Article:

Di Marzio, M, Panzera, A and Taylor, CC (2012) Non-parametric smoothing and prediction for nonlinear circular time series. Journal of Time Series Analysis, 33 (4). 620 - 630 . ISSN 0143-9782

https://doi.org/10.1111/j.1467-9892.2012.00794.x

Nonparametric circular time series analysis

Marco Di Marzio[a], Agnese Panzera[a], Charles C. Taylor[b,1]

[a]DMQTE, Università di Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy.
[b]Departement of Statistics, University of Leeds, Leeds LS2 9JT, UK.

Abstract

Although most circular datasets are in the form of time series, not much research has been done in the field of circular time series analysis. We propose a nonparametric theory for smoothing and prediction in the time domain for circular time series data. Our model is based on local polynomial fitting which minimizes an angular risk function. Both asymptotic arguments and empirical examples are used to describe the accuracy of our methods.

Key words: Angular risk, Kernel weights, Linear weights, Long range dependence, Short range dependence

2000 MSC: 62G07 - 62G08 - 62G20

---

1. Introduction

1.1. Motivation

Circular data arise whenever *directions* are measured, and are usually expressed as angles relative to some fixed reference point. Given that in the circular setting we have the identity $0 \equiv 2\pi$, it is immediately apparent that linear data methods are unsuitable for circular data. In last fifty years, however, *circular statistics* has greatly evolved, and now circular counterparts exist for several parametric data analysis techniques. For a comprehensive account, see the survey paper by Lee (2010), and the references therein.

Surely, not a great deal of theory exists for circular time series analysis, even though most datasets are in this form, obvious examples being wind and ocean directions. Note that, in some applications, time itself can be viewed as a circular measurement, but this is not the context of this paper in which we consider time as linear, with a circular observation taken at each time point. Breckling (1989) builds an autoregressive model by wrapping a linear AR(1) model around the circle. Four types of time series model are studied by Fisher & Lee (1994). The first is obtained by wrapping standard linear ARMA processes, and the second by projecting a bivariate linear process. Two further

Email addresses: mdimarzio@unich.it (Marco Di Marzio), agnesepanzera@yahoo.it (Agnese Panzera), charles@maths.leeds.ac.uk (Charles C. Taylor)

[1]Corresponding author

models are based on modifications of parametric regression models. Other recent approaches are provided by Artes et al. (2000), who target longitudinal data using generalized estimating equations, by Holzmann et al. (2006), who use hidden Markov models for fitting linear-circular and circular-circular time series, and by the thesis of Hughes (2007), who explores various adaptations of the Möbius model of Downs & Mardia (2002).

On the other hand, nonparametric statistical methods have become more and more popular over the last two decades for the analysis of non-linear time series, see the books of Fan & Yao (2003) and Gao (2007) and the references therein. See also Bosq (1998, chap. 7), who demonstrated that kernel-based estimators have the potential to improve on ARMA models in many situations even when the latter appears appropriate to use. Indeed, existing studies show that linearity assumptions – such as normality, symmetric cycles, unimodality, linearity between lagged variables, and homoscedasticity – are rather strong in most practical situations. Similarly, Box-Jenkins models and ARCH models often yield poor forecasts, particularly if the horizon is large, whereas a suitable nonparametric model behaves efficiently even if stronger assumptions are satisfied and the data are truly linear. Anyway a strong complementarity between parametric and nonparametric methods exists for time series: an obvious example lies in first performing nonparametric trend estimation, then imposing a parametric model for the residual stationary data.

The above arguments provide our motivation for constructing kernel based smoothers for circular time series, as done in the present paper. Specifically, we discuss trend estimation and prediction within a basic model which adds a stationary random noise process to a deterministic trend which is not restricted in its functional form, apart from some smoothness assumptions. In the subsequent Section 1.2 we formulate our model, whereas in Section 2 we present a couple of smoothers which optimize an angular risk: one being based on the idea of local constant fit, the other one employing local linear weights. In Section 3 we face boundary estimation, which in this context is essentially linked to the idea of prediction. Finally, in Section 4 empirical experiments illustrate the findings, using both simulated and real data.

1.2. The model

Let $\{\Theta_t\}_{t=1}^T$ be a time series of angles taking values in $\mathbb{T} := [-\pi, \pi]$. Assume the model

$$\Theta_t = [m(t/T) + \varepsilon_t](\text{mod} 2\pi) \tag{1}$$

where $m : [0, 1] \to \mathbb{T}$ is an unknown smooth function of time representing the trend, and $\{\varepsilon_t\}$ is a $\mathbb{T}$-valued stationary stochastic process such that $E[\sin(\varepsilon_t)] = 0$ and with autocovariance function regularly varying at infinity with exponent $\alpha > 0$, i.e., as ℓ goes to infinity, a1) $\gamma_{cc}(\ell) := Cov[\cos(\varepsilon_t), \cos(\varepsilon_{t+\ell})] \sim L_1|\ell|^{-\alpha}$;

a2) $\gamma_{sc}(\ell) := Cov[\cos(\varepsilon_t), \sin(\varepsilon_{t+\ell})] = E[\cos(\varepsilon_t)\sin(\varepsilon_{t+\ell})] \sim L_2|\ell|^{-\alpha}$;

a3) $\gamma_{ss}(\ell) := Cov[\sin(\varepsilon_t), \sin(\varepsilon_{t+\ell})] = E[\sin(\varepsilon_t)\sin(\varepsilon_{t+\ell})] \sim L_3|\ell|^{-\alpha}$;

where $L_i \in \mathbb{R} \setminus \{0\}$, for $i \in \{1, 2, 3\}$, $\ell \in \mathbb{Z}$, and $\sim$ means that the ratio of the left and right hand sides converges to 1. We define, moreover, $|\ell|^{\alpha} := 1$ if $\ell = 0$. We say that the case $0 < \alpha < 1$ indicates a *long-range dependence*, whereas the case $\alpha > 1$ implies so-called *short-range dependence*.

Note that here, when T goes to infinity, the number of observations become denser in the interval $[0, 1]$, with the dependence structure remaining the same. The fact that m depends on t/T makes the trend much more slowly varying than the noise. This is desirable, but involves trend dependence on the sample size T, which is slightly weird. The linear version of model (1) has been extensively used to model long-range dependent and non-stationary data. Of course, in the linear case the assumptions concern directly the errors. But, noting that $\sin(\theta) \simeq \theta$ for small θ, and that $E[\sin(\varepsilon_t)] = 0$ then a3) closely resembles the linear hypotheses, whereas a1) and a2) do not appear to have linear counterparts.

Concerning assumptions a1)–a3), they say that correlation goes to zero as the temporal lag ℓ diverges. This seems natural, otherwise trend identifiability problems would arise, becoming virtually impossible to distinguish between trend and autocorrelation signal. Nevertheless, it will be seen that an important rôle is played by the strength of the dependence, i.e. the value of α. In general, for a fixed sample size, we would expect that the variance of an estimator increases with the correlation. In fact, as this latter tends to one, the amount of information in the data diminishes until it coincides with that contained in a single observation. It will often be the case that the autocorrelation structure determines the optimal smoothing degree.

2. Trend estimation

2.1. A local estimator

If $m(t/T)$ is a predictor for the random angle Θ_t, a risk measure for it is

$$E[1 - \cos(\Theta_t - m(t/T))], \tag{2}$$

which is reminiscent of the L_2 linear risk, given that $\cos(\theta) \simeq 1 - \theta^2/2$ for small θ. Letting $m_1(t/T) := E[\sin(\Theta_t)]$, and $m_2(t/T) := E[\cos(\Theta_t)]$, the minimizer of (2) is

$$\text{atan2}[m_1(t/T), m_2(t/T)] = \begin{cases} \arctan\left(\dfrac{m_1(t/T)}{m_2(t/T)}\right), & if \;\; m_2(t/T) > 0; \\ \arctan\left(\dfrac{m_1(t/T)}{m_2(t/T)}\right) + \pi, & if \;\; m_2(t/T) < 0. \end{cases}$$

Therefore, the approach we suggest is to consider the sample statistics

$$\hat{m}_1(t/T) := 1/T \sum_{i=1}^{T} \sin(\Theta_i) W(i/T - t/T) \qquad \text{and} \qquad \hat{m}_2(t/T) := 1/T \sum_{i=1}^{T} \cos(\Theta_i) W(i/T - t/T) \tag{3}$$

with weights W defined in such a way that the ratio $\hat{m}_1(t/T)/\hat{m}_2(t/T)$ is an asymptotically unbiased estimator of $m_1(t/T)/m_2(t/T)$, and define as an estimator for the trend function at t/T,

$$\hat{m}(t/T) := \text{atan2}[\hat{m}_1(t/T), \hat{m}_2(t/T)]. \tag{4}$$

Note that $\hat{m}_i$ is not necessarily an unbiased estimator of m_i ($i = 1, 2$) since this will depend on the structure of the weights. Asymptotic unbiasedness will be a distinctive property of the $\hat{m}_i$s as formulated in next Section (see Lemma 1) but this will not hold any longer for the weights discussed in Section 2.3; see Lemma 2.

2.2. Using Kernel weights

Firstly we discuss the estimator (4) with weights $W(\cdot) = K_h(\cdot) = 1/hK(\cdot/h)$, where K, called a kernel, is a symmetric density function with maximum at 0 and scale parameter h, this latter being often called the bandwidth. We will give results for three different settings of α; see hypotheses a1)–a3) above. To derive the asymptotic properties of our estimator we need some notation which will also be used elsewhere.

Let $\mu_j(K) := \int v^j K(v)dv$, $R(K) := \int K^2(v)dv$, $\Gamma_1(\ell) := Cov[\sin(\Theta_i), \sin(\Theta_{i+\ell})]$, $\Gamma_2(\ell) := Cov[\cos(\Theta_i), \cos(\Theta_{i+\ell})]$, and $\Gamma_3(\ell) := Cov[\sin(\Theta_i), \cos(\Theta_{i+\ell})]$.

Now we are able to establish some preliminary results in the following

Lemma 1. *Given the time series $\{\Theta_t\}_{t=1}^T$ observed from the model* (1)*, consider the functions in* (3) *having K_h as the weight function. If*

i) the bandwidth h is such that $h \to 0$ and $Th \to \infty$ as $T \to \infty$;

ii) K is a bounded and compactly supported kernel;

iii) m_1'' and m_2'' exist and are continuous at t/T;

then, for $j \in \{1, 2\}$,

$$E[\hat{m}_j(t/T)] = m_j(t/T) + 2^{-1}h^2\mu_2(K)m_j''(t/T) + o(h^2),$$

$$Var[\hat{m}_j(t/T)] = \begin{cases} (Th)^{-\alpha} Q_j(u,v) \int\int K(x)K(y)|x-y|^{-\alpha} dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0,1); \\ 2(Th)^{-1} Q_j(u,v) R(K) \log(Th) + o((Th)^{-1}\log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1} \sum \Gamma_j(\ell) R(K) + o((Th)^{-1}), & if \;\; \alpha > 1; \end{cases}$$

with

$$Q_j(u,v) := \begin{cases} L_1 s_u s_v + L_2(c_u s_v + s_u c_v) + L_3 c_u c_v, & if \;\; j = 1; \\ L_1 c_u c_v - L_2(c_u s_v + s_u c_v) + L_3 s_u s_v, & if \;\; j = 2; \end{cases}$$

and

$$Cov[\hat{m}_1(t/T), \hat{m}_2(t/T)] = \begin{cases} (Th)^{-\alpha} U(u,v) \int\int K(x)K(y)|x-y|^{-\alpha} dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0,1); \\ 2(Th)^{-1} U(u,v) R(K) \log(Th) + o((Th)^{-1}\log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1} \sum \Gamma_3(\ell) R(K) + o((Th)^{-1}), & if \;\; \alpha > 1, \end{cases}$$

with $U(u,v) := s_u c_v L_1 - L_2(s_u s_v - c_u c_v) - L_3 c_u s_v$, *where* $u, v \in (0,h)$, $s_i := \sin(m(i))$ *and* $c_i := \cos(m(i))$.

Proof. See Appendix. □

Now, using the fact that $m_1(t/T)$ and $m_2(t/T)$ are the components of the first trigonometric moment of Θ_t, we can write $m_1(t/T) := C(t/T)f_s(t/T)$ and $m_2(t/T) := C(t/T)f_c(t/T)$, where $C(t/T) := \{m_1^2(t/T) + m_2^2(t/T)\}^{1/2}$ and $f_s^2(t/T) + f_c^2(t/T) = 1$, and we get the following

Theorem 1. *Given the time series* $\{\Theta_t\}_{t=1}^T$ *observed from the model* (1), *if assumptions* $i) - iii)$ *of Lemma 1 hold, for the estimator* $\hat{m}$ *having* K_h *as the weight function,*

$$E[\hat{m}(t/T)] - m(t/T) = h^2 \mu_2(K) \left[\frac{m''(t/T)}{2} + \frac{C'(t/T)}{C(t/T)} m'(t/T) \right] + O\left(\nu(T,h,\alpha)\right),$$

where

$$\nu(T,h,\alpha) = \begin{cases} (Th)^{-\alpha}, & if \;\; \alpha \in (0,1); \\ (Th)^{-1}\log(Th), & if \;\; \alpha = 1; \\ (Th)^{-1}, & if \;\; \alpha > 1; \end{cases}$$

moreover

$$Var[\hat{m}(t/T)] = \begin{cases} (Th)^{-\alpha} C(t/T)^{-2} V(t/T,u,v) \int\int K(x)K(y)|x-y|^{-\alpha} dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0,1); \\ 2(Th)^{-1} C(t/T)^{-2} V(t/T,u,v) R(K) \log(Th) + o((Th)^{-1}\log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1} C(t/T)^{-2} R(K) Z(t/T) + o((Th)^{-1}), & if \;\; \alpha > 1. \end{cases}$$

where

$$V(t/T,u,v) := f_s^2(t/T) Q_2(u,v) + f_c^2(t/T) Q_1(u,v) - 2 f_c(t/T) f_s(t/T) U(u,v), \tag{5}$$

and

$$Z(t/T) := f_s^2(t/T) \sum_\ell \Gamma_2(\ell) + f_c^2(t/T) \sum_\ell \Gamma_1(\ell) - 2 f_c(t/T) f_s(t/T) \sum_\ell \Gamma_3(\ell). \tag{6}$$

Proof. See Appendix. □

Note that for homoscedastic errors the function C will be a constant, and so C' will be zero, and in this case the first term in the bias is the usual form for a Nadaraya-Watson estimator when the observations have an equi-sapced design. We can see that, as in the linear theory, when the correlation decays at a slow rate, that rate dominates the asymptotic variance, whilst in the case of short range dependence we have rates which are the same as the *i.i.d.* case.

2.3. Using local linear weights

Let

$$W(i/T - t/T) = T^{-1}K_h(i/T - t/T)\left\{\sum_{k=1}^{T} K_h(k/T - t/T)(k/T - t/T)^2 - (i/T - t/T)\sum_{k=1}^{T} K_h(k/T - t/T)(k/T - t/T)\right\}, \tag{7}$$

which are the weights of local linear polynomial fitting. The quantities needed to derive the asymptotic properties of the resulting trend function estimator are provided by the following

Lemma 2. *Given the time series* $\{\Theta_t\}_{t=1}^{T}$ *from the model* (1), *if assumptions i)- iii) of Lemma* 1 *hold, then for the functions in* (3) *having* (7) *as the weight function, and* $j \in \{1, 2\}$,

$$E[\hat{m}_j(t/T)] = h^2\mu_2(K)\left\{m_j(t/T) + 2^{-1}h^2\mu_2(K)m_j''(t/T)\right\} + o(h^4),$$

$$Var[\hat{m}_j(t/T)] \sim \begin{cases} h^4\mu_2^2(K)(Th)^{-\alpha}Q_j(u,v)\int\int K(x)K(y)|x-y|^{-\alpha}dxdy + o(T^{-\alpha}h^{4-\alpha}), & if \;\; \alpha \in (0,1); \\ 2h^3\mu_2^2(K)T^{-1}Q_j(u,v)R(K)\log(Th) + o(T^{-1}h^3\log(Th)), & if \;\; \alpha = 1; \\ h^3\mu_2^2(K)T^{-1}\sum\Gamma_j(\ell)R(K) + o(T^{-1}h^3), & if \;\; \alpha > 1; \end{cases}$$

and

$$Cov[\hat{m}_1(t/T), \hat{m}_2(t/T)] \sim \begin{cases} h^4\mu_2^2(K)(Th)^{-\alpha}U(u,v)\int\int K(x)K(y)|x-y|^{-\alpha}dxdy + o(T^{-\alpha}h^{4-\alpha}), & if \;\; \alpha \in (0,1); \\ 2h^3\mu_2^2(K)T^{-1}U(u,v)R(K)\log(Th) + o(T^{-1}h^3\log(Th)), & if \;\; \alpha = 1; \\ h^3\mu_2^2(K)T^{-1}\sum\Gamma_3(\ell)R(K) + o(T^{-1}h^3), & if \;\; \alpha > 1. \end{cases}$$

Proof. See Appendix. □

The accuracy measures of the estimator (4) are provided by the following

Theorem 2. *Given the time series* $\{\Theta_t\}_{t=1}^{T}$ *observed from the model* (1), *if assumptions* $i) - iii)$ *of Lemma 1 hold, for the estimator* $\hat{m}$ *equipped with the weight function in* (7),

$$E[\hat{m}(t/T)] - m(t/T) = h^2\mu_2(K)\left\{\frac{m''(t/T)}{2} + \frac{C'(t/T)}{C(t/T)}m'(t/T)\right\} + O\left(\delta(T,h,\alpha)\right)$$

where

$$\delta(T,h,\alpha) = \begin{cases} T^{-\alpha}h^{4-\alpha}, & if \;\; \alpha \in (0,1); \\ T^{-1}h^3\log(Th), & if \;\; \alpha = 1; \\ T^{-1}h^3, & if \;\; \alpha > 1. \end{cases}$$

and

$$Var[\hat{m}(t/T)] = \begin{cases} (Th)^{-\alpha}C(t/T)^{-2}V(t/T,u,v)\int\int K(x)K(y)|x-y|^{-\alpha}dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0,1); \\ 2(Th)^{-1}C(t/T)^{-2}R(K)\log(Th)V(t/T,u,v) + o((Th)^{-1}\log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1}C(t/T)^{-2}R(K)Z(t/T) + o((Th)^{-1}), & if \;\; \alpha > 1; \end{cases}$$

where V and Z are the functions defined in (5) *and* (6) *respectively.*

Proof. See Appendix. □

3. Prediction and estimation at the boundary regions

We call prediction the task of estimating $m((T+r)/T)$, $r \in \mathbb{N}$, under the model (1). Our predictor implements the standard idea of using specific one-sided kernels, $\mathcal{K}(u) := aK(u)\mathbb{1}_{\{u\leq 0\}}$, with $a^{-1} := \int_{-\infty}^{0} K(u)du$, intended to make the contribution of Θ_t larger as t becomes closer to T. When kernel weights are employed, we can establish the following

Theorem 3. *Given the time series $\{\Theta_t\}_{t=1}^T$ observed from the model* (1)*, if assumptions i) − iii) of Lemma 1 hold, for the estimator $\hat{m}$ having K_h as the weight function, and each $t \in [0, T+r]$, then*

$$E[\hat{m}(t/T)] - m(t/T) = h\left\{\left[\mu_1(\mathcal{K}) + h\mu_2(\mathcal{K})\frac{C'(t/T)}{C(t/T)}\right]m'(t/T) + h\frac{\mu_2(\mathcal{K})}{2}m''(t/T)\right\} + O\left(\nu(T,h,\alpha)\right),$$

and

$$Var[\hat{m}(t/T)] = \begin{cases} (Th)^{-\alpha}C^{-2}(t/T)V(t/T,u,v)\int\int\mathcal{K}(x)\mathcal{K}(y)|x-y|^{-\alpha}dxdy + o((Th)^{-\alpha}), & if \quad \alpha \in (0,1); \\ (Th)^{-1}C^{-2}(t/T)V(t/T,u,v)R(\mathcal{K})\log(Th) + o((Th)^{-1}\log(Th)), & if \quad \alpha = 1; \\ (Th)^{-1}C^{-2}(t/T)R(\mathcal{K})Z(t/T) + o((Th)^{-1}), & if \quad \alpha > 1. \end{cases}$$

Proof. See Appendix. □

When linear weights are employed, we get

Theorem 4. *Given the time series $\{\Theta_t\}_{t=1}^T$ observed from the model* (1)*, if assumptions i) − iii) of Lemma 1 hold, for the estimator $\hat{m}$ equipped with the weight function* (7)*, with W in place of K_h, for each $t \in [0, T+r]$, then*

$$E[\hat{m}(t/T)] - m(t/T) = h^2\frac{\mu_2^2(\mathcal{K}) - \mu_1(\mathcal{K})\mu_3(\mathcal{K})}{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})}\left\{\frac{C'(t/T)}{C(t/T)}m'(t/T) + \frac{1}{2}m''(t/T)\right\} + O\left(\delta(T,h,\alpha)\right)$$

and

$$Var[\hat{m}(t/T)] = \begin{cases} \dfrac{V(t/T,u,v)\int\int\mathcal{K}(x)\mathcal{K}(y)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}\{\mu_2(\mathcal{K}) - y\mu_1(\mathcal{K})\}|x-y|^{-\alpha}dxdy}{(Th)^{\alpha}C^2(t/T)\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}^2} + o((Th)^{-\alpha}), & if \quad \alpha \in (0,1); \\ \dfrac{V(t/T,u,v)\log(Th)\int\mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx}{ThC^2(t/T)\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}^2} + o((Th)^{-1}\log(Th)), & if \quad \alpha = 1; \\ \dfrac{Z(t/T)\int\mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx}{ThC^2(t/T)\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}^2} + o((Th)^{-1}), & if \quad \alpha > 1. \end{cases}$$

Proof. See Appendix. □

As expected, linear weights improve on the kernel ones, and this is because prediction is substantially a boundary estimation problem, in which case, as is well-known, local linear estimation is superior to a local constant fit. Also note, that the quantities derived in Theorems 3 and 4 explain the behaviour of our estimator with weights based on full kernels at boundary regions. In fact, for $\omega \in [0,1)$, letting $\mu_{j,\omega}(K) := \int_{\omega}^{1} x^jK(x)dx$ and $R_{\omega}(K) := \int_{\omega}^{1} K^2(x)dx$, for the estimator (4) at the (left) boundary point ωh, in the bias expression $\mu_j(\mathcal{K})$ is replaced by $\mu_{j,\omega}(K)$, and all is divided by $\mu_{0,\omega}(K)$, when kernel weights are employed, while, for the case of linear weights, $\mu_{2,\omega}(K)$ at the denominator is multiplied by $\mu_{0,\omega}(K)$. Concerning the variance expression, $R(\mathcal{K})$ is replaced by $R_\omega(K)$, the integrals are taken on $[\omega, 1]$ and all appears to be divided by $\mu^2_{0,\omega}(K)$ when kernel weights are employed, while, for the case of linear weights, $\mu_{2,\omega}(K)$ at the denominator is multiplied by $\mu_{0,\omega}(K)$.

4. Numerical results

4.1. Simulated data

We now consider an example in which m is a simple function, but with an error structure which is a circular version of an AR(1) model. Specifically we consider (1) with $m(t/T) = 2\tan^{-1}(0.01t - 10), t = 1, \ldots, T = 1000$ and errors ε_t satisfying

$$\varepsilon_t = \text{atan2}(2\sin(\varepsilon_{t-1}), 0.5\cos(\varepsilon_{t-1}) + 1)) + \nu_t$$

where ν_t are independent and identically distributed von Mises random variables with mean zero, and concentration $\kappa = 5$. An example dataset is shown in Figure 1 with the function m superimposed.

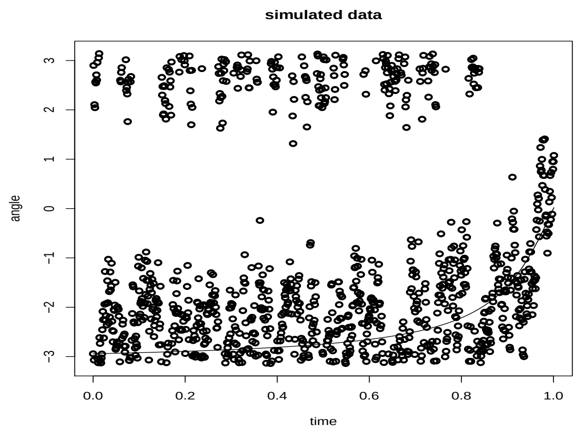

Figure 1: Simulated data from (1) with $m(t/1000) = 2\tan^{-1}(0.01t - 10)$ and $\varepsilon_t = \text{atan2}(2\sin(\varepsilon_{t-1}), 0.5\cos(\varepsilon_{t-1}) + 1)) + \nu_t$ and $\nu_t \sim vM(0, \kappa)$.

We simulated 1000 datasets from this model, and for each dataset we computed a one-step ahead forecast. In each simulation we store θ_{T+1} and $\hat{\theta}_{T+1}$, the latter of which is computed from a local constant estimator and a local linear estimator. In both cases, we choose the smoothing parameter by one-sided cross-validation (see next subsection for more detail) and, for reference purposes, by minimization of $\sum_{t=1}^{T-1}(1 - \cos(\hat{m}((t+1)/T) - m((t+1)/T)))$, in which $\hat{m}$ is estimated by a local constant, or local linear estimate (regression over full range and not a boundary estimate). The results are shown in Table 1. We can note that the smoothing parameter to optimally estimate m over the full range of t compares poorly when used for forecasting with that designed for one-step ahead prediction. We also note that for prediction of θ_{T+1} the local linar estimator gives a smaller bias than the local constant estimator, but with a higher variance (lower κ) and, overall, a slightly higher average error. Note also, that although we have used a maximum likelihood estimate of κ on the assumption of von Mises errors, we have observed that the distribution of $\hat{m}((t+1)/T)$ is bimodal.

	prediction of m									
smoothing selector	one-sided CV					"optimal" over full range				
forecast	bias (SE)		κ (SE)		error	bias (SE)		κ (SE)		error
local constant	0.029	(0.03)	1.79	(0.07)	0.339	1.019	(0.01)	9.25	(0.40)	0.505
local linear	0.014	(0.03)	1.59	(0.07)	0.380	0.681	(0.01)	9.18	(0.40)	0.266
	prediction of θ_{T+1}									
smoothing selector	one-sided CV					"optimal" over full range				
forecast	bias (SE)		κ (SE)		error	bias (SE)		κ (SE)		error
local constant	0.019	(0.02)	4.51	(0.18)	0.119	0.997	(0.03)	1.75	(0.07)	0.645
local linear	0.008	(0.02)	3.91	(0.16)	0.139	0.662	(0.03)	1.76	(0.07)	0.483

Table 1: Comparison of forecasts over 1000 simulations from model. Average error was taken relative to m and pairwise relative to θ_{T+1}, with bias and κ taken from maximum likelihood estimates assuming errors to be von Mises.

4.2. Wind direction data

We illustrate our methodology for real data using historic wind direction data recorded by the National Oceanic and Atmospheric Admistration's National Data Buoy Center (http://www.ndbc.noaa.gov/historical_data.shtml). We use standard meteorological data from 2009 monitored at station 41010 (120NM East of Cape Canaveral), which is automatically recorded every 30 minutes (at 20 and 50 past each hour). Wind direction (the direction the wind is coming from in degrees clockwise from true North) was converted to radians, and — for ease of presentation — taken to be on $(-\pi, \pi]$ from true South. Due to some missing values, we consider a contiguous period of records from 14th February to 31st August (199 days). The 9552 values were subsampled every 12th observation (4 times per day) leading to a time series of length 796. It was decided to obtain one-step ahead forecasts as if in the month of August (final 124 values). The data, together with autocorrelation plots, are shown in Figures 2 & 3. Note that fitting an AR model to the sin of the angles, and selecting the model complexity by AIC leads to a model of order 20, with cosines leading to a model of order 7, and it is clear from the plots that there are long-range correlations.

For real data, the function $m(\cdot)$ is unknown, so we take as our goal the problem of estimating θ_{T+1}, given previous observations $\theta_1, \ldots, \theta_T$ for $T = 672, \ldots, 795$. As a measure of accuracy we use

$$\sum_{T=672}^{795} (1 - \cos(\hat{\theta}_{T+1|T} - \theta_{T+1})). \tag{8}$$

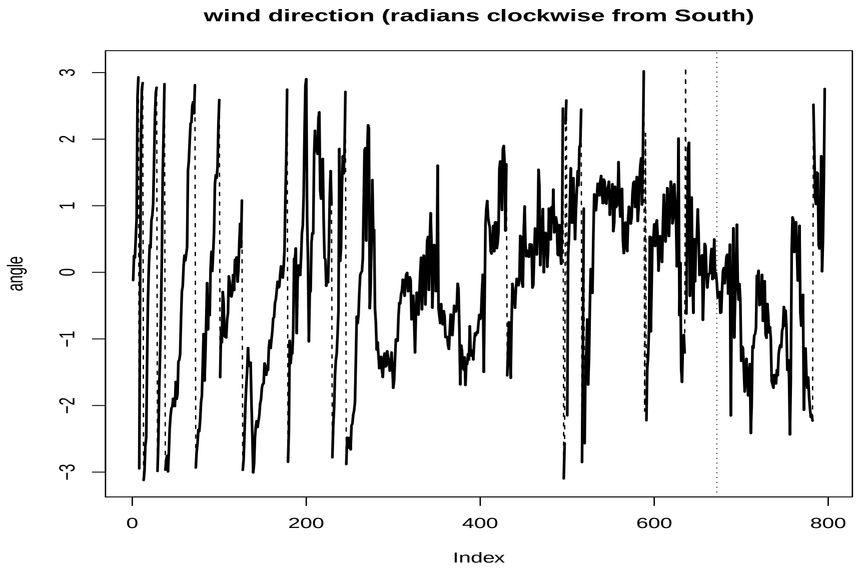

Figure 2: Wind directions (taken from South) 14th February – 31st August, 2009 sampled every 6 hours, starting at 2:50 a.m.. Dashed lines are used to join observations which are far apart on the circle (and so may be connected around the "back"). The vertical dotted line shows the period after which forecasting will be carried out.

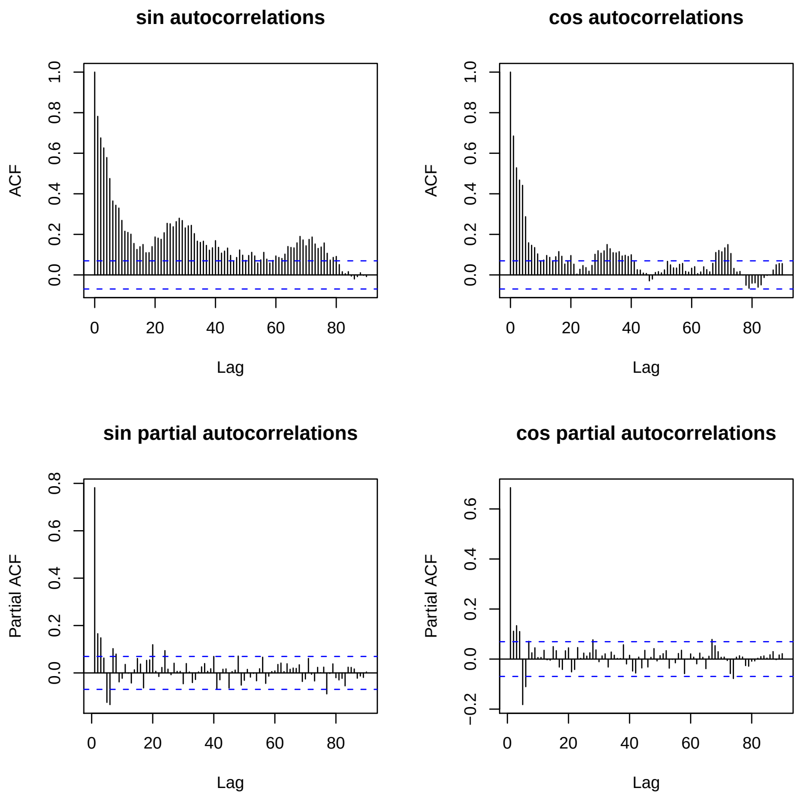

Figure 3: Autocorrelations and partial autocorrelations of sin and cosine for the data. Dashed lines show approximate confidence 95% intervals.

We suppose that for each value of T we can select a (different) smoothing parameter by cross-validation, which is chosen to minimize the prediction error of the one-step ahead forecasts over the second half of the data. That is, for each T we choose the smoothing parameter to minimize:

$$\sum_{t=T/2}^{T-1} (1 - \cos(\hat{\theta}_{t+1|t} - \theta_{t+1})). \tag{9}$$

We use $\hat{m}(\cdot)$ to obtain $\hat{\theta}$ in which the estimator in Equation (4) is equipped with one-sided gaussian kernels. Note that the normalizing constant a cancels in the atan2 function, which simplifies computations.

We compare the results for nonparametric local constant and local linear estimators. As a benchmark, we also consider a parametric model in which we use a Möbius transformation (Downs & Mardia, 2002) where, for each T, we predict θ_{T+1} using $\alpha_T + 2\tan^{-1}(w_T \tan(\theta_T - \beta_T)))$, with α_T, w_T and β_T chosen to minimize (9). Despite the fact that the time series itself is somewhat erratic the selected smoothing parameters did not vary much — the ranges of concentration were $(1.70, 1.76)$ and $(4.17, 4.26)$ for local constant and local linear estimates, respectively. The predicted values were also very similar at most time points; the sum of errors (8) was 26.0 for local constant, 26.5 for local linear and 28.8 for the Möbius model. Figure 4 indicates which of the nonparametric forecasts was better at each time point, and it can be seen that— as expected — the local linear prediction tended to do better when there was a local trend.

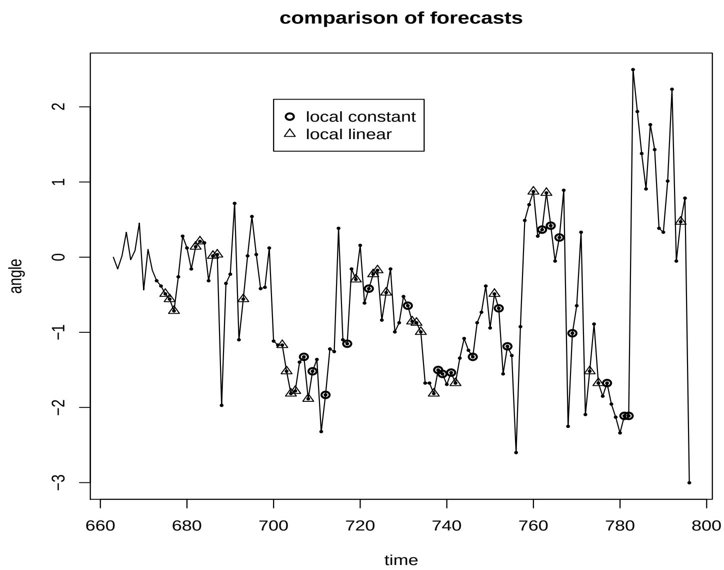

Figure 4: Data for August (forecasting period), with indication of method which gave more accurate prediction. Symbols indicate that the error of the inferior method was at least 50% more than the preferred method.

Appendix

Proof of Lemma 1. First observe that

$$E[\hat{m}_j(t/T)] = T^{-1} \sum_{i=1}^{T} K_h(i/T - t/T) m_j(i/T),$$

then, expand $m_j(i/T)$ in Taylor series around t/T, and, for each $\ell \geq 0$, use the approximation

$$T^{-1} \sum_{i=1}^{T} K_h(i/T - t/T)(i/T - t/T)^{\ell} = h^{\ell} \mu_{\ell}(K) + o(h^{\ell}), \tag{10}$$

and then use assumptions $i) - iii)$.

Concerning the variance, first note that

$$\Gamma_1(\ell) = s_{i/T} s_{(i+\ell)/T} \gamma_{cc}(\ell) + s_{i/T} c_{(i+\ell)/T} \gamma_{cs}(\ell) + c_{i/T} s_{(i+\ell)/T} \gamma_{sc}(\ell) + c_{i/T} c_{(i+\ell)/T} \gamma_{ss}(\ell),$$

and

$$\Gamma_2(\ell) = c_{i/T} c_{(i+\ell)/T} \gamma_{cc}(\ell) - c_{i/T} s_{(i+\ell)/T} \gamma_{cs}(\ell) - s_{i/T} c_{(i+\ell)/T} \gamma_{sc}(\ell) + s_{i/T} s_{(i+\ell)/T} \gamma_{ss}(\ell),$$

hence use conditions $a1) - a3)$ to get

$$\begin{aligned} Var[\hat{m}_j(t/T)] &= T^{-2} \sum_i \sum_k K_h(i/T - t/T) K_h(k/T - t/T) \Gamma_j(i-k) \\ &\sim T^{-2} \sum_i \sum_k K_h(i/T) K_h(k/T) |i-k|^{-\alpha} Q_j(i/T, k/T). \end{aligned}$$

Now, for $\alpha \in (0,1)$, by approximating the discrete sum by its integral, and after two Taylor series expansions we get

$$\begin{aligned} Var[\hat{m}_j(t/T)] &\sim (Th)^{-\alpha} \int \int K(x) K(y) |x-y|^{-\alpha} Q_j(hx, hy) dx dy \\ &= (Th)^{-\alpha} Q_j(u,v) \int \int K(x) K(y) |x-y|^{-\alpha} dx dy + o((Th)^{-\alpha}), \end{aligned}$$

and, for $\alpha = 1$ and each $c > 0$,

$$\begin{aligned} Var[\hat{m}_j(t/T)] &\sim (Th)^{-1} \int \int_{|x-y|>c/(Th)} K(x) K(y) |x-y|^{-1} Q_j(hx, hy) dx dy \\ &= (Th)^{-1} Q_j(u,v) \int K(x) dx \int_{|y|>c/(Th)} K(x+y) |y|^{-1} dy + O(hT^{-1}) \\ &\sim 2(Th)^{-1} Q_j(u,v) R(K) \log(Th) + o(T^{-1} h^{-1} \log(Th)). \end{aligned}$$

where $u \in (0, hx)$ and $v \in (0, hy)$. Finally, for $\alpha > 1$, we get

$$\begin{aligned} Var[\hat{g}_j(t/T)] &= T^{-2} \sum_{\ell=-\infty}^{+\infty} \Gamma_j(\ell) \sum_i K_h(i/T) K_h((i+\ell)/T) \\ &= (Th)^{-1} \sum_{\ell=-\infty}^{+\infty} \Gamma_j(\ell) R(K) + o((Th)^{-1}). \end{aligned}$$

Concerning the covariance, first observe that

$$Cov[\sin(\Theta_i), \cos(\Theta_k)] = s_{i/T} c_{k/T} \gamma_{cc}(i-k) - s_{i/T} s_{k/T} \gamma_{cs}(i-k) + c_{i/T} c_{k/T} \gamma_{sc}(i-k) - c_{i/T} s_{k/T} \gamma_{ss}(i-k),$$

then reason as in the previous paragraph. □

Proof of Theorem 1. For the bias, we start by expanding $\text{atan2}(\hat{m}_1, \hat{m}_2)$ in a Taylor series around (m_1, m_2), to get

$$\begin{aligned} \hat{m} = m &+ \frac{1}{C} [(\hat{m}_1 - m_1) f_c - (\hat{m}_2 - m_2) f_s] \\ &- \frac{1}{C^2} \left[\left\{ (\hat{m}_1 - m_1)^2 - (\hat{m}_2 - m_2)^2 \right\} f_c f_s + (\hat{m}_1 - m_1)(\hat{m}_2 - m_2)(f_c^2 - f_s^2) \right] \\ &+ O\left((\hat{m}_1 - m_1)^3 \right) + O\left((\hat{m}_2 - m_2)^3 \right). \quad (11) \end{aligned}$$

Now, taking expectations, and using the results in Lemma 1, from (11) it follows

$$\begin{aligned} E[\hat{m}(t/T)] - m(t/T) &= \frac{1}{C(t/T)} \{ E[\hat{m}_1(t/T) - m_1(t/T)] f_c(t/T) - E[\hat{m}_2(t/T) - m_2(t/T)] f_s(t/T) \} + O(\nu(T, h, \alpha)) \\ &= \frac{h^2 \mu_2(K)}{2C(t/T)} \{ m_1''(t/T) f_c(t/T) - m_2''(t/T) f_s(t/T) \} + O(\nu(T, h, \alpha)) \\ &= h^2 \mu_2(K) \left[\frac{m''(t/T)}{2} + \frac{C'(t/T) m'(t/T)}{C(t/T)} \right] + O(\nu(T, h, \alpha)). \end{aligned}$$

To derive the variance, use the expansion

$$\begin{aligned} \text{atan2}(\hat{m}_1, \hat{m}_2)^2 = \text{atan2}(m_1, m_2)^2 &- 2 \frac{m}{C} [(\hat{m}_2 - m_2) f_s - (\hat{m}_1 - m_1) f_c] \\ &- 2 \frac{m}{C^2} \left\{ [(\hat{m}_1 - m_1)^2 - (\hat{m}_2 - m_2)^2] f_c f_s + (\hat{m}_1 - m_1)(\hat{m}_2 - m_2)(f_c^2 - f_s^2) \right\} \\ &+ \frac{1}{C^2} \left\{ (\hat{m}_2 - m_2)^2 f_s^2 + (\hat{m}_1 - m_1)^2 f_c^2 - 2(\hat{m}_1 - m_1)(\hat{m}_2 - m_2) f_s f_c \right\} \\ &+ O\left((\hat{m}_1 - m_1)^3 \right) + O\left((\hat{m}_2 - m_2)^3 \right), \end{aligned}$$

then, after taking expectations, and noting that $E\left[(\hat{m}_j - m_j)^2 \right] = Var[\hat{m}_j] + \left(E[\hat{m}_j] - m_j \right)^2$, we obtain

$$Var[\hat{m}(t/T)] \approx \frac{1}{C^2(t/T)} \left\{ f_s^2(t/T)\, Var[\hat{m}_2(t/T)] + f_c^2(t/T)\, Var[\hat{m}_1(t/T)] - 2 f_s(t/T) f_c(t/T)\, Cov[\hat{m}_1(t/T), \hat{m}_2(t/T)] \right\},$$

and by Lemma 1 the result follows.

□

Proof of Lemma 2. Recalling assumption *ii*), and the approximation (10), the weight function in (7) can be approximated by

$$W(i/T - t/T) \sim T^{-1} K_h(i/T) h^2 \mu_2(K). \tag{12}$$

Consequently,

$$E[\hat{m}_j(t/T)] = T^{-1} h^2 \mu_2(K) \sum_{i=1}^{T} K_h(i/T - t/T) m_j(i/T).$$

Hence, expanding $m_j(i/T)$ in Taylor series around t/T, and using assumption *iii*) we get the result. To derive variance and covariance, use the approximation in (12), then reason as in the proof of Lemma 1. □

Proof of Theorem 2. To derive the bias and variance, follow the arguments used in the proof of Theorem 1, by expanding $\text{atan2}(\hat{m}_1, \hat{m}_2)$ and $\text{atan2}(\hat{m}_1, \hat{m}_2)^2$ in a Taylor series around $\left(h^2\mu_2(K)Cf_s, h^2\mu_2(K)Cf_c\right)$, then apply the results provided by Lemma 2. □

Proof of Theorem 3. By reasoning as in the proof of Lemma 1, we find that, for $j \in \{1, 2\}$,

$$E[\hat{m}_j(t/T)] = m_j(t/T) + h\mu_1(\mathcal{K}) m_j'(t/T) + \frac{h^2}{2}\mu_2(\mathcal{K}) m_j''(t/T),$$

$$Var[\hat{m}_j(t/T)] = \begin{cases} (Th)^{-\alpha} Q_j(u, v) \int\int \mathcal{K}(x)\mathcal{K}(y)|x - y|^{-\alpha} dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0, 1); \\ (Th)^{-1} Q_j(u, v) R(\mathcal{K}) \log(Th) + o((Th)^{-1} \log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1} \sum \Gamma_j(\ell) R(\mathcal{K}) + o((Th)^{-1}), & if \;\; \alpha > 1; \end{cases}$$

and

$$Cov[\hat{m}_1(t/T), \hat{m}_2(t/T)] = \begin{cases} (Th)^{-\alpha} U(u, v) \int\int \mathcal{K}(x)\mathcal{K}(y)|x - y|^{-\alpha} dxdy + o((Th)^{-\alpha}), & if \;\; \alpha \in (0, 1); \\ (Th)^{-1} U(u, v) R(\mathcal{K}) \log(Th) + o((Th)^{-1} \log(Th)), & if \;\; \alpha = 1; \\ (Th)^{-1} \sum \Gamma_3(\ell) R(\mathcal{K}) + o((Th)^{-1}), & if \;\; \alpha > 1, \end{cases}$$

Hence, the arguments used in the proof of Theorem 1 yield the results. □

Proof of Theorem 4. First of all, use the approximations

$$\sum_{j=1}^{T} \mathcal{K}_h(j/T - t/T)(j/T - t/T)^{\ell} = h^{\ell}\mu_{\ell}(\mathcal{K}) + o(h^{\ell})$$

in the weight function (7), and reason as in the proof of Lemma 2 to get, for $j \in \{1,2\}$,

$$E[\hat{m}_j(t/T)] = h^2\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}m_j(t/T) + \frac{h^4}{2}\left\{\mu_2^2(\mathcal{K}) - \mu_1(\mathcal{K})\mu_3(\mathcal{K})\right\}m_j''(t/T) + o(h^4),$$

$$Var[\hat{m}_j(t/T)] \sim$$
$$\begin{cases} h^4(Th)^{-\alpha}Q_j(u,v)\int\int \mathcal{K}(x)\mathcal{K}(y)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}\{\mu_2(\mathcal{K}) - y\mu_1(\mathcal{K})\}|x-y|^{-\alpha}dxdy + o(T^{-\alpha}h^{4-\alpha}), & if \quad \alpha \in (0,1); \\ h^3T^{-1}Q_j(u,v)\log(Th)\int \mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx + o(T^{-1}h^3\log(Th)), & if \quad \alpha = 1; \\ h^3T^{-1}\sum\Gamma_3(\ell)\int \mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx + o(T^{-1}h^3), & if \quad \alpha > 1; \end{cases} \tag{13}$$

and

$$Cov[\hat{m}_1(t/T), \hat{m}_2(t/T)] \sim$$
$$\begin{cases} \dfrac{h^4U(u,v)\int\int \mathcal{K}(x)\mathcal{K}(y)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}\{\mu_2(\mathcal{K}) - y\mu_1(\mathcal{K})\}|x-y|^{-\alpha}dxdy}{(Th)^{\alpha}} + o(T^{-\alpha}h^{4-\alpha}), & if \quad \alpha \in (0,1); \\ \dfrac{h^3U(u,v)\log(Th)\int \mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx}{T} + o(T^{-1}h^3\log(Th)), & if \quad \alpha = 1; \\ \dfrac{h^3\sum\Gamma_3(\ell)\int \mathcal{K}^2(x)\{\mu_2(\mathcal{K}) - x\mu_1(\mathcal{K})\}^2dx}{T} + o(T^{-1}h^3), & if \quad \alpha > 1, \end{cases} \tag{14}$$

then expand $\text{atan2}(\hat{m}_1, \hat{m}_2)$ and $\text{atan2}(\hat{m}_1, \hat{m}_2)^2$ in a Taylor series around $(h^2\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}Cf_s, h^2\{\mu_2(\mathcal{K}) - \mu_1^2(\mathcal{K})\}Cf_c)$, and reason as in the proof of Theorem 1. □

References

Artes, R., Paula, G.A. & Ranvaud, R. (2000). Analysis of circular longitudinal data based on generalised estimating equations. *Australian & New Zealand Journal of Statistics* **42**, 347–358.

Bosq, D. (1998). *Nonparametric Statistics For Stochastic Processes: Estimation And Prediction*. Springer; 2nd edition.

Breckling, J. (1989). *The Analysis of Directional time Serie: Applications to Wind Speed and Direction*, Lecture Notes in Statistics **61**. Berlin: Springer-Verlag.

Downs, T. D. & Mardia, K. V. (2002). Circular regression. *Biometrika*, **89**, 683–697.

Fan, J. & Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods* New York: Springer-Verlag.

Fisher, N. I. & Lee, A. J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society. Series B* **56**, 327–339.

Holzmann, H., Munk, A., Suster, M. & Zucchini, W. (2006). Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics* **13**, 325–347.

Hughes, G. (2007). *Multivariate and time series models for circular data with applications to protein conformational angles*. PhD thesis. University of Leeds, Leeds, UK.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. New York: Chapman and Hall/CRC.

Gijbels, I., Pope, A., Wand, M. P. (1999). Understanding exponential smoothing via kernel regression. *Journal of the Royal Statistical Society Series B* **61**, 39–50.

Lee, A. (2010). Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 477–486.