JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 808–829 (2020)
Published online 23 July 2020 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12549

ORIGINAL ARTICLE

MODELS FOR CIRCULAR DATA FROM TIME SERIES SPECTRA

MASANOBU TANIGUCHI,[a] SHOGO KATO,[b] HIROAKI OGATA[c*] AND ARTHUR PEWSEY[d]

[a] *Research Institute for Science & Engineering, Waseda University, Tokyo, Japan*
[b] *Department of Statistical Inference and Mathematics, Institute of Statistical Mathematics, Tokyo, Japan*
[c] *Faculty of Economics and Business Administration, Tokyo Metropolitan University, Tokyo, Japan*
[d] *Department of Mathematics, University of Extremadura, Cáceres, Spain*

Circular data are those for which the natural support is the unit circle and its toroidal extensions. Numerous constructions have been proposed which can be used to generate models for such data. We propose a new, very general, one based on the normalization of the spectra of complex-valued stationary processes. As illustrations of the new construction's application, we study models for univariate circular data obtained from the spectra of autoregressive moving average models and relate them to existing models in the literature. We also propose and investigate multivariate circular models obtained from the high-order spectra of stationary stochastic processes generated using linear filtering with an autoregressive moving average response function. A new family of distributions for a Markov process on the circle is also introduced. Results for asymptotically optimal inference for dependent observations on the circle are presented which provide a new paradigm for inference with circular models. The application of one of the new families of spectra-generated models is illustrated in an analysis of wind direction data.

Received 25 January 2019; Accepted 12 June 2020

Keywords: Joint circular distributions; high-order spectral density; stationary processes; optimal inference; local asymptotic normality test.

MOS subject classification: 37M10; 62H11.

1. INTRODUCTION

Circular data, the natural support for which is the unit circle, $\mathbb{S}^1$, arise in numerous scientific disciplines when observations are made on variables such as wind direction, river flow direction, migrating bird flight direction, and so on. The unit circle is also the natural support for variables such as the time of the day, week, month or year. When data are collected on $n \geq 2$ circular variables, the natural support becomes the n-torus, $\mathbb{T}^n = (\mathbb{S}^1)^n$. The field of statistics which deals with the analysis of circular data is referred to as circular statistics: itself a subfield of the wider discipline known as directional statistics (Mardia and Jupp, 1999). Certain key ideas underpinning both circular statistics and time series analysis can be traced to early work on random walks by Rayleigh (1880) and Pearson (1905).

Classical models for circular data include the wrapped normal (de Hars-Lorentz, 1913, pp. 24–25), von Mises (von Mises, 1918), wrapped Cauchy (Lévy, 1939) and wrapped stable (Wintner, 1947) distributions. In addition to the wrapping approach, in which a model defined on the real line, $\mathbb{R}$, is wrapped onto $\mathbb{S}^1$, historically there have been three other general methods that have been used to generate circular models: characterization through properties such as maximum likelihood and maximum entropy (von Mises, 1918; Mardia, 1972, pp. 65–66); radial projection of bivariate distributions defined on $\mathbb{R}^2$ (Mardia, 1972, p. 52); and stereographic projection of a distribution defined on $\mathbb{R}$ onto $\mathbb{S}^1$ (Jammalamadaka and SenGupta, 2001, p. 30).

---

* Correspondence to: Hiroaki Ogata, Faculty of Economics and Business Administration, Tokyo Metropolitan University, Tokyo, Japan.
E-mail: hiroakiogata@tmu.ac.jp

In recent years, the development of flexible models for circular data has received considerable attention in the literature, and more recently proposed general constructions include: conditional derivation (Jones and Pewsey, 2005); Möbius transformation (Kato and Jones, 2010); perturbation of existing circular models (Abe and Pewsey, 2011); transformation of argument (Jones and Pewsey, 2012; Abe *et al.*, 2013); Brownian motion (Kato and Jones, 2013) and parametric extension of the characteristic function of an existing circular model (Kato and Jones, 2015). Details of many of the models generated using these approaches are provided in Pewsey *et al.* (2013, Sec. 4.3).

A popular construction for bivariate circular distributions with specified marginal circular distributions is that of Wehrly and Johnson (1980) (see Section 4), studied recently in greater detail by Jones *et al.* (2013). A particularly appealing special case of it is the bivariate wrapped Cauchy model of Kato and Pewsey (2015). Alternative approaches used to generate models for toroidal data include wrapping (Johnson and Wehrly, 1977) and projection (Saw, 1983).

Spectral analysis has been used and developed in fields as varied as physics, electrical engineering, acoustics, geophysics, economics and medicine to identify and measure the strength of hidden periodicities in time series. Chapter 1 of Brillinger (2001) provides a brief historical review of developments. The finite Fourier transformation of a time series at frequency λ can be interpreted as a regression coefficient of the data at frequency λ when regressed on harmonic functions. The asymptotic variance and high-order moments of the finite Fourier transformation are referred to as the spectral density and high-order spectral densities respectively. For practical time series analysis, various parametric spectral models have been proposed: well-known ones including the autoregressive (AR(p)), moving average (MA(q)) and autoregressive moving average (ARMA(p, q)) models.

Recently, we observed that there are certain mathematical equivalences between ARMA(p, q) spectral densities and existing circular distributions: for example, between the AR(1) spectral density and the wrapped Cauchy distribution, and between the MA(1) spectral density and the cardioid distribution. Motivated by the discovery of these relations, here we propose a new construction for (multivariate) circular models based on the normalization of the high-order spectra of complex-valued stationary processes. Related to our approach, we note that Jupp and Mardia (1989) refer to the possibility of normalizing the spectral densities of stationary random fields on the n-dimensional lattice $\mathbb{Z}^n$ to obtain a very large class of distributions on $\mathbb{T}^n$. The new general construction is introduced in Section 2, together with results for the trigonometric moments of the resulting models. In Section 3 we propose models for univariate circular data obtained from the spectra of ARMA(p, q) models. In particular, we provide results for the circular models generated from the spectra of ARMA models with $p + q \leq 2$, and relate them to existing circular models. In Section 4 we propose multivariate circular models obtained from the high-order spectra of stationary stochastic processes generated using linear filtering with an ARMA response function. We illustrate the varied shapes such models can adopt, and propose a new family of distributions for a Markov process on the circle based on them. In Section 5 we address asymptotically optimal inference for dependent observations on the circle based on a local asymptotic normality property. Finally, in Section 6, models from an ARMA-generated family of circular densities are applied in an illustrative analysis of wind direction data. An appendix provides details of the use of the residue theorem in our spectral density-based approach, and the proofs of two theorems and a proposition.

2. GENERAL SPECTRAL DENSITY-BASED CONSTRUCTION

We propose a general approach to obtaining models for multivariate circular data derived from the high-order spectra of stationary processes.

Let $\{\boldsymbol{X}(t) = (X_1(t), \ldots, X_n(t))' : t \in \mathbb{Z}\}$ be a complex-valued stationary process with mean $\mathbf{0}$, autocovariance matrices $\{\Gamma(\ell) = (\Gamma_{jk}(\ell) : j, k = 1, \ldots, n), \ell \in \mathbb{Z}\}$ and spectral density matrix $\tilde{\boldsymbol{f}}(\omega) = \{\tilde{f}_{jk}(\omega) : j, k = 1, \ldots, n, \omega \in [-\pi, \pi]\}$. Then there exists an orthogonal-increment process $\mathbf{Z}(\omega) = (Z_1(\omega), \ldots, Z_n(\omega))'$ on $[-\pi, \pi]$ such that

$$\boldsymbol{X}(t) = \int_{-\pi}^{\pi} e^{-it\omega} d\mathbf{Z}(\omega),$$

DOI: 10.1111/jtsa.12549

where i is the imaginary unit,

$$E[d\mathbf{Z}(\omega)\overline{d\mathbf{Z}(\omega)}] = \tilde{\boldsymbol{f}}(\omega)d\omega,$$
$$E[d\mathbf{Z}(\omega)\overline{d\mathbf{Z}(\lambda)}] = \mathbf{0}, \quad \omega \neq \lambda,$$

(see, e.g., Hannan, 1970, p. 47), $\bar{z}$ denoting the complex conjugate of z. Furthermore, we make Assumption 1.

Assumption 1. For every $n \in \mathbb{N}$, the $2n$th moments

$$\Gamma_{11\ldots nn}(t_1, \ldots, t_n) \equiv E[X_1(0)\overline{X_1(t_1)} \ldots X_n(0)\overline{X_n(t_n)}]$$

exist, and satisfy

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_n=-\infty}^{\infty} |\Gamma_{11\ldots nn}(t_1, \ldots, t_n)| < \infty.$$

Then there exists a non-negative integrable function, $\tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n)$, such that

$$\begin{aligned} &E[dZ_1(\omega_1)\overline{dZ_1(\omega_1)} \ldots dZ_n(\omega_n)\overline{dZ_n(\omega_n)}] \\ &= \tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n)d\omega_1 \ldots d\omega_n \end{aligned} \tag{2.1}$$

and

$$\tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n) = \frac{1}{(2\pi)^n} \sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_n=-\infty}^{\infty} \Gamma_{11\ldots nn}(t_1, \ldots, t_n)e^{-\mathrm{i}(t_1\omega_1+\ldots+t_n\omega_n)}$$

(see, e.g., Brillinger, 2001, p. 101). The function $\tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n)$ is referred to as the $2n$th order spectral density of $\{\boldsymbol{X}(t)\}$. Whilst $\tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n)$ is a non-negative 2π-periodic function whose support is the n-torus, $\mathbb{T}^n$, it is not, in general, a probability density function (pdf). However, it can clearly be normalized, through multiplication by

$$c_n = \left(\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n)d\omega_1 ... d\omega_n\right)^{-1},$$

to obtain the periodic pdf,

$$f_{1\ldots n}(\omega_1, \ldots, \omega_n) = c_n\tilde{f}_{1\ldots n}(\omega_1, \ldots, \omega_n) \tag{2.2}$$

on $\mathbb{T}^n$. The normalizing constant c_n can be calculated using the residue theorem (see Appendix A.1). From its construction, the family with joint circular density (2.2) inherits the great flexibility of the high-order spectral densities underpinning it.

The characteristic function of a vector of circular random variables, $(\Omega_1, \ldots, \Omega_n)$, is made up from its trigonometric moments, defined as

$$\phi(p_1, \ldots, p_n) \equiv E[e^{\mathrm{i}(p_1\omega_1+\ldots+p_n\omega_n)}], \quad p_1, \ldots, p_n \in \mathbb{Z}. \tag{2.3}$$

The trigonometric moments of the distribution with joint pdf $f_{1\ldots n}(\omega_1, \ldots, \omega_n)$ can be calculated as

$$\phi(p_1, \ldots, p_n) = c_n\Gamma_{11\ldots nn}(p_1, \ldots, p_n).$$

DOI: 10.1111/jtsa.12549

In particular, for the univariate case $n = 1$, the trigonometric moments are given by

$$\phi(p_1) = c_1\Gamma_{11}(p_1), \quad p_1 \in \mathbb{Z}.$$

This implies that the elements of the autocovariance function of the base univariate stationary process are proportional to the trigonometric moments of the circular distribution obtained from the spectral density of that process.

When the components of $(Z_1(\omega), \dots, Z_n(\omega))$ are mutually independent, (2.1) reduces to

$$E[dZ_1(\omega_1)\overline{dZ_1(\omega_1)}] \dots E[dZ_n(\omega_n)\overline{dZ_n(\omega_n)}] = \tilde{f}_{11}(\omega_1) \dots \tilde{f}_{nn}(\omega_n) d\omega_1 \dots d\omega_n,$$

where $\tilde{f}_{jj}(\omega_j)d\omega_j = E[dZ_j(\omega_j)\overline{dZ_j(\omega_j)}], j = 1, \dots, n$, and $\tilde{f}_{jj}(\omega_j)$ is the spectral density of $\{X_j(t)\}$. Hence,

$$\tilde{f}_{1\dots n}(\omega_1, \dots, \omega_n) = \prod_{j=1}^{n} \tilde{f}_{jj}(\omega_j)$$

and

$$f_{1\dots n}(\omega_1, \dots, \omega_n) = \prod_{j=1}^{n} f_j(\omega_j),$$

where $f_j(\omega_j) = c_{jj}\tilde{f}_{jj}(\omega_j)$ is the pdf of ω_j and c_{jj} is its normalizing constant.

3. UNIVARIATE CIRCULAR DISTRIBUTIONS FROM ARMA SPECTRA

The popular Box–Jenkins approach to analysing non-seasonal univariate time series is founded on the ARMA(p, q) family of autoregressive moving average models (Box *et al.*, 2016). Using a reduced notation for simplicity, here we express the spectral density of a univariate stationary ARMA(p, q) process as

$$\tilde{f}(\omega) = \frac{\sigma^2}{2\pi} \frac{|1 + \sum_{k=1}^{q} \beta_k e^{\mathrm{i}k\omega}|^2}{|1 + \sum_{j=1}^{p} \alpha_j e^{\mathrm{i}j\omega}|^2}, \tag{3.1}$$

where σ^2 is the variance of the innovation process and the $\alpha_j, j = 1, \dots, p$, and $\beta_k, k = 1, \dots, q$, are its AR and MA coefficients respectively.

We consider a new, very general, family of univariate circular distributions with a density derived from the spectral density (3.1) with $\beta_k = \rho_k e^{-\mathrm{i}k\gamma_k}$ and $\alpha_j = r_j e^{-\mathrm{i}j\mu_j}$, namely

$$f(\omega) = \frac{c}{2\pi} \frac{\left|1 + \sum_{k=1}^{q} (\rho_k e^{-\mathrm{i}k\gamma_k}) e^{\mathrm{i}k\omega}\right|^2}{\left|1 + \sum_{j=1}^{p} (r_j e^{-\mathrm{i}j\mu_j}) e^{\mathrm{i}j\omega}\right|^2}, \tag{3.2}$$

where $\gamma_1, \dots, \gamma_q$, $\rho_1, \dots, \rho_q$, $\mu_1, \dots, \mu_p$ and $r_1, \dots, r_p$ are unknown parameters, and the normalizing constant c, which depends on them, is chosen so that $f(\omega)$ integrates to 1.

In the following subsections we focus on those special cases of density (3.2) for which $p+q \leq 2$. We provide their normalizing constants and trigonometric moments and, where possible, identify those existing circular models to which they correspond.

DOI: 10.1111/jtsa.12549

3.1. AR(1)-Generated Model

The spectral density of a first-order autoregressive (AR(1)) model with parameter $\alpha_1 = r_1 e^{-i\mu_1}$ is, from (3.1),

$$\tilde{f}(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 + \alpha_1 e^{i\omega}|^2}$$

and, from (3.2), the density of the corresponding univariate circular model is

$$f(\omega) = \frac{c}{2\pi} \frac{1}{1 + r_1^2 + 2r_1 \cos(\omega - \mu_1)},$$

the denominator of which can be recognized to be that of a wrapped Cauchy density (see, e.g., Mardia and Jupp, 1999, pp. 51–52) with mean direction $\mu_1 + \pi$ and mean resultant length r_1. It therefore follows that $c = 1 - r_1^2$ and the p_1th trigonometric moment is $\phi(p_1) = \{r_1 e^{i(\mu_1+\pi)}\}^{p_1}$, $p_1 \geq 1$. In addition it holds from general theory that $\phi(0) = 1$ and $\phi(-p_1) = \overline{\phi(p_1)}$ for this model and the subsequent circular distributions from ARMA spectra discussed in this section.

3.2. MA(1)-Generated Model

For a first-order moving average (MA(1)) process with parameter $\beta_1 = \rho_1 e^{-i\gamma_1}$, the circular density (3.2) becomes

$$f(\omega) = \frac{c}{2\pi} \left\{1 + \rho_1^2 + 2\rho_1 \cos(\omega - \gamma_1)\right\},$$

which can be re-expressed as

$$f(\omega) = \frac{c(1 + \rho_1^2)}{2\pi} \left\{1 + 2\rho \cos(\omega - \gamma_1)\right\},$$

where $\rho = \rho_1/(1 + \rho_1^2) \in [0, 1/2]$. The functional part of the density can be recognized as that of a cardioid distribution (see, e.g., Mardia and Jupp, 1999, Sec. 3.5.5) with mean direction γ_1 and mean resultant length ρ. It follows that $c = (1 + \rho_1^2)^{-1}$ and the trigonometric moments are $\phi(p_1) = \rho$ if $p_1 = 1$ and $\phi(p_1) = 0$ if $p_1 \geq 2$.

3.3. AR(2)-Generated Model

The case of (3.2) derived from the spectral density of an AR(2) process is

$$\begin{aligned} f(\omega) &= \frac{c}{2\pi} \frac{1}{|1 + r_1 e^{-i\mu_1} e^{i\omega} + r_2 e^{-i2\mu_2} e^{i2\omega}|^2} \\ &= \frac{c}{2\pi} \frac{1}{1 + r_1^2 + r_2^2 + r \cos(\omega - \mu) + 2r_2 \cos 2(\omega - \mu_2)}, \end{aligned} \tag{3.3}$$

where $\mu_1 \in [-\pi, \pi)$, $\mu_2 \in [0, \pi)$, $r_1, r_2 \geq 0$, $\mu = \arg[\cos\mu_1 + r_2 \cos(2\mu_2 - \mu_1) + i\{\sin\mu_1 + r_2 \sin(2\mu_2 - \mu_1)\}]$ and $r = 2r_1[1 + r_2^2 + 2r_2 \cos 2(\mu_1 - \mu_2)]^{1/2}$. The following result is essential to obtaining its normalizing constant c. A proof is given in Appendix A.2.

Proposition 1. The normalizing constant c of density (3.3) is given by

$$c = r_2^2 \left|1 - |c_+|^2\right| \left|1 - |c_-|^2\right| A, \tag{3.4}$$

DOI: 10.1111/jtsa.12549

where

$$c_+ = \frac{-r_1 e^{-i\mu_1} + \sqrt{r_1^2 e^{-2i\mu_1} - 4r_2 e^{-2i\mu_2}}}{2r_2 e^{2i\mu_2}}, \quad c_- = \frac{-r_1 e^{-i\mu_1} - \sqrt{r_1^2 e^{-2i\mu_1} - 4r_2 e^{-2i\mu_2}}}{2r_2 e^{2i\mu_2}}$$

and

$$A = \begin{cases} \dfrac{|1 - c_+\overline{c_-}|^2}{|1 - |c_+\overline{c_-}|^2|}, & |c_+|, |c_-| < 1 \text{ or } |c_+|, |c_-| > 1, \\ \dfrac{|c_+ - c_-|^2}{||c_+|^2 - |c_-|^2|}, & |c_+|^{-1}, |c_-| < 1 \text{ or } |c_+|^{-1}, |c_-| > 1. \end{cases}$$

It follows from expression (A.7) given in Appendix A.2 that, to ensure that $f(\omega) < \infty$ for any ω, $|c_+| \neq 1$ and $|c_-| \neq 1$.

The model with density (3.3) is related to one proposed by Kato and Jones (2013) with density

$$f_{KJ}(\omega) = \frac{C}{\{1 + \xi_1^2 - 2\xi_1 \cos(\omega - \eta_1)\}\{1 + \xi_2^2 - 2\xi_2 \cos(\omega - \eta_2)\}}, \tag{3.5}$$

where $\eta_1, \eta_2 \in [-\pi, \pi)$, $\xi_1, \xi_2 \in [0, 1)$, and C is the normalizing constant. The relationship between the two is clarified in Theorem 1, a proof of which is given in Appendix A.3.

Theorem 1. Let $\mathcal{F}_{AR2}$ be a set of the density functions defined by $\mathcal{F}_{AR2} = \{f_{AR2} : \mu_1, \mu_2 \in [-\pi, \pi),\ r_1, r_2 \geq 0\}$, where f_{AR2} is the density (3.3). Suppose $\mathcal{F}_{KJ} = \{f_{KJ} : \eta_1, \eta_2 \in [-\pi, \pi),\ \xi_1, \xi_2 \in [0, 1)\}$, where f_{KJ} is the density (3.5). Then $\mathcal{F}_{AR2} = \mathcal{F}_{KJ}$. In addition,

$$f_{AR2}(\theta; \tilde{\mu}_1, \tilde{\mu}_2, \tilde{r}_1, \tilde{r}_2) = f_{KJ}(\theta; \eta_1, \eta_2, \xi_1, \xi_2), \tag{3.6}$$

for any $\theta \in [-\pi, \pi)$, where $\tilde{\mu}_1 = \pi + \arg\{\xi_1 \cos\eta_1 + \xi_2 \cos\eta_2 + i(\xi_1 \sin\eta_1 + \xi_2 \sin\eta_2)\}$, $\tilde{\mu}_2 = (\eta_1 + \eta_2)/2$, $\tilde{r}_1 = \{\xi_1^2 + \xi_2^2 + 2\xi_1\xi_2 \cos(\eta_1 - \eta_2)\}^{1/2}$ and $\tilde{r}_2 = \xi_1\xi_2$.

Theorem 1 leads immediately to the following result concerning the trigonometric moments of the model with density (3.3). Assume that c_+ and c_- are defined as in Proposition 1. Then, if $|c_+|, |c_-| < 1$, then the p_1th trigonometric moment of the distribution with density (3.3) for $p_1 \geq 1$, $\phi(p_1; c_+, c_-)$, is given by

$$\phi(p_1; c_+, c_-) = \begin{cases} \dfrac{(1 - |c_-|^2)(1 - \overline{c_+}c_-)\, c_+^{p_1+1} - (1 - |c_+|^2)(1 - c_+\overline{c_-})\, c_-^{p_1+1}}{(c_+ - c_-)(1 - |c_+\overline{c_-}|^2)}, & c_+ \neq c_-, \\ \dfrac{1 + p_1 + (1 - p_1)|c_+|^2}{1 + |c_+|^2} c_+^{p_1}, & c_+ = c_-. \end{cases}$$

If $|c_+|$ or $|c_-|$ is greater than 1, then the p_1th trigonometric moment of (3.3) is

$$\phi(p_1; c_+, c_-) = \begin{cases} \phi(p_1; \overline{c_+}^{-1}, c_-), & |c_+| > 1,\ |c_-| < 1, \\ \phi(p_1; c_+, \overline{c_-}^{-1}), & |c_+| < 1,\ |c_-| > 1, \\ \phi(p_1; \overline{c_+}^{-1}, \overline{c_-}^{-1}), & |c_+|, |c_-| > 1. \end{cases}$$

Thus, the normalizing constant and trigonometric moments can be expressed in closed form.

DOI: 10.1111/jtsa.12549

3.4. MA(2)-Generated Model

The circular density corresponding to the spectral density of an MA(2) process is

$$
\begin{aligned}
f(\omega) &= \frac{c}{2\pi}\left|1+\rho_1 e^{-\mathrm{i}\gamma_1} e^{\mathrm{i}\omega}+\rho_2 e^{-\mathrm{i}2\gamma_2} e^{\mathrm{i}2\omega}\right|^2 \\
&= \frac{1}{2\pi}\left\{1+2\rho_1^* \cos(\omega-\gamma^*)+2\rho_2^* \cos 2(\omega-\gamma_2)\right\},
\end{aligned}
\tag{3.7}
$$

where $\gamma_1, \gamma_2 \in [-\pi, \pi)$, $\rho_1, \rho_2 \geq 0$, $\gamma^* = \arg[\cos\gamma_1 + \rho_2 \cos(2\gamma_2 - \gamma_1) + \mathrm{i}\{\sin\gamma_1 + \rho_2 \sin(2\gamma_2 - \gamma_1)\}]$, $\rho_1^* = c\rho_1[1 + \rho_2^2 + 2\rho_2 \cos 2(\gamma_1 - \gamma_2)]^{1/2}$, $\rho_2^* = c\rho_2$, and the normalizing constant is $c = 1/(1 + \rho_1^2 + \rho_2^2)$. This model appears to be a new one which extends the cardioid distribution obtained when $\rho_2 = 0$. As such, it provides an alternative to Papakonstantinou's extension of the cardioid distribution studied in detail by Abe *et al.* (2009). It is straightforward to see that the p_1th trigonometric moment of the model (3.7) for $p_1 \geq 1$ is

$$
\phi(p_1) = \begin{cases} \rho_1^* e^{\mathrm{i}\gamma^*} & p_1 = 1, \\ \rho_2^* e^{\mathrm{i}2\gamma_2}, & p_1 = 2, \\ 0, & \text{otherwise.} \end{cases}
$$

3.5. ARMA(1,1)-Generated Model

The circular density derived from the spectral density of an ARMA(1,1) process is

$$
f(\omega) = \frac{c}{2\pi}\frac{|1+\rho_1 e^{-\mathrm{i}\gamma_1} e^{\mathrm{i}\omega}|^2}{|1+r_1 e^{-\mathrm{i}\mu_1} e^{\mathrm{i}\omega}|^2} = \frac{c}{2\pi}\frac{1+\rho_1^2+2\rho_1\cos(\omega-\gamma_1)}{1+r_1^2+2r_1\cos(\omega-\mu_1)},
\tag{3.8}
$$

where $\mu_1, \gamma_1 \in [-\pi, \pi)$, $r_1 \in [0, 1) \cup (1, \infty)$ and $\rho_1 \geq 0$. Without loss of generality, henceforth we assume that $r_1 < 1$. Notice that $f(\omega; \mu_1, \gamma_1, r_1, \rho_1) = f(\omega; \mu_1, \gamma_1, r_1^{-1}, \rho_1)$.

Using the residue theorem (see Appendix A.1), it can be shown that the normalizing constant of density (3.8) is given by

$$
c = \frac{1-r_1^2}{1+\rho_1^2-2\rho_1 r_1\cos(\mu_1-\gamma_1)}
$$

and the p_1th trigonometric moment by

$$
\phi(p_1) = \frac{(-r_1 e^{\mathrm{i}\mu_1})^{p_1-1}\{1-r_1\rho_1 e^{\mathrm{i}(\mu_1-\gamma_1)}\}(\rho_1 e^{\mathrm{i}\gamma_1}-r_1 e^{\mathrm{i}\mu_1})}{1+\rho_1^2-2\rho_1 r_1\cos(\mu_1-\gamma_1)}, \quad p_1 \geq 1.
$$

The model with density (3.8) is related to that of Kato and Jones (2015) with density

$$
f(\omega) = \frac{1}{2\pi}\left[1+2\tau^2\frac{\tau\cos(\theta-\eta)-\bar{\alpha}_2}{\tau^2+\bar{\alpha}_2^2+\bar{\beta}_2^2-2\tau\{\bar{\alpha}_2\cos(\theta-\eta)+\bar{\beta}_2\sin(\theta-\eta)\}}\right],
\tag{3.9}
$$

where $\eta \in [-\pi, \pi)$, $\tau \in [0, 1)$, and $(\bar{\alpha}_2, \bar{\beta}_2) \neq (\tau, 0)$ satisfy $(\bar{\alpha}_2 - \tau^2)^2 + \bar{\beta}_2^2 \leq \tau^2(1-\tau)^2$. It is straightforward to show that density (3.8) reduces to density (3.9) if $\mu_1 = \eta + \arg(\bar{\alpha}_2 + \mathrm{i}\bar{\beta}_2) + \pi$, $r_1 = (\bar{\alpha}_2^2 + \bar{\beta}_2^2)^{1/2}/\tau$, $\gamma_1 = \eta + \arg(\tau^2 - \bar{\alpha}_2 - \mathrm{i}\bar{\beta}_2)$, $\rho_1 = \{a \pm (a^2-4)^{1/2}\}/2$ if $(\bar{\alpha}_2, \bar{\beta}_2) \neq (\tau^2, 0)$ and $\rho_1 = 0$ otherwise, where $a = \{(\bar{\alpha}_2 - \tau^2)^2 + \bar{\beta}_2^2\}^{1/2}/\tau + \tau(1 - \tau^2)/\{(\bar{\alpha}_2 - \tau^2)^2 + \bar{\beta}_2^2\}^{1/2}$.

DOI: 10.1111/jtsa.12549

The relationship between models (3.8) and (3.9) implies that the sine-skewed wrapped Cauchy distribution (Umbach and Jammalamadaka, 2009; Abe and Pewsey, 2011), for which the density is a three-parameter special case of density (3.9), is also another special case of the general model with density (3.2).

The ARMA(0,0) model corresponds to the circular uniform model with $c = 1$, density $f(\omega) = 1/(2\pi)$, and p_1th trigonometric moment

$$\phi(p_1) = \begin{cases} 1, & p_1 = 0, \\ 0, & p_1 \neq 0. \end{cases}$$

4. ARMA-GENERATED MULTIVARIATE CIRCULAR MODELS

In Section 2 we introduced a general approach to obtaining multivariate circular models whose densities are normalized spectral densities of complex-valued stationary processes. In this section we consider a particular subfamily of such models generated using linear filtering with an ARMA response function.

Suppose that each component $X_j(t)$ of the complex-valued stationary process $\boldsymbol{X}(t)$ can be expressed in linear form as

$$X_j(t) = \sum_{\ell=0}^{\infty} A_j(\ell) u_j(t-\ell), \tag{4.1}$$

where $\{A_j(\ell)\}$ is summable and $\{u_j(t)\}$ is a stationary innovation with

$$u_j(t) = \int_{-\pi}^{\pi} e^{\mathrm{i}t\omega} dY_j(\omega).$$

Here $\{Y_j(\omega)\}$ is an orthogonal-increment process with $2n$th order spectral density denoted by $\tilde{f}^{Y}_{1\ldots n}(\omega_1, \ldots, \omega_n)$. Let

$$A_j(\omega) \equiv \frac{\sigma_j}{\sqrt{2\pi}} \frac{1 + \sum_{k=1}^{q_j} (\rho_{jk} e^{-\mathrm{i}k\gamma_{jk}}) e^{\mathrm{i}k\omega}}{1 + \sum_{m=1}^{p_j} (r_{jk} e^{-\mathrm{i}m\mu_{jk}}) e^{\mathrm{i}m\omega}}$$

be the response function corresponding to (4.1). Recalling (2.1),

$$E[dX_1(\omega_1)\overline{dX_1(\omega_1)} \ldots dX_n(\omega_n)\overline{dX_n(\omega_n)}] = |A_1(\omega_1) \cdots A_n(\omega_n)|^2 \tilde{f}^{Y}_{1\ldots n}(\omega_1, \ldots, \omega_n) d\omega_1 \ldots d\omega_n,$$

say. We will refer to

$$f^{\mathrm{ARMA}}_{1\ldots n}(\omega_1, \ldots, \omega_n) \equiv c_n |A_1(\omega_1) \cdots A_n(\omega_n)|^2 \tilde{f}^{Y}_{1\ldots n}(\omega_1, \ldots, \omega_n), \tag{4.2}$$

where c_n is a normalizing constant, as an ARMA-generated joint circular density. If we assume that each response function $A_j(\omega_j), j = 1, ..., n$, satisfies the stationary condition, then $A_j(\omega_j)$ can be expressed as a linear summable Fourier series

$$\sum_{k=0}^{\infty} a_k^{(j)} e^{\mathrm{i}k\omega_j}$$

(see Example 2.9.7 of Brillinger, 2001, p. 37). Furthermore, if we assume that $\tilde{f}^{Y}_{1\cdots n}(\cdot)$ is of the form (4.3), then using Lemma 2.7.2 of Brillinger (2001, p. 29), it can be seen that (4.2) satisfies Assumption 1.

4.1. Shapes of ARMA-Generated Joint Circular Densities

To illustrate the varied shapes that density (4.2) can adopt, here we present graphical representations of some special cases of it for $n = 2, 3$ and

$$\tilde{f}^{Y}_{1\ldots n}(\omega_1, \ldots, \omega_n) \propto \exp\left[\sum_{j_1=0}^{p_1} \cdots \sum_{j_n=0}^{p_n} a_{j_1,\ldots,j_n} \cos(j_1\omega_1) \ldots \cos(j_n\omega_n)\right], \tag{4.3}$$

where the $a_{j_1,\ldots,j_n}$, $j_1 = 0, \ldots, p_1, \ldots, j_n = 0, \ldots, p_n$, are constants. The spectral density (4.3) is a multivariate extension of the exponential spectral density

$$\tilde{f}(\omega) = \sigma^2 \exp(1 + \theta_1 \cos\omega + \cdots + \theta_p \cos p\omega)$$

of Bloomfield (1973), the normalized counterpart of which is a generalized von Mises circular density (e.g., Maksimov, 1967; Gatto and Jammalamadaka, 2007). Specifically, when $n = 2$ we use

$$\tilde{f}^{Y}_{12}(\omega_1, \omega_2) = \exp\left[1 + \cos\omega_1 - \cos\omega_2 + \cos\omega_1 \cos\omega_2\right],$$

and when $n = 3$,

$$\begin{aligned}\tilde{f}^{Y}_{123}(\omega_1, \omega_2, \omega_3) = \exp[&1 + \cos\omega_1 - \cos\omega_2 + \cos\omega_3 \\ &+ \cos\omega_1\cos\omega_2 - \cos\omega_1\cos\omega_3 + \cos\omega_2\cos\omega_3 + \cos\omega_1\cos\omega_2\cos\omega_3].\end{aligned}$$

For the bivariate circular density

$$f^{\mathrm{ARMA}}_{12}(\omega_1, \omega_2) \propto |A_1(\omega_1)A_2(\omega_2)|^2 \tilde{f}^{Y}_{12}(\omega_1, \omega_2), \tag{4.4}$$

we illustrate cases with $|A_2(\omega_2)|^2 = |1 - 0.15e^{-\mathrm{i}\pi/2}e^{-\mathrm{i}\omega_2}|^{-2}$ and the following three alternatives for the squared response function $|A_1(\omega_1)|^2$:

1. $|A^{(1)}_1(\omega_1)|^2 = |1 - 0.3e^{-\mathrm{i}\pi}e^{-\mathrm{i}\omega_1}|^{-2}$;
2. $|A^{(2)}_1(\omega_1)|^2 = |1 - 0.9e^{-\mathrm{i}\pi}e^{-\mathrm{i}\omega_1}|^{-2}$;
3. $|A^{(3)}_1(\omega_1)|^2 = |1 - 0.3e^{-\mathrm{i}\pi}e^{-\mathrm{i}\omega_1} - 0.1e^{-\mathrm{i}3\pi/2}e^{-\mathrm{i}2\omega_1}|^{-2}$.

The first two correspond to AR(1) processes, the second being close to a unit root process, and the third is that of an AR(2) process. Graphical summaries for these selections are presented in Figure 1. The panels making up its three rows correspond to the choices $|A^{(1)}_1(\omega_1)|^2$, $|A^{(2)}_1(\omega_1)|^2$ and $|A^{(3)}_1(\omega_1)|^2$ respectively. The plots in its left-hand column are contour plots of the bivariate circular ARMA-generated density (4.4) unfolded from $\mathbb{T}^2$ to the plane . The plots in its right-hand column represent the corresponding conditional densities of $\Omega_1|\omega_2 = 4\pi/3$. Whilst, on $\mathbb{T}^2$, all three bivariate densities depicted in Figure 1 are unimodal, those in its first two rows are symmetric about $\omega_1 = \pi$ whilst that in its third row is skew. The conditional densities of $\Omega_1|\omega_2 = 4\pi/3$ range from being uni- or bimodal on $\mathbb{S}^1$ and differentiable or non-differentiable at the main mode.

To illustrate the flexibility of the trivariate circular density

$$f^{\mathrm{ARMA}}_{123}(\omega_1, \omega_2, \omega_3) \propto |A_1(\omega_1)A_2(\omega_2)A_2(\omega_3)|^2 \tilde{f}^{Y}_{123}(\omega_1, \omega_2, \omega_3), \tag{4.5}$$

we set $|A_1(\omega_1)|^2 = |1 - 0.3e^{-\mathrm{i}\pi}e^{-\mathrm{i}\omega_1}|^{-2}$, $|A_2(\omega_2)|^2 = |1 + 0.7e^{-\mathrm{i}\pi/2}e^{-\mathrm{i}\omega_2}|^{-2}$ and $|A_3(\omega_3)|^2 = |1 - 0.5e^{-\mathrm{i}\pi/3}e^{-\mathrm{i}\omega_3} - 0.2e^{-\mathrm{i}2\pi/3}e^{-\mathrm{i}2\omega_3}|^{-2}$. Figure 2 presents contour plots of the bivariate circular densities of $(\Omega_1, \Omega_2)|\omega_3 = \pi$ (left), $(\Omega_1, \Omega_3)|\omega_2 = \pi$ (middle) and $(\Omega_2, \Omega_3)|\omega_1 = \pi$ (right), unfolded from $\mathbb{T}^2$ to the plane. Despite all of them being unimodal on $\mathbb{T}^2$, otherwise their shapes are highly variable.

J. Time Ser. Anal. **41**: 808–829 (2020)
DOI: 10.1111/jtsa.12549

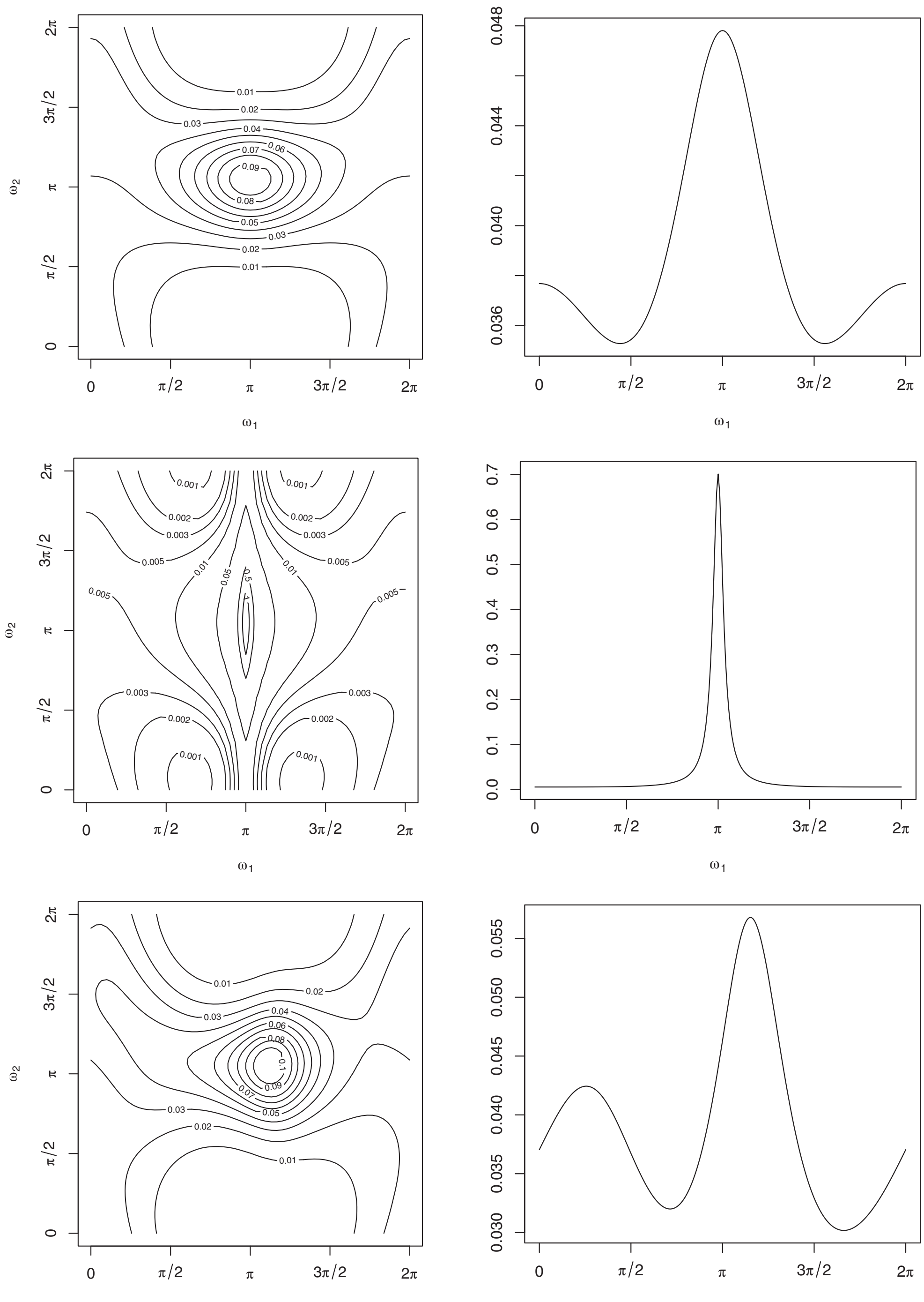

Figure 1. Contour plots (left-hand column) of the bivariate circular density (4.4) unfolded from $\mathbb{T}^2$ to the plane, with $\tilde{f}^{Y}_{12}(\omega_1, \omega_2) = \exp\left[1 + \cos\omega_1 - \cos\omega_2 + \cos\omega_1\cos\omega_2\right]$, $|A_2(\omega_2)|^2 = |1 - 0.15e^{-i\pi/2}e^{-i\omega_2}|^{-2}$ and $|A_1^{(1)}(\omega_1)|^2$ (top row), $|A_1^{(2)}(\omega_1)|^2$ (middle row) and $|A_1^{(3)}(\omega_1)|^2$ (bottom row). The panels in the right-hand column represent the corresponding marginal circular densities of $\Omega_1|\omega_2 = 4\pi/3$

DOI: 10.1111/jtsa.12549

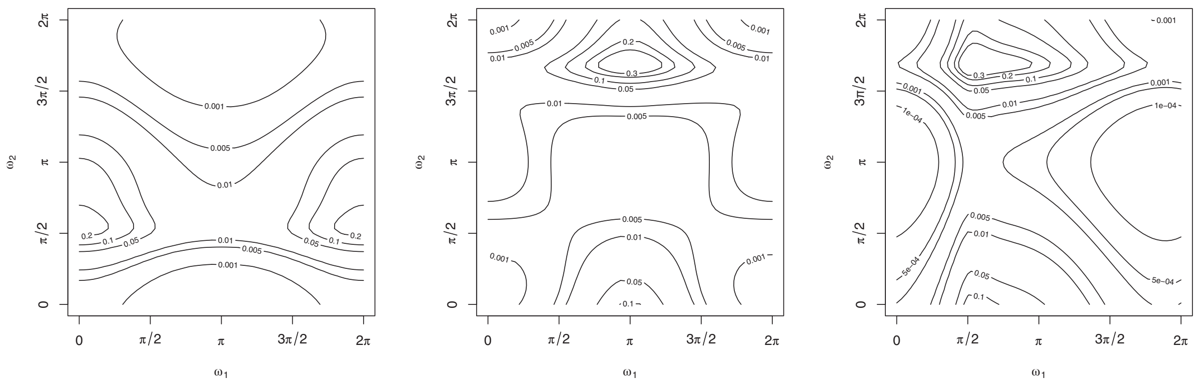

Figure 2. Contour plots of the bivariate circular densities of $(\Omega_1, \Omega_2)|\omega_3 = \pi$ (left), $(\Omega_1, \Omega_3)|\omega_2 = \pi$ (middle) and $(\Omega_2, \Omega_3)|\omega_1 = \pi$ (right) unfolded from $\mathbb{T}^2$ to the plane, for the trivariate circular density (4.5) with $\tilde{f}^Y_{123}(\omega_1, \omega_2, \omega_3) = \exp[1 + \cos\omega_1 - \cos\omega_2 + \cos\omega_3 + \cos\omega_1\cos\omega_2 - \cos\omega_1\cos\omega_3 + \cos\omega_2\cos\omega_3 + \cos\omega_1\cos\omega_2\cos\omega_3]$, $|A_1(\omega_1)|^2 = |1 - 0.3e^{-i\pi}e^{-i\omega_1}|^{-2}$, $|A_2(\omega_2)|^2 = |1 + 0.7e^{-i\pi/2}e^{-i\omega_2}|^{-2}$ and $|A_3(\omega_3)|^2 = |1 - 0.5e^{-i\pi/3}e^{-i\omega_3} - 0.2e^{-i2\pi/3}e^{-i2\omega_3}|^{-2}$

4.2. ARMA-Generated Markov Process

Wehrly and Johnson (1980) considered a family of bivariate circular distributions with density

$$f(\omega_1, \omega_2) = 2\pi g(2\pi\{F_1(\omega_1) - qF_2(\omega_2)\})f_1(\omega_1)f_2(\omega_2), \tag{4.6}$$

where $f_1(\omega_1)$ and $f_2(\omega_2)$ are specified marginal densities, $F_1(\omega_1)$ and $F_2(\omega_2)$ are their corresponding distribution functions, g is a circular density and $q \in \{-1, 1\}$ is fixed. Using density (4.6) with $f_1 = f_2$, and hence $F_1 = F_2$, Wehrly and Johnson (1980) proposed a family of distributions for a Markov process on the circle, $\{\Omega_t\}$, with $f(\omega_1) = f_1(\omega_1)$ as the density of the initial distribution and

$$f(\omega_n|\omega_{n-1}) = f(\omega_n|\omega_1, \dots, \omega_{n-1}) = 2\pi g(2\pi\{F_1(\omega_n) - qF_1(\omega_{n-1})\})f_1(\omega_n) \tag{4.7}$$

as the stationary transition density.

Here we propose an alternative family of distributions for a Markov process on the circle derived from the ARMA-generated joint circular density (4.2). Specifically, let the density of the initial distribution be $f^{\mathrm{ARMA}}(\omega_1) \propto |A_1(\omega_1)|^2\tilde{f}^Y_1(\omega_1)$ and the stationary transition density be

$$f^{\mathrm{ARMA}}(\omega_n|\omega_{n-1}) = f^{\mathrm{ARMA}}(\omega_n|\omega_1, \dots, \omega_{n-1}) \propto |A_n(\omega_n)|^2\varphi(\omega_n, \omega_{n-1}), \tag{4.8}$$

say, for some suitable function φ. Then the joint density for $\{\Omega_1, \dots, \Omega_n\}$,

$$\begin{aligned} f^{\mathrm{ARMA}}(\omega_1, \dots, \omega_n) &\propto f^{\mathrm{ARMA}}(\omega_n|\omega_{n-1}) \dots f^{\mathrm{ARMA}}(\omega_2|\omega_1)|A_1(\omega_1)|^2\tilde{f}^Y_1(\omega_1) \\ &\propto \tilde{f}^Y_1(\omega_1)\prod_n |A_n(\omega_n)|^2\varphi(\omega_n, \omega_{n-1}), \end{aligned}$$

provides a very general Markov-type joint circular density.

The stationary transition density (4.7) of Wehrly and Johnson (1980) can be seen to be a special case of (4.8) with $\varphi(\omega_n, \omega_{n-1}) = g(2\pi\{F_1(\omega_n) - qF_1(\omega_{n-1})\})$ and $|A_n(\omega_n)|^2 = f_1(\omega_n)$.

DOI: 10.1111/jtsa.12549

5. ASYMPTOTICALLY OPTIMAL INFERENCE FOR STOCHASTIC PROCESSES WITH CIRCULAR DISTRIBUTIONS

5.1. LAN Property

Suppose that $\boldsymbol{\Omega}_n = (\Omega_1, \dots, \Omega_n)'$ is a sequence of random variables forming a stochastic process with joint circular density $f_{1\dots n}(\omega_1, \dots, \omega_n)$ as in (2.2). In what follows we assume that $f_{1\dots n}(\omega_1, \dots, \omega_n)$ depends on an s-dimensional unknown parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)'$ in $\Theta \equiv \mathrm{Int}\{\mathcal{H}\} \subset \mathbb{R}^s$, where $\mathcal{H}$ is a compact subset of $\mathbb{R}^s$ and $\mathrm{Int}\{\cdot\}$ is the interior set of $\{\cdot\}$. Under this assumption, and for simplicity, we denote $f_{1\dots n}(\omega_1, \dots, \omega_n)$ as $p_{\theta}^n(\boldsymbol{\omega}_n)$.

In the field of directional statistics, the Fisher–von Mises–Langevin distribution is perhaps the best-known one after the spherical uniform distribution. It has density

$$f(\boldsymbol{x}) \propto \exp\{\kappa \boldsymbol{\mu}' \boldsymbol{x}\}, \tag{5.1}$$

where $\kappa > 0$ and $\boldsymbol{x}$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_q)'$ are unit vectors in $\mathbb{R}^q$. Because $\mu_1^2 + \cdots + \mu_q^2 = 1$, $\boldsymbol{\mu}$ is referred to as being a 'curved parameter'. Ley *et al.* (2013) developed LAN theory for such curved families. Paindaveine and Verdebout (2017) considered the case when $\kappa = \kappa_n \searrow 0$, and showed that this leads to non-regular LAN theory.

Because here we consider very general dependent observations with joint circular distribution $p_{\theta}^n(\boldsymbol{w}_n)$, in what follows we restrict ourselves to the case of joint circular distributions that are 'regular parametric families'. For (5.1), this corresponds to taking $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)' = (\mu_1, \dots, \mu_{q-1})'$, $(s = q - 1)$.

Let the conditional density of $\boldsymbol{\Omega}_k$ given $\boldsymbol{\Omega}_{k-1}$ be

$$p_{\theta}^k(\boldsymbol{\omega}_k | \boldsymbol{\omega}_{k-1}) \equiv \frac{p_{\theta}^k(\boldsymbol{\omega}_k)}{p_{\theta}^{k-1}(\boldsymbol{\omega}_{k-1})}, \tag{5.2}$$

where

$$p_{\theta}^{k-1}(\boldsymbol{\omega}_{k-1}) = \int p_{\theta}^k(\boldsymbol{\omega}_k)\, d\omega_k.$$

We assume that $p_{\theta}^k(\boldsymbol{\omega}_k)$, $k = 1, 2, \dots, n$, are differentiable with respect to $\boldsymbol{\theta}$ and that

$$\frac{\partial}{\partial \boldsymbol{\theta}} p_{\theta}^{k-1}(\boldsymbol{\omega}_{k-1}) = \int \frac{\partial}{\partial \boldsymbol{\theta}} p_{\theta}^k(\boldsymbol{\omega}_k)\, d\omega_k, \quad k = 1, \dots, n. \tag{5.3}$$

Then the log-likelihood function based on $\boldsymbol{\Omega}_n$ is

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log\{p_{\theta}^k(\boldsymbol{\omega}_k | \boldsymbol{\omega}_{k-1})\}, \tag{5.4}$$

where we take $p_{\theta}^0(\boldsymbol{\omega}_0) = 1$.

Now define

$$\boldsymbol{\theta}_n = \boldsymbol{\theta} + \frac{1}{\sqrt{n}} \boldsymbol{h}, \tag{5.5}$$

where $\boldsymbol{h} = (h_1, \dots, h_s)' \in \mathbb{R}^s$. For two values $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_n$, the log-likelihood ratio is

$$\Lambda_n(\boldsymbol{\theta}, \boldsymbol{\theta}_n) = \sum_{k=1}^{n} \log\left\{\frac{p_{\boldsymbol{\theta}_n}(\boldsymbol{\Omega}_k | \boldsymbol{\Omega}_{k-1})}{p_{\boldsymbol{\theta}}(\boldsymbol{\Omega}_k | \boldsymbol{\Omega}_{k-1})}\right\} = 2 \sum_{k=1}^{n} \log\{Y_{n,k} + 1\}, \tag{5.6}$$

DOI: 10.1111/jtsa.12549

where

$$Y_{n,k} = \frac{\sqrt{p^k_{\theta_n}(\mathbf{\Omega}_k|\mathbf{\Omega}_{k-1})} - \sqrt{p^k_{\theta}(\mathbf{\Omega}_k|\mathbf{\Omega}_{k-1})}}{\sqrt{p^k_{\theta}(\mathbf{\Omega}_k|\mathbf{\Omega}_{k-1})}}.$$

Let

$$\xi_{n,k} = \frac{1}{2\sqrt{n}}\boldsymbol{h}' \frac{\frac{\partial}{\partial\theta} p^k_{\theta_n}(\mathbf{\Omega}_k|\mathbf{\Omega}_{k-1})}{p^k_{\theta_n}(\mathbf{\Omega}_k|\mathbf{\Omega}_{k-1})} = \frac{1}{2\sqrt{n}}\boldsymbol{h}'\boldsymbol{\eta}_k \quad \text{(say)}.$$

The LAN property requires the following three assumptions to hold.

Assumption 2. There exist positive definite matrices

$$E[\boldsymbol{\eta}_k\boldsymbol{\eta}_k'] = \Gamma_k(\boldsymbol{\theta}), \quad k = 1, 2, \dots, n, \tag{5.7}$$

whose elements are uniformly bounded with respect to k, and the limit

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \Gamma_k(\boldsymbol{\theta}) \equiv \Gamma(\boldsymbol{\theta})$$

exists.

Assumption 3.

$$E[\{Y_{n,k} - \xi_{n,k}\}^2] = o\left(\frac{1}{n}\right), \tag{5.8}$$

where $o\left(\frac{1}{n}\right)$ is uniform with respect to $k = 1, \dots, n$.

For the ARMA-generated joint circular density (4.2), assume that the conditional density corresponding to $|A_j(\omega_j)|^2$ is of the following ARMA type:

$$\begin{aligned} f(\omega) &= c\,\frac{\prod_{k=1}^{q}(1 - a_k e^{\mathrm{i}b_k} e^{\mathrm{i}\omega})(1 - a_k e^{-\mathrm{i}b_k} e^{-\mathrm{i}\omega})}{\prod_{j=1}^{p}(1 - c_j e^{\mathrm{i}d_j} e^{\mathrm{i}\omega})(1 - c_j e^{-\mathrm{i}d_j} e^{-\mathrm{i}\omega})} \\ &= c\,\frac{\prod_{k=1}^{q}\{1 + a_k^2 - 2a_k\cos(b_k + \omega)\}}{\prod_{j=1}^{p}\{1 + c_j^2 - 2c_j\cos(d_j + \omega)\}}, \end{aligned} \tag{5.9}$$

where a_k, b_k, c_k, d_k are real constants, and all $|a_k| < 1$, and all $|c_k| < 1$, and $a_1, \dots, a_q, c_1, \dots, c_p$ are mutually different. Note that the density (5.9) can also be expressed in the form of (3.2). Let $\boldsymbol{\theta} = \mathrm{vec}[\{a_k\}, \{b_k\}, \{c_k\}, \{d_k\}]$. Assumptions 2 and 3 can be checked using the ideas in Taniguchi (1983, p. 166) and Appendix A.1, although the calculations are very complicated.

Assumption 4.

$$\sum_{k=1}^{n} \xi_{n,k}^2 \xrightarrow{p} \frac{1}{4}\boldsymbol{h}'\Gamma(\boldsymbol{\theta})\boldsymbol{h} = \tau^2 \quad \text{(say)}. \tag{5.10}$$

DOI: 10.1111/jtsa.12549

For the ARMA-generated joint circular density (4.2) with transition density (5.9), if we assume stationarity, that is, $f(\omega)$ is the same for all $j = 1, \ldots, n$, we can check Assumptions 2–4 using the residue theorem (see Appendix A).

Theorem 2 (LAN property)**.** Suppose that Assumptions 2–4 hold. Then, the log-likelihood ratio has the stochastic expansion

$$\Lambda_n(\boldsymbol{\theta}, \boldsymbol{\theta}_n) = \boldsymbol{h}'\Delta_n - \frac{1}{2}\boldsymbol{h}'\Gamma(\boldsymbol{\theta})\boldsymbol{h} + \eta_n(\boldsymbol{h}), \tag{5.11}$$

where $\Gamma(\boldsymbol{\theta})$ is an $s \times s$ positive definite matrix, the central sequence $\Delta_n \xrightarrow{d} N(\mathbf{0}, \Gamma(\boldsymbol{\theta}))$ and $\boldsymbol{\eta}_n(\boldsymbol{h}) \xrightarrow{p} \mathbf{0}$ for every $\boldsymbol{h} \in \mathbb{R}^s$.

5.2. Parameter Estimation

Here we consider the estimation of $\boldsymbol{\theta}$. In what follows, we denote the distribution law of a random vector $\boldsymbol{Y}_n$ under the probability measure $P_{\boldsymbol{\theta}}^{(n)}$ by $\mathcal{L}(\boldsymbol{Y}_n | P_{\boldsymbol{\theta}}^{(n)})$, and weak convergence to $\boldsymbol{\eta}$ by $\mathcal{L}(\boldsymbol{Y}_n | P_{\boldsymbol{\theta}}^{(n)}) \xrightarrow{d} \boldsymbol{\eta}$. Let the class $\mathcal{A}$ of estimators $\{\boldsymbol{S}_n\}$ of $\boldsymbol{\theta}$ be

$$\mathcal{A} = [\{\boldsymbol{S}_n\} : \mathcal{L}\{\sqrt{n}(\boldsymbol{S}_n - \boldsymbol{\theta}_n) | P_{\boldsymbol{\theta}_n}^{(n)}\} \xrightarrow{d} \boldsymbol{\eta}_{\boldsymbol{\theta}}],$$

where the probability distribution $\boldsymbol{\eta}_{\boldsymbol{\theta}}$ generally depends on $\{\boldsymbol{S}_n\}$. Denote by L the class of loss functions $\ell : \mathbb{R}^s \to [0, \infty)$ of the form $\ell(\boldsymbol{x}) = \tau(|\boldsymbol{x}|)$ which satisfies $\tau(0) = 0$ and $\tau(a) \le \tau(b)$ if $a \le b$. Typical examples are $\ell(\boldsymbol{x}) = I(|\boldsymbol{x}| > a)$ and $\ell(\boldsymbol{x}) = |\boldsymbol{x}|^p$, $p \ge 1$, where $I(\cdot)$ is the indicator function.

Under Assumptions 2–4, a sequence $\{\hat{\boldsymbol{\theta}}_n\}$ of estimators of $\boldsymbol{\theta}$ is said to be a sequence of asymptotically centering estimators if

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) - \Gamma^{-1}(\boldsymbol{\theta})\Delta_n = o_p(1) \text{ in } P_{\boldsymbol{\theta}}^{(n)}.$$

Following the arguments in Taniguchi and Kakizawa (2000, p. 69, see the proof of Theorem 3.1.9.), we get Proposition 2.

Proposition 2. Suppose Assumptions 2–4 hold and $\{\boldsymbol{S}_n\} \in \mathcal{A}$. Let Δ be a random vector distributed as $N(\mathbf{0}, \Gamma(\boldsymbol{\theta}))$. Then the following statements hold:

(i) For any $\ell \in L$ with $E\{\ell(\Delta)\} < \infty$,

$$\liminf_{n\to\infty} E[\ell\{\sqrt{n}(\boldsymbol{S}_n - \boldsymbol{\theta})\} | P_{\boldsymbol{\theta}}^{(n)}] \ge E\{\ell(\Gamma^{-1}(\boldsymbol{\theta})\Delta)\}.$$

(ii) If

$$\limsup_{n\to\infty} E[\ell\{\sqrt{n}(\boldsymbol{S}_n - \boldsymbol{\theta})\} | P_{\boldsymbol{\theta}}^{(n)}] \le E\{\ell(\Gamma^{-1}(\boldsymbol{\theta})\Delta)\},$$

for a non-constant $\ell \in L$ with $E\{\ell(\Delta)\} < \infty$, then $\boldsymbol{S}_n$ is a sequence of asymptotically centering estimators.

From Proposition 2 we can see that $\{\hat{\boldsymbol{\theta}}_n \in \mathcal{A}\}$ is asymptotically efficient if it is asymptotically centering.

Proposition 3. Under Assumptions 2–4, the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$, is asymptotically efficient.

DOI: 10.1111/jtsa.12549

Table I. Estimated size of the test of (5.13) based on (5.12), for a nominal size of 0.05, and, in brackets, the estimated standard error, calculated using 1000 samples of size n simulated from density (3.9) with $\eta = 0$, $\tau = 0.5$, $\bar{\alpha}_{2,0} = 0.25(1 + \psi/\sqrt{2})$ and $\bar{\beta}_{2,0} = 0.25\psi/\sqrt{2}$

	$n = 25$	$n = 50$	$n = 200$	$n = 1000$	$n = 2000$
$\psi = 0$	0.656 (0.015)	0.352 (0.015)	0.117 (0.010)	0.067 (0.008)	**0.058** (0.007)
$\psi = 0.3$	0.581 (0.016)	0.357 (0.015)	0.110 (0.010)	**0.060** (0.008)	**0.048** (0.007)
$\psi = 0.6$	0.582 (0.016)	0.335 (0.015)	0.124 (0.010)	**0.064** (0.008)	**0.048** (0.007)
$\psi = 0.9$	0.659 (0.015)	0.464 (0.016)	0.135 (0.011)	0.083 (0.009)	**0.060** (0.008)

5.3. Hypothesis Testing

Finally, we consider the inferential problem of testing for a specific case, $\boldsymbol{\theta}_0 \in \mathbb{R}^s$, of the vector $\boldsymbol{\theta}$. Let $\mathcal{M}(\boldsymbol{A})$ denote the linear space spanned by the columns of a matrix $\boldsymbol{A}$. Under the null hypothesis H_{θ_0}, $\boldsymbol{\theta} - \boldsymbol{\theta}_0 \in \mathcal{M}(\boldsymbol{A})$ for some given $s \times (s - r)$ matrix $\boldsymbol{A}$ of full rank. Then, proceeding similarly as in Strasser (1985, Sec. 82) and Taniguchi and Kakizawa (2000, p. 78), we observe that, under H_{θ_0}, the test statistic

$$n(\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0)'[\Gamma(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}) - \Gamma(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})\boldsymbol{A}\{\boldsymbol{A}'\Gamma(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})\boldsymbol{A}\}^{-1}\boldsymbol{A}'\Gamma(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})](\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0) \tag{5.12}$$

is asymptotically distributed as a chi-squared random variable with r degrees of freedom, χ_r^2, and the test based on it is locally asymptotically optimal.

5.4. Simulation Study

To investigate the finite sample size performance of the test based on (5.12), we carried out two Monte Carlo experiments: the first to investigate its true size, and the second its power under a simultaneous two-sided alternative hypothesis. In both experiments, random samples of size $n = 25,\ 50,\ 200,\ 1000$ and 2000 were simulated from the ARMA(1,1)-generated model, under the reparametrization of the distribution of Kato and Jones (2015) given in (3.9), for specified values of the parameters. For each combination of the sample size and parameter values, 1000 samples were simulated using the algorithm given in the Supplementary Material to Kato and Jones (2015). We investigated the case when $A = (1, 0, 0, 0)^T$ in (5.12), corresponding to the testing scenario in which the value of the mean direction is considered irrelevant. The values of $\Gamma(\boldsymbol{\theta})$ in (5.12) were calculated using the expected Fisher information matrix given in the Supplementary Material to Kato and Jones (2015).

In the first experiment, we tested

$$H_0 : (\tau, \bar{\alpha}_2, \bar{\beta}_2) = (0.5, \bar{\alpha}_{2,0}, \bar{\beta}_{2,0}) \text{ against } H_1 : (\tau, \bar{\alpha}_2, \bar{\beta}_2) \neq (0.5, \bar{\alpha}_{2,0}, \bar{\beta}_{2,0}), \tag{5.13}$$

using samples simulated under H_0 for the values of the kurtosis and skewness parameters, $\bar{\alpha}_{2,0}$ and $\bar{\beta}_{2,0}$, specified in Table I. That table gives the estimated size of the simultaneous two-sided test based on (5.12) for a nominal size of 0.05. The case $\psi = 0$ corresponds to an AR(1)-generated model, and the greater the value of ψ, the greater the skewness and kurtosis of the underlying distribution. For a true size of 0.05 and 1000 samples, a 95% confidence interval for the size of the test calculated using normal approximation is $(0.036, 0.064)$ to 3 decimal places. The estimated size values lying inside that interval are highlighted in bold in Table I. Clearly, when $n \leq 200$ the test is liberal, its true size only approaching the nominal size when $n \geq 1000$. In particular, when $n = 2000$, the test appears to maintain the nominal size whatever the value of ψ. In general, the convergence of the true size of the test to the nominal size is slowest when $(\bar{\alpha}_2, \bar{\beta}_2)$ is close to the boundary of the parameter space.

In the second experiment, we tested

$$H_0 : (\tau, \bar{\alpha}_2, \bar{\beta}_2) = (0.5, 0.25, 0) \text{ against } H_1 : (\tau, \bar{\alpha}_2, \bar{\beta}_2) \neq (0.5, 0.25, 0), \tag{5.14}$$

DOI: 10.1111/jtsa.12549

Table II. Estimated power of the test of (5.14) based on (5.12), for a nominal size of 0.05, and, in brackets, the estimated standard error, calculated using 1000 samples of size n simulated from density (3.9) with $\eta = 0$, $\tau = 0.5$, $\bar{\alpha}_2 = 0.25(1 + \psi/\sqrt{2})$ and $\bar{\beta}_2 = 0.25\psi/\sqrt{2}$

	$n = 25$	$n = 50$	$n = 200$	$n = 1000$	$n = 2000$
$\psi = 0.3$	0.656 (0.015)	0.515 (0.016)	0.563 (0.016)	0.991 (0.003)	**1.000** (0.000)
$\psi = 0.6$	0.806 (0.013)	0.838 (0.012)	0.999 (0.001)	1.000 (0.000)	**1.000** (0.000)
$\psi = 0.9$	0.987 (0.004)	0.996 (0.002)	1.000 (0.000)	1.000 (0.000)	**1.000** (0.000)

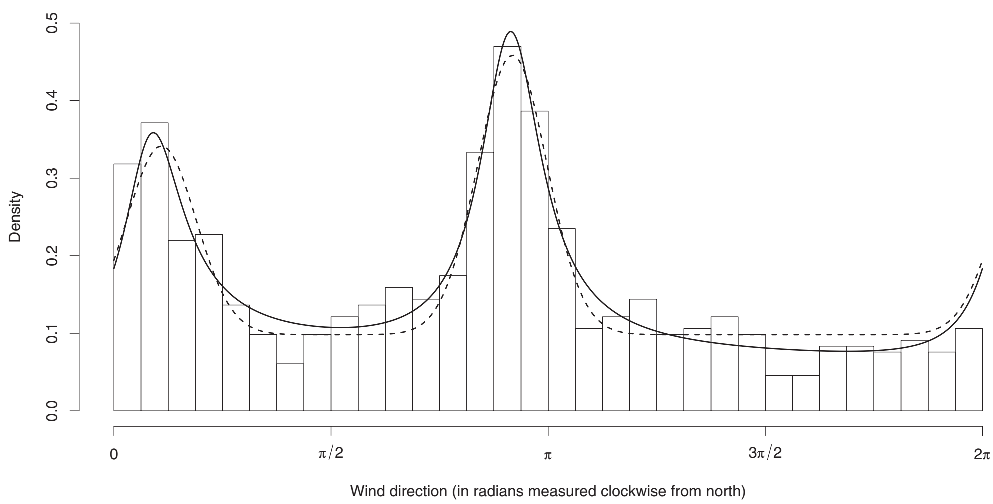

Figure 3. Histogram of the $n = 672$ wind direction observations with the ML fitted densities for the ARMA(2,2) model (continuous curve) and reduced von Mises mixture model (6.1) with $\kappa_1 = \kappa_2$ (dashed curve) superimposed

for samples simulated under H_1. The null hypothesis corresponds to the underlying model being AR(1)-generated with arbitrary mean direction and mean resultant length 0.5, whilst the alternative corresponds to it being full ARMA(1,1)-generated. Table II displays the estimated power of the test of (5.14) based on (5.12) for a nominal size of 0.05, and the estimated standard error. Apart from when $\psi = 0.3$, the power increases with n and ψ, as expected. In particular, when $\psi = 0.9$, the power of the test is high even for small n. Given the results in Table I, the powers highlighted in bold in Table II can be considered to closely approximate the true power of the size 0.05-test. Clearly, for those settings, the test based on (5.12) maintains the nominal significance level well and its power is essentially 1.

6. ILLUSTRATIVE APPLICATION

We present results obtained from fitting the ARMA-generated circular models of Section 3 to wind direction data recorded at the Agricultural Meteorology weather station in Afton, Wyoming, USA. The data are available from the United States Bureau of Reclamation website https://www.usbr.gov/pn/agrimet/webagdayread.html, the wind direction having been recorded each quarter of an hour to the nearest degree measured clockwise from north. We analysed the $n = 672$ observations corresponding to the time period of a week between January 1 and January 7, 2015. A histogram of the data converted to radians in $[0, 2\pi)$ is presented in Figure 3. A visual inspection of that histogram reveals a data distribution with two main modes: a smaller one centred around the NNE direction (0.3 radians) and a larger around SSE (2.9 radians), the separation between the two being around 2.5 radians (145 degrees).

DOI: 10.1111/jtsa.12549

Table III. AIC values for maximum likelihood fitted ARMA(p, q)-type models with $p + q \leq 4$

q	p	AIC
0	0	2470.11
0	1	2427.47
1	0	2419.02
0	2	2282.10
1	1	2347.54
2	0	2336.47
0	3	2263.74
1	2	2269.72
2	1	2308.15
3	0	2296.48
0	4	2258.69
1	3	2260.56
2	2	**2257.23**
3	1	2283.04
4	0	2297.35

As potential models for the data we considered ARMA(p, q)-type densities of the form (5.9) with $p + q \leq 4$ and $a_k \in [0.01, 0.99], b_k \in (-\pi, \pi]$ $(k = 1, \ldots, q)$ and $c_k \in [0.01, 0.99], d_k \in (-\pi, \pi]$ $(k = 1, \ldots, p)$, the ranges for the a_k's and c_k's imposed so as to avoid computational problems associated with values close to zero or unity. We fitted the models via maximum likelihood (ML) using `R`'s `constrOptim` function and, for each model, 100 randomly chosen starting values for the parameters. The ML solution was taken as that one with the maximum value of the log-likelihood amongst the solutions obtained using the 100 randomly chosen starting values. At least for the AR(1)- and MA(1)-type models, an alternative would have been to use method of moments estimates as starting values.

Table III presents the values of the Akaike information criterion (AIC) for the 15 fitted ARMA(p, q)-type models, the best model according to the AIC being the ARMA(2, 2) model. The ML point estimates of the parameters of that model, together with the values to add and subtract to obtain nominally 95% confidence intervals based on normal approximation, are, to 3 decimal places, $\hat{a}_1 = 0.510 \pm 0.006$, $\hat{a}_2 = 0.577 \pm 0.005$, $\hat{b}_1 = -2.798 \pm 0.010$, $\hat{b}_2 = -0.105 \pm 0.013$, $\hat{c}_1 = 0.768 \pm 0.003$, $\hat{c}_2 = 0.763 \pm 0.003$, $\hat{d}_1 = -0.248 \pm 0.006$ and $\hat{d}_2 = -2.863 \pm 0.004$. The ML-fitted density is superimposed on the histogram in Figure 3.

As a comparison, we also considered the fit of a mixture model with two von Mises components and a circular uniform background component, the density of which is

$$f(\omega) = p_1 f_{\mathrm{vM}}(\omega; \mu_1, \kappa_1) + p_2 f_{\mathrm{vM}}(\omega; \mu_2, \kappa_2) + (1 - p_1 - p_2)\frac{1}{2\pi}, \tag{6.1}$$

where $p_1, p_2, (1 - p_1 - p_2) \in [0, 1]$, $f_{\mathrm{vM}}(\omega; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\omega - \mu)\}$, $\mu \in [0, 2\pi)$ is the mean direction, $\kappa > 0$ is a concentration parameter and I_0 denotes the modified Bessel function of the first kind and order 0. After inspecting the ML estimates of its parameters, we also considered the fit of the reduced model with $\kappa_1 = \kappa_2$. With an AIC value of 2267.27, the reduced model was found to be the superior of the two. The ML estimates of its parameters were, to 2 decimal places, $\hat{p}_1 = 0.15$, $\hat{p}_1 = 0.23$, $\hat{\mu}_1 = 0.34$, $\hat{\mu}_2 = 2.88$ and $\hat{\kappa}_1 = \hat{\kappa}_2 = 15.86$. The density for this solution is also superimposed on the histogram in Figure 3.

Comparing the two superimposed densities, both appear to model the data well around the two main modes, but the ARMA(2,2) fit appears to do slightly better in the remainder of the data range. Their AIC values also corroborate the superiority of the ARMA(2,2) fit. To formally investigate their goodness-of-fit, we used a parametric bootstrap approach analogous to that employed in Pewsey *et al.* (2013, Chapter 6) incorporating the Kuiper, Rayleigh and Watson U^2 tests of circular uniformity. The p-values of the three tests obtained using $B = 999$ parametric bootstrap samples were 0.259, 0.519 and 0.240 for the ARMA(2,2) fit, and 0.022, 0.031 and 0.023 for the reduced mixture fit. Hence, none of the tests finds significant evidence against the ARMA(2,2) fit being a potential

DOI: 10.1111/jtsa.12549

model for the data, but all three reject, at the 5% significance level, the reduced mixture fit as the model from which the data arose.

Here we have conducted model selection by fitting a range of plausible models, identifying the best-fitting ones using an information criterion, and applying formal testing to investigate their goodness-of-fit. In wider practice, this approach can be complemented by the inclusion of models known to be generated by constructions relevant to the data context under consideration. See, for example, Mardia and Jupp (1999, Chapter 3) for constructions leading to many of the classical models for circular data. We note that parametric bootstrap goodness-of-fit tests for data on $\mathbb{T}^2$ were developed in Pewsey and Kato (2016).

7. CONCLUSION

In Section 2 we proposed a general spectral density-generated family of distributions, with density (2.2), for data on $\mathbb{T}^n$. As a particular case of that general family when $n = 1$, in Section 3 we introduced the family of circular distributions with density (3.2), generated by the spectral density of a univariate stationary ARMA(p, q) process. We also considered five of its special cases and their relationship with existing circular models. As a particular case of the general construction of Section 2 when $n \geq 2$, in Section 4 we proposed a family of multivariate circular distributions with density (4.2) generated using linear filtering with an ARMA response function. We illustrated the varied shapes that such densities can adopt, and proposed Markov-type joint circular densities derived from them. In Section 5 we provided results for asymptotic optimal inference for the general family with density (2.2) and its particular case with density (4.2). There we also presented the results from a simulation study designed to explore the finite sample size performance of a local asymptotically optimal hypothesis testing approach. Finally, in Section 6 we illustrated the application of a particular case of the ARMA(p, q)-generated family of circular distributions with density (3.2) in the modelling of wind direction data.

ACKNOWLEDGEMENTS

Financial support for this research was received from Japan Society for the Promotion of Science in the form of grants 18H05290 (M. T.), JP25400218, JP17K05379 and JP20K03759 (S. K.), 18K11193 and 26870655 (H. O.), and from the Junta de Extremadura, the Spanish Ministry of Science, Innovation and Universities, and the European Union in the form of grants GR15013, GR18016 and PGC2018-097284-B-I00 (A. P.).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the United States Bureau of Reclamation (USBR) at https://www.usbr.gov/pn/agrimet/webagdayread.html.

REFERENCES

Abe T, Pewsey A. 2011. Sine-skewed circular distributions. *Statistical Papers* **52**: 683–707.

Abe T, Pewsey A, Shimizu K. 2009. On Papakonstantinou's extension of the cardioid distribution. *Statistics and Probability Letters* **79**: 2138–2147.

Abe T, Pewsey A, Shimizu K. 2013. Extending circular distributions through transformation of argument. *Annals of the Institute of Statistical Mathematics* **65**: 833–858.

Bloomfield P. 1973. An exponential model for the spectrum of a scalar time series. *Biometrika* **60**: 217–226.

Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2016. *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: Wiley.

Brillinger DR. 2001. *Time Series: Data Analysis and Theory*, expanded ed. San Francisco: Holden-Day.

de Hars-Lorentz GL. 1913. *Die Brownsche Bewegung und einige verwandte Erscheinungen*: Brunswick, Frieder, Vieweg und Sohn.

Garel B, Hallin M. 1995. Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics* **47**: 551–579.

Gatto R, Jammalamadaka SR. 2007. The generalized von Mises distribution. *Statistical Methodology* **4**: 341–353.

Hannan EJ. 1970. *Multiple Time Series*. New York: Wiley.

DOI: 10.1111/jtsa.12549

Jammalamadaka SR, SenGupta A. 2001. *Topics in Circular Statistics*. London: World Scientific.
Jeganathan P. 1995. Some aspects of asymptotic theory with applications to time series models. *Economic Theory* **11**: 818–887.
Johnson RA, Wehrly TE. 1977. Measures and models for angular correlation and angular-linear correlation. *Journal of the Royal Statistical Society: Series B* **39**: 222–229.
Jones MC, Pewsey A. 2005. A family of symmetric distributions on the circle. *Journal of the American Statistical Association* **100**: 1422–1428.
Jones MC, Pewsey A. 2012. Inverse Batschelet distributions for circular data. *Biometrics* **68**: 183–193.
Jones MC, Pewsey A, Kato S. 2013. On a class of circulas: copulas for circular distributions. *Annals of the Institute of Statistical Mathematics* **67**: 843–862.
Jupp PE, Mardia KV. 1989. A unified view of the theory of directional statistics, 1975–1988. *International Statistical Review* **57**: 261–294.
Kato S, Jones MC. 2010. A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *Journal of the American Statistical Association* **105**: 249–262.
Kato S, Jones MC. 2013. An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* **19**: 154–171.
Kato S, Jones MC. 2015. A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika* **102**: 181–190.
Kato S, Pewsey A. 2015. A Möbius transformation-induced distribution on the torus. *Biometrika* **102**: 359–370.
Lévy P. 1939. L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société Mathématique de France* **67**: 1–41.
Ley C, Swan Y, Thiam B, Verdebout T. 2013. Optimal R-estimation of a spherical location. *Statistica Sinica* **23**: 305–333.
Mardia KV. 1972. *Statistics of Directional Data*. London: Academic Press.
Mardia KV, Jupp PE. 1999. *Directional Statistics*. Chichester: Wiley.
Maksimov VM. 1967. Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups (in Russian). *Theoria Verojatna* **12**: 307–321.
Paindaveine D, Verdebout T. 2017. Inference on the mode of weak directional signals: a LeCam perspective on hypothesis testing near singularities. *Annals of Statistics* **45**: 800–832.
Pearson K. 1905. The problem of the random walk. *Nature* **72**: 294–342.
Pewsey A, Kato S. 2016. Parametric bootstrap goodness-of-fit testing for Wehrly–Johnson bivariate circular distributions. *Statistics and Computing* **26**: 1307–1317.
Pewsey A, Neuhäuser M, Ruxton GD. 2013. *Circular Statistics in R*. Oxford: Oxford University Press.
Rayleigh L. 1880. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine* **10**: 73–78.
Saw JG. 1983. Dependent unit vectors. *Biometrika* **70**: 655–671.
Strasser H. 1985. *Mathematical Theory of Statistics*. Berlin: Walter de Gruyter.
Swensen AE. 1985. The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16**: 54–70.
Taniguchi M. 1983. On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Annals of Statistics* **11**: 157–169.
Taniguchi M, Kakizawa Y. 2000. *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.
Umbach D, Jammalamadaka SR. 2009. Building asymmetry into circular distributions. *Statistics and Probability Letters* **79**: 659–663.
von Mises R. 1918. Über die "Ganzzahligkeit" der Atomgewichte und verwandete Fragen. *Physikalische Zeitschrift* **19**: 490–500.
Wehrly TE, Johnson RA. 1980. Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **67**: 255–256.
Wintner A. 1947. On the shape of the angular case of Cauchy's distribution curves. *Annals of Mathematical Statistics* **18**: 589–593.

APPENDIX A

A.1. Use of the Residue Theorem

Let $f(\omega)$ be the function in (5.9). First we show that, for $p_1 \in \mathbb{Z}$, the trigonometric moment $\phi(p_1)$ can be expressed as

$$\int_{-\pi}^{\pi} e^{\mathrm{i}p_1\omega} f(\omega)d\omega = (-1)^{q-p}\, 2\pi c \frac{\prod_{k=1}^{q} A_k}{\prod_{k=1}^{p} C_k} \left\{ \mathrm{Res}(0) I(p-q+p_1 < 1) + \sum_{m=1}^{p} \mathrm{Res}(\overline{C_m}) \right\}, \tag{A.1}$$

DOI: 10.1111/jtsa.12549

where $I(\cdot)$ is the indicator function, and $\text{Res}(\overline{C_m})$ and $\text{Res}(0)$ are given in (A.3) and (A.4).
Let $A_k = a_k e^{\mathrm{i}b_k}$ and $C_k = c_k e^{\mathrm{i}d_k}$. Setting $z = e^{\mathrm{i}\omega}$, we have $d\omega = dz/(\mathrm{i}z)$, and

$$\begin{aligned}\int_{-\pi}^{\pi} e^{\mathrm{i}p_1\omega} f(\omega)d\omega &= \int_{-\pi}^{\pi} e^{\mathrm{i}p_1\omega} c\,\frac{\prod_{k=1}^{q}(1-a_k e^{\mathrm{i}b_k}e^{\mathrm{i}\omega})(1-a_k e^{-\mathrm{i}b_k}e^{-\mathrm{i}\omega})}{\prod_{k=1}^{p}(1-c_k e^{\mathrm{i}d_k}e^{\mathrm{i}\omega})(1-c_k e^{-\mathrm{i}d_k}e^{-\mathrm{i}\omega})}d\omega \\ &= \frac{c}{\mathrm{i}}\int_C z^{p_1-1}\frac{\prod_{k=1}^{q}(1-A_k z)\left(1-\frac{\overline{A_k}}{z}\right)}{\prod_{k=1}^{p}(1-C_k z)\left(1-\frac{\overline{C_k}}{z}\right)}dz \\ &= \frac{c}{\mathrm{i}}\frac{\prod_{k=1}^{q}(-A_k)}{\prod_{k=1}^{p}(-C_k)}\int_C z^{p-q+p_1-1}\frac{\prod_{k=1}^{q}\left(z-\frac{1}{A_k}\right)(z-\overline{A_k})}{\prod_{k=1}^{p}\left(z-\frac{1}{C_k}\right)(z-\overline{C_k})}dz, \end{aligned} \tag{A.2}$$

where $C = \{z \in \mathbb{C} : |z| = 1\}$. If $p - q + p_1 - 1$ is a nonnegative integer, the integrand in (A.2) has p poles of order 1 at $z = \overline{C_m}$ $(m = 1, 2, \ldots, p)$. The residue at $z = \overline{C_m}$ is

$$\text{Res}(\overline{C_m}) = \overline{C_m}^{\,p-q+p_1-1}\frac{\prod_{k=1}^{q}\left(\overline{C_m}-\frac{1}{A_k}\right)(\overline{C_m}-\overline{A_k})}{\left(\overline{C_m}-\frac{1}{C_m}\right)\prod_{k=1,k\neq m}^{p}\left(\overline{C_m}-\frac{1}{C_k}\right)(\overline{C_m}-\overline{C_k})}. \tag{A.3}$$

If $p - q + p_1 - 1$ is a negative integer, the integrand in (A.2) has one more pole of order $p - q + p_1 - 1$ at $z = 0$, in addition to the above p poles at $z = \overline{C_m}$ $(m = 1, 2, \ldots, p)$. In this case, the residue at $z = 0$ is

$$\text{Res}(0) = \frac{1}{(q-p-p_1)!}\left[\left(\frac{d}{dz}\right)^{q-p-p_1}\left\{\frac{\prod_{k=1}^{q}\left(z-\frac{1}{A_k}\right)(z-\overline{A_k})}{\prod_{k=1}^{p}\left(z-\frac{1}{C_k}\right)(z-\overline{C_k})}\right\}\right]_{z=0}. \tag{A.4}$$

From the residue theorem, we obtain (A.1).
The normalizing constant c is calculated using (A.1) with $p_1 = 0$.
In practice, calculation of the $(q - p - p_1)$th derivative in (A.4) can be done as follows. Let

$$h(z) = \frac{\prod_{k=1}^{q}\left(z-\frac{1}{A_k}\right)(z-\overline{A_k})}{\prod_{k=1}^{p}\left(z-\frac{1}{C_k}\right)(z-\overline{C_k})}.$$

By taking logarithms of both sides, we have

$$\ln h(z) = \sum_{k=1}^{q}\left\{\ln\left(z-\frac{1}{A_k}\right)+\ln(z-\overline{A_k})\right\} - \sum_{k=1}^{p}\left\{\ln\left(z-\frac{1}{C_k}\right)+\ln(z-\overline{C_k})\right\}.$$

Differentiating both sides, we have

$$\frac{\frac{d}{dz}h(z)}{h(z)} = \sum_{k=1}^{q}\left\{\frac{1}{z-\frac{1}{A_k}}+\frac{1}{z-\overline{A_k}}\right\} - \sum_{k=1}^{p}\left\{\frac{1}{z-\frac{1}{C_k}}+\frac{1}{z-\overline{C_k}}\right\} = g(z) \quad \text{(say)}.$$

DOI: 10.1111/jtsa.12549

Then, we obtain

$$\frac{d}{dz}h(z) = h(z)g(z)$$

and the nth order derivative of $h(z)$ is

$$\frac{d^n}{dz^n}h(z) = \sum_{j=0}^{n-1} \binom{n-1}{j} \cdot \frac{d^{n-1-j}}{dz^{n-1-j}}h(z) \cdot \frac{d^j}{dz^j}g(z). \tag{A.5}$$

The jth order derivative of $g(z)$ is given by

$$\frac{d^j}{dz^j}g(z) = (-1)^j j! \left[\sum_{k=1}^{q} \left\{ \left(z - \frac{1}{A_k}\right)^{-(j+1)} + \left(z - \overline{A_k}\right)^{-(j+1)} \right\} - \sum_{k=1}^{p} \left\{ \left(z - \frac{1}{C_k}\right)^{-(j+1)} + \left(z - \overline{C_k}\right)^{-(j+1)} \right\} \right]$$

and its value at $z = 0$ is

$$\left.\frac{d^j}{dz^j}g(z)\right|_{z=0} = j! \left\{ \sum_{k=1}^{q} \left(C_k^{j+1} + \overline{C_k}^{-(j+1)} \right) - \sum_{k=1}^{p} \left(A_k^{j+1} + \overline{A_k}^{-(j+1)} \right) \right\}. \tag{A.6}$$

Combining (A.5) and (A.6), we obtain

$$\left.\frac{d^n}{dz^n}h(z)\right|_{z=0} = \sum_{j=0}^{n-1} \frac{(n-1)!}{(n-1-j)!} \cdot \left\{ \sum_{k=1}^{q} \left(C_k^{j+1} + \overline{C_k}^{-(j+1)} \right) - \sum_{k=1}^{p} \left(A_k^{j+1} + \overline{A_k}^{-(j+1)} \right) \right\} \cdot \left.\frac{d^{n-1-j}}{dz^{n-1-j}}h(z)\right|_{z=0}$$

and we can recursively calculate the $(q - p - p_1)$th derivative in (A.4).

A.2. Proof of Proposition 1

For convenience, write $z = e^{\mathrm{i}\omega}$, $\phi_1 = r_1 e^{-\mathrm{i}\mu_1}$ and $\phi_2 = r_2 e^{-2\mathrm{i}\mu_2}$. Then it follows from the fundamental theorem of algebra that the equation

$$1 + \phi_1 z + \phi_2 z^2 = 0$$

always has two solutions in $\mathbb{C}$, which are given by $z = c_+$ and $z = c_-$. This implies that density (3.3) can be expressed as

$$f(\omega) = \frac{c}{2\pi} \frac{1}{|\phi_2|^2 |z - c_+|^2 |z - c_-|^2}. \tag{A.7}$$

If $|c_+|, |c_-| < 1$, it is easy to see that density (A.7) is equivalent to that of the distribution proposed by Kato and Jones (2013) apart from parametrization. Therefore the constant c is given by

$$c = \frac{|\phi_2|^2 \left|1 - |c_+|^2\right| \left|1 - |c_-|^2\right| |1 - c_+\overline{c_-}|^2}{|1 - |c_+\overline{c_-}|^2|}.$$

Next, consider the case $|c_+| > 1$ and $|c_-| < 1$. For this case we note that density (A.7) can be expressed as

$$f(\omega) = \frac{c}{2\pi} \frac{1}{|\phi_2|^2 |c_+|^2 |z - \overline{c_+}^{-1}|^2 |z - c_-|^2}.$$

DOI: 10.1111/jtsa.12549

This implies that $f(\omega; c_+, c_-) = f(\omega; \overline{c_+}^{-1}, c_-)$. Since $|\overline{c_+}^{-1}| < 1$, it follows that density (A.7) with $|c_+| > 1$ and $|c_-| < 1$ is also equivalent to the density of Kato and Jones (2013). Hence the constant c of density (A.7) with $|c_+| > 1$ and $|c_-| < 1$ is

$$
\begin{aligned}
c &= \frac{|\phi_2|^2 |c_+|^2 \left|1 - |\overline{c_+}^{-1}|^2\right| \left|1 - |c_-|^2\right| |1 - \overline{c_+}^{-1}\overline{c_-}|^2}{|1 - |\overline{c_+}^{-1}\overline{c_-}|^2|} \\
&= \frac{|\phi_2|^2 (|c_+|^2 - 1)(1 - |c_-|^2)|c_+ - c_-|^2}{|c_+|^2 - |c_-|^2}.
\end{aligned}
$$

Similarly, it can be seen that $f(\omega; c_+, c_-) = f(\omega; c_+, \overline{c_-}^{-1}) = f(\omega; \overline{c_+}^{-1}, \overline{c_-}^{-1})$. This implies that the constant c of density (A.7) with $|c_+| < 1$ and $|c_-| > 1$ and that with $|c_+| > 1$ and $|c_-| > 1$ are given by (3.4). □

A.3. Proof of Theorem 1

As seen in the proof of Proposition 1, density (3.3) can be expressed as (A.7), which is the density of the distribution proposed by Kato and Jones (2013). It follows that $\mathcal{F}_{AR2} \subset \mathcal{F}_{KJ}$. On the other hand, the density of Kato and Jones (2013) can be expressed as

$$
\begin{aligned}
f_{KJ}(\omega) &= \frac{C}{|1 - \xi_1 e^{-\mathrm{i}\eta_1} e^{\mathrm{i}\omega}|^2 |1 - \xi_2 e^{\mathrm{i}\eta_2} e^{\mathrm{i}\omega}|^2} \\
&= \frac{C}{|1 - (\xi_1 e^{-\mathrm{i}\eta_1} + \xi_2 e^{-\mathrm{i}\eta_2}) e^{\mathrm{i}\omega} + \xi_1 \xi_2 e^{-\mathrm{i}(\eta_1+\eta_2)} e^{2\mathrm{i}\omega}|^2}.
\end{aligned} \tag{A.8}
$$

Since density (A.8) is equal to density (3.3) with

$$
r_1 e^{-\mathrm{i}\mu_1} = -\xi_1 e^{-\mathrm{i}\eta_1} - \xi_2 e^{-\mathrm{i}\eta_2} \quad \text{and} \quad r_2 e^{-2\mathrm{i}\mu_2} = \xi_1 \xi_2 e^{-\mathrm{i}(\eta_1+\eta_2)}, \tag{A.9}
$$

it holds that $\mathcal{F}_{KJ} \subset \mathcal{F}_{AR2}$. Hence we have $\mathcal{F}_{AR2} = \mathcal{F}_{KJ}$. The equation (3.6) is clear from the fact that the density (3.3) is equal to the density (A.8) when the parameters of the two densities have the relationship (A.9). □

A.4. Proof of Theorem 2

It suffices to check Swensen's (1985) conditions (S1)–(S6) (see also Garel and Hallin, 1995; Taniguchi and Kakizawa, 2000). The arguments are similar to those used in Taniguchi and Kakizawa (2000, pp. 32–36). (S1) $E(\xi_{n,k}|\mathcal{F}_{k-1}) = 0$ a.e. follows from, for example, Taniguchi and Kakizawa (2000, p. 7). (S2) $E[\sum_{k=1}^n \{Y_{n,k} - \xi_{n,k}\}^2] \to 0$, follows from Assumption 3. (S3) $\sup_n E[\sum_{k=1}^n \xi_{n,k}^2] < \infty$, and (S4) $\max_k |\xi_{n,k}| \xrightarrow{p} 0$, follow from Assumption 2. Assumption 4 implies (S5) $\sum_{k=1}^n \xi_{n,k}^2 \xrightarrow{p} \tau^2$. (S6) $\sum_{k=1}^n E\{\xi_{n,k}^2 \chi(|\xi_{n,k}| > \delta)|\mathcal{F}_{k-1}\} \xrightarrow{p} 0$ for some $\delta > 0$, follows from Assumption 2. □