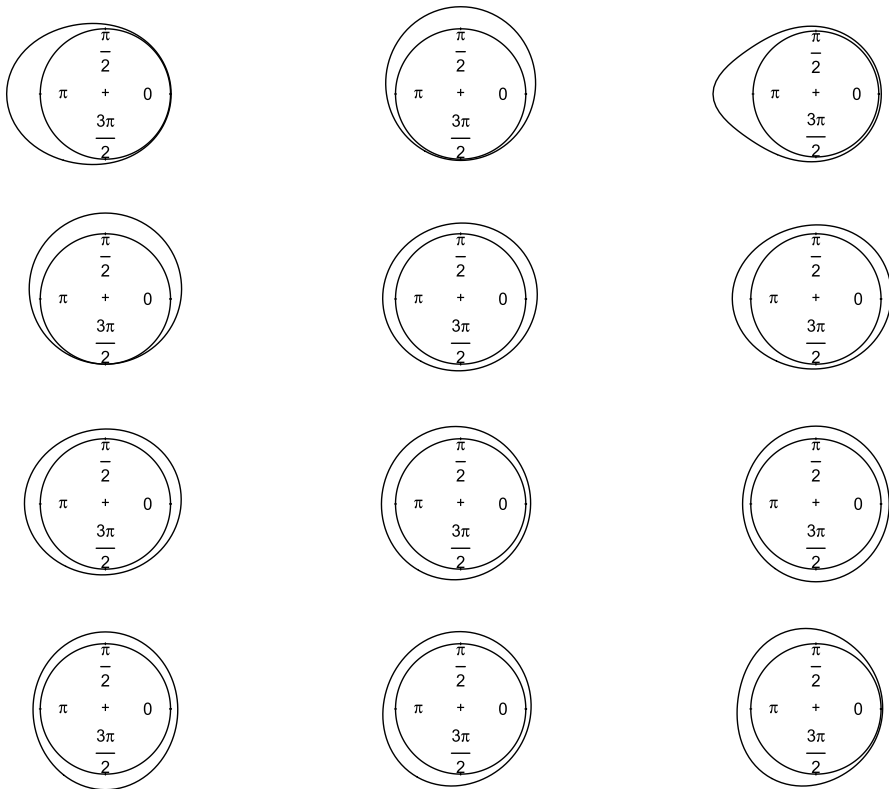
**Fig. 6** Graphs of distributions in the simulation for models M1–M12, in order from left to right, top to bottom

long-distance modes (i.e., models M5 and M9, see their respective ISEs). Although the pilot estimates by themselves are satisfactory in most cases, there is no much room for significant improvement.

In summary, the adaptive procedures significantly improve errors in the LSCV, LCV and PI methods with models M5, M9, M10 and M11 at both $n = 100$ and $n = 250$. More satisfactory behaviour is achieved with smaller sample size; it is therefore recommended that these methods are applied to small to mid-sized datasets. Selection of the weight function does not appear to be crucial, and $\alpha = 1/2$ proves to be the optimal sensitivity parameter.

## 6 Application to real data

The proposed methods are applied to two real datasets: the first contains data on the orientations of dragonflies with respect to solar azimuth, the second contains the wind direction data. Each represents an example of a dataset with different numbers of modes.

All the methods presented in this study, i.e., fixed bandwidths (RT, LSCV, LCV, PI, BT and SCV) and their adaptive counterparts with the geometric mean (*g*) and arithmetic mean (*a*) as factors for variable bandwidth, are applied to two datasets.

**Example 1** *Dragonfly orientation*. This dataset introduced by Batschelet (1981) contains data for dragonfly behaviour and the tendency of these insects to sit in a direction perpendicular to the sun; the data describe 214 such orientations (in degrees of arc) measured in relation to solar azimuth. Figure 4 clearly indicates bimodality in the dataset.

In this example, almost all of the methods exhibit similar performance, see Fig. 4. The exceptions are RT, which produces an estimate that approximates uniform distribution, and PI, which handles bimodality better than RT, but produces an oversmoothed estimate. Methods based on cross-validation and BT handle bimodality relatively well, and their performance can be considered as adequate. The adaptive counterparts of the fixed-bandwidth methods tend to slightly oversmooth the estimate. The proposed SCV method, however, produces a very good resulting estimate compared with the LSCV and LCV methods, whose estimates are partly undersmoothed (see respective estimates near points 0, $\pi$ and $2\pi$).

**Example 2** *Wind directions*. Collected by the Czech Hydrometeorological Institute (CHMI), the dataset contains wind direction measurement performed over approximately ten years (at 10-minute intervals for a total 144 recordings per day) at the Brno-Tuřany airport. For the aims of the current study, the first days of May, July, September and December from the year 2011 were selected for analysis (Fig. 5).

Nearly all of the methods exhibit similar performance. As with the previous example, the adaptive methods produce slightly better estimates and resulting curves which are not excessively undersmoothed, especially around the modes.

Although the RT method produces a flatter estimate with the May data than the other methods, it is still able to discover both modes for wind direction. Using the July data, the SCV method is smoother near the modes than other methods. Using the data for September, a month with changing wind conditions, the estimated densities indicate two groups, one for the RT, PI and SCV methods and the other for the LSCV and LCV methods, which are slightly oversmoothed. Using the December data, RT and PI undersmooth the resulting estimate; the other estimated densities are not as obviously bimodal, although a trace of two modes is still observable.

The selected examples describe two typical instances: the dataset for the first example above is frequently used and ideal for testing purposes; the dataset for the second example was produced in cooperation with the CHMI. Hydrometeorologists study these measurement outputs in a limited, wind rose format only, with a slice arc of 22.5 degrees (i.e., circles with 16 sections). More accurate than the rose diagram, kernel estimates and the resulting densities provide a more detailed view at a resolution of one degree. Consultation with CHMI scientists confirmed that the resulting estimates (the SCV method for July, September and December and the RT method

for May), while being slightly oversmoothed from a statistical point of view, are sufficient for their needs. The kernel density estimates could be used instead of the wind roses to check measurement accuracy at individual measuring stations.

## 6.1 Additional comments regarding computation time

Compared to the commonly used methods, the proposed SCV method for bandwidth selection in circular density estimation produced comparable, and in many cases, even significantly improved results on real datasets and on simulated datasets as well, albeit at a higher computational cost. Computation of SCV bandwidth using the wind direction dataset required 9.57 s on average, whereas with other methods, it was less than 1 s (e.g., BT required 0.91 s to run the code multiple times in a loop). This difference was probably due to the necessity for numerical integration when evaluating the ISB formula (3.8) (application of a simple trapezoidal rule with partitions of 100 equally spaced intervals). To minimize the objective function (3.6) and also the objective functions of other methods (maximization was applied to LCV), the optimize function available in R was used. The computational cost for SCV in the simulations was 9 s for $n = 100$ (similar size as the above mentioned dataset), 45 s for $n = 250$ on average, which was a clear yet unsurprising indication that computational time increases with larger sample sizes. By contrast, comparable results in terms of average ISE were achieved with lower failure rates during the optimization process: with model M1, SCV did not fail in any sample, whereas LSCV failed in 7 of total 200 samples, and BT failed in 12. Similar results were obtained also with the other simple models, with the exception of model M3, in which SCV sometimes failed and overall exhibited poor performance.

The adaptive methods for bandwidth selection discussed in the second part of Sect. 5 behaved satisfactorily in each simulation case and in both real datasets at a computational cost comparable to fixed bandwidth methods. Compared to fixed bandwidth methods, Steps 2 and 3 of the algorithm (see p. 7) for adaptive methods are supplemental but should not be computationally too expensive, at least on paper. For example, model M5 required 5.4 s with the (fixed) RT method at $n = 250$ (0.83 s at $n = 100$); the adaptive method with the RT selector required 8.32 s (1.52 s at $n = 100$). The failure rates of the adaptive methods equalled the failure rates of their fixed bandwidth counterparts, as both depended on the same optimization process (Step 1 of the adaptive algorithm).

## 7 Conclusion

The current study presented a novel smoothed cross-validation SCV (3.7) method for circular kernel density estimation and investigated common methods based on cross-validation. The study also presented a method to transform adaptive methods, typically used with linear data, for use with circular data (see (4.1)) and discussed the potential to improve on the results produced by their fixed-bandwidth counterparts.

Simulations of the presented methods using real datasets yielded competitive results in relation to the common estimators applied in the field. In some cases, the simulations which used small samples produced unsatisfactory results, but with larger samples, the distance from the true density in the resulting estimators decreased. The simulation results also demonstrated that the proposed adaptive method which applied an arithmetic mean often outperformed existing methods. The SCV method performed well with larger datasets, although with small datasets, a dependence on the sample size and the shape of the true density was evident. Small to mid-sized datasets are therefore recommended for mixture models; for simple models and real datasets, the SCV method produced more satisfactory results with larger datasets.

Future research directions include extending the proposed methods for application in the multivariate settings, i.e., for spherical data. Other research paths would be closer collaboration with CHMI, combining the circular and linear approaches, and investigating methods for the so-called directional-linear data, which are useful in hydrometeorological studies.

## 8 Appendix

### 8.1 Properties of Bessel functions

The properties of Bessel functions are useful in deriving the methods introduced in this study.

The modified Bessel function of the first kind and order $r$,

$$I_r(v) = \frac{1}{2\pi} \int_0^{2\pi} e^{v \cos \alpha} \cos(r\alpha) d\alpha, \tag{8.1}$$

especially for $r = 0$, provides a normalizing constant for the von Mises distribution (2.2). Substituting $a = v \cos \theta$, $b = v \sin \theta$ into the following integral and using the relationship (8.1), we obtain

$$\frac{1}{2\pi} \int_0^{2\pi} e^{a \cos \alpha + b \sin \alpha} d\alpha$$
$$= \frac{1}{2\pi} \int_0^{2\pi} e^{v \cos(\alpha - \theta)} d\alpha \tag{8.2}$$
$$= I_0(v) = I_0(\sqrt{a^2 + b^2}).$$

Other useful properties of modified Bessel function are:

$$I_{-r}(v) = I_r(v),$$
$$I_{r-1}(v) + I_{r+1}(v) = 2I_r'(v),$$
$$I_{r-1}(v) - I_{r+1}(v) = \frac{2r}{v}I_r(v),$$
$$\frac{\mathrm{d}}{\mathrm{d}v}I_0(v) = I_0'(v) = I_1(v),$$

where $I_r'(v)$ is the derivative of the Bessel function.

Let us suppose that $\theta_1 \sim VM(\mu_1, \kappa_1)$ and $\theta_2 \sim VM(\mu_2, \kappa_2)$. Using (8.2), the convolution of two von Mises densities is

$$g(\theta) = \frac{1}{4\pi^2 I_0(\kappa_1)I_0(\kappa_2)}$$
$$\times \int_0^{2\pi} \exp\{\kappa_1 \cos(\alpha - \mu_1) + \kappa_2 \cos(\theta - \alpha - \mu_2)\}\mathrm{d}\alpha$$
$$= \frac{1}{4\pi^2 I_0(\kappa_1)I_0(\kappa_2)} \int_0^{2\pi} \exp\{[\kappa_1 \cos\mu_1 + \kappa_2 \cos(\theta - \mu_2)]\cos\alpha$$
$$+ [\kappa_1 \sin\mu_1 + \kappa_2 \sin(\theta - \mu_2)]\sin\alpha\}\mathrm{d}\alpha$$
$$= \frac{1}{2\pi I_0(\kappa_1)I_0(\kappa_2)}I_0\left(\sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2 \cos(\theta - (\mu_1 - \mu_2))}\right)$$

(Jammalamadaka and SenGupta 2001).

## 8.2 Models

The models applied in the simulation discussed in Sect. 5 are described below. The corresponding graph for each model is given in Fig. 6. The simulation used a combination of simple models M1–M4 and two-component mixture models M5–M12.

*Von Mises* distribution $VM(\mu, \kappa)$ is a symmetric and unimodal distribution characterized by two parameters: $\mu$ describes the mean direction and $\kappa$ is a concentration parameter (Mardia and Jupp 2009, Section 3.5.4). Model M1 used $VM(\pi, 2)$, but the compositions of two von Mises distributions were also considered, i.e.,

M5: $\frac{1}{2}VM(0.5, 1) + \frac{1}{2}VM(\pi, 1)$,
M6: $\frac{1}{2}VM(0.5, 1) + \frac{1}{2}VM(\pi, 2)$,
M7: $\frac{1}{2}VM(0.5, 1) + \frac{1}{2}VM(\pi, 1.5)$,
M8: $\frac{1}{2}VM(2, 1) + \frac{1}{2}VM(4, 1)$,
M9: $\frac{1}{2}VM(\pi/2, 1) + \frac{1}{2}VM(3\pi/2, 1)$,
M10: $\frac{1}{2}VM(\pi/2, 1) + \frac{1}{2}VM(3\pi/2, 1.5)$,
M11: $\frac{1}{2}VM(\pi/2, 1) + \frac{1}{2}VM(4, 1.5)$,
M12: $\frac{1}{2}VM(2, 2) + \frac{1}{2}VM(4, 2)$.

*Wrapped normal* distribution $WN(\mu, \rho)$ is obtained by wrapping the normal distribution $N(\mu, \sigma^2)$ onto a circle, where $\rho = e^{\sigma^2/2}$ ( Mardia and Jupp 2009, Section 3.5.7). It is unimodal and symmetric around its mean $\mu$. Model M2 used $WN(\pi/2, 0.5)$.

*Wrapped Cauchy* distribution $WC(\mu, \rho)$ is unimodal and symmetric around $\mu$ (Mardia and Jupp 2009, Section 3.5.7). Model M3 used $WC(\pi, 0.6)$.

*Cardioid* distribution $C(\mu, \rho)$ is produced as a perturbation of the uniform density by a cosine function. It is a symmetric and unimodal distribution with mode at $\mu$ (Mardia and Jupp 2009, Section 3.5.5). Model M4 used $C(\pi/2, 0.5)$.

# References

Batschelet E (1981) Circular statistics in biology, mathematics in biology. Academic Press, London

Breiman L, Meisel W, Purcell E (1977) Variable kernel estimates of multivariate densities. Technometrics 19(2):135–144. https://doi.org/10.2307/1268623

Cao R, Cuevas A, González Manteiga W (1994) A comparative study of several smoothing methods in density estimation. Comput Stat Data Anal 17(2):153–176. https://doi.org/10.1016/0167-9473(92)00066-Z

Demir S (2018) Adaptive Kernel density estimation with generalized least square cross-validation. Hacettepe J Math Stat 48:616–625. https://doi.org/10.15672/HJMS.2018.623

Di Marzio M, Fensore S, Panzera A, Taylor CC (2019) Kernel density classification for spherical data. Stat Probab Lett 144:23–29. https://doi.org/10.1016/j.spl.2018.07.018

Di Marzio M, Panzera A, Taylor CC (2011) Density estimation on the torus. J Statist Plann Inference 141:2156–2173. https://doi.org/10.1016/j.jspi.2011.01.002

Fisher NI (1995) Statistical analysis of circular data. Cambridge University Press, Cambridge

García-Portugués E (2013) Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. Electron J Stat 7:1655–1685. https://doi.org/10.1214/13-EJS821

García-Portugués E, Crujeiras RM, Wenceslao GM (2013) Kernel density estimation for directional-linear data. J Multivar Anal 121:152–175. https://doi.org/10.1016/j.jmva.2013.06.009

Hall P, Marron J, Park BU (1992) Smoothed cross-validation. Probab Theory Relat Fields 92(1):1–20. https://doi.org/10.1007/BF01205233

Hall P, Watson G, Cabrera J (1987) Kernel density estimation with spherical data. Biometrika 74(4):751–762. https://doi.org/10.1093/biomet/74.4.751

Härdle WK, Müller M, Sperlich S, Werwatz A (2012) Nonparametric and semiparametric models. Springer, Heidelberg

Horová I, Koláček J, Zelinka J (2012) Kernel smoothing in MATLAB: theory and practice of Kernel smoothing. World Scientific, Singapore

Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics, of series on multivariate analysis, vol 5. World Scientific, Singapore

Jones M, Kappenman R (1992) On a class of Kernel density estimate bandwidth selectors. Scand J Stat 19:337–349

Ley C, Verdebout T (2017) Modern directional statistics. Chapman and Hall, New York

Ley C, Verdebout T (2018) Applied directional statistics: modern methods and case studies. Chapman and Hall, New York

Mardia K (2021) Comments on: recent advances in directional statistics. TEST 30:1–5. https://doi.org/10.1007/s11749-021-00760-4

Mardia KV, Jupp PE (2009) Directional statistics, of Wiley Series in Probability and Statistics. Wiley, Chichester

Marron JS, Ruppert D (1994) Transformations to reduce boundary bias in kernel density estimation. J R Stat Soc Series B (Methodological) 56(4):653–671. https://doi.org/10.2307/2346189

Oliveira M, Crujeiras RM, Rodríguez-Casal A (2012) A plug-in rule for bandwidth selection in circular density estimation. Comput Stat Data Anal 56(12):3898–3908. https://doi.org/10.1016/j.csda.2012.05.021

Oliveira M, Crujeiras RM, Rodríguez-Casal A (2013) Nonparametric circular methods for exploring environmental data. Environ Ecol Stat 20:1–17. https://doi.org/10.1007/s10651-012-0203-6

Park BU, Marron JS (1990) Comparison of data-driven bandwidth selectors. J Am Stat Assoc 85(409):66–72. https://doi.org/10.2307/2289526

Pewsey A (2000) The wrapped skew-normal distribution on the circle. Commun Stat-Theory Methods 29(11):2459–2472. https://doi.org/10.1080/03610920008832616

Pham Ngoc TM (2019) Adaptive optimal Kernel density estimation for directional data. J Multivar Anal 173:248–267. https://doi.org/10.1016/j.jmva.2019.02.009

R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Scott DW (1992) Multivariate density estimation: Theory, practice, and visualization. Wiley, New York

Scott DW, Terrell GR (1987) Biased and unbiased cross-validation in density estimation. J Am Stat Assoc 82(400):1131–1146. https://doi.org/10.2307/2289391

Silverman BW (1986) Density estimation for statistics and data analysis, of monographs on statistics and applied probability, vol 26. Chapman and Hall, London

Simonoff JS (2012) Smoothing methods in statistics. Springer, New York

Taylor CC (2008) Automatic bandwidth selection for circular density estimation. Comput Stat Data Anal 52(7):3493–3500. https://doi.org/10.1016/j.csda.2007.11.003

Tenreiro C (2022) Kernel density estimation for circular data: a Fourier series-based plug-in approach for bandwidth selection. J Nonparametric Stat 34(2):377–406. https://doi.org/10.1080/10485252.2022.2057974

Tsuruta Y, Sagae M (2017) Higher order kernel density estimation on the circle. Stat Probab Lett 131:46–50. https://doi.org/10.1016/j.spl.2017.08.003

Wand MP, Jones MC (1994) Kernel smoothing. Chapman and Hall, London