On nonparametric regression for bivariate circular long-memory time series

Jan Beran · Britta Steffens · Sucharita Ghosh

Received: 16 June 2020 / Revised: 8 February 2021 / Accepted: 23 February 2021 / Published online: 20 March 2021
© The Author(s) 2021

Abstract
We consider nonparametric regression for bivariate circular time series with long-range dependence. Asymptotic results for circular Nadaraya–Watson estimators are derived. Due to long-range dependence, a range of asymptotically optimal bandwidths can be found where the asymptotic rate of convergence does not depend on the bandwidth. The result can be used for obtaining simple confidence bands for the regression function. The method is illustrated by an application to wind direction data.

Keywords Circular time series · Circular circular kernel regression · Long-range dependence · Gaussian subordination · Confidence interval

1 Introduction
Directional or circular time series arise in many scientific fields such as meteorology, oceanography, biology, neuroscience, bioinformatics, geoscience and cosmology. An overview of directional data analysis can be found for instance in books by Mardia (1972), Fisher (1993), Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001) and Ley and Verdebout (2017). Most literature deals with iid observations. More generally, for cases where more than one circular variable is observed, parametric and nonparametric circular-circular regression with iid residuals is discussed for instance in Gould (1969), Fisher and Lee (1992), Johnson and Wehrly (1978), Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), Kato et al. (2008), Kim and SenGupta (2016) and Polsen and Taylor (2015). Nonparametric density estimation on the circle is discussed for instance by Watson (1983), Hall et al. (1987), Bai et al. (1988), Fisher (1989, 1993) and Taylor (2008). Also see Tsurata and Sagae (2017) for properties of higher order circular kernels. In practice, one often observes circular time series that
exhibit serial dependence. Circular processes with weak serial dependence are considered for instance in Wehrly and Johnson (1980), Breckling (1989), Kato (2010), Modlin et al. (2012) and Wang and Gelfand (2014). Circular time series with long-range dependence are considered in Di Marzio et al. (2012) and Beran and Ghosh (2019). Di Marzio et al. (2012) discuss nonparametric trend estimation, Beran and Ghosh (2019) define models with long-range dependence using Gaussian subordination, and derive asymptotic results for parametric estimators where the mean direction depends on deterministic explanatory variables. In this paper, we extend the results in Beran and Ghosh (2019) to a nonparametric circular-circular regression model of the form

$$\vartheta_t = [\mu(\psi_t) + Z_t] \mod 2\pi$$

where $\vartheta_t$, $\psi_t \in [-\pi, \pi)$ ($t = 1, 2, \ldots$), $\psi_t$ and $Z_t$ are stationary long-memory processes defined by Gaussian subordination, and $Z_t$ is such that

$$E(\sin Z_t) = 0, E(\cos Z_t) = R > 0$$

(see e.g. Gould 1969). Equations (1) and (2) mean that, given $\psi_t = w^0$, the mean direction of $\vartheta_t$ is $\mu(w^0)$, i.e.

$$E \left[ \exp(i\vartheta_t) \mid \psi_t = w^0 \right] = R \exp \left( i\mu(w^0) \right) \quad (w^0 \in [-\pi, \pi]).$$

We consider asymptotic properties of circular kernel estimators of $\mu(w^0)$. Limit theorems are derived using in particular general results in Mielniczuk and Wu (2004). Due to long-range dependence, a range of asymptotically optimal bandwidths can be found where the asymptotic rate of convergence does not depend on the bandwidth. The results can be used for obtaining simple simultaneous confidence bands for $\mu(w_{1}^0), \ldots, \mu(w_{m}^0)$ ($w_{1}^0, \ldots, w_{m}^0 \in [-\pi, \pi], m \in \mathbb{N}$).

The paper is organized as follows. A general introduction and definitions are given in Sect. 2. Limit theorems are discussed in Sect. 3. In Sect. 4, asymptotic results from Sect. 3 are used to obtain confidence bands for $\mu(\psi)$. An application to wind direction data is discussed in Sect. 5. Proofs are given in the Appendix.

### 2 Definitions

#### 2.1 Definition of the model

We extend the model introduced in Beran and Ghosh (2019) to bivariate circular time series with long-range dependence. First we recall the definition of long memory in the sense of second order dependence. A real-valued second order stationary time series with autocovariance function $\gamma(k)$ is said to exhibit long-range (or strong) dependence, if $\sum \gamma(k) = \infty$. Often it is assumed that

$$\gamma(k) \sim c\gamma k^{2H-2} \quad (k \to \infty),$$

where $H \in (0, 1)$.
or that the spectral density has a pole at the origin characterized by

\[
f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda} \sim c_f \left| \lambda \right|^{-2H}
\]

with \( H \in (\frac{1}{2}, 1) \) and \( 0 < c_f < \infty \). Here, \( \sim \) means that the ratio of the left and right hand side converges to 1. References to the extended literature on long-memory processes can be found for instance in Beran (1994), Giraitis et al. (2012) and Beran et al. (2013). For generating schemes for long memory processes also see e.g. Davidson and Sibbertsen (2005).

To obtain bivariate circular time series \((\psi_t, \vartheta_t) \in [0, 2\pi)^2 (t \in \mathbb{Z})\) with long-range dependence, the following assumptions will be used:

- (A1) Let \( Z_t, X_t (t \in \mathbb{Z}) \) be stationary Gaussian processes with \( E(Z_t) = E(X_t) = 0, 0 < \sigma_Z^2 = \text{var}(Z_t) < \infty, \sigma_X^2 = \text{var}(X_t) = 1, \) and autocovariance functions \( \gamma_Z, \gamma_X \) and spectral densities \( f_Z, f_X \) such that

\[
\gamma_Z(k) \sim \frac{c_{\gamma,Z}}{k \to \infty} e^{-ik\lambda} \sim c_{f,Z} \left| \lambda \right|^{-2H_Z},
\]

\[
f_Z(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Z(k) e^{-ik\lambda} \sim c_{f,Z} \left| \lambda \right|^{-2H_Z},
\]

\[
\gamma_X(k) \sim \frac{c_{\gamma,X}}{k \to \infty} e^{-ik\lambda} \sim c_{f,X} \left| \lambda \right|^{-2H_X},
\]

\[
f_X(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-ik\lambda} \sim c_{f,X} \left| \lambda \right|^{-2H_X}
\]

for some constants \( 0 < c_{\gamma,Z}, c_{f,Z}, c_{\gamma,X}, c_{f,X} < \infty \) and \( \frac{1}{2} < H_Z, H_X < 1 \). Moreover, assume that the two processes are independent from each other.

- (A2) Let \( G_\psi (u) = \int_{-\pi}^u g_\psi (v) dv \) \((u \in [-\pi, \pi])\) be an absolutely continuous circular cumulative distribution function with density \( g_\psi = G_\psi' \), and denote by \( \Phi \) the standard normal distribution function. Define the circular time series \( \psi_t (t \in \mathbb{Z}) \) by

\[
\psi_t = G_\psi^{-1} (\Phi (X_t)).
\]

- (A3) Let \( \mu : [-\pi, \pi) \rightarrow \mathbb{R} \) be a twice continuously differentiable function. The circular time series \( \vartheta_t (t \in \mathbb{Z}) \) is defined by

\[
\vartheta_t = ([\mu(\psi_t) + Z_t] \mod 2\pi).
\]

**Remark 1** Assumption (A1) implies

\[
E(\sin Z_t) = 0, E(\cos Z_t) = R_Z > 0,
\]

\[
E(\sin X_t) = 0, E(\cos X_t) = R_X > 0.
\]
Note also that $Z_t$ and $X_t$ can be written in the Wold representation with iid innovations $\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$ and $\eta_t \sim N(0, \sigma^2_\eta)$ respectively, as

$$Z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad X_t = \sum_{j=0}^{\infty} c_j \eta_{t-j} \quad (t \in \mathbb{Z})$$

(14)

where $a_0 = c_0 = 1$ and

$$a_j \sim C_{a_j} H^{3/2}, \quad c_j \sim C_{c_j} H^{3/2} \quad (j \to \infty)$$

(15)

for some suitable constants $0 < C_a, C_c < \infty$.

**Remark 2** The circular time series $\psi_t = G^{-1}_\psi(\Phi(X_t))$ is said to be subordinated to the Gaussian process $X_t$ (see e.g. Rosenblatt 1961, 1979; Taqqu 1975, 1979; Dobrushin and Major 1979; Dobrushin 1980). By definition, the marginal distribution of the series is the circular distribution $G_\psi$. Note that the circular density function $g_\psi = G'_\psi$ is arbitrary, i.e. using (10) any circular density function can be obtained as a marginal density via Gaussian subordination.

Assumption (8) means that $X_t$ exhibits long-range dependence. For the circular time series $\psi_t$, a different definition of autocovariance and autocorrelation function has to be used. Let $\nu_\psi$ denote the mean direction of $\psi_t$, i.e. $E[\exp(i \psi_t)] = R \exp(i \nu_\psi)$. Jammalamadaka and Sarma (1988) proposed the circular autocorrelation function

$$\rho_{\text{circular}}(k) = \frac{E \left[ \sin \left( \psi_t - \nu_\psi \right) \sin \left( \psi_{t+k} - \nu_\psi \right) \right]}{E \left[ \sin^2 \left( \psi_1 - \nu_\psi \right) \right]}.$$  

(16)

To see whether long memory of $X_t$ is inherited by $\psi_t$, one needs to introduce the notion of Hermite rank. Denote by $\varphi(x) = (2\pi)^{-1/2} \exp(-1/2 x^2)$ the standard normal density function, and by $L^2(\mathbb{R}; \varphi)$ the space of real valued functions $h$ with $||h||^2 = \int |h(z)|^2 \varphi(z) dz < \infty$. Equipped with the scalar product $<h, \tilde{h}> = \int h(x)\tilde{h}(x)\varphi(x)dx$, $L^2(\mathbb{R}; \varphi)$ is a Hilbert space. Hermite polynomials

$$\mathcal{H}_q(z) = (-1)^q \exp \left( \frac{1}{2} z^2 \right) \frac{d^q}{dz^q} \exp \left( -\frac{1}{2} z^2 \right) \quad (q = 0, 1, \ldots)$$

build an orthogonal basis in $L^2(\mathbb{R}; \varphi)$. Thus, every function $h \in L^2(\mathbb{R}; \varphi)$ has an orthogonal $L^2$-representation

$$h(z) = \sum_{q=0}^{\infty} \frac{a_q}{q!} \mathcal{H}_q(z),$$

where the $q$th Hermite coefficient $a_q$ is given by

$$a_q = \langle h, \mathcal{H}_q \rangle = \int_{-\infty}^{\infty} h(z) \mathcal{H}_q(z) \varphi(z) dz.$$
A function $h$ with $a_0 = 0$ is said to have Hermite rank $m \geq 1$, if $a_m \neq 0$ and $a_j = 0$ ($j < m$) (Taqqu 1975). The following Lemma is derived in Beran and Ghosh (2019):

**Lemma 1** Let $\psi_t$ ($t \in \mathbb{Z}$) be defined in (10) with $\gamma_X$ characterized by (8). Denote by $\nu_\psi$ the mean direction of $\psi_t$, and define

$$h(x) = \sin \left\{ G^{-1}_\psi (\Phi (x)) - \nu_\psi \right\}.$$ 

Suppose that the Hermite rank of $h$ is $m$. Furthermore assume that

$$H_X > 1 - \frac{1}{2m}. \quad (17)$$

Then, setting

$$H_m := 1 + (H_X - 1) m, \quad (18)$$

we have

$$\rho_{\text{circular}} (k) \sim c_m k^{2H_m - 2} \quad (19)$$

where

$$c_m = \frac{a_m^2 c^{m}_{\gamma Z}}{m! E \left[ \sin^2 (\vartheta_1 - \mu) \right]}.$$ 

**Remark 3** Note that, by definition of the mean direction $\nu_\psi$,

$$E [h (X_t)] = E \left[ \sin \left\{ G^{-1}_\psi (\Phi (X_t)) - \nu_\psi \right\} \right] = 0.$$ 

Lemma 1 means that, under assumption (17), the directional series $\psi_t$ has long-range dependence in the sense that its circular autocorrelation $\rho_{\text{circular}} (k)$ has a hyperbolic decay and is not summable.

As for the second circular series $\vartheta_t$, (11) implies

$$\vartheta_t = \left[ \mu \left( G^{-1}_\psi (\Phi (X_t)) \right) + Z_t \right] \mod 2\pi \quad (20)$$

$$= \left[ \tilde{\mu} (X_t) + Z_t \right] \mod 2\pi \quad (21)$$

where

$$\tilde{\mu} (x) = \mu \left( G^{-1}_\psi (\Phi (x)) \right).$$

Since the processes $X_t$ and $Z_t$ are stationary, $\vartheta_t$ is stationary with a constant mean direction $\nu_\vartheta$. By definition, the conditional mean direction of $\vartheta_t$ given $\psi_t = w^0$ is equal to $\mu (w^0)$ (see (3)). Moreover, (20) implies that long-range dependence in $Z_t$ leads to long-range dependence in $\vartheta_t$. 

\( \Box \) Springer
2.2 Kernel estimation of the conditional mean direction

Let \( w^0 \in [0, 2\pi) \). The conditional mean direction \( \mu(w^0) \) of \( \vartheta_t \) can be estimated by a circular kernel estimator. In the context of iid observations, circular kernel estimators have been discussed for instance by Hall et al. (1987), Fisher (1989), Jammalamadaka and SenGupta (2001), Taylor (2008), Tsurata and Sagae (2017); Tsurata and Sagae (2020), Bedouhene and Zougab (2020). Here, we will consider Nadaraya–Watson estimators of the form

\[
\hat{\mu}_n(w^0) = \frac{\sum_{s=1}^{n} K_b \left( \psi_s - w^0 \right) \vartheta_s}{\sum_{s=1}^{n} K_b \left( \psi_s - w^0 \right)}.
\]  

(22)

For the kernel function \( K_b \) we adopt the definition introduced in Hall et al. (1987) and Tsurata and Sagae (2017) among others.

**Definition 1** Let \( L_\kappa(u) = L(\kappa(1 - \cos u)) \) and \( C_\kappa(L) = \int_{-\pi}^{\pi} L_\kappa(u) du \). Then, setting \( \kappa = b^{-2} (b > 0) \), we define circular kernels

\[
K_b(u) = C_\kappa^{-1}(L) L_\kappa(u) \ (u \in [-\pi, \pi]).
\]  

(23)

Moreover, the \( l \)th moment of \( L \) is defined by

\[
m_l(L) = \int_0^\infty L(r) r^{(l+1)/2} dr \ (l = 2k, k \in \{0, 1, 2, \ldots\} \).
\]  

(24)

**Remark 4** Note that by definition \( \int_{-\pi}^{\pi} K_b(u) du = 1 \). Thus, if \( L \geq 0 \), then \( K_b \) is a circular density function.

The following assumptions on \( L \) are used in Tsurata and Sagae (2017):

- (L1) \( L'(r) = dL/dr \) is continuous.
- (L2)
  \[
  \int_0^\infty L^2 \left( \frac{x^2}{2} \right) dx < \infty, \ \int_0^\infty L^2 \left( \frac{x^2}{2} \right) x^2 dx < \infty.
  \]
- (L3) For \( j_0 = 2k_0 \), we have \( m_l(L) < \infty \ (0 \leq l \leq j_0) \) and
  \[
m_l(L) = \int_0^y L(r) r^{(l-1)/2} dr + O \left( y^{-(j_0+1)/2} \right) \ (0 \leq l \leq j_0).
\]
- (L4)
  \[
  \lim_{r \to \infty} L(r) r^{(j_0+1)/2} = 0.
  \]
Note that, under assumptions (L1) to (L4), $K_b$ has a Fourier series representation

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(b) \exp(-isu) \quad (u \in [-\pi, \pi))$$  \hspace{1cm} (25)

with

$$\alpha_{-s}(b) = \alpha_s(b) = \int_{-\pi}^{\pi} K_b(u) \exp(isu) du.$$  

The order of a circular kernel is defined as follows (Tsurata and Sagae 2017):

**Definition 2** Let $p = 2k$ for some $k \in \{1, 2, \ldots\}$. Then $K_b(u) = C_k^{-1}(L) L_\kappa(u)$ with $\kappa = b^{-2}$ is called a circular kernel of order $p$, if (L1) to (L4) hold with $j_0 \geq p + 2$, and

$$m_0(L) \neq 0, m_l(L) = 0 \quad (l = 2, 4, \ldots, p - 2), \quad m_p(L) \neq 0.$$  

**Remark 5** Tsurata and Sagae (2017) discuss methods for constructing higher order circular kernels from lower order kernels. For instance, let

$$L_2(r) = \exp(-r), \quad L_2^{(1)}(r) = \frac{d}{dr} L_2(r),$$  

$$r_\kappa(u) = \kappa (1 - \cos u),$$  

$$L_{\kappa,2}(u) = L_2(r_\kappa(u)) = \exp(-\kappa (1 - \cos u)),$$

and denote by $C_k(L_{k,2})$ the corresponding normalizing constant. Then

$$m_l(L) = \int_{0}^{\infty} L_2(r) r^{(l-1)/2} dr = \Gamma\left(l + \frac{1}{2}\right) \quad (l = 2k, k \in \{0, 1, 2, \ldots\}).$$  

(Tsurata and Sagae 2017). In particular, $m_0(L_2) = \Gamma(\frac{1}{2}) = \sqrt{\pi}$, and

$$m_2(L_2) = \Gamma\left(2 + \frac{1}{2}\right) = \frac{3}{4} \sqrt{\pi}.$$

Thus, $K_{b,2}(u) = C_k^{-1}(L_{k,2}) L_{\kappa,2}(u)$ with $\kappa = b^{-2}$ is a circular kernel of order 2. Now, let

$$L_4(r) = \frac{3}{2} L_2(r) + r L_2^{(1)}(r),$$

set $L_{\kappa,4}(u) = L_4(r_\kappa(u))$ and denote by $C_k(L_{k,4})$ the corresponding normalizing constant. Then

$$K_{b,4}(u) = C_k^{-1}(L_{k,4}) L_{\kappa,4}(u)$$
is a kernel of order 4. Similarly, kernels $K_{b,p+2}$ of order $p+2$ can be obtained recursively by setting

$$L_{p+2}(r) = \frac{p+1}{p} L_p(r) + \frac{2}{p} r L_p^{(1)}(r) \quad (p = 2, 4, 6, \ldots)$$

$$L_{\kappa,p+2}(u) = L_{p+2}(r_k(u)) \quad \text{and}$$

$$K_{b,p+2}(u) = C_k^{-1} \left( L_{\kappa,p+2} \right) L_{\kappa,p+2}(u).$$

### 3 Asymptotic results

The asymptotic rate of convergence of $\hat{\mu}_n(w^0) - \mu(w^0)$ is characterized by the following decomposition:

**Theorem 1** Let $w^0 \in [-\pi, \pi)$, suppose that (A1) to (A3) and (L1) to (L4) hold and $K_b$ is a kernel of order $p = 2k$. Furthermore assume that $\tilde{\mu}(x) = \mu \left( G^{-1}_\psi(\Phi(x)) \right)$ is $p + 2$ times continuously differentiable, the derivatives $\tilde{\mu}^{(s)}$ and $g_\psi^{(s)} (s = 1, 2, \ldots, p)$ are square integrable, $g_\psi(w) > 0$ ($w \in [-\pi, \pi)$) and

$$n \to \infty, b \to 0, nb \to \infty.$$

Then,

$$\hat{\mu}_n(w^0) - \mu(w^0) = O \left( b^{2k} \right) + O_p \left( \frac{1}{\sqrt{nb}} \right) + O_p \left( n^{H_2-1} \right) + O_p \left( b^2 n^{H_2-1} \right).$$

(26)

**Remark 6** The first term in (26) is deterministic. It is due to the difference $\mu_n(w^0) - \mu(w^0)$ where $\mu_n(w^0) = E[K_b(\psi - w^0) \delta]/E[K_b(\psi - w^0)]$. The order of this error is smaller the higher the order of the kernel $K_b$ is.

Equation (26) can be used to address the question of optimal bandwidth choice. Note first that the third term in (26) is the only one that does not depend on the bandwidth $b$. This means that an error of order $O_p(n^{H_2-1})$ is always there, independently of the choice of $b$. A sequence of bandwidths $b_n$ may therefore be called optimal, if

$$\max \left\{ b_n^{2k}, \frac{1}{\sqrt{nb_n}}, b_n^2 n^{H_2-1} \right\} = o \left( n^{H_2-1} \right).$$

(27)

The first question is whether (27) can be achieved. The answer is given by the following Lemma:

**Lemma 2** Under the assumptions of Theorem 1 the following statements are equivalent:
(i) The set of bandwidth sequences such that \(b_n > 0, b_n \to 0, nb_n \to \infty\) and (27) holds is not empty.

(ii) The set of bandwidth sequences such that \(b_n > 0, b_n \to 0, nb_n \to \infty\) and

\[
\lim_{n \to \infty} n^{(1-H_Z)/2k} b_n = 0, \quad \lim_{n \to \infty} n^{(H_X-H_Z)/2} b_n = 0, \quad \lim_{n \to \infty} n^{2H_Z-1} b_n = \infty
\]

(28) is not empty.

(iii) \(H_Z > \max\left\{ \frac{H_X + 2}{5}, \frac{2k + 1}{4k + 1} \right\} \) (29).

As a corollary we obtain:

**Corollary 1** Suppose that the assumptions of Theorem 1, and (29) hold. Then the set of optimal sequences is not empty. Moreover, for any such sequence \(b_n\) we have

\[
\hat{\mu}_n \left( w^0 \right) - \mu \left( w^0 \right) = O_p \left( n^{H_Z-1} \right) \quad \text{for } w^0 \in [0, 2\pi).
\]

(30)

A closer analysis of (30) leads to convergence in distribution:

**Theorem 2** Suppose that the assumptions of Theorem 1, and (29) hold. Then

\[
n^{1-H_Z} \left[ \hat{\mu}_n \left( w^0_j \right) - \mu \left( w^0_j \right) \right] \to_d c\mu \xi
\]

(31)

where

\[
c\mu = \sqrt{\frac{\sin(\pi H_Z - \frac{\pi}{2}) \Gamma(2 - 2H_Z) c_{f.Z}}{H_Z \left( H_Z - \frac{1}{2} \right)}},
\]

\(c_{f.Z}\) is defined in (7) and \(\xi\) is a standard normal random variable. More generally, for \(w^0_1 < w^0_2 < \cdots < w^0_m \) \((w^0_j \in [0, 2\pi))\),

\[
n^{1-H_Z} \left[ \hat{\mu}_n \left( w^0_j \right) - \mu \left( w^0_j \right) \right] \to_d \left[ c\mu, \ldots, c\mu \right] \cdot \xi.
\]

(32)

**Remark 7** Equation (32) means in particular that the standardized random deviations of estimates at different values \(w^0_j\), \(w^0_l\) are asymptotically perfectly correlated. This is very much in contrast to usual behaviour in nonparametric regression. The degenerate limit theorem simplifies the construction of simultaneous confidence intervals for \(\mu(w^0_1), \ldots, \mu(w^0_m)\).

**Remark 8** It can also be shown that, if \(H_X \leq \frac{1}{2}\), then (32) holds for any \(H_Z > (2k + 1)/(4k + 1)\). Note that in this case we also have \(\max\{(H_X + 2)/5, (2k + 1)/(4k + 1)\} = (2k + 1)/(4k + 1)\).
Remark 9  For \( k = 1 \), we have \((2k + 1)/(4k + 1) = 3/5\). Now \( H_X < 1 \) implies \((H_X + 2)/5 < 3/5\) so that condition (29) reduces to

\[
H_Z > \frac{3}{5}.
\] (33)

On the other hand, for \( k \to \infty \), the ratio \((2k + 1)/(4k + 1)\) decreases monotonically to \(1/2\). Therefore, in the limit (as \( k \to \infty \)), under the assumption that \( H_X > 1/2\) condition (29) reduces to \( H_Z > (H_X + 2)/5\).

Under simple additional conditions, uniform convergence of \( \hat{\mu}_n(w^0) \) can be obtained:

Theorem 3  Suppose that the assumptions of Theorem 1, and (29) hold. Moreover, assume

\[
\lim_{n \to \infty} \left( n^{H_Z+H_X-2} \sum_{s=-\infty}^{\infty} |\alpha_s(b_n)| \right) = 0, \tag{34}
\]

where \( \alpha_s(b) \) are the coefficients in the Fourier series representation (25). Then

\[
\sup_{w^0 \in [0,2\pi)} \left| \hat{\mu}_n \left( w^0 \right) - \mu \left( w^0 \right) \right| \to 0.
\]

Remark 10  The conditions in Theorem 3 are related to Lemma 3.2 and Corollary 3.2 in Ghosh (2014).

Typically, \( \sum_{s=-\infty}^{\infty} |\alpha_s(b_n)| \) is asymptotically proportional to \( b_n^{-\beta} \) for some \( \beta > 0\). Theorem 3 then simplifies as follows.

Corollary 2  Suppose that the assumptions of Theorem 1, and (29) hold. Moreover, assume that

\[
\sum_{s=-\infty}^{\infty} |\alpha_s(b)| = O \left( b^{-\beta} \right) \ (b \to 0) \tag{35}
\]

for some \( \beta > 0\), and \( b_n \) is a sequence of bandwidths such that \( b_n > 0, b_n \to 0, nb_n \to \infty, \) (27) holds and

\[
\lim_{n \to \infty} n^{2-H_Z-H_X} b_n^\beta = \infty. \tag{36}
\]

Then

\[
\sup_{w^0 \in [0,2\pi)} \left| \hat{\mu}_n \left( w^0 \right) - \mu \left( w^0 \right) \right| \to 0.
\]

Example 1  Denote by \( I_s(\kappa) \) \( (s = 0, 1, 2, \ldots) \) modified Bessel functions of order \( s \). Consider a von Mises kernel with \( \kappa = b^{-2}, \)

\[
K_b \left( u \right) = \frac{1}{2\pi I_0(\kappa)} \exp \left( \kappa \cos u \right).
\]
The Fourier series representation of $K_b$ is given by

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(\kappa) \exp(-isu)$$

where

$$\alpha_s(\kappa) = \frac{I_s(\kappa)}{I_0(\kappa)}.$$  

Since $I_s(\kappa) > 0$, we have

$$\sum_{s=-\infty}^{\infty} |\alpha_s(\kappa)| = 2\pi K_b(0) = \frac{\exp(\kappa)}{I_0(\kappa)}.$$  

Then, $I_0(\kappa) \sim \exp(\kappa)/\sqrt{2\pi\kappa} (\kappa \to \infty)$ (see e.g. Robert 1990) implies

$$\sum_{s=-\infty}^{\infty} |\alpha_s(\kappa)| \sim \sqrt{2\pi\kappa} = \sqrt{2\pi b^{-1}}.$$  

Thus, condition (36) with $\beta = 1$ holds. Similar results can be obtained for the generalized von Mises distribution introduced by Kim et al. (2013).

**Example 2** For a wrapped normal kernel with $\sigma = b$,

$$K_b(u) = \frac{1}{\sqrt{2\pi b}} \sum_{j=-\infty}^{\infty} \exp\left(-\frac{(u + 2\pi j)^2}{2b^2}\right),$$

we have the Fourier series representation

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp\left(-\frac{b^2}{2s^2}\right) \exp(-isu).$$

Thus,

$$\sum_{s=-\infty}^{\infty} |\alpha_s(b)| = 2\pi K_b(0)$$

$$= \frac{\sqrt{2\pi}}{b} \sum_{j=-\infty}^{\infty} \exp\left(-\frac{2\pi^2}{b^2} j^2\right).$$
and, for $b \leq 1$,
\[
\sum_{s=-\infty}^{\infty} |\alpha_s (b)| \leq b^{-1} \sqrt{2\pi} \sum_{j=-\infty}^{\infty} \exp \left(-2\pi^2 j^2\right) = C b^{-1}.
\]

Therefore, condition (36) with $\beta = 1$ holds.

4 Asymptotic confidence intervals

Theorem 2 can be used to obtain asymptotic confidence sets for $\mu(w_0^0), \ldots, \mu(w_m^0)$. Since condition (29) has to be checked first, a two-step procedure has to be applied. Suppose for instance that we use a kernel of order 2. Then we have to test the null hypothesis $H_0 : H_Z \leq 3/5$ against the alternative $H_1 : H_Z > 3/5$. If $H_0$ is rejected, then (32) implies that confidence intervals with asymptotic coverage probability $1 - \alpha$ can be defined by
\[
\left[\hat{\mu}_n \left( w_1^0 \right), \ldots, \hat{\mu}_n \left( w_m^0 \right) \right] \pm n^{H_Z - 1} q_{1-\alpha/2} \left[ c_{\mu} \left( w_1^0 \right), \ldots, c_{\mu} \left( w_m^0 \right) \right]
\]
where $q_{1-\alpha/2}$ is the $(1 - \alpha/2)$-quantile of the standard normal distribution (see (32)).

Testing $H_Z \leq 3/5$ versus $H_Z > 3/5$ can be done as follows. Under the given conditions, $S_{Z,t} = \sin Z_t \mod 2\pi$ has the same long-memory parameter $H_Z$ as $Z_t$. Therefore a consistent estimator of $H_Z$ can be based on the periodogram of the series $S_{Z,t} = \sin \hat{Z}_t$ $(t = 1, \ldots, n)$ where $\hat{Z}_t = [\vartheta_t - \hat{\mu}_n(\psi_t)] \mod 2\pi$. There is a vast range of consistent methods for estimating the long-memory parameter (see e.g. Beran et al. 2013, chapter 5, and references therein). For instance, let $\hat{H}_Z$ be a local Whittle estimator based on the periodogram of $S_{Z,t}$ at the $m$ lowest frequencies. Then, under mild regularity conditions (see Robinson 1995), as $m \to \infty, m/n \to 0$,
\[
\sqrt{m} \left( \hat{H}_Z - H_Z \right) \to_d \frac{1}{2} \xi_1
\]
where $\xi_1$ is a standard normal random variable. Thus, given a level of significance $\alpha \in (0, 1)$, we reject $H_0 : H_Z \leq 3/5$, if
\[
\hat{H}_Z > \frac{3}{5} + \frac{1}{2\sqrt{m}} q_{1-\alpha}
\]
where $q_{1-\alpha}$ is the $(1 - \alpha)$-quantile of the standard normal distribution.

Remark 11 Confidence intervals (37) are conditional, since they are computed under the condition that $H_0 : H_Z \leq 3/5$ is rejected. In principle this would call for an adjustment of quantiles, for instance using a Bonferroni correction. However, note that (32) and (37) are applicable only, if $H_Z > 3/5$. If this is the case (i.e. $H_Z > 3/5$),
then then $H_0$ is rejected with asymptotic probability one, provided that we use a consistent estimator $\hat{H}_Z$ of $H_Z$. Therefore, no correction for multiple testing needs to be applied asymptotically.

**Remark 12** In practice, the constants $c_{\mu}(w_j^0) (j = 1, \ldots, m)$ in (37) have to be estimated. This means that $H_Z$ and $c_{f, Z}$ have to be replaced by consistent estimates.

**Remark 13** If one uses a kernel of order $p = 2k$ with $k \geq 2$ (instead of $k = 1$), a test of

$$H_0 : H_Z \leq \max \left\{ \frac{H_X + 2}{5}, \frac{2k + 1}{4k + 1} \right\}$$

against

$$H_1 : H_Z > \max \left\{ \frac{H_X + 2}{5}, \frac{2k + 1}{4k + 1} \right\}$$

has to be carried out. This is slightly more complicated, because the right hand side of the inequality involves the unknown parameter $H_X$. Since $\sin \psi_t$ has the same long-memory parameter $H_X$ as $\psi_t$, $H_X$ can be estimated based on the periodogram of the series $S_{\psi,t} = \sin \psi_t (t = 1, \ldots, n)$. Suppose for instance that $\hat{H}_Z$ and $\hat{H}_X$ are local Whittle estimators based on $S_{\hat{Z},t}$ (see above) and $S_{\psi,t}$ respectively. Then

$$\sqrt{m} \left( \hat{H}_Z - H_Z, \hat{H}_X - H_X \right) \xrightarrow{d} \frac{1}{2} (\zeta_1, \zeta_2)$$

where $\zeta_1, \zeta_2$ are independent standard normal random variables and $m$ is the number of lowest frequencies used for estimation. Let

$$T = \hat{H}_Z - \max \left\{ \frac{\hat{H}_X + 2}{5}, \frac{2k + 1}{4k + 1} \right\}.$$  

A rejection region at an asymptotic level of significance $\alpha$ is then defined by the condition $T > q_{1-\alpha}$ where $q_{1-\alpha}$ is the $(1 - \alpha)$-quantile of the random variable

$$W = W(H_Z, H_X) = H_Z + \frac{1}{2\sqrt{m}} \xi_1 - \max \left\{ \frac{1}{5} H_X + \frac{2}{5} + \frac{1}{10\sqrt{m}} \xi_2, \frac{2k + 1}{4k + 1} \right\}.$$  

(38)

The distribution of $W$ depends on $H_Z$ and $H_X$. For each pair $(H_Z, H_X) \in (\frac{1}{2}, 1)^2$, $(1 - \alpha)$-quantiles $q_{1-\alpha} = q_{1-\alpha}(H_Z, H_X)$ of $W(H_Z, H_X)$ can be obtained by numeric integration or by simulations. If $H_Z$ and $H_X$ are replaced by their consistent estimates, approximate $(1 - \alpha)$-quantiles are given by $q_{1-\alpha}(\hat{H}_Z, \hat{H}_X)$.  

 Springer
5 Data example

We consider daily average wind directions recorded in Chicago and Milwaukee between 1 January and 31 December, 2015 (data source https://www.glerl.noaa.gov/). Figure 1a–d show windrose plots of wind directions in Chicago (here given in degrees between 0° to 360°), split into four three-months periods, i.e. (a) January to March, (b) April to June, (c) July to September and (d) October to December. The same plots are displayed for Milwaukee in Fig. 2a–d. These figures indicate typical seasonal patterns. In a first step we therefore remove a deterministic seasonal trend from both series. This is done using the deseasonalization method for circular data described in Beran and Ghosh (2019). Figure 3a shows the two series together with the fitted seasonal trend. The residual series are displayed in Fig. 3b. Note that for better visibility, the Milwaukee series was moved down vertically by 360°.

Define now $\vartheta_t$ to be the residual series for Milwaukee (lower series in Fig. 3b), and $\psi_t$ the residual series for Chicago. The plot of $\vartheta_t$ vs. $\psi_t$ in Fig. 4 indicates a possible relationship between the two wind directions, though the circular nature of
Fig. 2 Windrose plots of daily wind directions in Milwaukee in 2015, split into four three-months periods: a January to March, b April to June, c July to September and d October to December.

Fig. 3 a Daily wind directions in Chicago and Milwaukee, in 2015. Also plotted are fitted seasonal trends. For better visibility, the Milwaukee series is shifted vertically by $-360^\circ$. b Residual series obtained after removing the seasonal trend.
the variables makes a visual inspection more difficult. The Nadaraya–Watson estimate of $\mu(\psi)$, based on a von Mises kernel, is shown in Fig. 6. Since we are using a kernel of order 2, the value of $H_X$ does not play any role (see Remark 9). With respect to estimation of $H_Z$, we consider a local Whittle estimator with $m = \lfloor n^{0.6} \rfloor$ frequencies. For $S_{\hat{Z}_t} = \sin \hat{Z}_t$ where $\hat{Z}_t = \vartheta_t - \hat{\mu}_n(\psi_t) \mod 2\pi$, we obtain $\hat{H}_Z = 0.75$ with a 95%-confidence interval of [0.58, 0.92]. Thus, there is significant evidence for long memory in $Z_t$. The estimate does not change much, when other estimation methods are used. For instance, fitting fractional autoregressive models of orders $p = 0, 1, \ldots, 3$, and selecting the best model using the BIC (see e.g. Beran et al. 1998), leads to
\( \hat{H}_Z = 0.67 \ (0.59, 0.75) \). Figure 5 shows the log-log-periodogram \( \sin \hat{Z}_t \), together with a fitted spectral density.

The next step is to test \( H_Z \leq 3/5 \) against \( H_Z > 3/5 \). For the local Whittle estimator, the critical limit for the 5\% level of significance is \( 3/5 + 1.645 \cdot \frac{1}{2} m^{-1/2} = 0.74 \). Since \( \hat{H}_Z = 0.75 \) is above this threshold, there is evidence at the 5\%-level of significance for condition (29). The same conclusion is reached when using estimates based on the BIC method. Confidence bands for \( \mu(\psi) \) based on Theorem 2 and (37) are given in Fig. 6. As a cautionary remark one should bear in mind that, strictly speaking, the band is simultaneous for a finite - though possibly large - number of \( \psi \)-values only. Note also that due to the circular nature of the data, there is an apparent jump around \( \psi = 17.6^\circ \). This is not a jump in the function \( \mu \), but rather due to plotting modulo 360\°. The same comment applies to the confidence band.

Acknowledgements Open Access funding enabled and organized by Projekt DEAL. The data set was obtained from the homepage of the NOAA Great Lakes Environmental Research Laboratory (https://www.glerl.noaa.gov/). We would like to thank the associate editor and the referees for their insightful remarks.

Funding Open Access funding enabled and organized by Projekt DEAL.
7 Appendix: proofs, figures

7.1 Proofs

Proof (of Theorem 1) The method of the proof of Theorem 1 follows similar lines as Proposition 1 and Section 3.2 in Mielniczuk and Wu (2004). For $\psi^0 \in [0, 2\pi)$, we have

$$\hat{\mu}_n(\psi^0) = \frac{n^{-1} \sum_{s=1}^{n} K_b (\psi_s - \psi^0) \vartheta_s}{\hat{f}_n(\psi^0)}$$

where

$$\hat{f}_n(\psi^0) = n^{-1} \sum_{s=1}^{n} K_b (\psi_s - \psi^0)$$

is a kernel density estimator of the density $f_\psi$ at $\psi^0$. Using the notation

$$\mu_n(\psi^0) = \frac{E \left[ K_b \left( \psi_t - \psi^0 \right) \vartheta_t \right]}{E \left[ K_b \left( \psi_t - \psi^0 \right) \right]}, \quad D_n(\psi^0) = \hat{\mu}_n(\psi^0) - \mu_n(\psi^0)$$

we have

$$\hat{\mu}_n(\psi^0) - \mu(\psi^0) = \mu_n(\psi^0) - \mu(\psi^0) + D_n(\psi^0).$$

The first term is deterministic and does not depend on the dependence structure. Following arguments in Tsurata and Sagae (2017) (for the special case of $k = 1$ also see Taylor 2008), leads to

$$\mu_n(\psi^0) - \mu(\psi^0) = O \left( b^{2k} \right).$$

To study asymptotic properties of $D_n$, the following notation is useful:

$$\xi_t = \xi_t(\psi^0) = K_b (\psi^0 - \psi_t) - E \left[ K_b \left( \psi^0 - \psi(X_t) \right) \mid \eta_s, s \leq t - 1 \right],$$

$$A_n(\psi^0) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t K_b(\psi^0 - \psi(X_t)) + \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{j=1}^{\infty} a_j \epsilon_{t-j} \right) \xi_t,$$

and

$$J_t(\psi^0) = Z_t K_b(\psi^0 - \psi_t).$$
Consider the decomposition

\[ D_n \left( \psi^0 \right) \hat{f}_n \left( \psi^0 \right) = n^{-1} \sum_{s=1}^{n} K_b \left( \psi - \psi^0 \right) \left[ \varphi_s - \mu_n \left( \psi^0 \right) \right] \]

\[ = A_n \left( \psi^0 \right) + B_n \left( \psi^0 \right) + C_n \left( \psi^0 \right) + r_n \]

with

\[ A_n \left( \psi^0 \right) = \frac{1}{n} \sum_{t=1}^{n} \left\{ J_t \left( \psi^0 \right) - E \left[ Z_t \mid \varepsilon_s, s \leq t - 1 \right] \times E \left[ K \left( \psi^0 - \psi_t \right) \mid \eta_s, s \leq t - 1 \right] \right\} , \]

\[ B_n \left( \psi^0 \right) = \frac{1}{n} \sum_{t=1}^{n} E \left[ Z_t \mid \varepsilon_s, s \leq t - 1 \right] \times E \left[ K_b \left( \psi^0 - \psi_t \right) \mid \eta_s, s \leq t - 1 \right] \]

and

\[ C_n \left( \psi^0 \right) = n^{-1} \sum_{t=1}^{n} K_b \left( \psi_t - \psi^0 \right) \left[ \mu \left( \psi^0 \right) - \mu_n \left( \psi^0 \right) \right] . \]

The asymptotic results in Mielniczuk and Wu (2004) are obtained for kernel regression with non-circular kernels. Taking into account the definition and properties of circular kernels, the arguments in Mielniczuk and Wu (2004) can be reformulated accordingly, to yield

\[ A_n = O_p \left( \frac{1}{\sqrt{nb}} \right) , B_n = O_p (n^{H_Z-1}) , C_n \left( \psi^0 \right) = O_p \left( b^2 n^{H_X-1} \right) \]

and \( r_n = o_p(\max\{A_n, B_n, C_n\}) \).

**Proof** (of Lemma 2) From Theorem 1 and (26) we have

\[ \hat{\mu}_n \left( \psi^0 \right) - \mu \left( \psi^0 \right) = O \left( b^{2k} \right) + O_p \left( \frac{1}{\sqrt{nb}} \right) + O_p \left( n^{H_Z-1} \right) + O_p \left( b^2 n^{H_X-1} \right) . \]

In order that the term \( O_p (n^{H_Z-1}) \) dominates we need

\[ \max \left\{ b^{2k} , (nb)^{-1/2} , b^2 n^{H_X-1} \right\} = o \left( n^{H_Z-1} \right) . \]

This leads to

\[ b = o \left( \min \left\{ n^{(H_Z-1)/2k} , n^{(H_Z-H_X)/2} \right\} \right) \] (39)
and
\[ n^{1-2Hz} = o(b). \]  \hfill (40)

In order that (39) and (40) hold simultaneously, we need
\[ n^{1-2Hz} = o\left(\min\left\{ n^{(Hz-1)/2k}, n^{(Hz-Hx)/2}\right\}\right), \]
which leads to
\[ Hz > \max\left\{ \frac{Hx + 2}{5}, \frac{2k + 1}{4k + 1}\right\}. \]

Proof (of Corollary 1) The result follows directly from Lemma 2 and Theorem 1.

Proof (of Theorem 2) Writing
\[ \nu_t = E\left[ K_b \left( \psi_0 - \psi_t \right) \mid \eta_s, s \leq t - 1 \right] \]
we have
\[ B_n(\psi_0) = \frac{1}{n} \sum_{t=1}^{n} E\left[ Z_t \mid \epsilon_s, s \leq t - 1 \right] E\left[ K_b \left( \psi_0 - \psi_t \right) \mid \eta_s, s \leq t - 1 \right] \]
\[ = \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{j=1}^{\infty} a_j \epsilon_{t-j} \right) \nu_t \]
\[ = \frac{1}{n} \sum_{t=1}^{n} Z_t \nu_t + o_p(B_n). \]

Now, define
\[ \tilde{B}_n(\psi_0) = \left( K_b * f_\psi \left( \psi_0 \right) \right) \frac{1}{n} \sum_{t=1}^{n} Z_t \]
where
\[ K_b * f_\psi \left( \psi_0 \right) = \int_{-\pi}^{\pi} K_b \left( \psi_0 - y \right) f_\psi \left( y \right) dy. \]

Since
\[ E(\nu_t) = K_b * f_\psi \left( \psi_0 \right). \]
we have
\[
\frac{1}{n} \sum_{t=1}^{n} Z_t v_t - \tilde{B}_n (\psi^0) = \frac{1}{n} \sum_{t=1}^{n} Z_t [v_t - E (v_t)]
\]
and hence also
\[
E \left[ \frac{1}{n} \sum_{t=1}^{n} Z_t v_t - \tilde{B}_n (\psi^0) \right] = 0.
\]
Denote by \( \gamma_v (k) \) the autocovariance function of \( v_t \). Then
\[
\text{var} \left( \frac{1}{n} \sum_{t=1}^{n} Z_t v_t - \tilde{B}_n (\psi^0) \right) = n^{-2} \sum_{t,s=1}^{n} \gamma_Z (t-s) \gamma_v (t-s)
= n^{-1} \sum_{|k| < n} \left( 1 - \frac{|k|}{n} \right) \gamma_Z (k) \gamma_v (k).
\]
Now, as \( k \to \infty \),
\[
|\gamma_v (k)| \leq c_{\gamma,v} \cdot k^{2H_Z-2}, \quad \gamma_Z (k) \sim c_{\gamma,Z} k^{2H_Z-2}
\]
where \( c_{\gamma,v}, c_{\gamma,Z} \) are suitable constants. Therefore,
\[
\text{var} \left( \frac{1}{n} \sum_{t=1}^{n} Z_t v_t - \tilde{B}_n (\psi^0) \right) \leq \text{const} \cdot n^{2H_X + 2H_Z-4} \sum_{|k| < n} \left( 1 - \frac{|k|}{n} \right)
\times \left( \frac{|k|}{n} \right)^{2H_X+2H_Z-4} \frac{1}{n}
\sim C \cdot n^{2H_X + 2H_Z-4}.
\]
Now \( H_X < 1 \) implies \( n^{2H_X + 2H_Z-4} = o(n^{2H_Z-2}) \). Thus, \( B_n \) may be replaced asymptotically by \( \tilde{B}_n \). For \( \tilde{B}_n \), the asymptotic distribution follows from \( K_b * f_{\psi} (\psi^0) \to f_{\psi} (\psi^0) \) and well known limit theorems for sample means under long-range dependence (see e.g. Beran et al. 2013; Chapter 4.2 and Corollary 1.2). More specifically,
\[
n^{H_Z-1} \left[ \hat{\mu}_n (\psi^0) - \mu (\psi^0) \right] = n^{H_Z-1} \tilde{B}_n (\psi^0) + o_p \left( n^{H_Z-1} \right)
= n^{H_Z-1} \frac{1}{n} \sum_{t=1}^{n} Z_t + o_p \left( n^{H_Z-1} \right)
= c_{\mu} W + o_p \left( n^{H_Z-1} \right).
\]
Proof (of Theorem 3) The line of proof is similar to Ghosh (2014). Under the given conditions, it is sufficient to prove that

\[
S_n(\psi^0) = \frac{1}{n} \sum_{t=1}^{n} E \left[ K_b \left( \psi^0 - \psi_t \right) \mid \eta_s, s \leq t - 1 \right] Z_t
\]

converges uniformly to zero in probability. It is then enough to show

\[
\lim_{n \to \infty} E \left[ \sup_{\psi^0} \left| S_n(\psi^0) \right| \right] = 0.
\]

By assumption,

\[
K_b(w) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(b) \exp(-isw)
\]

and \(Z_t\) is independent of \(\eta_s (s \in \mathbb{Z})\). Thus,

\[
S_n(\psi^0) = \frac{1}{n} \sum_{t=1}^{n} E \left[ K_b(\psi_t - \psi^0) \mid \eta_j, j \leq t - 1 \right] Z_t
\]

\[
= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(b) \left\{ \frac{1}{n} \sum_{t=1}^{n} Z_t E \left[ \exp(-is\psi_t) \mid \eta_j, j \leq t - 1 \right] \right\} \exp(i\psi^0),
\]

and

\[
\left| S_n(\psi^0) \right| \leq \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |\alpha_s(b)| \left\{ \frac{1}{n} \sum_{t=1}^{n} Z_t E \left[ \exp(-is\psi_t) \mid \eta_j, j \leq t - 1 \right] \right\}.
\] (41)

Now the right-hand side of (41) does not depend on \(\psi^0\). Hence,

\[
\sup_{\psi^0} \left| S_n(\psi^0) \right| \leq \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |\alpha_s(b)| \left\{ \frac{1}{n} \sum_{t=1}^{n} Z_t E \left[ \exp(-is\psi_t) \mid \eta_j, j \leq t - 1 \right] \right\}.
\]

Using the notation

\[
\nu_{t,s,\cos} = E \left[ \cos(s\psi_t) \mid \eta_j, j \leq t - 1 \right], \nu_{t,s,\sin} = E \left[ \sin(s\psi_t) \mid \eta_j, j \leq t - 1 \right],
\]

we then have, for a suitable constant \(0 < C < \infty\),

\[
E \left[ \sup_{\psi^0} \left| S_n(\psi^0) \right| \right] \leq \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |\alpha_s(b)|
\]

Springer
\[
\times \left[ \text{var} \left( \frac{1}{n} \sum_{t=1}^{n} Z_{t} \nu_{t,s,\cos} \right) + \text{var} \left( \frac{1}{n} \sum_{t=1}^{n} Z_{t} \nu_{t,s,\sin} \right) \right] \leq C n^{H_{X}+H_{Y}-2} \sum_{s=-\infty}^{\infty} |\alpha_{s}(b)|.
\]

Thus, assumption (34) implies

\[
\lim_{n \to \infty} E \left[ \sup_{\psi_{0}} \left| S_{n} \left( \psi_{0} \right) \right| \right] = 0.
\]

\[\square\]

**Proof** (of Corollary 2) The result follows directly from Theorem 3, (35) and (36). \[\square\]

**References**

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

Bedouhene K, Zougab N (2020) Nonparametric multiplicative bias correction for von Mises kernel circular density estimator. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2019.1659358

Beran J (1994) Statistics for long-memory processes. Monographs on statistics and applied probability (Vol. 61). Chapman and Hall/CRC, New York

Beran J, Ghosh S (2019) Estimating the mean direction of strongly dependent circular time series. J Time Ser Anal. https://doi.org/10.1111/jtsa.12500

Beran J, Bhansali RJ, Ocker D (1998) On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. Biometrika 85(4):921–934

Beran J, Feng Y, Ghosh S, Kulik R (2013) Long memory processes: probabilistic properties and statistical methods. Springer, Berlin/Heidelberg

Breckling J (1989) The analysis of directional time series: application to wind speed and direction. Springer, Berlin

Davidson J, Sibbertsen P (2005) Generating schemes for long memory processes: regimes, aggregation and linearity. J Econ 128(2):253–282

Di Marzio M, Panzera A, Taylor CC (2012) Non-parametric smoothing and prediction for nonlinear circular time series. J Time Ser Anal 33(4):620–630

 Dobrushin RL (1980) Gaussian random fields—Gibbsian point of view. In: Dobrushin RL, Sinai YG (eds) Multicomponent random systems: advances in probability and related topics, vol 6. Dekker, New York, pp 119–152

 Dobrushin RL, Major P (1979) Non-central limit theorems for nonlinear functionals of Gaussian fields. Zeitchrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 50(1):27–52

 Fisher NI (1989) Smoothing a sample of circular data. J Struct Geol 11:775–778

 Fisher NI (1993) Statistical analysis of circular data. Cambridge University Press, Cambridge

 Fisher NI, Lee A (1992) Regression models for an angular response. Biometrics 48(3):665–677

 Ghosh S (2014) On local slope estimation in partial linear models under Gaussian subordination. J Stat Plann Infer 155:42–53

 Giraitis L, Koul HL, Surgailis D (2012) Large sample inference for long memory processes. Imperial College Press, London

 Gould AL (1969) A regression technique for angular variates. Biometrics 25:683–700

 Hall P, Watson GS, Cabrera J (1987) Kernel density estimation with spherical data. Biometrika 74(4):751–762

 Jammalamadaka SR, Sarma YR (1988) A correlation coefficient for angular variables. In: Matusita K (ed) Theory statistical, analysis data. North Holland, New York, pp 349–364

 Springer
Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. World Scientific, Singapore
Johnson RA, Wehrly TE (1978) Some angular-linear distributions and related regression models. J Am Stat Assoc 73(363):602–606
Kato S (2010) A Markov process for circular data. J R Stat Soc B 72:655–672
Kato S, Shimizu K, Shieh GS (2008) A circular-circular regression model. Stat Sin 18(2):633–645
Kim S, SenGupta A (2013) A three-parameter generalized von Mises distribution. Stat Pap 54:685–693
Kim S, SenGupta A (2016) Multivariate-multiple circular regression. J Stat Comput Simul 87:1–15
Ley C, Verdebout T (2017) Modern directional statistics. Chapman & Hall/CRC, Boca Raton
Mardia KV (1972) Statistics of directional data. Academic Press, London
Mardia KV, Jupp PE (2000) Directional statistics. Wiley, New York
Mielniczuk J, Wu WB (2004) On random-design model with dependent errors. Stat Sin 14:1105–1126
Modlin D, Fuentes M, Reich B (2012) Circular conditional autoregressive modeling of vector fields. Environmetrics 23:46–53
Polsen O, Taylor C (2015) Parametric circular–circular regression and diagnostic analysis. In: Dryden IL, Kent JT (eds) Geometry driven statistics. Leeds University Press, Leeds, pp 115–128
Robert R (1990) Modified Bessel functions and their applications in probability and statistics. Stat Prob Lett 9(2):155–161
Robinson PM (1995) Gaussian semiparametric estimation of long range dependence. Ann Stat 23(5):1630–1661
Rosenblatt M (1961) Independence and dependence. In Proceedings of the 4th Berkeley sympos. math. statist. and prob. (vol II, pp 431–443). University of California Press, Berkeley
Rosenblatt M (1979) Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 49(2):125–132
Taqqu MS (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31:287–302
Taqqu MS (1979) Convergence of integrated processes of arbitrary Hermite rank. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 50:53–83
Taylor CC (2008) Automatic bandwidth selection for circular density estimation. Comput Stat Data Anal 52(7):3493–3500
Tsuruta Y, Sagae M (2017) Higher order kernel density estimation on the circle. Stat Prob Lett 131:46–50
Tsuruta Y, Sagae M (2020) Theoretical properties of bandwidth selectors for kernel density estimation on the circle. Ann Inst Stat Math 72:511–530
Wang F, Gelfand AE (2014) Space and space-time directional data using projected Gaussian processes. J Am Stat Assoc 109(508):1565–1580
Watson GS (1983) Statistics on spheres. Wiley, New York
Wehrly TE, Johnson RA (1980) Bivariate models for dependence of angular observations and a related Markov process. Biometrika 66:255–256

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.