Chilean Journal of Statistics
Vol. 3, No. 2, September 2012, 129–141

Probabilistic and Inferential Aspects of Skew-Symmetric Models
Special Issue: "IV International Workshop in honour of Adelchi Azzalini's 60th Birthday"

A wrapped flexible generalized skew-normal model for a bimodal circular distribution of wind directions

Estefanía Hernández-Sánchez[1] and Bruno Scarpa[2,*]

[1]Faculty of Statistical Sciences, University of Padua, Padua, Italy
[2]Department of Statistical Sciences, University of Padua, Padua, Italy

(Received: 03 October 2011 · Accepted in final form: 24 April 2012)

Abstract

Motivated by the analysis of wind directions, in this paper we consider skew-symmetric circular distributions generated by perturbation of a symmetric circular distribution. This class of models is able to describe different distribution shapes, including symmetric, skewed, and bimodal, which are often observed in circular data, as in our motivating example of wind directions at a site in Spain. We propose a wrapped version of the flexible generalized skew-normal distribution to fit these data. The model is presented and the parameters of the proposed model are estimated by the maximum likelihood method. We also note that in the considered area, typical sea breeze and land breeze directions are observed. Thus, we consider a subdivision of the data to fit these directions separately. The likelihood ratio test shows that the proposed model outperforms the wrapped skew-normal, and the Akaike information criterion reveals that it fits our wind-direction data with comparable performances with respect to other well-known bimodal circular distributions.

Keywords: Circular model · Flexible generalized skew-normal distribution · Wind directions · Wrapped distributions · Wrapped skew-normal distribution.

Mathematics Subject Classification: Primary 62E99 · Secondary 62P12.

1. Introduction

Circular data are observed in many sciences and a number of statistical models have been proposed to describe data related to directions, that is, represented by mathematical angles. Mardia (1972) is probably the first book related to statistics for circular data, which relied substantially on Rao Jammalamadaka's PhD dissertation (see Rao, 1969), but this book has been followed by many others; see, e.g., Fisher (1993), Mardia and Jupp (1999), and Jammalamadaka and SenGupta (2001). In addition to these books, many authors have proposed models and statistical methodology for analyzing circular data.

*Corresponding author. Bruno Scarpa, Dipartimento di Scienze Statistiche, Università di Padova, via Cesare Battisti 241, 35121, Padova, Italia. Email: scarpa@stat.unipd.it

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

As pointed out by Pewsey (2006), among others, most of the classical statistical models for circular data involve symmetric distributions, and since this kind of data is very often asymmetric, they are rarely applied in practice. However, a number of models that include asymmetry have recently been proposed; see Jammalamadaka and Kozubowski (2003), Fernandez-Duran (2004), Gatto and Jammalamadaka (2007), and Umbach and Jammalamadaka (2009). Pewsey (2000a) obtained a skewed circular model by wrapping the Azzalini skew-normal (SN) distribution. Other wrapped distributions have been obtained more recently by Jammalamadaka and Kozubowski (2003), by wrapping a skewed Laplace distribution, and by Pewsey (2008), by wrapping a four-parameter stable family of densities.

Often, real circular data exhibit not only skewness but also bimodality. Recently, models have been proposed to fit multimodal data; see Kato and Jones (2010) and Abe and Pewsey (2011). In this paper, motivated by the analysis of wind data collected at a station in the Alicante region of Spain, we propose wrapping what Ma and Genton (2004) called a flexible generalized skew-normal (FGSN) distribution, thus obtaining a model that can deal with skewness and bimodality.

In Section 2, we describe the motivating problem and the available data. In Section 3, we review the SN distribution and its generalization. In Section 4, we propose the new model, and in Section 5, this model is fitted to the available data. In Section 6, we compare the proposed model with other bimodal models and in Section 7 a discussion of the results is included.

2. Wind Directions in Villena

We are interested in modeling wind directions observed at a single point. Specifically, we consider wind directions registered at a meteorological station in Villena, a town in the province of Alicante, Spain. This site is along the river Vinalopó at a height of 504 meters above mean sea level. The town is in a valley which forms part of the Prebaetic System, a chain of mountains in the south of Spain, covering more than 600 km between Gibraltar and Cabo de la Nao, running roughly south-west to north-east. The weather in Villena is typical Mediterranean, but due to its height and its proximity to the mountains, weather oscillations are more pronounced. In addition, the Vinalopó River runs in a valley that goes directly from the Villena region to the coast, providing a corridor for winds coming from the sea to reach directly to the town.

The sea breeze and mountain breeze dynamics are observed daily in the area of Villena; see Azorin-Molina et al. (2011). During the daytime, a wind (sea breeze) is expected from the sea moving overland in a south-easterly direction, formed by increasing temperature differences between the land and water. Over night and up until late morning, a mountain breeze is also observed in the north-west direction. Mountain air cools and causes wind to move down the valley.

Data were collected using an Oregon Scientific WMR928NX automatic weather station (AWS) during June of 2009. The anemometer of the station was positioned at a height of 10 meters above ground. Figure 1 represents the topography of the Villena area and the location of the station. The software recorded observations every 20 minutes, collecting 2160 measurements. Occasionally, the data failed to be recorded, so a small number of observations were missing. In addition, when no wind was observed, the data were considered to be missing. Therefore, we discarded 327 measurements, leaving us with 1833 available observations. Figure 2 shows a rose diagram of the wind directions observed in June of 2009. The external points show the observed angles, while inside the circle, we have plotted the rose diagram with classes of 20° intervals.

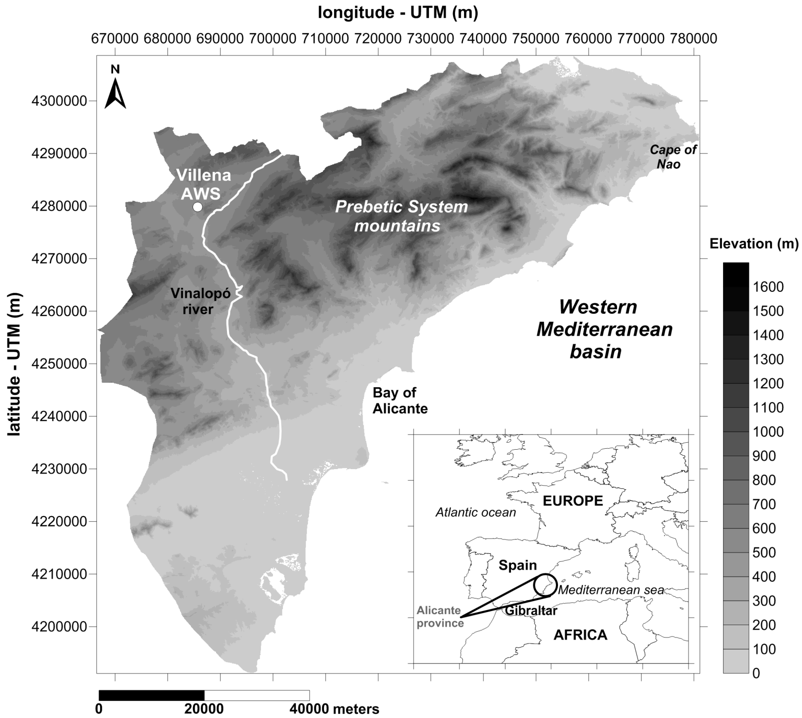

Figure 1. Topographic map of the Villena area with the location of the Villena AWS station.

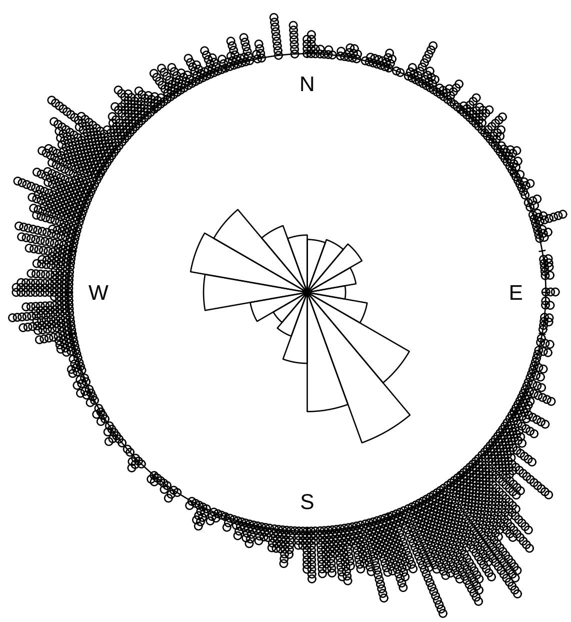

Figure 2. Rose diagram and raw circular data plot of the wind direction data collected in Villena in June 2009.

The rose diagram reveals that the distribution is bimodal with modes made up from the directions associated with the sea and mountain breezes. Furthermore, at least one of the modes, the one in the west/north-west direction, is clearly skewed. Our goal is to model such a bimodal skewed distribution.

3. Skew-Normal and Flexible Generalized Skew-Normal Distributions

Pewsey (2000a, 2006) proposed a circular distribution by wrapping a SN distribution onto the unit circle. The univariate SN distribution, as defined by Azzalini (1985), has density

$$f(x;\xi,\omega,\alpha) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right), \tag{1}$$

where $x \in \mathbb{R}$ and ϕ and Φ denote the standard normal density and distribution functions, respectively, $\xi \in \mathbb{R}$, $\omega > 0$, and $\alpha \in \mathbb{R}$ are location, scale, and shape parameters (usually referred to as its 'direct parameters'). Let us denote this distributional relationship as $X \sim \mathrm{SN}(\xi,\omega,\alpha)$. Moments for the SN distribution are available (see, e.g., Genton, 2004; Azzalini, 2005), and it is possible to parameterize the SN distribution by using the mean, variance, and skewness index, using its so-called centered parameterization. The centered parameters, μ, σ^2, and γ_1 (which are the mean, variance, and third cumulant of X, respectively), are related to the direct ones according to

$$\mu = \mathrm{E}[X] = \xi + \sqrt{\frac{2}{\pi}}\omega\delta, \qquad \sigma^2 = \omega^2\left(1-\frac{2}{\pi}\delta^2\right), \qquad \gamma_1 = \frac{(4-\pi)\mu^3}{2(1-\mu^2)^{3/2}},$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. As pointed out by Azzalini (1985), and discussed in Azzalini and Capitanio (1999) and Pewsey (2000b), the direct parameterization may present some estimation problems in particular cases, which can be circumvented by using the centered version.

The circular random variable $\Theta = X(\mathrm{mod}\ 2\pi)$, corresponding to wrapping X onto the unit circle, has density

$$f(\theta;\xi,\omega,\alpha) = \frac{2}{\omega}\sum_{r=-\infty}^{\infty}\phi\left(\frac{\theta+2\pi r-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{\theta+2\pi r-\xi}{\omega}\right)\right), \tag{2}$$

for $0 \le \theta \le 2\pi$, and is denoted as $\Theta \sim \mathrm{WSN}(\xi,\omega,\alpha)$. The properties of the wrapped SN distribution are derived in Pewsey (2000a), where it is shown that there is not an equivalent with the centered parameterization for the circular SN distribution. However, Pewsey (2000a) proposed to use a so-called circular parameterization by working with the mean direction, one minus the circular variance, and the circular skewness index introduced by Batschelet (1981) instead of ξ, ω, and α. Ma and Genton (2004) showed that density in Equation (1) is unimodal, a characteristic inherited by the wrapped version in Equation (2), while Figure 2 clearly shows that our motivating data have more than a single mode, suggesting that model given in Equation (2) is not appropriate for our aim.

A bimodal circular random variable can be obtained by generalizing the SN distribution using the results obtained by Azzalini and Capitanio (2003), Ma and Genton (2004), and Wang et al. (2004). They showed that for any symmetric density function f_0 and distribution function G with symmetric density, the function

$$f(x) = 2f_0(x)G(w(x))$$

is a density function for any odd function w. Clearly, when $f_0 = \phi$, $G = \Phi$, and $w(x) = \alpha x$, we get Equation (1). Under this framework, setting $w(x) = \alpha x + \beta x^3$, we obtain the FGSN distribution, with density

$$f(x;\xi,\omega,\alpha,\beta) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)+\beta\left(\frac{x-\xi}{\omega}\right)^3\right). \tag{3}$$

The same authors proved that the density given in Equation (3) has at most two modes and noted that, in general, if w is an odd polynomial, the number of modes is positively related to its degree.

4. Wrapped Flexible Generalized Skew-Normal Distribution

We propose a bimodal skewed model by wrapping the FGSN distribution given in Equation (3) onto the unit circle. If X has a FGSN distribution, then $\Theta = X(\text{mod } 2\pi)$ is its wrapped version with density

$$f(\theta;\xi,\omega,\alpha,\beta) = \frac{2}{\omega}\sum_{r=-\infty}^{\infty}\phi\left(\tfrac{\theta+2\pi r-\xi}{\omega}\right)\Phi\left(\alpha\left(\tfrac{\theta+2\pi r-\xi}{\omega}\right)+\beta\left(\tfrac{\theta+2\pi r-\xi}{\omega}\right)^3\right), \tag{4}$$

for $0 \leq \theta \leq 2\pi, \xi \in \mathbb{R}, \omega > 0, \alpha \in \mathbb{R}$, and $\beta \in \mathbb{R}$, and is denoted by $\Theta \sim \text{WFGSN}(\xi,\omega,\alpha,\beta)$.

As for the linear FGSN distribution, circular moments of the WFGSN distribution exist, but are not available in closed form. Therefore, centered or circular parameterizations are not explicitly available for this distribution. Figure 3 shows the shape of the density given in Equation (4) for various choices of α and β, and $\xi = 0$ and $\omega = 1$. Mazzuco and Scarpa (2011) performed a simulation based study on how the values of α and β affect the shape of the FGSN distribution. However, their findings can also be generalized to the WFGSN distribution. They found that: (a) if α and β have the same sign, the resulting density has only one mode (in some cases, the density shows a small "bulge", but this is never an additional mode); (b) if α and β have opposite signs, the resulting density has two modes; (c) if the absolute value of β increases, the height of the second mode increases; and (d) the higher the absolute value of α, the more distant the two modes between them are.

Inference for the parameters of the WFGSN distribution can be performed by the maximum likelihood method. When n independent observations $\theta_1,\ldots,\theta_n$ from this distribution are available, the log-likelihood function is

$$\begin{aligned} l(\xi,\omega,\alpha,\beta;\theta) &= n\log(2) - n\log(\omega) \qquad (5)\\ &+\sum_{i=1}^{n}\log\left(\sum_{r=-\infty}^{\infty}\phi\left(\frac{\theta_i+2\pi r-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{\theta_i+2\pi r-\xi}{\omega}\right)+\beta\left(\frac{\theta_i+2\pi r-\xi}{\omega}\right)^3\right)\right), \end{aligned}$$

which can be maximized using a numerical algorithm such as, for example, Nelder and Mead's algorithm; see Nelder and Mead (1965).

We assume that the different observations are independent although there may be dependence between the wind directions over time. Breckling (1989) and Fisher and Lee (1994), among others, proposed a temporal model that takes into account autocorrelation by wrapping time series processes. In discussing these procedures, Coles (1998) pointed out that, even for a Gaussian model, the wrapping operation makes inference complicated. Working directly with the wrapped version of a Gaussian process, to compute, for example, the maximum likelihood estimates is intractable. The use of skew-symmetric distributions clearly adds an additional level of complexity to the procedure.

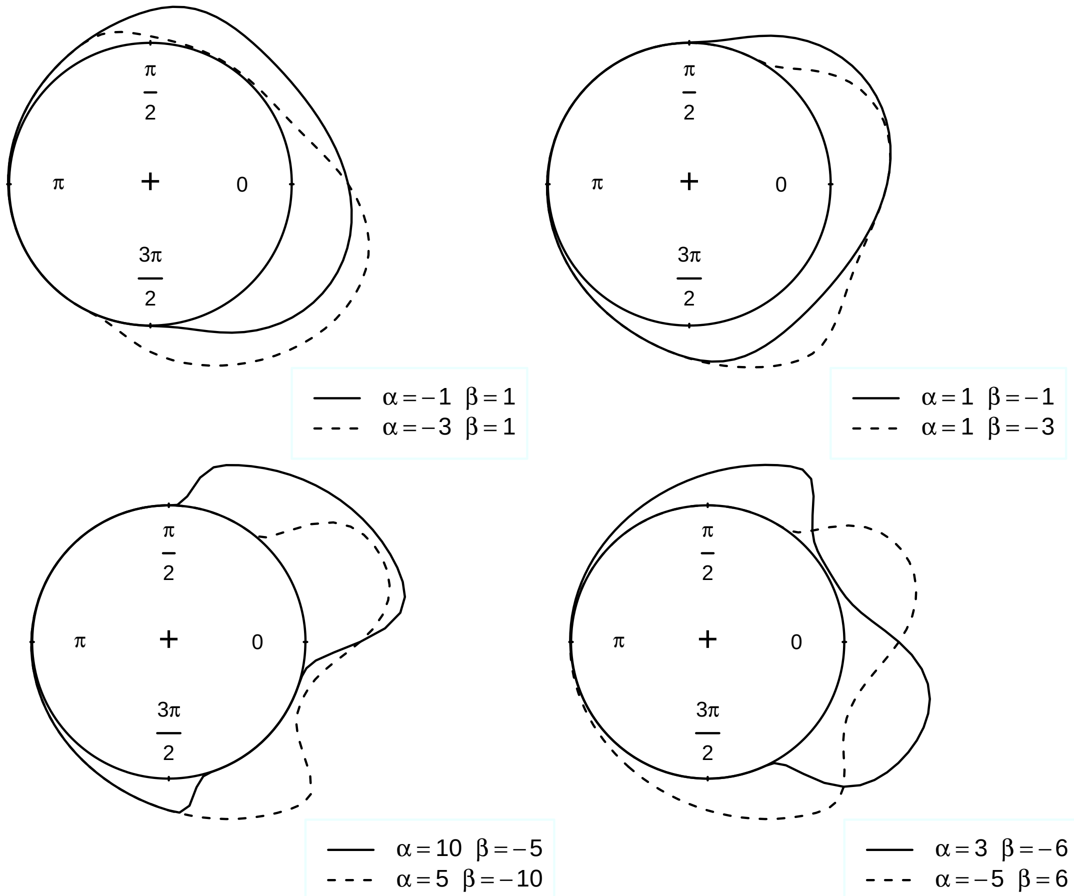

Figure 3. Examples of wrapped flexible generalized skew-normal (WFGSN) densities with $\xi = 0$, $\omega = 1$ and various combinations of α and β.

However, as pointed out, for example by Chandler and Bate (2007) who discussed more general dependence structures, estimates obtained assuming independence usually exhibit little bias, although they often lack efficiency. In our case, this means, for example, that profile likelihood based confidence intervals obtained under the assumption of independence among observations are likely to be narrower than the true ones. Some authors, such as Chandler and Bate (2007) and Pace et al. (2011), proposed corrections in order to obtain better approximations of the intervals. We could use similar modifications for our estimates of the Villena wind direction data. Here, we concentrate on assessing the improvement in fit of the WFGSN distribution over the WSN distribution, and leave the analysis of dependence structures for future research.

5. Models for Wind Direction Data

5.1 A model for the full data set

Model given in Equation (4) is a reasonable candidate for wind direction data in Villena, which, as we have seen in Section 2, are skewed and bimodal. Estimation of parameters was obtained using the maximum likelihood method, and the corresponding density is shown in Figure 4. To avoid local maxima, we carry out an optimization algorithm with several different starting points, widely scattered in the effective part of the parameter space. Starting from many different points, we arrive at a small number of local maxima, of which we select the one corresponding to the maximum value of the likelihood function.

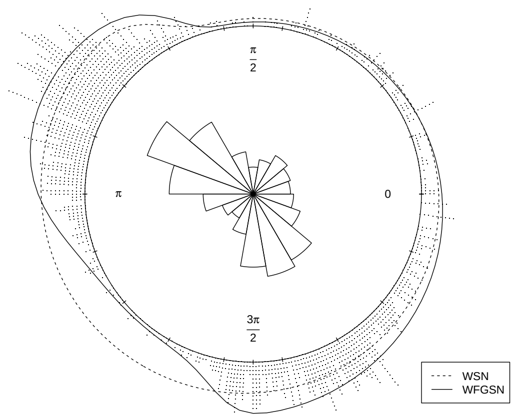

Figure 4. Raw circular data plot and rose diagram for the Villena wind direction data together with wrapped skew-normal (WSN) and wrapped flexible generalized skew-normal (WFGSN) densities fitted by maximum likelihood estimation.

Also in Figure 4, we plot the fit of the Pewsey wrapped SN distribution. This fit was obtained by maximizing the log-likelihood function for the circular parameters, as suggested by Pewsey (2000a), so as to avoid the known problems of maximum likelihood estimation under the direct parameterization; see, e.g., Azzalini and Capitanio (1999) and Pewsey (2000a). Table 1 presents point estimates and 95% profile confidence intervals for the direct parameters of the WSN and WFGSN distributions.

Table 1. Point estimates and 95% profile confidence intervals (in parenthesis) for the parameters of the WSN and WFGSN distributions fitted to the Villena wind direction data.

	Parameter			
Distribution	ξ	ω	α	β
WSN	2.09	3.06	23.41	—
	(2.04, 2.14)	(2.90, 3.25)	(17.79, 33.28)	
WFGSN	−3.01	1.92	−6.59	16.21
	(−3.04, −2.99)	(1.88, 1.96)	(−6.84, −6.33)	(15.47, 17.00)

For the FGSN distribution, and consequently for the WFGSN distribution, there is no closed forms for centered parameterization, that is, explicit expressions for the cumulants are not available. Therefore, we maximize the log-likelihood function in Equation (5). However, in the case of Villena wind direction data, their distribution was not expected to be symmetric or even similar to a wrapped normal distribution, so that the typical saddle point of the log-likelihood function seems to be far away from the expected maximum, not affecting our estimates. Because the WSN distribution is nested within the WFGSN distribution, a likelihood ratio test can be performed to detect the significance of the extra parameter. The value of the test statistic is 585.3, giving a p-value of approximately zero, indicating that the improvement in fit is highly significant for these data. Formal assessment of goodness-of-fit for these models is complicated due to the complexity of the wrapping procedure, by the uncertainty of the values of the wrapping coefficients (see Coles, 1998), and by the complexity of the SN densities given in Equations (2) and (4).

Table 2. Point estimates and 95% confidence intervals (in parenthesis) for the parameters of the WSN and WFGSN distributions fitted to the sea breeze and mountain breeze subsamples of the the Villena wind direction data.

		Parameter			
Breeze	Distribution	ξ	ω	α	β
Sea	WSN	2.00	2.17	11.70	—
		(1.94, 2.04)	(2.07, 2.27)	(9.21, 15.07)	
Sea	WFGSN	3.20	1.49	−5.02	6.56
		(3.17, 3.23)	(1.45, 1.52)	(−4.81, −5.27)	(6.26, 6.99)
Mountain	WSN	4.49	3.09	32.99	—
		(4.43, 4.55)	(2.84, 3.42)	(17.40, 120.25)	
Mountain	WFGSN	0.15	1.83	−2.94	2.41
		(0.08, 0.21)	(1.78, 1.90)	(−3.12, −2.76)	(2.23, 2.60)

5.2 Sea breeze and mountain breeze

As was discussed in Section 2, winds in Villena are characterized by the phenomena of a sea breeze during daytime and a mountain breeze during nighttime. The specification of these types of breezes for the region under analysis is typical in the analysis of wind direction data; see, e.g., Breckling (1989, Chapter 4). As discussed in Azorin-Molina et al. (2011), in the Villena area (see the map in Figure 1), the Vinalopó River connects the town to the sea in a south-easterly direction, explaining the presence of the typical sea breeze from this direction during the warm hours of the day. In addition, on the north-west/west side of the town, the mountain chain of the Subsistema Prebético creates winds during the colder hours of the day. To allow for these specific directions in our analysis, one possibility is to pursue the logic of other authors (see, e.g., Neumann and Mahrer, 1971; Mak and Walsh, 1976) who fitted separate models for daytime and nighttime. Therefore, we divide the hours in a day in two, namely, between 0:30 and 12:30 (nighttime), where we expect a mountain breeze, and between 12:30 and 00:30 (daytime), where we expect a sea breeze. There were a total of 752 nighttime observations and 1081 daytime observations. Clearly, not only sea and mountain breezes are observed in the considered time spots, so we may expect skewed or even bimodal distributions. For each of the two separate time periods, we fit the wrapped distributions described in Sections 3 and 4. Here too, we compute the maximum likelihood estimates using the Nelder and Mead algorithm. For both time periods, we plot in Figure 5 a rose diagram of the observations within it, as well as the fitted densities for the wrapped SN and WFGSN distributions.

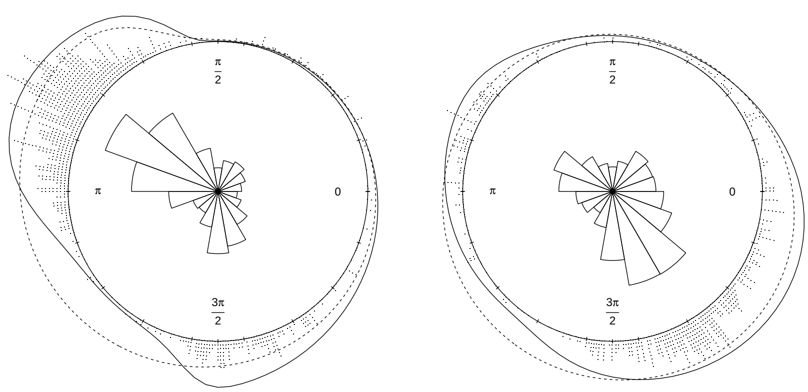

Figure 5. Raw circular data plots and rose diagrams with wrapped skew-normal (dashed) and WFGSN (solid) densities fitted using maximum likelihood estimation to the sea breeze (left) and mountain breeze (right) subsamples of the Villena wind direction data.

Table 2 presents point estimates and 95% profile likelihood based confidence intervals of the parameters for both fitted models in each of the two subsamples.

The parameters of the WFGSN model are estimated much more precisely than the corresponding ones for the WSN, especially if we consider the estimation of α. The likelihood ratio test on the extra parameter β was performed in both cases and gave the values 291.6 and 131.5 for the sea and mountain breeze subsamples, respectively. In both cases, the p-value is approximately 0, indicating that the WFGSN model provides a highly significant improvement in fit over the WSN one.

6. Comparisons with Other Bimodal Models

We compare the fit to our data of the WFGSN distribution and some of the several families of distributions recently proposed as models for circular data. We use the Akaike information criterion (AIC) as the measure with which to conduct this comparison.

Gatto and Jammalamadaka (2007) considered a generalization of the von Mises distribution obtained by re-expressing a particular case of the family of distributions proposed by Maksimov (1967) with density

$$f(\theta) = \frac{1}{2\pi G_0(\delta, \kappa_1, \kappa_2)} \exp\left(\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2)\right),$$

where $0 \leq \theta < 2\pi$, $0 \leq \mu_1 < 2\pi$, $0 \leq \mu_2 < \pi$, $\kappa_1 > 0$, and $\kappa_2 > 0$ are the four parameters of the distribution, and $\delta = (\mu_1 - \mu_2) \mod \pi$. The normalizing constant is

$$G_0(\delta, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta)\right) d\theta.$$

More recently, Kato and Jones (2010) applied the Möbius transformation (see, e.g., Rudin, 1987) to a von Mises random variable, in order to obtain a new random variable with density

$$f(\theta) = \frac{1 - r^2}{2\pi \mathcal{I}_0(\kappa)} \exp\left(\frac{\kappa\{\xi \cos(\theta - \eta) - 2r\cos\nu\}}{1 + r^2 - 2r\cos(\theta - \gamma)}\right) \frac{1}{1 + r^2 - 2r\cos(\theta - \gamma)},$$

where $0 \leq \theta < 2\pi$, $\gamma = \mu + \nu$, $\xi = \sqrt{r^4 + 2r^2\cos(2\nu) + 1}$, $\eta = \mu + \arg\left(r^2\{\cos(2\nu) + i\sin(2\nu)\} + 1\right)$, and $\mathcal{I}_0(\kappa)$ is the modified Bessel function of the first kind and order 0. The four parameters of this distribution are $\mu \geq 0$, $\nu \leq 2\pi$, $0 \leq r < 1$, and $\kappa > 0$. Finally, Abe and Pewsey (2011) proposed a perturbation of a symmetric base density different from the FGSN distribution in Equation (3) by introducing the family of sine-skewed circular distributions. In particular, they suggested using the sine-skewed Jones-Pewsey circular distribution, which can be multimodal and has density

$$f(x) = \frac{(\cosh(\kappa\psi) + \sinh(\kappa\psi)\cos(\theta - \mu))^{1/\psi}(1 + \lambda\sin\theta)}{2\pi P_{1/\psi}(\cosh(\kappa\psi))},$$

where $0 \leq \theta < 2\pi$ and $P_{1/\psi}$ is the associated Legendre function of the first kind, degree $1/\psi$, and order zero. As with the other models, this family has four parameters: $-\pi \leq \mu < \pi$, $\kappa > 0$, $\psi \in \mathbb{R}$, and $-1 \leq \lambda \leq 1$.

The same perturbation was performed on the von Mises distribution, obtaining the so-called sine-skewed von Mises distribution introduced by Umbach and Jammalamadaka

(2009), but so named by Abe and Pewsey (2011), with density

$$f(\theta) = \frac{\exp(\kappa\cos(\theta-\mu))}{2\pi\mathcal{I}_0(\kappa)}[1+\lambda\sin(\theta-\mu)],$$

where $-\pi \leq \theta < \pi$, $\kappa \geq 0$, $-\pi \geq \mu < \pi$, $-1 \leq \lambda \leq 1$, and $\mathcal{I}_0(\kappa)$ is the modified Bessel function of the first kind and order zero. This distribution is a special case of the sine-skewed Jones-Pewsey distribution of Abe and Pewsey (2011), obtained when $\psi = 0$. Note that this model only has three parameters and is seldom bimodal.

We fit these models to the complete set of wind direction data, as in Section 5.1, and to each of the sea breeze and mountain breeze time periods, as in Section 5.2. Figure 6 shows the fitted distributions for the three datasets. It seems clear that different models fit different portions of data well, while all of the models present some degree of lack of fit.

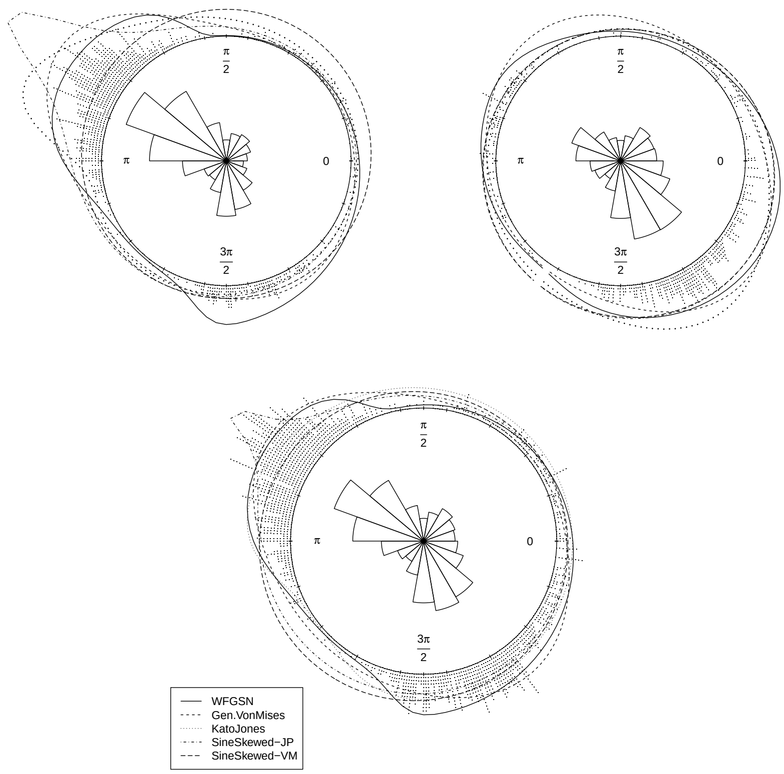

Figure 6. Raw data plots, rose diagrams and fitted densities for the full set of Villena wind direction data (bottom) and its sea breeze (top-left) and mountain breeze (top-right) subsamples

We compare the fits of the different models using the AIC (see Akaike, 1973), based on the maximized value of the log-likelihood function and the number of parameters estimated. Most of the models require the estimation of four parameters, whilst the WSN and sine-skewed von Mises distributions require the estimation of just three. The AIC is calculated as $-2(\log(L)-p)$, where $\log(L)$ is the maximized value of the log-likelihood function and p is the number of parameters to be estimated. The maximized log-likelihood and AIC values are presented in Table 3. For the complete dataset and for the mountain breeze period, the WFGSN distribution seems to fit best the data, while for the sea breeze period, the Jones and Pewsey sine-skewed distribution fits best. However, from a joint consideration of Table 3 and Figure 6, it is evident that, whilst the competing distributions display various attractive features, there is not one of them that dominates overall.

Table 3. Maximized log-likelihood (MLL) and AIC values for six different distributions fitted to the complete dataset of Villena wind direction data and its sea breeze and mountain breeze subsamples. The minimum AIC values for each sample are identified using bold type.

	Complete		Sea breeze		Mountain breeze	
Distribution	MLL	AIC	MLL	AIC	MLL	AIC
WSN	−3106.7	6219.4	−1640.0	3288.0	−1272.2	2550.5
WFGSN	−2814.0	5636.0	−1493.7	2995.3	−1201.6	2411.2
Generalized von Mises	−2919.1	5846.2	−1474.5	2957.3	−1245.6	2499.3
Kato and Jones	−3011.4	6030.8	−1498.1	3004.3	−1227.7	2463.4
Sine-skewed Jones-Pewsey	−3027.3	6062.5	−1434.6	2877.2	−1272.6	2553.2
Sine-skewed von Mises	−3265.2	6536.5	−1953.3	3914.7	−1273.7	2553.3

7. Conclusions

Wrapping the flexible generalized skew-normal distribution around the circle is a natural method of constructing models for bimodal and skewed circular data. In this paper, we have explored how well this model fits wind direction data collected in the Spanish town of Villena. Separate models for daytime and nighttime seem to describe the shape of the observed distribution quite well. We have compared the new model with others capable of modelling the shape of bimodal circular data. A comparison based on the AIC has identified the WFGSN distribution to fit the Villena wind data best. However, some of the other distributions considered have been found to fit some of the features of the data distribution better, so we are unable to conclude that there is one model that fits best globally. As observed in Section 4, wind directions are typically autocorrelated and this was indeed observed for the Villena data. We have considered the possibility of correcting our inferences for this autocorrelation, so as to obtain more reliable confidence intervals and a more accurate p-value for the likelihood ratio test for the improvement in fit of the wrapped generalized flexible skew-normal model over the wrapped skew-normal model. Nevertheless, the observed p-value of the likelihood ratio test without any such correction have been very small, so that we expect that the modified test would not change the final decision, namely that the wrapped generalized flexible skew-normal distribution fits better than the wrapped skew-normal one.

Acknowledgements

The authors thank Arthur Pewsey and Adelchi Azzalini, the editors, and two anonymous referees for their helpful comments and suggestions, and Cesar Azorin-Molina for his climatology advice and map provision. This research was partially supported by grant CPDA097208/09 from the University of Padua, Italy.

References

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

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In Petrov, B.N., Csaki, F., (eds.). Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Azorin-Molina, C., Chen, D., Tijmd, S., Baldie, M., 2011. A multi-year study of sea breezes in a Mediterranean coastal site: Alicante (Spain). International Journal of Climatology, 31, 468–486.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, 171–178.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families. Scandinavian Journal of Statistics, 32, 159–188.

Azzalini, A., Capitanio, A. 1999. Statistical applications of the multivariate skew normal distributions. Journal of The Royal Statistical Society Series B - Statistical Methodology, 61, 579602.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of The Royal Statistical Society Series B - Statistical Methodology, 65, 367–389.

Batschelet, E., 1981. Circular Statistics in Biology. Academic Press, London.

Breckling, J., 1989. The Analysis of Directional Time Series: Applications to Wind Speed and Direction. Springer, London.

Chandler, R.R., Bate, S., 2007. Inference for clustered data using the independence loglikelihood. Biometrika, 94, 167–183.

Coles, S., 1998. Inference for circular distributions and processes. Statistics and Computing, 8, 105–113.

Fernández-Durán, J.J., 2004. Circular distributions based on nonnegative trigonometric sums. Biometrics, 60, 499–503.

Fisher, N.I., 1993. Statistical Analysis of Circular Data. Cambridge University Press, Cambridge.

Fisher, N.I., Lee, A.J., 1994. Time series analysis of circular data. Journal of The Royal Statistical Society Series B - Statistical Methodology, 56, 327–339.

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

Genton, M.G., 2004. Skew-symmetric and generalized skew-elliptical distributions. In Genton, M.G., (ed.). Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman & Hall / CRC, Boca Raton, FL, pp. 81–100.

Jammalamadaka, S.R., Kozubowski, T.J., 2003. A new family of circular models: the wrapped Laplace distributions. Advances and Applications in Statistics, 3, 77–103.

Jammalamadaka, S.R., SenGupta, A., 2001. Topics in Circular Statistics. World Scientific, Singapore.

Kato, S., Jones, M.C., 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.

Ma, Y., Genton, M.G., 2004. A flexible class of skew-symmetric distributions. Scandinavian Journal of Statistics, 31, 459–468.

Mak, M.K., Walsh, J.E., 1976. On the relative intensities of land and sea breezes. Journal of the Athmospheric Sciences, 33, 242–351.

Maksimov, V.M., 1967. Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups. Theoria Verojatna, 12, 307–321.

Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, London.

Mardia, K.V., Jupp, P.E., 1999. Directional Statistics. Academic Press, London.

Mazzuco, S., Scarpa, B., 2011. Fitting age-specific fertility rates by a skew-symmetric probability density function. Working Paper. Department of Statistical Sciences, University of Padua, Italy. `http://www.stat.unipd.it/ricerca/fulltext?wp=436`

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Computer Journal, 7, 308–313.

Neumann, J., Mahrer, Y., 1971. A theoretical study of the land and sea breeze circulation. Journal of the Athmospheric Sciences, 28, 532–542.

Pace, L., Salvan, A., Sartori, N., 2011. Adjusting composite likelihood ratio statistics.

Statistica Sinica, 21, 129–148.
Pewsey, A., 2000a. The wrapped skew-normal distribution on the circle. Communications in Statistics - Theory and Methods, 29, 2459–2472.
Pewsey, A., 2000b. Problems of inference for Azzalini's skew-normal distribution. Journal of Applied Statistics, 27, 859–870.
Pewsey, A., 2006. Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. Environmental and Ecological Statistics, 13, 257–269.
Pewsey, A., 2008. The wrapped stable family of distributions as a flexible model for circular data. Computational Statistics and Data Analysis, 52, 1516–1523.
Rao, J.S., 1969. Some contributions to the analysis of circular data. Ph.D. Dissertation. Indian Statistical Institute, India.
Rudin, W., 1987. Real and Complex Analysis. Third edition. McGraw-Hill, New York.
Umbach, D., Jammalamadaka, S.R., 2009. Building asymmetry into circular distributions. Statistics and Probability Letters, 79, 659–663.
Wang, J., Boyer, J., Genton, M.G., 2004. A skew-symmetric representation of multivariate distributions. Statistica Sinica, 14, 1259–1270.