Ann Inst Stat Math (2015) 67:843–862
DOI 10.1007/s10463-014-0493-6

On a class of circulas: copulas for circular distributions

M. C. Jones · Arthur Pewsey · Shogo Kato

Received: 16 September 2013 / Accepted: 10 July 2014 / Published online: 26 October 2014

Abstract This article is concerned with the analogue of copulas for circular distributions, which we call 'circulas'. We concentrate on one particular class of circulas, which is pre-existing but not studied in such explicit form or detail before. This class is appealing in many ways but does not necessarily result in especially attractive bivariate circular models for arbitrary non-uniform marginals. A major exception to this is an elegant bivariate wrapped Cauchy distribution previously proposed and developed by two of the current authors. We look both at properties of the circulas themselves, including their density behaviour, distribution function, and dependence measures, and at properties of various distributions based on these circulas by transformation to non-uniform marginal distributions. We consider inference for the latter distributions and present two applications of them to modelling data. We concentrate mostly on the bivariate case, but also briefly consider extension to the multivariate case.

Keywords Bivariate · Circular distributions · Copula · Dependence · Multivariate · Torus

---

Electronic supplementary material The online version of this article (doi:10.1007/s10463-014-0493-6) contains supplementary material, which is available to authorized users.

M. C. Jones (✉)
Department of Mathematics and Statistics, The Open University, Walton Hall,
Milton Keynes MK7 6AA, UK
e-mail: m.c.jones@open.ac.uk

A. Pewsey
Department of Mathematics, Escuela Politécnica, Universidad de Extremadura,
Avenida de la Universidad s/n, 10003 Cáceres, Spain
e-mail: apewsey@unex.es

S. Kato
Institute of Statistical Mathematics, Tachikawa, Tokyo 190–8562, Japan
e-mail: skato@ism.ac.jp

1 Introduction

This paper concerns the circular analogues of copulas. The latter are, of course, bivariate and multivariate distributions for linear data, whose defining property is that they have uniform univariate marginal distributions. Concentrating, until Sect. 6, on the important bivariate case for simplicity, any bivariate distribution for linear data can be decomposed into its copula—which contains dependence information—and its marginals. This is Sklar's theorem. In terms of densities, a general bivariate density f on $\mathbb{R}^2$ can be written in terms of its copula density c_R on $(0, 1)^2$ and its marginal density and distribution functions f_X, f_Y, F_X and F_Y, all on $\mathbb{R}$, as

$$f(x, y) = f_X(x) f_Y(y)\, c_R(F_X(x), F_Y(y)).$$

Joe (1997) and Nelsen (2010) are excellent introductions to this subject.

We are concerned with bivariate distributions for circular data, and especially with the circular analogue of copulas which we propose to call 'circulas': these are bivariate, and later, multivariate, distributions for circular data whose marginals are circular uniform distributions. Let $\mathbb{C}$ denote the circle. As above, a general bivariate circular density f on $\mathbb{C}^2$—or a density on the unit torus—can be written in terms of its circula density c on $\mathbb{C}^2$ and its marginal circular density and distribution functions f_1, f_2, F_1 and F_2, all on $\mathbb{C}$, as

$$f(\theta_1, \theta_2) = 4\pi^2 f_1(\theta_1) f_2(\theta_2)\, c(2\pi F_1(\theta_1), 2\pi F_2(\theta_2)). \tag{1}$$

In general, the marginal distribution functions can be defined from arbitrary starting points on $\mathbb{C}$. Circulas differ from rescaled linear copulas in also requiring periodicity:

$$c(\theta_1 \pm 2k\pi, \theta_2 \pm 2l\pi) = c(\theta_1, \theta_2), \quad (\theta_1, \theta_2) \in \mathbb{C}^2, \quad k, l = 1, 2, \ldots.$$

This paper is actually concerned with one particular construction of circulas, which can be found elsewhere (see below) but which has not previously been given a unified explicit treatment. In the course of this paper, we will point out both its advantages and its limitations. This circula construction is extremely simple. Let Θ_1 follow the circular uniform distribution. Then, for any constant angle ω, $(\Theta_1 + \omega) \pmod{2\pi}$ also follows the circular uniform distribution. Now, let Ω follow a circular distribution with density g, say, independently of Θ_1. Then, by dint of the previous result, $\Theta_2 = (\Theta_1 + \Omega) \pmod{2\pi}$ also follows the circular uniform distribution: see also Mardia and Jupp (1999, p. 36). It follows that (Θ_1, Θ_2) follows a bivariate circular distribution with circular uniform marginals, that is, a circula. Moreover, the conditional density of $\Theta_2|\Theta_1 = \theta_1$ is $g(\theta_2 - \theta_1)$ which, combined with the uniform marginal distribution of Θ_1, means that the circula density is

$$c_1(\theta_1, \theta_2) = \frac{1}{2\pi} g(\theta_2 - \theta_1). \tag{2}$$

A similar argument based on $\Theta_2 = (\Omega - \Theta_1)\ (\mathrm{mod}\ 2\pi)$ yields the complementary circula density

$$c_{-1}(\theta_1, \theta_2) = \frac{1}{2\pi} g(\theta_2 + \theta_1) = c_1(2\pi - \theta_1, \theta_2). \tag{3}$$

The two cases can be combined as

$$c_q(\theta_1, \theta_2) = \frac{1}{2\pi} g(\theta_2 - q\theta_1), \tag{4}$$

where $q \in \{-1, 1\}$ is non-random; this is the density of the joint distribution of Θ_1 and $\Theta_2 = (\Omega + q\Theta_1)\ (\mathrm{mod}\, 2\pi)$.

When used in (1), (4) yields

$$f(\theta_1, \theta_2) = 2\pi f_1(\theta_1) f_2(\theta_2)\, g(2\pi(F_2(\theta_2) - qF_1(\theta_1))). \tag{5}$$

So, by construction, (5) has marginals with densities $f_1(\theta_1)$ and $f_2(\theta_2)$ and conditional densities which can be immediately written down, e.g.

$$f_{2|1}(\theta_2|\theta_1) = 2\pi f_2(\theta_2)\, g(2\pi(F_2(\theta_2) - qF_1(\theta_1))).$$

We will sometimes call g the 'binding' density.

Such distributions for bivariate circular data can first be found in four papers in the late 1970s: as models under which a proposed angular correlation measure is calculated, first for g von Mises and then for general g in form (2) in Johnson and Wehrly (1977); when g is cardioid, again in form (2), as the transition density for the angular part of a bivariate Markov point process expressed in polar co-ordinates, in Isham (1977); in a hybrid version of form (5) with $q = 1$ in Johnson and Wehrly (1978), where Θ_1 is replaced by a linear random variable; while (5) itself appears in Wehrly and Johnson (1980), where its role in Markov processes is suggested, without reference to Isham (1977), and some properties are given when g is the von Mises density. So long ago, the term 'copula' was not in vogue so was not used, but the copula-like role of (4) has been explicitly recognised in much more recent publications looking at special cases, in both g and marginals, of (5): Shieh and Johnson (2005), Fernández-Durán (2007), Kato (2009), Shieh et al. (2011), García-Portugués et al. (2013) and Kato and Pewsey (2013). Alfonsi and Brigo (2005) utilise much the same 'periodic copula' construction, but for use as ordinary copulas for linear data; Perlman and Wellner (2011) 'circular copulas' are also ordinary copulas, derived from distributions supported on the disc (Jones 2013). Our purpose in this paper is to give a more focussed account of the circulas with density (4), the distributions with densities of form (5) arising from them, and their extensions, per se.

Properties of circulas themselves constitute Sect. 2. These include their density behaviour, distribution function, and dependence measures. We move on, in Sect. 3, to consider properties of various distributions based on these circulas by transformation to non-uniform marginal distributions. In that section, the following notation will be used for special cases of the distribution with density (5):

$$f_1 - f_2 - g(q, \mu_1, \rho_1 \text{ or } \kappa_1, \mu_2, \rho_2 \text{ or } \kappa_2, \mu_g, \rho_g \text{ or } \kappa_g). \tag{6}$$

There, each of f_1, f_2 or g will be replaced by abbreviations such as wC for wrapped Cauchy or vM for von Mises, leading to condensed descriptions such as vM–vM–wC. The μ's are the corresponding location parameters and the ρ's, as mean resultant lengths, or κ's, in the von Mises case, are the corresponding concentration parameters. In Sect. 3.1, disadvantages of *some* of these distributions will become apparent. In the remainder of Sect. 3, we consider the distribution function and dependence measures for these distributions, together with random variate generation. Section 4 is devoted to maximum likelihood estimation of parameters and goodness-of-fit testing. Section 5 gives two applications of these distributions to data. In Sect. 6, consideration is given to extending bivariate circulas to the multivariate case, and the paper closes with a brief discussion in Sect. 7.

2 Properties of the circula with density (4)

2.1 Circula densities

If g is itself chosen to be the circular uniform density, then the circula of interest reduces to the independence circula for which

$$c_I(\theta_1, \theta_2) = \frac{1}{4\pi^2}.$$

Otherwise, the circula densities (4) have linear contours parallel to the $q\pi/4$, or $q \times 45°$, diagonal. If the polar representation of the density g is unimodal with mode at μ_g, the circula density is maximal at every point of the diagonals $\theta_2 = \mu_g + q\theta_1 \pm 2\pi k$, $k = 0, 1, \ldots$. Let ρ_g denote the mean resultant length of the distribution with density g. Then, 'tightness' to the diagonal is determined by the value of ρ_g. This is illustrated in linearised form in Fig. 1 when g is the wrapped Cauchy density with $\mu_g = 0$. For much consideration of what happens to density contours on transformation to non-uniform marginals, see Sect. 3.1.

2.2 Circula distribution functions

For $-2\pi \leq \omega \leq 4\pi$, define

$$W(\omega) = \int_0^\omega \int_0^b g(a)\,\mathrm{d}a\,\mathrm{d}b = \int_0^\omega (\omega - a) g(a)\mathrm{d}a.$$

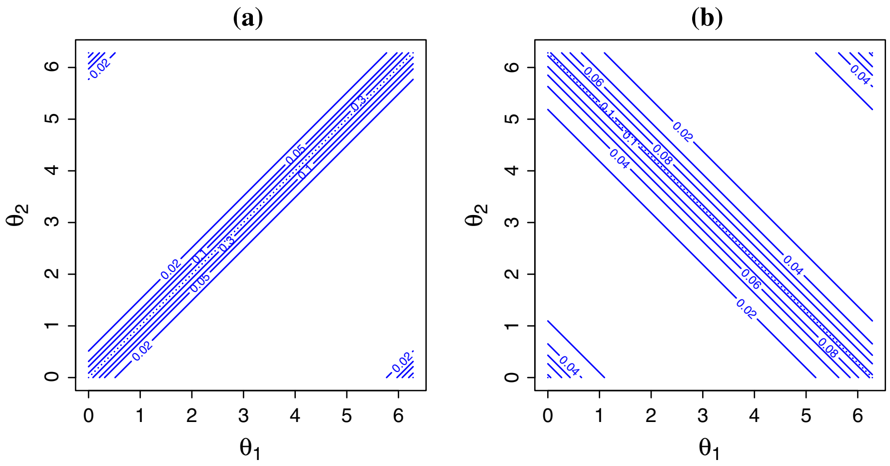

Fig. 1 Examples of c_q using wrapped Cauchy g with $\mu_g = 0$ **a** $\rho_g = 0.9, q = 1$; **b** $\rho_g = 0.6, q = -1$. The *dotted diagonal line* in each plot identifies those (θ_1, θ_2) combinations for which the density is maximal. In both cases, the *contours* in the corners are parts of the periodic repetitions of the central bands

Setting the origin of the circula distribution function C_q to (0, 0), we find, after a certain amount of manipulation, that, rather beautifully,

$$C_q(\theta_1, \theta_2) = \frac{q}{2\pi} \{W(\theta_2) + W(-q\theta_1) - W(\theta_2 - q\theta_1)\}, \quad 0 \leq \theta_1, \theta_2 \leq 2\pi. \quad (7)$$

For a derivation of (7) see the online Supplementary Material; it is straightforward to confirm that C_q has the correct margins and $\partial^2 C_q(\theta_1, \theta_2)/(\partial\theta_1 \partial\theta_2) = c_q(\theta_1, \theta_2) \geq 0$ given by (4).

Most non-trivial g's do not have tractable W functions. An exception is the cardioid density employed by Isham (1977), for which $g(\omega) = (2\pi)^{-1}(1 + 2\rho \cos \omega)$, $0 \leq \rho \leq \frac{1}{2}$, $W(\omega) = (2\pi)^{-1} \left\{\frac{1}{2}\omega^2 + 2\rho(1 - \cos \omega)\right\}$ and

$$C_q(\theta_1, \theta_2) = \frac{1}{4\pi^2} [\theta_1\theta_2 + 2q\rho\{1 - \cos \theta_1 - \cos \theta_2 + \cos(\theta_1 - q\theta_2)\}].$$

Of course, this reduces to the independence case when $\rho = 0$ so that g is itself uniform and

$$C_I(\theta_1, \theta_2) = \frac{\theta_1\theta_2}{4\pi^2}.$$

The cardioid-based circula distribution functions corresponding to $\rho = 1/2$, $q = \pm 1$ are shown in Fig. 2a, b; the wrapped Cauchy-based circula distribution functions corresponding to the circula density functions shown in Fig. 1, calculated using one-dimensional numerical integration of the explicit wrapped Cauchy distribution function, are shown in Fig. 2c, d, respectively.

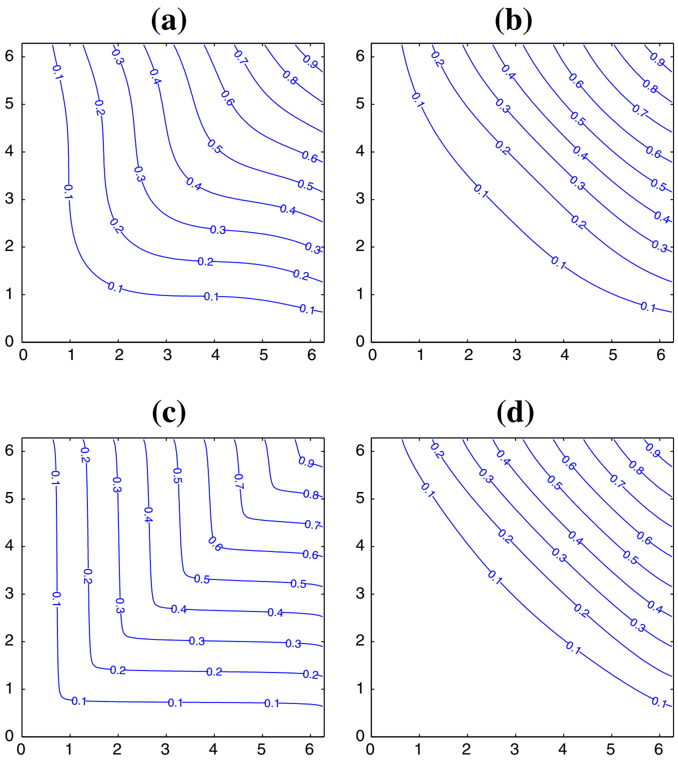

Fig. 2 Examples of C_q using: cardioid g with **a** $\rho = 0.5$, $q = 1$, **b** $\rho = 0.5$, $q = -1$; wrapped Cauchy g with **c** $\rho_g = 0.9$, $q = 1$, **d** $\rho_g = 0.6$, $q = -1$

2.3 The dependence parameter

It is conceptually clear that the concentration of g controls the dependence of c_q: when g is highly concentrated, circula dependence is high; when g is more diffuse, circula dependence is low. The mean resultant length, $\rho_g = \sqrt{\alpha_g^2 + \beta_g^2}$ where $\alpha_g = E_g(\cos\Omega)$ and $\beta_g = E_g(\sin\Omega)$, measures the concentration of g; the following paragraphs quantify the role of ρ_g as the dependence parameter of c_q.

We consider five pre-existing dependence measures for circular data, namely those of Johnson and Wehrly (1977), Jupp and Mardia (1980), Rivest (1982), Fisher and Lee (1983) and Jammalamadaka and Sarma (1988). For circulas with density c_q, the formulae for all five come out straightforwardly. Partly following Sect. 3.8 of Kato and Pewsey (2013) and noting that $E(\cos\Theta_1) = E(\sin\Theta_1) = E(\cos\Theta_2) = E(\sin\Theta_2) = 0$ by circular uniformity, all but one of the dependence measures depend on functions of the 2×2 matrices $\Sigma_{kl} = E(X_k X_l^T)$, $k, l = 1, 2$, where $X_1 = (\cos\Theta_1, \sin\Theta_1)^T$, $X_2 = (\cos\Theta_2, \sin\Theta_2)^T$. Because Σ_{11} and Σ_{22} depend only on the circular uniform marginals, it is easy to see that $\Sigma_{11} = \Sigma_{22} = \frac{1}{2}\mathcal{I}_2$ where $\mathcal{I}_2$ is the 2×2 identity matrix. Only slightly more difficult calculations involving basic trigonometric identities and the general relation $\Theta_2 = \Omega + q\Theta_1$ result in

$$\Sigma_{12} = \frac{1}{2}\begin{pmatrix} \alpha_g & \beta_g \\ -q\beta_g & q\alpha_g \end{pmatrix}.$$

A first signed dependence measure for bivariate circular data is that of Rivest (1982), ρ_R. For the circula with density c_q, $\rho_R = 2\lambda_2$ where λ_2 denotes the smallest singular value of Σ_{12} multiplied by $\mathrm{sgn}(\det\Sigma_{12})$. Most gratifyingly, this simplifies to $\rho_R = q\rho_g$. A second signed dependence measure is that of Fisher and Lee (1983), essentially an analogue of Spearman's rho. For a circula, it is given by

$$\rho_{FL} = \det\Sigma_{12}/\sqrt{\det\Sigma_{11}\det\Sigma_{22}},$$

and so, for the circula with density c_q, is

$$\rho_{FL} = q(\alpha_g^2 + \beta_g^2) = q\rho_g^2.$$

This result is Example 2 of Fisher and Lee (1983). A third signed dependence measure is that of Jammalamadaka and Sarma (1988) which, for a circula, is given by their (2.3) in slightly simplified form:

$$\rho_{JS} = \left|E\left(e^{i(\Theta_1-\Theta_2)}\right)\right| - \left|E\left(e^{i(\Theta_1+\Theta_2)}\right)\right|.$$

For the circula with density g, ρ_{JS} also reduces to $q\rho_g$. This result is given in Sect. 2.3 of Jammalamadaka and Sarma (1988) for the special case with g von Mises.

In addition, the unsigned dependence measures of Johnson and Wehrly (1977) and Jupp and Mardia (1980) depend on

$$S = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$$

which reduces to $S = \rho_g^2\mathcal{I}_2$. Thus, Johnson and Wehrly's dependence measure, ρ_{JW}, which is the square root of the largest eigenvalue of S, is ρ_g, as Johnson and Wehrly (1977, Example 7.2) obtain for the case of $q = 1$. Also, Jupp and Mardia's dependence measure, ρ_{JM}, which is the trace of S, is $2\rho_g^2$.

The key observations here, of course, are the ways in which all five dependence measures relate to ρ_g, justifying its role as the dependence parameter of the circula with density c_q.

2.4 Local dependence

Justifications for the *local* dependence function $\gamma_f(x, y) = \partial^2 \log f(x, y)/\partial x\partial y$ (Holland and Wang 1987; Jones 1996) transfer immediately to the bivariate circular case. The local dependence function is particularly simple for the class of circulas under consideration:

$$\gamma_{c_q}(\theta_1, \theta_2) = -q(\log g)''(\theta_2 - q\theta_1).$$

Local dependence therefore follows the circula's contours. For unimodal g, $(\log g)''$ is typically negative at and near the mode, and the local dependence function correspondingly, and reasonably, shares its sign with q at and near the parts of the circula with highest density.

If desired, another, signed, scalar dependence measure can be obtained by averaging γ_{c_q} with respect to the circula. The result is readily seen to be

$$\gamma = q \int_0^{2\pi} \frac{\{g'(\phi)\}^2}{g(\phi)} \mathrm{d}\phi = q I_g, \tag{8}$$

say. Since I_g, the Fisher information for location of g, is positive, like ρ_{FL}, γ has the sign of q. Its magnitude depends on ρ_g, and not μ_g, and can be expected to increase with increasing ρ_g (higher concentration $\Rightarrow$ more information). For example, for cardioid g, $I_g = 1 - \sqrt{1 - 4\rho^2}$, for wrapped Cauchy g, $I_g = 2\rho_g^2/(1 - \rho_g^2)^2$, and for von Mises g with concentration parameter κ and mean resultant length $A(\kappa)$ (Mardia and Jupp 1999, (3.5.31)), $I_g = \kappa A(\kappa)$.

3 Properties of densities of the form (5)

3.1 Density shapes

First, if the binding density is uniform, the marginals are independent and, of course, $f(\theta_1, \theta_2) = f_1(\theta_1) f_2(\theta_2)$.

The point (μ_1, μ_2), where μ_i is the location parameter of the marginal distribution $F_i, i = 1, 2$, will be of particular interest in the rest of this subsection. For clarity, concreteness and convenience, for non-uniform marginals, we will specifically associate with (5) the particular circular distribution function definition

$$F_1(\theta_1) = \int_{\mu_1}^{\theta_1} f_1(\phi)\mathrm{d}\phi \tag{9}$$

where μ_1 denotes the location parameter of f_1 and $\mu_1 \le \theta_1 \le \mu_1 + 2\pi$, and similarly for F_2. In general, it proves most convenient to specify μ_1 as the mode of f_1, this coinciding with other specifications of location for symmetric unimodal distributions. Then, $F_1(\mu_1) = F_2(\mu_2) = 0$ and the argument of g in (5) when $\theta_i = \mu_i, i = 1, 2$, will also be zero whatever the value of q. Now, if g is maximal at $\mu_g = 0$ and f_1 and f_2 are maximal at μ_1 and μ_2, respectively, then (5) will be maximal at (μ_1, μ_2). Note that Wehrly and Johnson (1980) employed the versions of F_1 and F_2 starting from 0 rather than μ_1 and μ_2.

Using notation (6), this is illustrated for the wC–wC–wC$(q, \mu_1, \rho_1, \mu_2, \rho_2, \mu_g, \rho_g)$, vM–vM–vM$(q, \mu_1, \kappa_1, \mu_2, \kappa_2, \mu_g, \kappa_g)$ and vM–vM–wC$(q, \mu_1, \kappa_1, \mu_2, \kappa_2, \mu_g, \rho_g)$ families by the first column of Fig. 3. All of the distributions portrayed in Fig. 3 have $\mu_g = 0$, $\rho_g = 0.6$ and marginal distributions with equal concentration values.

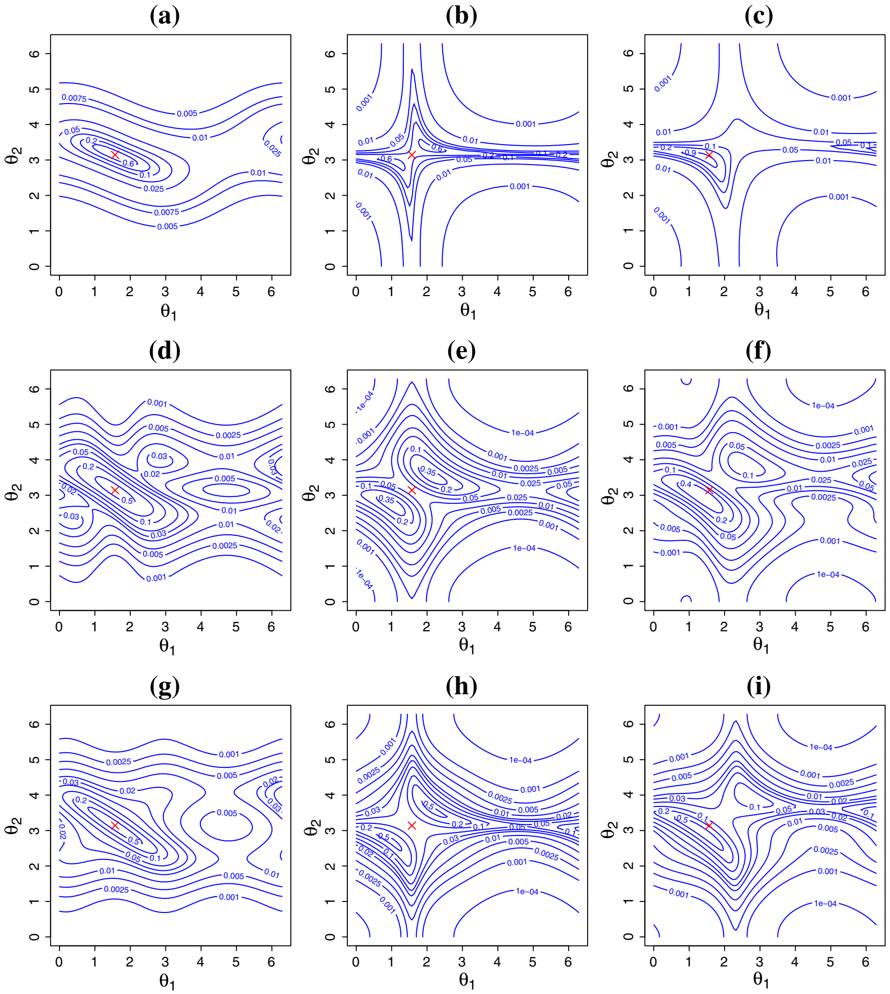

Fig. 3 Contour plots of **a**–**c** wC–wC–wC$(-1, \pi/2, 0.6, \pi, 0.8, \mu_g, 0.6)$, **d**–**f** vM–vM–vM $(-1, \pi/2, 1.509, \pi, 2.862, \mu_g, 1.509)$, **g**–**i** vM–vM–wC$(-1, \pi/2, 1.509, \pi, 2.862, \mu_g, 0.6)$ densities, with: first column, $\mu_g = 0$; second column, $\mu_g = \pi$; third column, $\mu_g = 5$. All densities shown have $\rho_g = 0.6$. The *cross* in each panel identifies $(\mu_1 = \pi/2, \mu_2 = \pi)$

For comparison with densities with $\rho_g = 0.3$ and $\rho_g = 0.9$, see Figures S1–S3 in the online Supplementary Material. The density in Fig. 3a is an example of the bivariate wrapped Cauchy (bwC) distribution proposed by Kato and Pewsey (2013); see the supplementary material of that paper for many more examples. These densities can be proved to be unimodal, with mode at (μ_1, μ_2). They also have many other attractive properties including, remarkably, wrapped Cauchy conditional distributions, in addition to the wrapped Cauchy marginal distributions and binding distribution provided by construction.

Except for the case wC–wC–wC$(q, \mu_1, \rho_1, \mu_2, \rho_2, 0, \rho_g)$, and vM–vM–wC $(q, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \rho_g)$ the distributions in Fig. 3 are bimodal. This unattractive feature is not marked in the wC–wC–wC (bwC) case, but can be marked in the vM–vM–vM case, with vM–vM–wC a little less so. Nevertheless, as can be seen in Fig. 3d for the vM–vM–vM family, for families of distributions with $\mu_g = 0$ and more than one mode, (μ_1, μ_2) is the major mode. Henceforth, we will refer to the vM–vM–vM distribution as being bivariate von Mises (bvM).

So, using (9) to define the marginal distribution functions, and with the choice $\mu_g = 0$, the roles played by the parameters q, μ_1, ρ_1 or κ_1, μ_2, ρ_2 or κ_2 and ρ_g or κ_g are all clear-cut. As is evident from the first column of Fig. 3, the densities obtained are twofold symmetric when rotated (through $2\pi/2 = \pi$ radians) about (μ_1, μ_2). Indeed, it can be shown that if f_1, f_2 and g are symmetric about μ_1, μ_2 and 0, respectively, then density (5) is twofold symmetric when rotated about $(\mu_1, \mu_2 + \pi)$, $(\mu_1 + \pi, \mu_2)$ and $(\mu_1 + \pi, \mu_2 + \pi)$ as well as (μ_1, μ_2). The densities are not, in general, reflectively symmetric.

Setting $\mu_g = \pi$, instead of $\mu_g = 0$, results in densities that are still twofold symmetric when rotated about (μ_1, μ_2), but which, if $\rho_g \neq 0$, are bimodal; (μ_1, μ_2) is not one of the modes, but appears to be at a saddlepoint in between them. This is illustrated by the second column of Fig. 3. Using choices of μ_g other than 0 or π produces densities that, when $\rho_g \neq 0$, are no longer twofold symmetric when rotated about (μ_1, μ_2). Moreover, (μ_1, μ_2) is not a mode nor saddlepoint, but some apparently arbitrary point between modes (when $\rho_g \neq 0$). The value of μ_g also determines the orientation of asymmetry. These features are illustrated by the panels in the third column of Fig. 3.

Figure S4 in the online Supplementary Material shows the rather unappealing densities of the extensions of the bivariate von Mises distributions of Shieh and Johnson (2005) to include $q = -1$. In that figure, we vary μ_g and κ_g while keeping the marginal von Mises parameters fixed. The panels in the top row of Figure S4 are of the same densities as those in Figure 2 of Shieh and Johnson (2005) and show that their contour plots are insufficiently detailed to fully represent the true forms of the densities. They also reveal that the condition $\mu_g = \mu_1 - \mu_2$ conjectured by Shieh and Johnson (2005) does not, in fact, assure unimodality. Without going into further detail, the message from Figure S4 is clear: bvM distributions can be unimodal, bimodal or even trimodal (but not, we think, more than trimodal). It seems that von Mises marginals are accommodated in a little less multimodal manner by combining them with a wrapped Cauchy binding density; i.e. through the vM–vM–wC model.

Here is another consequence for the role of (μ_1, μ_2) when $\mu_g = 0$. Suppose that in the bwC$(q, \mu_1, \rho_1, \mu_2, \rho_2, 0, \rho_g)$ model, we let $\rho_1, \rho_2 \to 0$. For small ρ_1, ρ_2, this model is close to the circula with wrapped Cauchy g, yet (μ_1, μ_2) must be a point at which the circular density is maximal; but, from Sect. 2.1, when $\mu_g = 0$ the circula density is maximal on the diagonal $\theta_2 = q\theta_1$. The limiting circular density is, however, not the one with $\mu_g = 0$ but the one with $\mu_g = \mu_2 - q\mu_1$. This effect is illustrated in Fig. 4. So, we see an implicit effect akin to that of taking $\mu_g \neq 0$ in the circula case after all.

The above considerations lead us to a considerable preference for setting $\mu_g = 0$ in model (5).

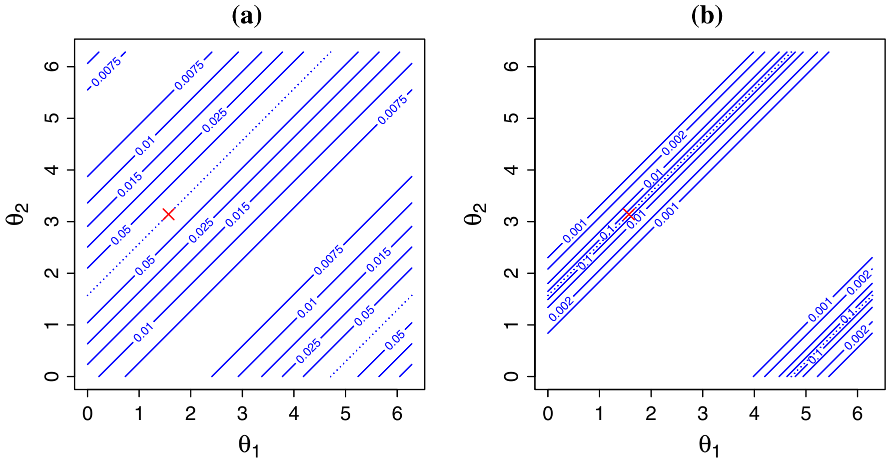

Fig. 4 Contour plots of bwC$(1, \pi/2, \rho_1 \to 0, \pi, \rho_2 \to 0, 0, \rho_g)$ densities with: **a** $\rho_g = 0.6$; **b** $\rho_g = 0.99$. The *dotted diagonal line* in each plot identifies those (θ_1, θ_2) combinations for which the density is maximal. The *cross* in each panel identifies the point $(\pi/2, \pi)$

3.2 Distribution function in terms of circula

Provided we parallel the univariate marginal case and define

$$F(\theta_1, \theta_2) = \int_{\mu_1}^{\theta_1} \int_{\mu_2}^{\theta_2} f(\phi_1, \phi_2) \mathrm{d}\phi_2 \mathrm{d}\phi_1, \quad \mu_1 \le \theta_1 \le \mu_1 + 2\pi, \mu_2 \le \theta_2 \le \mu_2 + 2\pi,$$

then it is easily seen that

$$F(\theta_1, \theta_2) = C_q(F_1(\theta_1), F_2(\theta_2)), \quad \mu_1 \le \theta_1 \le \mu_1 + 2\pi, \mu_2 \le \theta_2 \le \mu_2 + 2\pi.$$

3.3 Dependence, global and local

The scalar dependence measures of Sect. 2.3 and the averaging of the local dependence function as in Sect. 2.4 can both be applied to densities of form (5) directly. Formulae for the bwC distribution can be found in Kato and Pewsey (2013). Alternatively, as in the linear case, one can define the values of the dependence measures obtained from the circula to apply to densities (5) too, providing alternative 'margin-free' dependence measures for those distributions.

Regarding local dependence, we observe that patterns of *signs* of $\gamma_f(x, y)$ are reflected in patterns of signs of $\gamma_c(u, v)$, albeit distorted by marginal transformation. In particular, for example, $\gamma_f(x, y) > 0$ for all x, y if and only if $\gamma_c(u, v) > 0$ for all $0 < u, v < 1$. In the ordinary copula case, this corresponds to a TP$_2$ density; see, e.g. Joe (1997, Sect. 2.1.5).

3.4 Random variate generation

Random variate generation for the circula (4) is immediate using the construction given in the Introduction: generate Θ_1 from the circular uniform distribution, Ω from the distribution with density g, and set $\Theta_2 = (\Omega + q\Theta_1) \pmod{2\pi}$.

A basic algorithm for random variate generation from the density (5) is also immediate if marginal distributions allow generation by inversion of the distribution function: given (Θ_1, Θ_2) generated from (4) as above, then $\Theta_1^* = F_1^{-1}(\Theta_1/2\pi) \pmod{2\pi}$, $\Theta_2^* = F_2^{-1}(\Theta_2/2\pi) \pmod{2\pi}$ follow (5).

Minor modifications of this algorithm allow speedups in some situations, by avoiding one of the distribution function inversions. A first version is:

Algorithm A1 simulate Θ_1^* from f_1 and Ω from g, independently;
set $\Theta_2^* = F_2^{-1}\left\{\left(qF_1(\Theta_1^*) + \frac{\Omega}{2\pi}\right) \pmod 1\right\} \pmod{2\pi}$.

This is essentially the algorithm used by Shieh and Johnson (2005), Sect. 2.2, in the bvM case, although we can implement this algorithm much more efficiently. A second version is:

Algorithm A2 simulate Θ_2^* from f_2 and Ω from g, independently;
set $\Theta_1^* = F_1^{-1}\left[\left\{q\left(F_2(\Theta_2^*) - \frac{\Omega}{2\pi}\right)\right\} \pmod 1\right] \pmod{2\pi}$.

Examples of f_1 or f_2 for which these algorithms would be advantageous include the sine-skewed wrapped Cauchy distribution (Umbach and Jammalamadaka 2009; Abe and Pewsey 2011) and many wrapped distributions.

4 Inference

4.1 Maximum likelihood estimation

As explained in Sect. 3.1, we take $\mu_g = 0$. Let $\tau = (\tau_1, \tau_2, \tau_g)$, where τ_1 is the vector of, typically two, parameters of f_1, τ_2 that of f_2, and τ_g is the concentration parameter of g. For a random sample of size n from the distribution with density (5), $(\theta_{1,1}, \theta_{2,1}), \cdots, (\theta_{1,n}, \theta_{2,n})$, the log-likelihood function is given by

$$\ell(\tau) = n\log(2\pi) + \sum_{i=1}^{n}\log(f_1(\theta_{1,i})) + \sum_{i=1}^{n}\log(f_2(\theta_{2,i})) + \sum_{i=1}^{n}\log(g(2\pi(F_2(\theta_{2,i}) - qF_1(\theta_{1,i})))). \tag{10}$$

In general, the first two summations in (10) will be functionally related to the third in terms of the parameters, so there will be no closed-form solutions for the maximum likelihood estimates and numerical methods must be used to maximise (10). The constant $q = \pm 1$ determines whether the dependence between Θ_1 and Θ_2 is positive ($q = 1$) or negative ($q = -1$). Thus, q is a model choice indicator rather than a conventional parameter. In most applications, the form of any dependence, and hence the value of q, should be obvious from a consideration of a scatter-plot of the data. If not, (10) can be maximised twice, with $q = 1$ and $q = -1$, respectively, and the maximised values compared to identify the maximum likelihood solution.

Our experience of maximising (10) has been based on the use of `R`'s `optim` function together with its `L-BFGS-B` implementation of the optimisation method of Byrd et al. (1995) which allows for box constraints. The `hessian` argument of `optim` can be used to obtain a numerical approximation to the Hessian matrix. We also employ multiple starting values in an attempt to ensure that the global maximum likelihood solution is identified. Shieh and Johnson (2005) and Kato and Pewsey (2013) discuss maximum likelihood based inference for the bvM and bwC models discussed in Sect. 3.1. Their approaches can be extended to other cases of (10) in obvious ways. Alternatively, profile log-likelihood and parametric bootstrap methods can be used to construct confidence intervals.

4.2 Goodness-of-fit testing

The independence of $\Theta_1 = 2\pi F_1(\Theta_1^*)$ and $\Omega = 2\pi(F_2(\Theta_2^*) - qF_1(\Theta_1^*)) \pmod{2\pi}$ provides a means of exploring the goodness of fit of density (5) to a random sample, $(\theta_{1,1}, \theta_{2,1}), \cdots, (\theta_{1,n}, \theta_{2,n})$, of bivariate circular data.

Suppose, first, that, under the null hypothesis, the density is fully specified. Write $\omega_i = 2\pi(F_2(\theta_{2,i}) - qF_1(\theta_{1,i})) \pmod{2\pi}$, $i = 1, \ldots, n$. As Θ_1 and Ω are independent, if the data do come from the case of (5) specified under the null hypothesis, then the values of $\{2\pi F_1(\theta_{1,i}), 2\pi G(\omega_i)\}$, $i = 1, \ldots, n$, will be uniformly distributed on the torus; here, G is the distribution function associated with g. Various tests for toroidal uniformity have been proposed in the literature (see Jupp 2005, 2009, and references therein), the simplest being Wellner's (1979) extension of the Rayleigh test for isotropy.

In practice, the parameters of (5) will be unknown and must be estimated from the data. If the maximum likelihood approach of Sect. 4.1 is employed, goodness-of-fit tests can be based instead on the values of $\{2\pi \hat{F}_1(\theta_{1,i}), 2\pi \hat{G}(\hat{\omega}_i)\}$, $i = 1, \ldots, n$, where the hats denote evaluation at the maximum likelihood solution. Such values can be tested for toroidal uniformity using the tests referred to in the previous paragraph. However, the sampling distributions of those tests will no longer be the same as under the fully specified scenario. As a generally applicable method, the p value of a chosen test can be estimated using a parametric bootstrap approach. A large number, B, of parametric bootstrap samples of size n are simulated from the distribution fitted to the original sample. For each such sample, the parameters of (5) are estimated using maximum likelihood, resulting in tildes instead of hats, and the values of $\{2\pi \tilde{F}_1(\theta_{1,i}), 2\pi \tilde{G}(\tilde{\omega}_i))\}$, $i = 1, \ldots, n$, and the test statistic computed. The p-value of the test is estimated by the proportion of the $(B + 1)$ values of the test statistic that are at least as extreme as that for the original data. This approach incorporating Wellner's (1979) test is applied in the illustrative examples of the next section.

5 Examples

5.1 Texas wind data

Kato (2009) considered a data set of $n = 30$ pairs of wind directions measured each day at 6:00 and 7:00 from June 1, 2003 to June 30, 2003, in radians, at a weather station in Texas coded as C28-1. We treat these measurements as a set of independent bivariate data. Within pairs, one would expect the measurements to be strongly related as the time between the two measurements is just an hour. It is, though, natural to think of these data as a bivariate time series. However, time series plots and sample autocorrelation functions (Fisher and Lee 1983, 1994) for the series of wind directions at 6:00 and 7:00 separately (not shown) provide little evidence of dependence between successive observations in the separate series. The sample autocorrelations at lag 10 for wind directions at 6:00 and lags 1 and 2 for 7:00 are significantly different from zero, according to 95 % confidence bounds obtained using 1,000 randomisations of the original data. Nevertheless, all these autocorrelations remain very small and arguably not practically significant; for instance, the lag 1 autocorrelation is just 0.206. It seems therefore that the time gap of 24 h between pairs of recordings makes the assumption of independence of the pairs reasonable. A more sophisticated analysis might allow for any slight dependence. A scatterplot of the measurements appears in each of the panels of Fig. 5. Most of the points in the scatterplot indeed indicate a fairly strong positive relationship between pairs of observations. One might also contemplate potential bimodality but any suggestion of such in this small dataset is far from conclusive. Moreover, other data and other interests might, of course, relate to a full univariate time series of wind directions at C28-1 at all times through the day.

Tacitly assuming independence between distinct pairs of observations, Kato (2009) fitted three six-parameter bivariate circular distributions with von Mises marginals to these data, one of them being the bvM model with $\mu_g \neq 0$ and the classical definition of the distribution function starting at zero; i.e. not (9). The other two distributions were proposed in Kato (2009) and SenGupta (2004), respectively. Kato (2009) did not consider formal approaches to assessing the goodness of fit of the three fitted bivariate

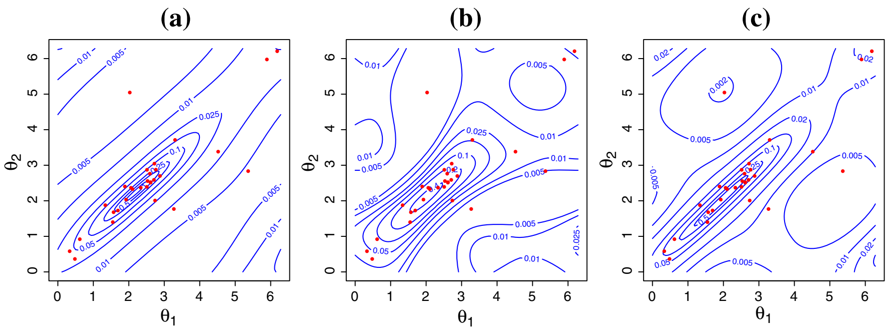

Fig. 5 Contour plots for the bwC (*left*), bvM (*centre*) and vM–vM–wC (*right*) densities fitted using maximum likelihood to the 30 pairs of wind directions measured at 6:00 (θ_1) and 7:00 (θ_2) at the C28-1 Texan weather station

Table 1 Maximum likelihood estimates, maximised log-likelihood value ($\ell_{\max}$), and the p-value for the bootstrap version of the goodness-of-fit test based on the use of Wellner's (1979) test for toroidal uniformity and $B = 99$ parametric bootstrap samples ($p_{\text{g-o-f}}$), for the fits to the C28-1 Texan wind direction data of the bwC, bvM and vM-vM-wC models with $\mu_g = 0$, distribution function as defined in (9), and $q = 1$

Model	$\hat{\mu}_1$	$\hat{\kappa}_1/\hat{\rho}_1$	$\hat{\mu}_2$	$\hat{\kappa}_2/\hat{\rho}_2$	$\hat{\kappa}_g/\hat{\rho}_g$	$\ell_{\max}$	$p_{\text{g-o-f}}$
bwC$(1, \mu_1, \rho_1, \mu_2, \rho_2, 0, \rho_g)$	2.22	0.48	2.27	0.52	0.73	−64.93	0.30
bvM$(1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \kappa_g)$	2.00	1.12	2.10	1.33	2.18	−71.13	0.03
vM–vM–wC $(1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \rho_g)$	1.93	1.05	2.01	1.16	0.75	−65.99	0.02

von Mises models. We fitted the bwC, bvM and vM–vM–wC models, with $\mu_g = 0$ and the distribution function as defined in (9), to the data. The results obtained for the three fits are presented in Table 1. Contour plots of the fitted densities are superimposed upon scatterplots of the data in the panels of Fig. 5. The $\ell_{\max}$ and $p_{\text{g-o-f}}$ values indicate that the bwC model provides the superior fit to the underlying distribution of the data. The vM–vM–wC model is quite close to the bwC model in terms of $\ell_{\max}$ but not $p_{\text{g-o-f}}$; this disparity can be explained by the vM–vM–wC density apparently overfitting by adapting its shape too closely to the outlying points, an effect that does not withstand bootstrapping the data. The $\ell_{\max}$ value of the bwC model is also higher than those of all three six-parameter bivariate von Mises models considered by Kato (2009).

5.2 Japanese earthquake data

In our second example, we consider data introduced by Hamada and O'Rourke (1992) and analysed in Rivest (1997) on the pre-earthquake direction of steepest descent (Θ_1) and the direction of lateral ground movement (Θ_2) before and after, respectively, an earthquake in Noshiro, Japan. Originally, observations at 763 different locations were recorded with a number of the lateral ground movement measurements being rounded to 90° and 270°, and a few to 0° and 180°. Removing the cases with rounded θ_2-values reduces the sample size to 678. A scatterplot of the data converted to radians appears in each of the panels of Fig. 6. The points in the scatterplot suggest that the underlying distribution is bimodal. For geotectonic reasons, it is probably doubtful that distinct pairs of measurements are independent. Nevertheless, here we analyse them assuming they are independent.

We first fitted single-component bwC, bvM and vM–vM–wC models, with $\mu_g = 0$ and the distribution function as defined in (9), to the data. The results obtained were, as expected, inadequate. The vM–vM–wC model was identified as providing the best fit of the three. However, visual inspection of the corresponding contour plots superimposed on the data (not shown) suggested that none of the models provides adequate fits to the underlying distribution of the data; the major lack of fit corresponds to θ_2-values in, roughly, the interval (2,4) radians, where a considerable amount of density seems to occur away from the main mode in the data.

In a search for a better-fitting model, we next explored the fits of two-component bvM, bwC and vM–vM–wC mixture models with mixing probability p as the multiple

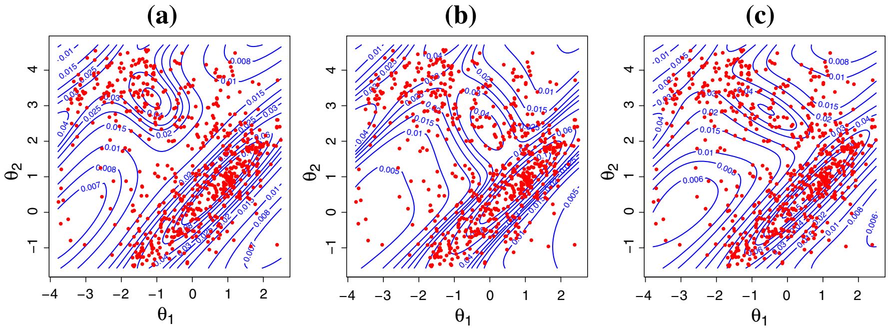

Fig. 6 Contour plots for the two-component bwC (*left*), bvM (*centre*) and vM–vM–wC (*right*) mixture densities fitted using maximum likelihood to the 678 unrounded pairs of pre-earthquake direction of steepest descent (θ_1) and direction of lateral ground movement (θ_2)

Table 2 Maximum likelihood estimates and maximised log-likelihood value ($\ell_{\max}$) for the fits to the 678 unrounded pairs of pre-earthquake direction of steepest descent (θ_1) and direction of lateral ground movement (θ_2) values of the two-component bwC, bvM and vM–vM–wC mixture models

Model	$\hat{\mu}_1$	$\hat{\kappa}_1/\hat{\rho}_1$	$\hat{\mu}_2$	$\hat{\kappa}_2/\hat{\rho}_2$	$\hat{\kappa}_g/\hat{\rho}_g$	$\hat{p}$	$\ell_{\max}$
bwC$(1, \mu_1, \rho_1, \mu_2, \rho_2, 0, \rho_g)_1$	0.73	0.27	0.72	0.25	0.57	0.78	
bwC$(-1, \mu_1, \rho_1, \mu_2, \rho_2, 0, \rho_g)_2$	5.16	0.50	3.20	0.48	0.14		−2206.71
bvM$(1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \kappa_g)_1$	0.70	0.37	0.67	0.37	3.45	0.57	
bvM$(-1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \kappa_g)_2$	6.19	0.90	2.37	0.51	0.31		−2202.71
vM–vM–wC $(1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \rho_g)_1$	0.53	0.51	0.52	0.50	0.59	0.79	
vM–vM–wC $(-1, \mu_1, \kappa_1, \mu_2, \kappa_2, 0, \rho_g)_2$	5.70	0.95	2.88	1.84	0.25		−2199.72

of the density for the first component. All three mixture models have a total of 11 parameters. The results obtained from fitting them are presented in Table 2. Contour plots of the fitted densities are superimposed upon scatterplots of the data, shifted to a linear scale on which they can be most fully appreciated, in the panels of Fig. 6. The $\ell_{\max}$ values in Table 2 identify the two-component vM–vM–wC mixture model as providing the best fit, and a visual inspection of panel (c) of Fig. 6 suggests that it does provide a respectable fit to the data. However, we have not been able to formally assess the goodness-of-fit of this more complicated model. According to that fitted model, around 80 % of the data arise from a distribution centred around the location (0.53, 0.52), i.e. almost equal individual location parameter values, with von Mises marginals with similar low concentrations that are moderately positively correlated. The remaining 20 % arise from a second distribution, centred around the location (5.70, 2.88) with more concentrated von Mises marginals that are weakly negatively correlated. In the best fitting model(s), the second mode is not especially pronounced against a non-negligible 'background' level, the whole reflecting a distribution of data with a major mode together with something of a low-peaked 'plateau' towards the northwest of the mode in the representations of Fig. 6.

6 Multivariate extension

As in the linear case, the value of direct multivariate extensions of the circula is not especially clear, given the attraction of pair copula constructions (Bedford and Cooke 2002; Kurowicka and Cooke 2006; Aas et al. 2009) to more meaningfully model highly multivariate situations. There would appear to be no impediment to employing the same techniques in the circula case.

Nonetheless, here is our best suggestion for a direct d-variate, $d \geq 3$, extension of the circula of interest. It has d separate, and hence somewhat constrained when $d \geq 4$, dependence parameters. Start from the joint density of Φ and $\Theta_k = (\Omega_k + q_k\Phi) \pmod{2\pi}$, $k = 1, \ldots, d$, where Ω_k follows density g_k, $k = 1, \ldots, d$, independently of each other and of Φ which is circular uniformly distributed, and $q_k \in \{-1, 1\}$, $k = 1, \ldots, d$; this is

$$c_{d+1}(\phi, \theta_1, \ldots, \theta_d) = \frac{1}{2\pi} \prod_{k=1}^{d} g_k(\theta_k - q_k\phi).$$

By construction, this density clearly has circular uniform univariate marginals, as does the d-dimensional marginal distribution of $\Theta_k = (\Omega_k + q_k\Phi) \pmod{2\pi}$, $k = 1, \ldots, d$, which has the proposed multivariate circula density

$$c_d(\theta_1, \ldots, \theta_d) = \frac{1}{2\pi} \int_0^{2\pi} \prod_{k=1}^{d} g_k(\theta_k - q_k\phi)\, \mathrm{d}\phi. \tag{11}$$

The (k, l)th bivariate marginal of c_d, $k = 1, \ldots, d$, $l = 1, \ldots, d$, $k \neq l$, is the joint distribution of

$$\Theta_k \text{ and } \Theta_l = (q_{kl}\Theta_k + \Omega_l - q_{kl}\Omega_k) \pmod{2\pi},$$

where $q_{kl} = q_k q_l \in \{-1, 1\}$. This has circula density of form (4) given by

$$\frac{1}{2\pi} h_{kl}(\theta_l - q_{kl}\theta_k),$$

where

$$h_{kl}(\omega) = \int_0^{2\pi} g_k(\psi) g_l(\omega + q_{kl}\psi) \mathrm{d}\psi \tag{12}$$

is the density of $(\Omega_l - q_{kl}\Omega_k) \pmod{2\pi}$.

If g_k is symmetric about 0 with mean resultant length ρ_k, $k = 1, \ldots, d$, then the (k, l)th marginal copula density has mean resultant length and hence dependence parameter

$$\rho_{kl} = E\{\cos(\Omega_l - q_{kl}\Omega_k)\} = E(\cos\Omega_l)E(\cos\Omega_k) = \rho_k\rho_l,$$

$k = 1, \ldots, d$, $l = 1, \ldots, d$, $k \neq l$. This is the dependence structure to which we referred at the start of the second paragraph of this section.

Particularly attractive versions of this construction arise when the g's are all of the same form and are closed under convolution, so that the h's are of the same form as the g's as well. Wrapped stable distributions (Pewsey 2008) and a new family of circular distributions due to Kato and Jones (2014) are amongst those with the required property; both include the wrapped Cauchy distribution, and the latter the cardioid distribution, as special cases. So, for example, a d-dimensional version of the circula underlying a multivariate extension of the bivariate wrapped Cauchy distribution of Kato and Pewsey (2013) would be based on g_k being a wrapped Cauchy density with location zero and mean resultant length ρ_k so that h_{kl} is the wrapped Cauchy density with location zero and mean resultant length $\rho_k \rho_l$. In this case, the integration in (11) can be performed explicitly and it is possible to express the density in closed form.

There is an analogy between the construction above—particularly its correlation structure—and the one-factor Gaussian model on $\mathbb{R}^d$ introduced in the financial literature by Li (2000); there, $W_k = \rho_k Z + \sqrt{1-\rho_k^2}\, Z_k, -1 < \rho_k < 1, k = 1, \ldots, d$, and $Z, Z_1, \ldots, Z_d$ are mutually independent standard normal random variables. See, for example, Oh and Patton (2012) and Krupskii and Joe (2013) for access to the factor copula literature stemming from this idea. Inspired by this work, a suggestion—for future research—is an interesting structured 'two-factor' multivariate extension in which

$$\Theta_k = \begin{cases} (\Omega_k + q_k \Phi_1) \,(\mathrm{mod}\, 2\pi), & k = 1, \ldots, d_1, \\ (\Omega_k + q_k \Phi_2) \,(\mathrm{mod}\, 2\pi), & k = d_1 + 1, \ldots, d, \end{cases}$$

where, as before, the Ω_k's independently follow density g_k, $k = 1, \ldots, d$, but now Φ_1, Φ_2 follow a dependent bivariate circula.

7 Discussion

In this paper, we have concentrated on a particular class of circulas not because arguments for its use are entirely compelling but because, unlike the linear copula case, attractive alternative constructions seem difficult to come by. The current class of circulas is certainly attractive in its simplicity and tractability, but does not necessarily result in especially attractive bivariate circular models for arbitrary non-uniform marginals. A major exception to this arises in the case of wrapped Cauchy g binding wrapped Cauchy marginals, the elegant bivariate wrapped Cauchy model of Kato and Pewsey (2013).

We envisage using distributions based on circulas directly in unimodal situations. Where cluster structure is apparent, as in the example of Sect. 5.2, we naturally advocate using mixtures of distributions based on circulas. A particularly important application in which multimodal bivariate circular data arise is in understanding the structure of proteins (see Mardia 2013). The example in Kato and Pewsey (2013) confirms that their bwC distribution can appropriately model a component of such a mixture distribution. Further evidence for whether there may be a role for mixtures of these circula-based distributions as alternatives to the models currently employed (Mardia 2013, Sect. 3.3) is a question for future work.

We emphasise again that, for non-uniform marginals, it is recommended that the location parameter of g be set to zero and that the marginal distribution functions be defined as at (9). Given the positivity or negativity of the dependence in the data (reflected in $q = \pm 1$), the resulting five-parameter models afford parsimony and interpretability; like bivariate normal distributions on $\mathbb{R}^2$, their parameters consist of two location parameters, two concentration parameters and one parameter controlling the strength of the relationship between the two variables.

Finally, extension in a direction different to that of Sect. 6 might be to 'sphericas': copulas linking spherical marginals. Rivest (1984) discussed such a class under the name '$O(d)$-symmetric distributions'. Their densities are of the form $f(u, v) = g(u^T Qv)$, where g is a rotationally symmetric density, u and v are d-dimensional unit vectors and Q is a $d \times d$ orthogonal matrix. The circulas (4) with wrapped Cauchy or von Mises binding density considered in this paper are examples of $O(2)$-symmetric distributions.

Acknowledgments We are grateful to the editorial team for suggestions, including two of the five pre-existing dependence measures to which we refer and the multivariate analogy with factor models, that have improved the paper in revision. Financial support for this work was received by Arthur Pewsey in the form of grant MTM2010-16845 from the Spanish Ministry of Science and Education and grant GR10064 from the Junta de Extremadura.

We would like to dedicate this paper to the memory of Arthur's father who passed away around the time work for the paper first began.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Mathematics and Economics*, *44*, 182–198.

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

Alfonsi, A., Brigo, D. (2005). New families of copulas based on periodic functions. *Communication in Statistics: Theory and Methods*, *34*, 1437–1447.

Bedford, T., Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *Annals of Statistics*, *30*, 1031–1068.

Byrd, R. H., Lu, P., Nocedal, J., Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190–1208.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, *63*, 579–585.

Fisher, N. I., Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika*, *70*, 327–332.

Fisher, N. I., Lee, A. J. (1994). Times series analysis of circular data. *Journal of the Royal Statistical Society Series B*, *21*, 327–339.

García-Portugués, E., Crujeiras, R. M., González-Manteiga, W. (2013). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stochastic Environmental Research and Risk Assessment*, *27*, 1055–1067.

Hamada, M., O'Rourke, T. (1992). Case Studies of Liquefaction and Lifeline Performance During Past Earthquake. Vol. 1. Japanese Case Studies. Buffalo, NY: National Center for Earthquake Engineering Research.

Holland, P. W., Wang, Y. J. (1987). Dependence function for continuous bivariate densities. *Communications in Statistics: Theory and Methods*, *16*, 863–876.

Isham, V. (1977). A Markov construction for a multidimensional point process. *Journal of Applied Probability*, *14*, 507–515.

Jammalamadaka, S. R., Sarma, Y. R. (1988). A correlation coefficient for angular variables. In K. Matusita (Ed.), *Statistical Theory and Data Analysis II* (pp. 349–364). Amsterdam: Elsevier.
Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.
Johnson, R. A., Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *Journal of the Royal Statistical Society Series B*, *39*, 222–229.
Johnson, R. A., Wehrly, T. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, *73*, 602–606.
Jones, M. C. (1996). The local dependence function. *Biometrika*, *83*, 899–904.
Jones, M. C. (2013). Perlman and Wellner's circular and transformed circular copulas are particular beta and *t* copulas. *Symmetry*, *5*, 81–85.
Jupp, P. E. (2005). Sobolev tests of goodness of fit of distributions on compact Riemannian manifolds. *Annals of Statistics*, *33*, 2957–2966.
Jupp, P. E. (2009). Data-driven tests of uniformity on product manifolds. *Journal of Statistical Planning and Inference*, *139*, 3820–3829.
Jupp, P. E., Mardia, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika*, *67*, 163–173.
Kato, S. (2009). A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli*, *15*, 898–921.
Kato, S., Jones, M. C. (2014). A tractable and interpretable four-parameter family of unimodal distributions on the circle. Biometrika, *To appear*.
Kato, S., Pewsey, A. (2013). A Möbius transformation-induced distribution on the torus. *To appear*.
Krupskii, P., Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, *120*, 85–101.
Kurowicka, D., Cooke, R. M. (2006). *Uncertainty Analysis With High Dimensional Dependence Modelling*. New York: Wiley.
Li, D. X. (2000). On default correlation: a copula approach. *Journal of Fixed Income*, *9*, 43–54.
Mardia, K. V. (2013). Statistical approaches to three key challenges in protein structural bioinformatics. *Applied Statistics*, *62*, 487–514.
Mardia, K. V., Jupp, P. E. (1999). *Directional Statistics*. New York: Wiley.
Nelsen, R. B. (2010). *An Introduction to Copulas* (2nd ed.). New York: Springer.
Oh, D. H., Patton, A. J. (2012). Modelling dependence in high dimensions with factor copulas. Working paper (revised version). Duke University.
Perlman, M. D., Wellner, J. A. (2011). Squaring the circle and cubing the sphere: Circular and spherical copulas. *Symmetry*, *3*, 574–599.
Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics and Data Analysis*, *52*, 1516–1523.
Rivest, L.-P. (1982). Some statistical methods for bivariate circular data. *Journal of the Royal Statistical Society Series B*, *44*, 81–90.
Rivest, L.-P. (1984). Symmetric distributions for dependent unit vectors. *Annals of Statistics*, *12*, 1050–1057.
Rivest, L.-P. (1997). A decentred predictor for circular-circular regression. *Biometrika*, *84*, 717–726.
SenGupta, A. (2004). On the construction of probability distributions for directional data. *Bulletin of the Calcutta Mathematical Society*, *96*, 139–154.
Shieh, G. S., Johnson, R. A. (2005). Inference based on a bivariate distribution with von Mises marginals. *Annals of the Institute of Statistical Mathematics*, *57*, 789–802.
Shieh, G. S., Zheng, S., Johnson, R. A., Chang, Y. F., Shimizu, K., Wang, C. C., et al. (2011). Modeling and comparing the organization of circular genomes. *Bioinformatics*, *27*, 912–918.
Umbach, D., Jammalamadaka, S. R. (2009). Building asymmetry into circular distributions. *Statistics and Probability Letters*, *79*, 659–663.
Wehrly, T., Johnson, R. A. (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, *66*, 255–256.
Wellner, J. A. (1979). Permutation tests for directional data. *Annals of Statistics*, *7*, 929–943.