Afrika Matematika (2025) 36:86
https://doi.org/10.1007/s13370-025-01306-9

A Power-Cardioid candidate for wind direction modelling motivated by two South African case studies

Delene van Wyk-de Ridder[1,2] · Najmeh Nakhaei Rad[1] · Mohammad Arashi[1,3] · Johan Ferreira[1] · Andriette Bekker[1,4]

Received: 13 January 2023 / Accepted: 13 March 2025

Abstract
Wind energy claims a positive image globally; therefore, accurate modelling of wind direction at generation sites accurately can enhance the potential of this green energy source. The uncertain nature of wind direction can be modelled through probability distributions; in this paper, we propose a flexible yet simple distribution, namely the Power-Cardioid distribution, as an alternative and implementable candidate to model wind direction. After discussing some characteristics, the performance of the Power-Cardioid distribution is evaluated via a simulation study and applied to datasets of two wind farms in South Africa. The numerical results demonstrate that this distribution is a promising and exciting new candidate compared to well-known models within circular statistics.

Keywords Generator · Kato and Jones distribution · Mixture distribution · von Mises distribution · Wind direction

Mathematics Subject Classification 60E05 · 62E17 · 62F10 · 62H11

✉ Johan Ferreira
johan.ferreira@up.ac.za

Delene van Wyk-de Ridder
delene.vanwyk@uct.ac.za

Najmeh Nakhaei Rad
najmeh.nakhaeirad@up.ac.za

Mohammad Arashi
arashi@um.ac.ir

Andriette Bekker
andriette.bekker@up.ac.za

1 Department of Statistics, University of Pretoria, Pretoria, South Africa

2 Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

3 Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

4 Department of Geography, Geoinformatics and Meteorology, Centre for Environmental Studies, University of Pretoria, Pretoria, South Africa

Published online: 18 April 2025

1 Introduction

Harvesting wind energy can solve ongoing energy shortfalls, for example in South Africa due to the ongoing loadshedding (rolling power blackout) crisis. Wind energy is one of the fastest-growing energy sources in the world. This is partly due to its cost-effectiveness; indeed, the land-based utility-scale wind is one of the lowest-priced energy sources available today. South Africa has great potential for considering wind energy as an alternative resource, especially along the coastal areas of the Eastern and Western Capes. It is projected that by 2030, 22.7% of the required electricity in South Africa will be generated from wind energy. However, the success of any wind energy program lies in the accuracy of the assessment of wind flow patterns based on available data at a potential site for a wind farm. As [1] highlighted, wind statistics is the science that describes the patterns of the wind regime over a considerable period. Since the 1940 s, numerous papers have been published proposing and implementing different continuous probability distributions to describe wind speed, including the Weibull, gamma, normal, Rayleigh, log-normal, inverse Gaussian and logistic distributions as commonly considered models [2–4]. What makes wind different from other climate variables (e.g., temperature and precipitation) is its inherent nature, which necessitates investigating wind direction and not only wind speed. An accurate statistical analysis of the wind direction can provide important information about changes in the amount of wind at a specific location. Considering direction in this context naturally gives rise to the consideration and extension of suitable parametric directional probability distributions to better understand the directional nature of this essential energy factor.

Because most parametric inferential approaches for directional data are obtained from only a handful of proposed models, there are still ongoing discussions surrounding the development of approaches that will increase the robustness of most existing directional models [5]. Methods for deciphering directional data have been developed towards characterizing population distributions from which circular data may arise, and most of the tools that are generally implemented rely heavily on the von Mises- (vM), Cardioid-, and wrapped Cauchy distributions [5]. As a result, many existing methods appropriate for directional statistics are parametric and reliant on restricted distributional assumptions, such as having a unimodal and symmetric spread.

In this paper, we model the wind direction at two operational wind farms in South Africa: Jeffreys Bay (Humansdorp), located in the Eastern Cape (onshore wind farm) and Noupoort, located in the Northern Cape (offshore wind farm). Interested readers may be referred to studies on wind direction such as [6–11]. The contribution of this paper is not limited to the modelling of wind direction at two South African wind farms, but includes the introduction of a new two-parameter distribution, namely the Power-Cardioid ($\mathcal{P}-\mathcal{C}$) distribution which emanates from a proposed Cardioid-generator class. Due to the nature of wind direction, finite mixture distributions are often considered for modelling this type of heterogeneous wind data. For example, mixtures of von Mises (vM) have been widely implemented to model wind direction for different locations [12–14]. Finite mixtures of the proposed $\mathcal{P}-\mathcal{C}$ distribution are therefore of particular interest and are applied and compared to other often considered candidates when wind direction is under consideration.

First, we provide details of the datasets in Sect. 2 that form part of the performance investigation of the new distribution. In Sect. 3, we focus on the definition of the Cardioid-generator class of distributions, specifically the $\mathcal{P}-\mathcal{C}$ distribution and some of its characteristics, along with the estimation of its parameters. A simulation study is conducted for parameter recovery on synthetic data in Sect. 4, and the impact of this new distribution is investigated by its

Fig. 1 Jeffreys Bay wind farm https://jeffreysbaywindfarm.co.za (left) and Noupoort wind farm https://noupoortwind.co.za (right)

well-known circular models using the two South African wind datasets in Sect. 5. Section 6 contains final thoughts and conclusions.

2 Site location and wind data

The first dataset reflects the wind direction of the Jeffreys Bay wind farm recorded every 10 minutes at 20 metres height between 2017 − 2019. Jeffreys Bay (Humansdorp) wind farm, located in the Eastern Cape, is one of the largest wind farms in South Africa, spanning 3700 hectares with 138 MW capacity. The second dataset shows the wind direction of the Noupoort wind farm recorded every 10 minutes at 60 metres height between 2017 − 2019. Noupoort is located in the Northern Cape, comprising 7500 hectares and providing an 80 MW capacity. Figure 1 shows aerial views of these two wind farms, respectively. Rose plots of the datasets in Fig. 2 confirm that these wind direction datasets reveal multimodal patterns, whereas Table 1 reports the respective datasets' summary statistics.

3 A Cardioid-generator class

3.1 Class specification

The Cardioid distribution is mainly used as a small-concentration approximation to the von Mises (vM) distribution and was introduced by [15] with the probability density function (pdf):

$$f(\theta) = \frac{1}{2\pi}(1 + \rho\cos(\theta - \mu)), \quad 0 < \theta \leq 2\pi, \tag{1}$$

where $0 \leq \rho < 1$ and $0 < \mu \leq 2\pi$. When ρ is restricted to be non-negative, there can be no difficulties resulting from non-identifiability. Inspired by (1) and eq. 2.20 of [16], a more flexible class of distributions for circular data is proposed in Proposition 1, which we call the Cardioid-generator class of distributions.

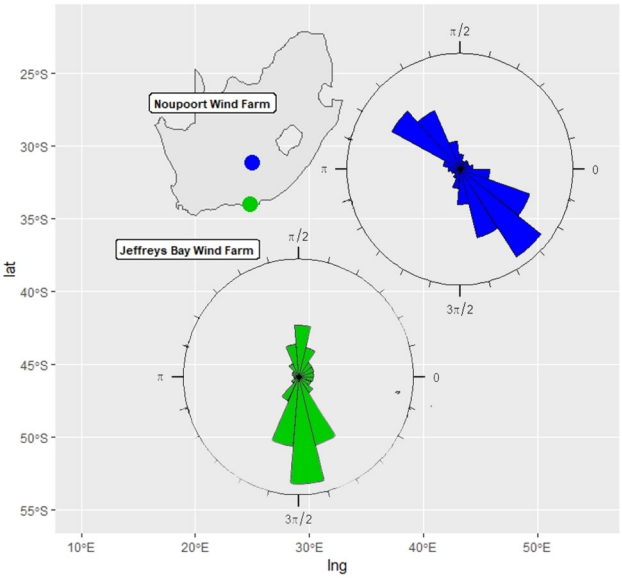


Fig. 2 Map of South Africa with the locations of Jeffreys Bay and Noupoort wind farms and rose plots of the wind direction (created by R programming language version 4.1.3 https://www.r-project.org)

Table 1 Descriptive statistics for the wind direction data

Location	Duration	n	Mean	Variance	Mean resultant length	Skewness	Kurtosis
Jeffreys Bay	2017–2019	166,315	5.1212	0.7223	0.2777	−0.6020	0.6899
Noupoort	2017–2019	163,526	5.4571	0.8222	0.1778	−0.0373	0.8093

Proposition 1 *The distribution of a random variable* Θ *belongs to the class of symmetric distributions on the circle (here called Cardioid-generator class) if it has a pdf of the following form:*

$$f(\theta) = c.g(\cos(\theta - \mu)), \quad 0 < \theta \leq 2\pi, \tag{2}$$

where $g(\cdot)$ *is a Borel measurable function which admits a Taylor series expansion and* $g(\cdot) \neq 1$. *Here,* $g(\cdot)$ *is absolutely continuous and may be referred to as a generator function with a normalizing constant c.*

In this case, it is clear that the generator function, $g(\cos(\theta - \mu))$, relates to the Cardioid distribution in (1) through

$$g(\cos(\theta - \mu)) = (1 + \rho \cos(\theta - \mu)).$$

Subsequently, a general formula for the appropriate normalizing constant for (2) is obtained. The generalized formula proposed below only requires the derivatives of the function $g(cos(\theta - \mu))$ evaluated at 0 in each case. To arrive at such a formula, the definition of a probability distribution of an absolutely continuous distribution must be used [5]. Since

$$\int_0^{2\pi} f(\theta) \mathrm{d}\theta = 1,$$

it follows that

$$c = \left(\int_0^{2\pi} g(\cos(\theta - \mu)) \mathrm{d}\theta \right)^{-1}. \tag{3}$$

For the function $g(\cos(\theta - \mu))$, the Maclaurin series can be written as

$$g(\cos(\theta - \mu)) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} \cos^k(\theta - \mu), \tag{4}$$

where $g^{(k)}(0)$ represents the k^{th} derivative of $g(\cdot)$ around zero. Note that this series can also be represented as

$$\begin{aligned} g(\cos(\theta - \mu)) &= \tfrac{g^{(0)}(0)}{0!} \cos^0(\theta - \mu) + \textstyle\sum_{k=1}^{\infty} \tfrac{g^{(k)}(0)}{k!} \cos^k(\theta - \mu) \\ &= g(0) + \textstyle\sum_{k=1}^{\infty} \tfrac{g^{(k)}(0)}{k!} \cos^k(\theta - \mu), \end{aligned}$$

which will be referred to for a specific distribution under consideration. Substituting (4) into (3), it follows that

$$\begin{aligned} c &= \left(\textstyle\int_0^{2\pi} \left(\sum_{k=0}^{\infty} \tfrac{g^{(k)}(0)}{k!} \cos^k(\theta - \mu) \right) \mathrm{d}\theta \right)^{-1} \\ &= \left(\textstyle\sum_{k=0}^{\infty} \tfrac{g^{(k)}(0)}{k!} \int_0^{2\pi} \cos^k(\theta - \mu) \mathrm{d}\theta \right)^{-1}. \end{aligned} \tag{5}$$

Two approaches to compute the normalising constant in (5) are discussed in Remarks 1 and 2 respectively.

Remark 1 Using [17] for even values of k and $(k-1)!! = (k-1)(k-3)(k-5)\ldots 1$, the integral in (5) can be evaluated at 0 and 2π to solve for

$$\begin{aligned} \textstyle\int_0^{2\pi} \cos^k(\theta - \mu) \mathrm{d}\theta = \tfrac{\sin(\theta-\mu)}{k} \Bigg(& \cos^{k-1}(\theta - \mu) + \textstyle\sum_{b=1}^{\frac{k}{2}-1} \tfrac{(k-1)(k-3)\ldots(k-2b+1)}{2^b(\frac{k}{2}-1)(\frac{k}{2}-2)\ldots(\frac{k}{2}-b)} \\ & \times \cos^{(k-2b-1)}(\theta - \mu) \Bigg) + \tfrac{(k-1)!!}{2^{\frac{k}{2}} \frac{k}{2}!} (\theta - \mu). \end{aligned}$$

For odd values of k, the integral the integral in (5) can be evaluated at 0 and 2π to solve for

$$\begin{aligned} \textstyle\int_0^{2\pi} \cos^k(\theta - \mu) \mathrm{d}\theta = \tfrac{\sin(\theta-\mu)}{k} \Bigg(& \cos^{k-1}(\theta - \mu) + \textstyle\sum_{b=0}^{\frac{k-1}{2}-1} \tfrac{2^{b+1}(\frac{k-1}{2})(\frac{k-1}{2}-1)\ldots(\frac{k-1}{2}-b)}{(k-2)(k-4)\ldots((k-1)-2b-1)} \\ & \times \cos^{((k-1)-2b-2)}(\theta - \mu) \Bigg). \end{aligned}$$

Remark 2 Another approach to obtain c in (5) could be to make use of Euler's formula [17]. Instead of substituting $\cos^k(\theta - \mu)$ in (4), Euler's formula can be included in the integral to find

$$\begin{aligned} \textstyle\int_0^{2\pi} f(\theta) \mathrm{d}\theta &= c \textstyle\sum_{k=0}^{\infty} \tfrac{g^{(k)}(0)}{k!} \int_0^{2\pi} \left(\tfrac{\exp(i(\theta-\mu)) + \exp(-i(\theta-\mu))}{2} \right)^k \mathrm{d}\theta \\ &= \tfrac{c}{2^k} \textstyle\sum_{k=0}^{\infty} \tfrac{g^{(k)}(0)}{k!} \int_0^{2\pi} (\exp(i(\theta - \mu)) + \exp(-i(\theta - \mu)))^k \mathrm{d}\theta \\ &= \tfrac{c}{2^k} \textstyle\sum_{k=0}^{\infty} \tfrac{g^{(k)}(0)}{k!} \int_0^{2\pi} \exp(-ik(\theta - \mu))(\exp(2i(\theta - \mu)) + 1)^k \mathrm{d}\theta. \end{aligned}$$

3.2 Pdf and some characteristics

For our specific $\mathcal{P} - \mathcal{C}$ candidate, we assume a power function of the form $g(\cos(\theta - \mu) = (1 + w.\cos(\theta - \mu))^{-m}$ for some constants $w, m > 0$ for consideration in (2). The following proposition describes this proposed candidate.

Proposition 2 *A random variable Θ is distributed as a $\mathcal{P}-\mathcal{C}$ distribution if it has a pdf of the following form:*

$$f_{\mathcal{P}-\mathcal{C}}(\theta) = c\left((1+\frac{\kappa}{m}\cos(\theta-\mu))\right)^{-m}, \quad 0<\theta\leq 2\pi, \tag{6}$$

where $0\leq\kappa<m$ and m is a positive integer.

The normalizing constant c in (6) can then be obtained as

$$c^{-1} = \sum_{i=0}^{\infty}\binom{m+i-1}{i}\left(\frac{-\kappa}{m}\right)^{i}\int_0^{2\pi}\cos^{i}(\theta-\mu)\mathrm{d}\theta. \tag{7}$$

The pdf (6) is illustrated in Fig. 3a for $m=2$ while varying κ, in Fig. 3b for $m=3$ while varying κ, and in Fig. 4a and 4b for $\kappa=2$ while varying m. In particular, Fig. 4b highlights the added value of m for capturing peakedness of data (when m is small) compared to elevated levels of kurtosis (when m is large).

The following proposition highlights three key characteristics from the proposed $\mathcal{P}-\mathcal{C}$ distribution with pdf in (6).

Proposition 3 *If a random variable Θ follows the $\mathcal{P}-\mathcal{C}$ distribution with pdf in (6), then:*

1. *The p^{th} trigonometric moment of Θ is given by*

$$\varphi_p = ce^{ip\mu}\sum_{j=0}^{\infty}\binom{m+j-1}{j}\left(\frac{-\kappa}{m}\right)^{j}\int_0^{2\pi}\cos(p\theta)\cos^{j}(\theta)d\theta, \quad p=0,\pm1,\pm2,\ldots,$$

2. *The distribution in (6) is unimodal with the maximum (thus, the mode) at $\theta=\mu+\pi$.*
3. *The pdf (6) is symmetric about μ, that is, $f_{\mathcal{P}-\mathcal{C}}(\mu+\theta)=f_{\mathcal{P}-\mathcal{C}}(\mu-\theta)$.*

Regarding point 2 in Proposition 3, the first derivative of the kernel of (6) is determined to be $\kappa\sin(\theta-\mu)\left(1+\frac{\kappa}{m}\cos(\theta-\mu)\right)^{-m-1}$, which is equal to 0 if and only if $\theta=\mu+\pi$, indicating the presence of a single global maximum. Since the support is $0<\theta\leq 2\pi$, this implies only one maximum point. Following from point 3 in Proposition 3, it is clear that the skewness coefficient is zero. The circular kurtosis coefficient of a variable with a $\mathcal{P}-\mathcal{C}$ distribution for $\mu=0$ (without loss of generality) is obtained from [18] as follows:

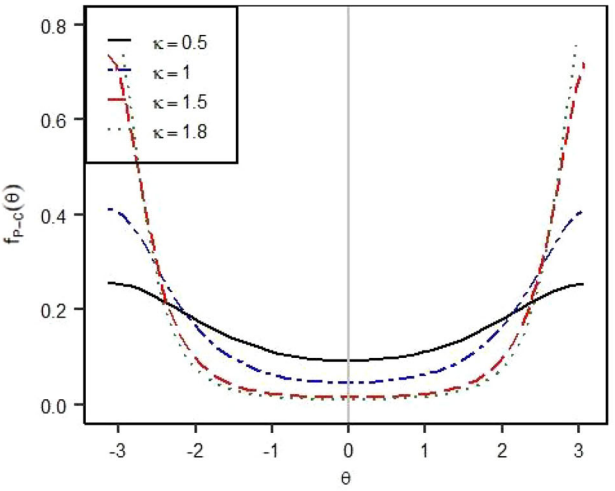
$$\gamma_2 = \frac{(\alpha_2-\alpha_1^4)}{(1-\alpha_1)^2}, \tag{8}$$

where

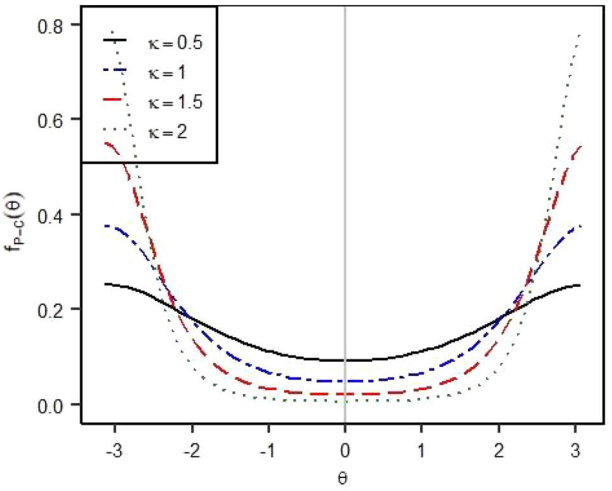
$$\alpha_1 = c\sum_{j=0}^{\infty}\binom{m+j-1}{j}\left(\frac{-\kappa}{m}\right)^{j}\int_0^{2\pi}\cos^{j+1}(\theta)\mathrm{d}\theta,$$

$$\alpha_2 = c\sum_{j=0}^{\infty}\binom{m+j-1}{j}\left(\frac{-\kappa}{m}\right)^{j}\left[2\int_0^{2\pi}\cos^{j+2}(\theta)\mathrm{d}\theta-\int_0^{2\pi}\cos^{j}(\theta)\mathrm{d}\theta\right].$$

Figure 5 indicates how the kurtosis coefficient (8) converges for relatively small values of j, depending on the chosen parameter values.



(a) For $m = 2$ and different values of κ



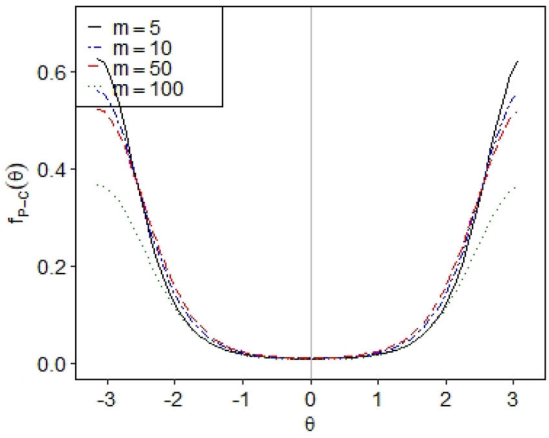
(b) For $m = 3$ and different values of κ

Fig. 3 Pdf of the $\mathcal{P} - \mathcal{C}$ distribution (6) for different parameter values

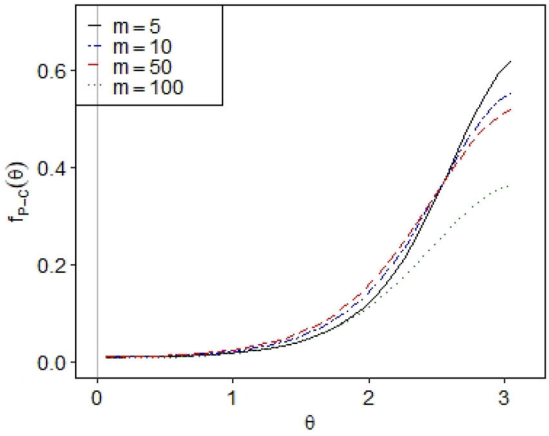
3.3 Parameter estimation

For a random sample of size n from the $\mathcal{P} - \mathcal{C}$ distribution in (6) $\theta_1, \theta_2, \ldots, \theta_n$, the log-likelihood function can be expressed as

$$\ln L(\mu, \kappa|\theta) = n \ln(c) - m \sum_{i=1}^{n} \ln\left(1 + \frac{\kappa}{m} \cos(\theta_i - \mu)\right). \tag{9}$$



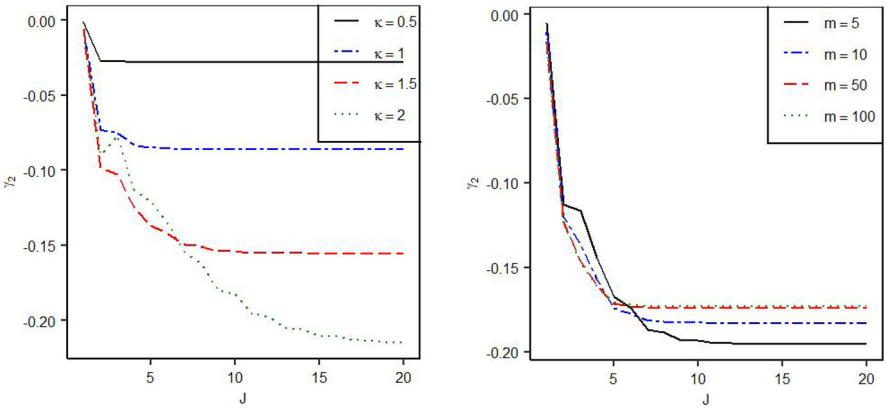
(a) For $\kappa = 2$ and different values of m



(b) For $\kappa = 2$ and different values of m: magnified view

Fig. 4 Pdf of the $\mathcal{P} - \mathcal{C}$ distribution (6) for different parameter values

By equating the partial derivatives of the loglikelihood function with respect to the parameters to zero, the maximum likelihood estimates (MLEs) can be obtained. Since no closed form exists, numerical methods are used to obtain the estimates. The `DEoptim` package [19] in `R` software which is based on the Differential Evolution (DE) algorithm [20] is used to obtain the MLEs. Its significant performance as a global optimization algorithm on continuous numerical minimization problems has been extensively studied [21].



(a) Plot of the converging kurtosis when $\mu = 0$ and $m = 3$

(b) Plot of the converging kurtosis when $\mu = 0$ and $\kappa = 2$

Fig. 5 Plots of the converging kurtosis coefficient (8) for different parameter values

4 Simulation study

Simulation studies were conducted to assess the accuracy of the selected approach for obtaining the MLEs of parameters of the $\mathcal{P} - \mathcal{C}$ distribution. To this effect, we generated 200 samples of size $n = 30, 100, 500$ from the $\mathcal{P} - \mathcal{C}$ distribution with $\mu = 3$, $\kappa = 0.5$ and $m = 1, 2, 3, 4, 5$. The biases and the mean squared errors (MSEs) of ML,Es are calculated with the results summarized in Table 2. The MLE values in Table 2 represent the average of the obtained MLEs across 200 iterations. From Table 2 we observe that the values of bias and MSE decrease as n increases consistently across different values of m.

The box plots in Fig. 6 a) and b) show the shrinking bias when estimating μ (left) and κ (right) for $n = 30, 100, 500$, whereas c) and d) show the shrinking MSE when estimating μ (left) and κ (right) for the same sample sizes. As the sample size increases, both the bias and MSE tend towards zero.

Table 3 reflects the value of the log-likelihood (9) for 2000 samples of size $n = 30$ when $m = 1, 3, 5$ (rows) for fixed $\mu = 3$ and $\kappa = 0.5$. Each sampled set was fitted via maximum likelihood estimation to the $\mathcal{P} - \mathcal{C}$ distribution (6) for with varying $m = 1, 3, 5$ in order to assess the behaviour of best fit of the model when m differs from the "true" m used in each sampling process. The largest loglikelihood (or, equivalently, smallest Akaike Information Criterion - AIC - or Bayesian Information Criterion - BIC) value indicates a better fit (in bold). In this case, we observe that the true value of m yields the correct result in the fit in each case. Different $\mathcal{P} - \mathcal{C}$ models were fitted to the sampled data, varying for $m = 1, 3, 5$. Such an investigation can guide the practitioner to determine a suitable value of m which maximizes the loglikelihood (or, effectively, minimises the AIC or BIC) in practice.

5 Real data

To demonstrate the performance of the $\mathcal{P} - \mathcal{C}$ distribution (6) for the wind direction data from the South African wind farms, we analyze the two real datasets of Sect. 2 (see Table 1) and conduct a model performance comparison with the well-known vM distribution, and

Table 2 MLEs and corresponding biases and MSEs for the $\mathcal{P}-\mathcal{C}$ distribution

$\mathcal{P}-\mathcal{C}$		Actual value	m	MLE	Bias	MSE
$n=30$	μ	3	1	3.1389	0.1389	1.1113
	κ	0.5		0.7170	0.2170	0.0968
			2	3.30932	0.0932	1.2562
				0.8662	0.3662	0.3255
			3	3.0976	0.0960	1.2718
				0.9446	0.4446	0.4991
			4	3.0976	0.0976	1.2813
				0.9847	0.4847	0.6164
			5	3.0986	0.0986	1.2875
				1.0086	0.5086	0.7018
$n=100$	μ	3	1	3.0892	0.0892	0.5378
	κ	0.5		0.5647	0.0647	0.0409
			2	3.0835	0.0835	0.5378
				0.6642	0.1642	0.1057
			3	3.0809	0.0809	0.5544
				0.6642	0.1642	0.1057
			4	3.0796	0.0796	0.5567
				0.6713	0.1713	0.1125
			5	3.0788	0.0788	0.5582
				0.6746	0.1746	0.1159
$n=500$	μ	3	1	2.9997	−0.0003	0.0863
	κ	0.5		0.4906	−0.0094	0.0122
			2	2.9986	−0.0014	0.0817
				0.5359	0.0359	0.0198
			3	2.9985	−0.0015	0.0804
				0.5460	0.0460	0.0223
			4	2.9984	−0.0016	0.0797
				0.5498	0.0498	0.0233
			5	2.9984	−0.0016	0.0794
				0.5516	0.0516	0.0237

Table 3 Log-likelihood values (9) for 2000 samples of size $n=30$ when $m=1$, 3 and 5 for fixed $\mu=3$ and $\kappa=0.5$

	m	Model with specified m		
		1	3	5
Chosen m for sampling	1	**−50.8275**	−51.5562	−51.6347
	3	−47.3465	**−47.1067**	−47.1247
	5	−53.7567	−53.5197	**−53.4604**

Table 4 MLEs and corresponding values of the loglikelihood, AIC and BIC for the Jeffreys Bay wind farm dataset

Model	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\tilde{\alpha}}_2$	$\hat{\tilde{\beta}}_2$	$\hat{w}_i$	Log-likelihood	AIC	BIC
Mixture of vM ($M = 2$)	1.4676	1.2611	–	–	0.4533	−245040.8	490091.6	490141.7
	7.6259	4.7698	–	–	0.5467			
Mixture of vM ($M = 3$)	19.9322	1.5327	–	–	0.1386			
	0.6634	0.7691	–	–	0.3638	−239795.1	479606.2	479686.3
	9.3181	4.7613	–	–	0.4976			
Mixture of vM ($M = 4$)	9.2694	4.7618	–	–	0.4991			
	0.8151	0.7921	–	–	0.1901	−239794.9	479611.7	479722.0
	0.5379	0.7908	–	–	0.1748			
	20.7177	1.5341	–	–	0.1360			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 2$)	0.9547	1.6489	–	–	0.6631	−243287.5	486584.9	486635.1
	0.9371	4.6127	–	–	0.3369			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 3$)	0.9653	4.6666	–	–	0.2724			
	0.9624	1.5773	–	–	0.6369	−239768.1	479552.3	479632.5
	0.8975	3.4969	–	–	0.0907			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 4$)	0.9939	1.9455	–	–	0.0860			
	0.9669	4.6781	–	–	0.2601	−237170.6	**474,363.1**	**474,473.4**
	0.8746	3.5433	–	–	0.1379			
	0.9676	1.5767	–	–	0.5160			

Table 4 continued

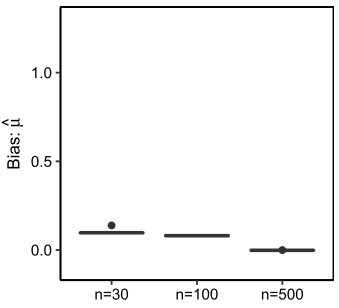
Model	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{w}_i$	Log-likelihood	AIC	BIC
Mixture of KJ ($M = 2$)	0.7972	4.8479	0.5790	−0.0689	0.6154	−241430.4	482878.8	482969.1
	0.6123	1.3794	0.4203	0.0822	0.3846			
Mixture of KJ ($M = 3$)	0.4910	4.5436	0.3612	0.2184	0.2408			
	0.7960	1.3424	0.5706	0.1471	0.2633	−240220.5	480468.9	480609.2
	0.7158	4.9520	0.5020	−0.1862	0.5041			
Mixture of KJ ($M = 4$)	0.7711	1.3222	0.5656	0.1674	0.2450			
	0.4854	4.4479	0.3157	0.2361	0.2097	−240481.3	481000.7	481191.1
	0.5446	4.6553	0.4273	−0.2093	0.0451			
	0.6520	4.9791	0.4672	−0.1804	0.5002			

Table 5 MLEs and corresponding values of the loglikelihood AIC and BIC for the Noupoort wind farm dataset

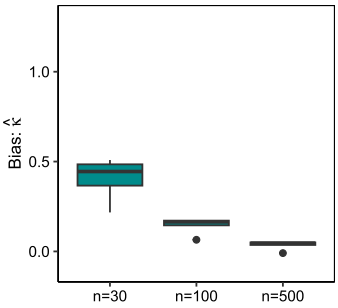
Model	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\tilde{\alpha}}_2$	$\hat{\tilde{\beta}}_2$	$\hat{w}_i$	Log-likelihood	AIC	BIC
Mixture of vM ($M = 2$)	3.0801	5.4363	–	–	0.6226	−246268.0	492546.0	492596.0
	4.3314	2.2836	–	–	0.3774			
Mixture of vM ($M = 3$)	5.6456	5.4269	–	–	0.5045			
	0.5644	2.1033	–	–	0.2880	−239767.9	479551.8	479631.8
	22.7218	2.3287	–	–	0.2075			
Mixture of vM ($M = 4$)	2.3274	2.1558	–	–	0.2133			
	1.5909	5.3446	–	–	0.2960			
						−239011.8	478045.5	478155.6
	34.6464	2.3536	–	–	0.1604			
	8.7451	5.4612	–	–	0.3303			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 2$)	0.9334	2.2993	–	–	0.6539			
						−239978.0	479966.0	480016.0
	0.9748	5.4684	–	–	0.3461			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 3$)	0.9717	5.4673	–	–	0.3589			
	0.9900	2.7039	–	–	0.0675	−239375.2	478766.4	478846.4
	0.9414	2.2349	–	–	0.5736			
Mixture of $\mathcal{P} - \mathcal{C}$ ($M = 4$)	0.9455	2.2257	–	–	0.5401			
	0.9354	5.2191	–	–	0.1249			
						−238877.8	477777.6	477887.7
	0.9860	2.6640	–	–	0.0906			
	0.9851	5.5182	–	–	0.2444			

Table 5 continued

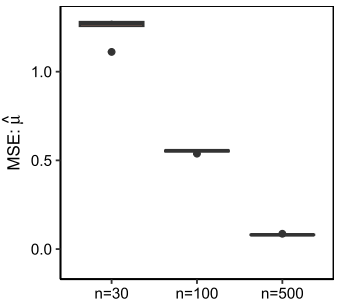
Model	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{w}_i$	Log-likelihood	AIC	BIC
Mixture of KJ ($M = 2$)	0.7579	2.2064	0.5827	0.0989	0.3894			
						−239290.6	478599.3	478689.3
	0.7736	5.3831	0.5088	0.0501	0.6106			
Mixture of KJ ($M = 3$)	0.7563	2.2002	0.5926	0.1076	0.3673			
	0.7245	5.3271	0.5146	0.0311	0.5176	−238792.6	**477,613.2**	**477753.3**
	0.7238	5.6180	0.5878	0.1527	0.1151			
Mixture of KJ ($M = 4$)	0.6990	2.1039	0.5411	0.1649	0.2894			
	0.5145	5.4610	0.4236	0.1838	0.0592			
						−239099.8	478237.6	478427.7
	0.7677	2.5643	0.5249	−0.1661	0.1056			
	0.7637	5.4149	0.5034	−0.0016	0.5458			



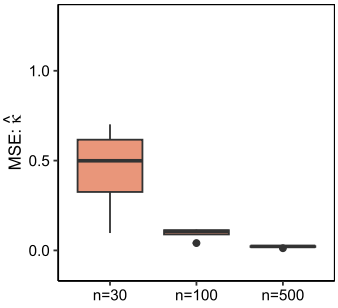
(a) Bias of estimating μ for different sample sizes



(b) Bias of estimating κ for different sample sizes



(c) MSE of estimating μ for different sample sizes



(d) MSE of estimating κ for different sample sizes

Fig. 6 Boxplots of simulation study for bias and MSE for different sample sizes

Kato and Jones (KJ) distribution [22] with the following pdf:

$$f_{KJ}(\theta) = \frac{1}{2\pi}\left(1 + 2\kappa^2 \frac{\kappa\cos(\theta-\mu) - \bar{\alpha_2}}{\kappa^2 + \bar{\alpha_2}^2 + \bar{\beta_2}^2 - 2\kappa\left(\bar{\alpha_2}\cos(\theta-\mu) + \bar{\beta_2}\sin(\theta-\mu)\right)}\right), \quad (10)$$

where $0 < \theta \leq 2\pi, 0 \leq \kappa < 1, \bar{\alpha_2} \neq \kappa, \bar{\beta_2} \neq 0$ and $(\bar{\alpha_2} - \kappa^2)^2 + \bar{\beta_2}^2 \leq \kappa^2(1-\kappa)^2$.

Consider a set of pdfs $f_1(\theta), \cdots, f_M(\theta)$ and weights $w_1, \cdots, w_M$ such that $w_i \geq 0$ and $\sum_{i=1}^{M} w_i = 1$. A finite mixture distribution $f(\theta)$ is then represented by

$$f(\theta) = \sum_{i=1}^{M} w_i f_i(\theta). \quad (11)$$

Due to the multimodal pattern of the datasets observed in Fig. 2 a finite mixture approach is proposed for the illustration of the results. The following distributions were considered:

- Finite mixture of vM distributions (see [5]) with $M = 2, 3, 4$ components,
- Finite mixture of $\mathcal{P} - \mathcal{C}$ distributions (6) with $M = 2, 3, 4$ components, and

Table 6 Execution time for obtaining MLEs for finite mixtures of vM-, $\mathcal{P}-\mathcal{C}$-, and KJ distributions

Distribution	$M=2$	$M=3$	$M=4$
vM	1084.044	4841.181	9149.872
$\mathcal{P}-\mathcal{C}$	1141.127	5291.676	9532.749
KJ	49421.678	144756.621	275405.372

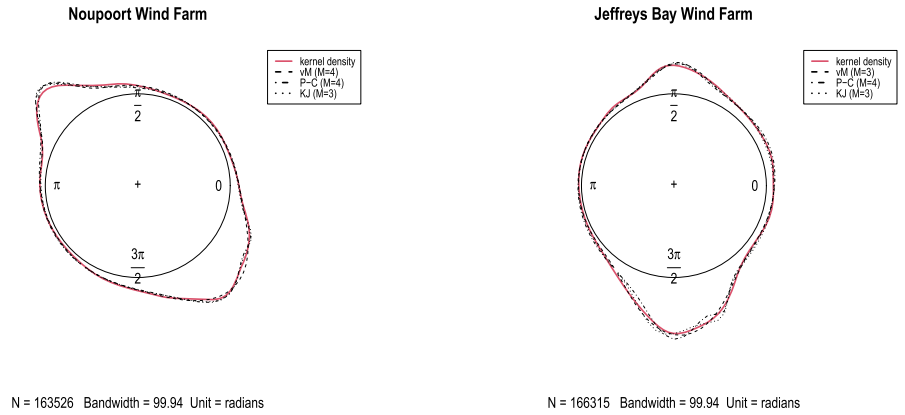


Fig. 7 Kernel density plots of datasets and fitted curves

- Finite mixture of KJ distributions (10) with $M=2,3,4$ components.

The MLEs of parameters for each distribution were obtained using the `DEoptim` package in `R`. The AIC, given by $AIC=2k-2\ln(\hat{L})$ and the BIC, given by $BIC=k\ln(n)-2\ln(\hat{L})$, where k is the total number of parameters estimated, $\hat{L}$ is the maximized value of the likelihood function and n is the number of data points, were used for model selection. The MLEs and corresponding loglikelihood, AIC, and BIC results are reported in Tables 4 and 5. Here, a model with the maximum loglikelihood and minimum values of AIC and BIC provides a better fit for the data.

For the first dataset from Jeffreys Bay wind farm, the mixture of $\mathcal{P}-\mathcal{C}$ distributions with $m=1$ and $M=4$ provides the best fit, followed by a mixture of vM distributions as the second best model. For the second dataset from the Noupoort wind farm, the mixture of KJ distributions with $M=3$ provides the best fit. In this case, the mixture of $\mathcal{P}-\mathcal{C}$ distributions is the second best model with $M=4$. It seems the mixture of $\mathcal{P}-\mathcal{C}$ distributions has a better performance for modelling wind direction in the offshore Noupoort wind farms, while the mixture of KJ distributions provides a better fit for wind direction in the onshore Jeffreys Bay wind farms. The execution time for obtaining MLEs for mixtures of vM-, $\mathcal{P}-\mathcal{C}$-, and KJ distributions are reported in Table 6. The mixture of KJ distributions has a higher running time, which could be attributed to the constraints for parameters and also the higher number of parameters; this particularly highlights the added value of considering the $\mathcal{P}-\mathcal{C}$ distributions as a viable candidate in this environment. The kernel density plots of the datasets and the fitted curves are shown in Fig. 7.

6 Conclusion

Due to the importance of wind direction modelling in addressing ongoing energy crises, we proposed a new flexible unimodal two-parameter distribution on the circle, called the Power-Cardioid ($\mathcal{P}-\mathcal{C}$) distribution. This distribution emanates from an introduced Cardioid-generator class, and some characteristics of the $\mathcal{P}-\mathcal{C}$ distribution are discussed. A simulation study illustrated accurate parameter recovery for different sample sizes, and finite mixtures of the $\mathcal{P}-\mathcal{C}$ distribution were applied in modeling the wind direction at two sites in South Africa and compared with two other well-known finite mixture competitors in the context of circular data. It is meaningful to note that the proposed distribution adhere closely to the guidelines of desirability for flexible distributions as outlined by [23]: the Cardioid-generator class (1) is flexible due to the generator function $g(\cdot)$ admitting a Taylor series expansion; the investigated $\mathcal{P}-\mathcal{C}$ candidate (3) presents as tractable and amenable to calculations; simulated variates from the $\mathcal{P}-\mathcal{C}$ distribution is readily possible; and parameter estimation for the $\mathcal{P}-\mathcal{C}$ distribution presents as feasible. Finally, the Cardioid-generator class naturally nests an existing distribution of popular use, namely the Cardioid distribution.

It is well known that practitioners should treat different generation sites individually to optimize wind power generation. Here, our proposed distribution outperformed the KJ distribution at a major wind generation site in South Africa. This is significant since the KJ distribution is considered a flexible circular distribution and is extensively used in the literature to model wind direction based on AIC and BIC measures. In addition, the performance of our proposed distribution was also close to the performance of the KJ distribution for the offshore wind farm dataset (Noupoort). Given the high computational cost associated with the KJ distribution, the new distribution proposed in this paper is a valuable and exciting contribution within the framework circular statistics and finite mixtures thereof.

Acknowledgements The authors thank the chief and associate editors and the anonymous reviewers for comments that led to an improved paper.

Funding Open access funding provided by University of Pretoria. This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa (SA), grant RA201125576565, nr 145681; grant RA210106581084, nr 150170; grant SRUG2204203865 nr 120839; grant RA211204653274, nr 151035; STATOMET at the Department of Statistics at the University of Pretoria; the Department of Research and Innovation at the University of Pretoria (SA), as well as DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS) based at the University of the Witwatersrand, Johannesburg, South Africa. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. Mohammad Arashi's work is based on research supported in part by the Iran National Science Foundation (INSF) grant nr 4015320.

Data Availability The datasets used and/or analyzed during the current study are available from the corresponding author upon request.

References

1. Zhang, M.H.: Wind Resource Assessment and Micro-siting: Science and Engineering. John Wiley & Sons, New York (2015)
2. Carta, J.A., Ramirez, P., Velazquez, S.: A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands. Renew. Sustain. Energy Rev. **13**(5), 933–955 (2009)
3. Gugliani, G.K.: Comparison of different multi-parameters probability density models for wind resources assessment. J. Renew. Sustain. Energy **12**(6), 063303 (2020)
4. Gugliani, G.K., Ley, C., Nakhaei Rad, N., Bekker, A.: Comparison of probability distributions used for harnessing the wind energy potential: a case study from India. Stochast. Environ. Res. Risk Assess. **38**, 2213–2230 (2024)
5. Mardia, K.V., Jupp, P.E.: Directional Statistics. John Wiley & Sons, New York (2009)
6. Qin, X., Zhang, J., Yan, X.: A new circular distribution and its application to wind data. J. Math. Res. **2**(3), 12–17 (2010)
7. Gatto, R., Jammalamadaka, S.R.: The generalized von Mises distribution. Stat. Methodol. **4**(3), 341–353 (2007)
8. Belu, R., Koracin, D.: Statistical and spectral analysis of wind characteristics relevant to wind energy assessment using tower measurements in complex terrain. J. Wind Energy **2013**(1), 739162 (2013)
9. Quill, R., Sharples, J.J., Wagenbrenner, N.S., Sidhu, L.A., Forthofer, J.M.: Modeling wind direction distributions using a diagnostic model in the context of probabilistic fire spread prediction. Front. Mechan. Eng. **5**, 5 (2019)
10. Gugliani, G., Sarkar, A., Ley, C., Mandal, S.: New methods to assess wind resources in terms of wind speed, load, power and direction. Renew. Energy **129**, 168–182 (2018)
11. Rad, N.N., Bekker, A., Arashi, M.: Enhancing wind direction prediction of South Africa wind energy hotspots with Bayesian mixture modeling. Sci. Rep. **12**(1), 1–15 (2022)
12. Gugliani, G., Sarkar, A., Mandal, S., Agrawal, V.: Location wise comparison of mixture distributions for assessment of wind power potential: A parametric study. Int. J. Green Energy **14**(9), 737–753 (2017)
13. Carta, J.A., Bueno, C., Ramírez, P.: Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study. Energy Conver. Manag. **49**(5), 897–907 (2008)
14. Masseran, N., Razali, A.M., Ibrahim, K., Latif, M.T.: Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia. Energy Conver. Manag. **72**, 94–102 (2013)
15. Jeffreys, H.: The Theory of Probability. OUP Oxford, London (1998)
16. Ley, C., Verdebout, T.: Modern Directional Statistics. CRC Press, New York (2017)
17. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products. Academic press, New York (2014)
18. Mardia, K.V., Jupp, P.E.: Distributions on spheres. Direct. Stat. **898**, 159–192 (2000)
19. Mullen, K., Ardia, D., Gil, D.L., Windover, D., Cline, J.: DEoptim: An R package for global optimization by differential evolution. J. Stat. Softw. **40**(6), 1–26 (2011)
20. Storn, R., Price, K.: Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J. Global Opt. **11**(4), 341–359 (1997)
21. Price, K., Storn, R.M., Lampinen, J.A.: Differential Evolution: a Practical Approach to Global Optimization. Springer, New York (2006)
22. Kato, S., Jones, M.: A tractable and interpretable four-parameter family of unimodal distributions on the circle. Biometrika **102**(1), 181–190 (2015)
23. Ley, C., Babić, S., Craens, D.: Flexible models for complex data with applications. Ann. Rev. Stat. Appl. **8**(1), 369–391 (2021)