Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-021-00515-2

Analysis of correlated circular and extremal data with a flexible cylindrical distribution

Zeynep Kalaylioglu[1]

Received: 29 October 2020 / Revised: 20 May 2021 / Accepted: 31 July 2021

Abstract
In this article, we introduce a flexible cylindrical distribution for modeling and analysis of dependent extremal and directional observations. The distribution can be used to investigate the connection between two related phenomena, such as the daily fastest wind speed and its direction. The proposed model is applicable for the analysis of a wide variety of cylindrical data, including datasets with asymmetrically distributed directional observations. The model enjoys the advantages of interpretable model parameters, known marginal and conditional distributions, and a practical test for independence. Our simulation study shows that maximum likelihood estimators of the model parameters maintain desired finite sample properties. The distribution is then used to characterize the joint behavior of atmospheric variables in the context of wildfires or bushfires.

Keywords Gumbel distribution · Von Mises distribution · Copula · Wind direction · Maximum temperature · Bushfire simulation

1 Introduction

Circular-extremal data are observations that consist of a circular and an extremal component. Circular data are observations that can be regarded as points on a circle of unit radius. They include observations measured by compass such as wind directions,

Handling Editor: Luiz Duczmal

This research was carried out at School of Science, Mathematical Sciences, RMIT University, Melbourne, VIC, Australia, where the author was a visiting associate professor on a scholarship by the Scientific and Technological Research Council of Turkey (TUBITAK) grant 2219.

✉ Zeynep Kalaylioglu
kzeynep@metu.edu.tr

[1] Department of Statistics, Middle East Technical University, Ankara, Turkey

Published online: 21 August 2021

directions of sea currents, and orientation of migratory birds or observations measured by clock such as time of the day and month of the year. Extremal data mostly occur as a result of extreme environmental and atmospheric events, e.g. maximum wind speed, maximum precipitation, and maximum rainfall. There are several examples for dependent circular-extremal events such as extreme wind speed and wind direction (Coles and Walsaw (1994)), wave-induced extreme water level and wave direction (Coles and Walsaw (2016)), extreme groundwater level and time (Yozgatligil and Turkes (2018); Collet et al. (2017)). The specific example we focus on in this study is the speed and direction of the fastest gust as well as daily maximum temperature and morning wind direction, the main meteorological characteristics used in Australian fire spread simulation models.

Cylindrical distributions are used to model bivariate data with a linear and a circular component (Mardia and Sutton (1978); Johnson and Wehrly (1978); Abe and Ley (2017)). When the linear part is extreme value, modeling exerts the use of a cylindrical distribution function with extreme value marginal. In this paper, we propose a novel cylindrical distribution for modeling and analyzing data with an extremal and circular component. Our distribution is based on a combination of Gumbel distribution and sine-skewed von Mises distribution and denoted by GSSvM. Gumbel distribution is also called Type I Extreme Value distribution and is frequently applied for the analysis of extreme values observed in various research areas such as earth sciences, environmental sciences, geological sciences as well as economics, finance, and insurance. Efficient joint modeling of circular and extremal data depend on cylindrical distribution that is flexible enough to account for different distributional shapes and degree of dependencies. Another desired property for an efficient cylindrical distribution is that the conditional and/or marginal probability density functions have known forms.

Our proposed distribution accounts for a wide variety of dependencies and shapes. The real benefit is that it can be used to reveal general characteristics of the correlated linear-circular data, such as the mean direction of animal orientation given extreme climatic conditions. The rest of the article is organized as follows. Section 2 gives the specification of our distribution and its properties. Parameter estimation is given in Section 3. A simulation study illustrating the finite sample performances of the estimators is given in Section 4. Our model is then applied in Section 5 to characterize the joint behavior of meteorological extremes and wind direction in the context of Australian bushfires. Finally, section 6 gives concluding remarks and a critique.

2 GSSvM model and its properties

Our distribution is based on the family of maximum entropy distributions introduced by Johnson and Wehrly (1978). In their seminal article, they give a general method that is based on copulas to obtain circular-linear distributions with specified marginals. Their method is further generalized by the inclusion of an index for the direction of the dependence between the circular and the linear variates, and also used in modeling multi-modal cylindrical time series data (Lagona (2019)). Here we initially consider their copula based construction with marginals specified as von Mises and exponential distributions and extend it to develop a more flexible cylindrical distribution that is

specifically designed to fit cylindrical data where one of the components is extremal variable, and the other one is circular. Our approach is based on perturbing the distribution of the circular component of Johnson-Wehrly construction and transforming the exponential component. A perturbation is applied on the circular part to sine-skew von Mises distribution with mean μ and concentration κ via multiplication with the term $(1 + \lambda \sin(\theta - \mu))$ for $\lambda \in [-1, 1]$ following Abe and Pewsey (2011). This adjustment allows the analysis of symmetric as well as asymmetric distributions using the model given in (1). For the exponential part, we consider the following transformation. Let $Y = a - b^{-1} \log(Xb)$ where $X \sim Expo(1)$, $a \in \mathbb{R}$ and $b > 0$ are some constants and log is natural logarithmic function. Resulting Y has a Gumbel distribution, i.e. $Y \sim Gumbel(a, b^{-1})$ where a and b^{-1} are location and scale parameters respectively. This altogether leads to a more flexible model, as seen in later sections that is specifically designed to analyze a wide variety of extremal data that co-occur with circular data in nature. Letting (Θ,X) and (θ,x) denote the circular and extremal random variables and their realized values respectively, formal definition of GSSvM is given below.

Definition 2.1 A random cylinder (Θ,X) is said to follow the GSSvM distribution, if its density is given by

$$f(\theta, x) = \frac{\beta}{2\pi \cosh(\kappa)}(1 + \lambda \sin(\theta - \mu)) \exp[-(x - a)\beta(1 - \tanh(\kappa) \cos(\theta - \mu))] \times \exp[-\exp[-(x - a)\beta(1 - \tanh(\kappa) \cos(\theta - \mu))]] \quad (1)$$

where $-\pi \leq \theta \leq \pi$, $x \in \mathbb{R}$, $-\pi \leq \mu \leq \pi$, $a = 0$ for identifiability, $\kappa > 0$, $\beta > 0$, $\lambda \in [-1, 1]$, and $cosh$ and $tanh$ are hyperbolic cosine and hyperbolic tangent functions respectively.

The distribution has four parameters where (μ,κ,λ) are associated with the circular variate and β is associated with the extremal variate. μ, κ and λ are respectively circular location, concentration and skewness parameters while β is linear scale parameter. The parameter κ also accounts for the dependence between Θ and X, they are independent if $\kappa = 0$ in which case the joint density is the product of Gumbel and Cardioid densities. For κ and λ both equal to zero, the joint density is the product of Gumbel and circular uniform densities. The two dimensional contours of GSSvM distribution are given in Fig. 1 for different values of κ and λ illustrating various shapes of the model. Further properties are given in following sections.

2.1 Marginal distributions

Proposition 2.2 *Let* (Θ, X) *follow GSSvM with pdf given in (1). Then circular marginal pdf, i.e. pdf of* Θ*, and marginal pdf of* X *are given respectively by*

$$f_\Theta(\theta) = \frac{(1 + \lambda \sin(\theta - \mu))}{2\pi \cosh(\kappa)(1 - \tanh(\kappa) \cos(\theta - \mu))} \quad (2)$$

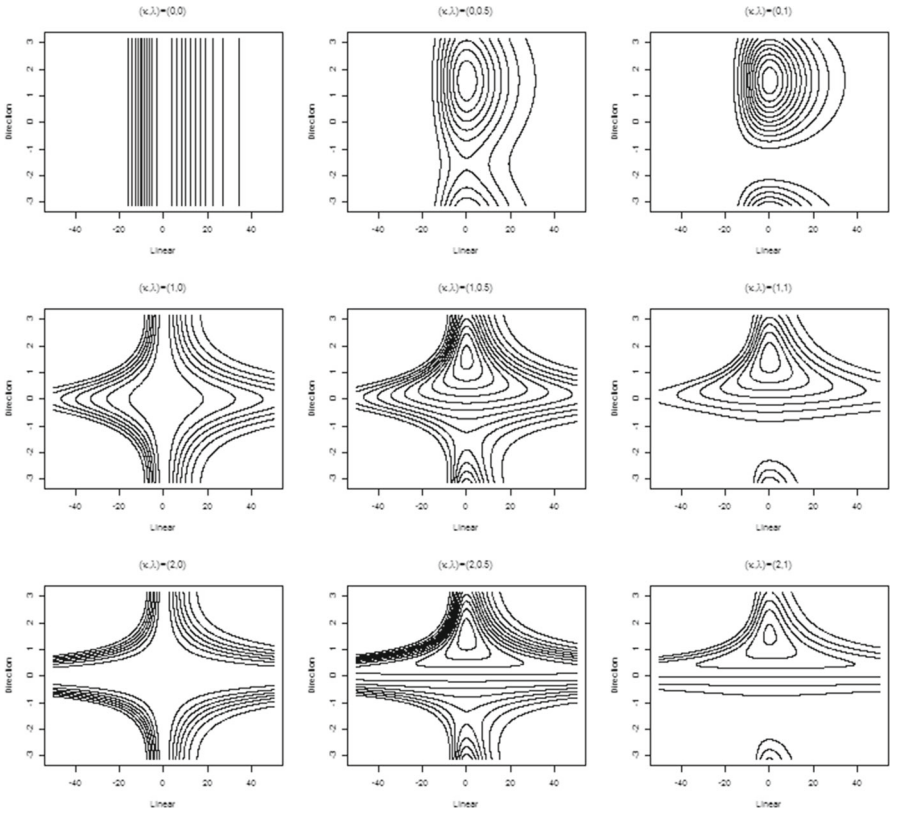

Fig. 1 Contour of GSSvM($\mu = 0$,κ,$\beta = 2$,λ) distribution with various κ and λ values.

which is the sine-skewed wrapped Cauchy (wC) distribution given by Abe and Pewsey (2011), and

$$f_X(x) = \frac{\beta A_0(\beta^*)}{\cosh(\kappa)} \exp[-\beta x] \exp[-\exp[-\beta x]] \tag{3}$$

where $\beta^* = x\beta \tanh(\kappa)$ *and* $A_0(\beta^*)$ *is the normalizing constant that is defined in Proposition 2.4. Note that the marginal distribution of X simplifies to Gumbel-maximum for* $\kappa = 0$.

Proof of (2) follows from integrating (1) with respect to x and the fact that $\int_{-\infty}^{\infty} b \exp[-bx] \exp[-\exp[-bx]]dx = 1$ where $b = \beta(1 - \tanh(\kappa)\cos(\theta - \mu))$.

Proof of (3) is given as follows. Integrating GSSvM pdf with respect to θ leads to

$$f_X(x) = \frac{\beta^*}{\cosh(\kappa)} \exp[-\beta x] \exp[-\exp[-\beta x]] \tag{4}$$
$$* \int_{-\pi}^{\pi} (1 + \lambda \sin(\theta - \mu))h(\theta)d\theta$$

where $h(\theta) = \frac{1}{2\pi}\exp[-\beta^*\cos(\theta-\mu)]\exp[-\exp[-\beta^*\cos(\theta-\mu)]]$. Define $g(\theta;\beta^*,\mu) = \frac{1}{A_0(\beta^*)}h(\theta)$. Note that $g(\theta;\beta^*,\mu) \geq 0$ is periodic and $\int_{-\pi}^{\pi} g(\theta;\beta^*,\mu)\, d\theta = 1$ and therefore it is a circular pdf. Hence by Abe and Pewsey (2011), the perturbated version, i.e. $(1+\lambda\sin(\theta-\mu))g(\theta;\beta^*,\mu)$ is a sine-skewed circular pdf. Then the integral in (4) is unity and completes the proof.

2.2 Conditional distributions

Proposition 2.3 *Conditional distribution of $\Theta|X=x$ is given by the following pdf*

$$f_{\Theta|X}(\theta|x) = \frac{1+\lambda\sin(\theta-\mu)}{2\pi A_0(\beta^*)}\exp[\beta^*\cos(\theta-\mu)] \\ *\exp[-\exp[-\beta^*\cos(\theta-\mu)]] \tag{5}$$

which follows from (1) and (3), and conditional distribution of $X|\Theta=\theta$ is Gumbel $-$ max$(0, b^{-1})$ with the following pdf that follows from (1) and (2)

$$f_{X|\Theta}(x|\theta) = b\exp[-bx]\exp[-\exp[-bx]] \tag{6}$$

where $b = \beta(1-\tanh(\kappa)\cos(\theta-\mu))$.

Note that, we use the notations b and β^* to reparameterize the distributions of X and θ respectively. Here, b is a function of (β,κ,μ) whereas β^* is a function of (β,κ).

2.3 Normalizing constant

Proposition 2.4 *The normalizing constant is given by*

$$A_0(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp[z\cos(\theta)]\exp[-\exp[z\cos(\theta)]]d\theta \tag{7}$$

The proof follows from Fourier transformation of a periodic and real analytic function. Let $-\pi \leq \alpha \leq \pi$ and $z \in \mathbb{R}$. We define $h(\alpha,z) = \exp[z\cos(\alpha)]\exp[-\exp[z\cos(\alpha)]]$. Then $h(\alpha,z)$ is a bounded, periodic and real analytic function in α and z and its Fourier series representation is

$$h(\alpha,z) = \sum_{n=0}^{\infty} A_n(z)\cos(n\alpha) + \sum_{m=0}^{\infty} B_m(z)\sin(m\alpha)$$

where

$$A_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h(\alpha,z)\cos(n\alpha)d\alpha$$
$$B_m(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h(\alpha,z)\sin(m\alpha)d\alpha$$

for $n, m = 0, 1, 2, \ldots$. Since $h(-\alpha, z) = h(\alpha, z)$ and $\sin(m\alpha)$ is an odd function, $B_m(z) = 0 \; \forall \; m = 0, 1, 2, \ldots$.

Note that for $\alpha = \theta - \mu$,

$$A_n(z) = \frac{1}{2\pi} \int_{-\pi+\mu}^{\pi+\mu} h(\theta - \mu, z) \cos(n(\theta - \mu)) d\theta$$
$$= \frac{1}{2\pi} \int_{-\pi+\mu}^{\pi+\mu} \exp[z \cos(\theta - \mu)] \exp[-\exp[z \cos(\theta - \mu)]] \cos(n(\theta - \mu)) d\theta$$

and $h(\theta - \mu, z) = \sum_{n=0}^{\infty} A_n(z) \cos(n(\theta - \mu))$. The normalizing constant, $A_0(z)$, is the order zero of this Fourier expansion.

2.4 Moments

Lemma 2.5 below gives a compact form for moments of general $Gumbel - max(a, b^{-1})$ distribution. Then, it is used in Proposition 2.6 to construct moments of the GSSvM distribution.

Lemma 2.5 *Let $X \sim Gumbel - max(a, b^{-1})$ where $Gumbel - max$ is Gumbel-maximum distribution. Then the nth. moment of X, i.e. $E(X^n)$, is given by*

$$E(X^n) = \sum_{k=0}^{n} \binom{n}{k} \frac{d^k}{dt^k} \Gamma\left(1 + \frac{t}{b}\right) a^{n-k}$$

which is evaluated at $t = 0$, where $\Gamma(.)$ is Gamma function and $\frac{d^k}{dt^k} \Gamma\left(1 + \frac{t}{b}\right) = \frac{E(U^k)}{b}$ at $t = 0$ where $U \sim Gumbel - max(0, 1)$.

Proof is given as follows. Note that moment generating function (mgf) of Gumbel-maximum distribution is $M(t) = \Gamma\left(1 + \frac{t}{b}\right) e^{at}$. The nth moment can be written in a compact form as follows. Let $D = \frac{d}{dt}$, $(D + a) = \frac{d}{dt} + a$, and $(D + a)w = w' + aw$ where $w(.)$ is a function of t and $(D + a)$ is a convolution applied on w. Then $D^n = \frac{d^n}{dt^n}$ and $(D + a)^n w(t) = \sum_{k=0}^{n} \binom{n}{k} \frac{d^k}{dt^k} w(t) a^{n-k}$ where $D^0 = \frac{d^0}{dt^0} = 1$. Now, let $w(t) = \Gamma\left(1 + \frac{t}{b}\right)$. Then, the mgf is $M(t) = w(t)e^{at}$. Then, first two derivatives of mgf are $\frac{d}{dt} M(t) = e^{at}(D + a)w(t)$ and $\frac{d^2}{dt^2} M(t) = e^{at}\left[(D + a)^2 w(t)\right]$ which generalizes to $\frac{d^n}{dt^n} M(t) = e^{at}\left[(D + a)^n w(t)\right]$. Then the result is straightforward.

Proposition 2.6 *Let (Θ, X) follow the GSSvM distribution with pdf given in (1). Then the joint moments are given by*

$$E(X^n \cos m\theta) = W_n I_n^m(\cosh \kappa) \tag{8}$$

and

$$E(X^n \sin m\theta) = W_n \frac{\lambda}{2} (I_n^{m-1}(\cosh \kappa) - I_n^{m+1}(\cosh \kappa)) \tag{9}$$

where

$$I_n^m(\cosh\kappa) = \frac{1}{\pi}\int_{-\pi}^{0} \frac{\cos m\theta}{(\cosh\kappa - \sqrt{\cosh^2\kappa - 1}\cos\theta)^{n+1}} d\theta$$

and $W_n = \frac{(\cosh\kappa)^n E(U^n)}{\beta^n}$ *where* $U \sim Gumbel - max(0,1)$ *for* $n, m = 0, 1, \cdots$.

Proof follows from Lemma 2.5 and the identity $(1 - \tanh\kappa\cos\theta) = \frac{\cosh\kappa - \sqrt{\cosh^2\kappa - 1}\cos\theta}{\cosh\kappa}$ and is straightforward. The integral, $I_n^m(\cosh\kappa)$, can be satisfactorily approximated by Monte Carlo integration for intermediate values of κ ($\kappa \leq 5$), larger number of iterations is needed for $\kappa > 5$. Marginal linear and trigonometric kth moments, i.e. $E(X^k)$, $E(\sin k\theta)$ and $E(\cos k\theta)$, are also obtained from equations (8) and (9).

3 Parameter estimation and test for independence

The natural logarithm of the GSSvM likelihood function, $\mathcal{L}(\beta, \mu, \kappa, \lambda)$, is given below to guide likelihood based inference.

$$\mathcal{L}(\beta, \mu, \kappa, \lambda) \propto n(\log(\beta) - \log(2\pi\cosh\kappa)) - \sum_{i=1}^{n} x_i\beta(1 - \tanh\kappa\cos(\theta_i - \mu))$$
$$- \sum_{i=1}^{n} \exp[-x_i\beta(1 - \tanh\kappa\cos(\theta_i - \mu)] + \sum_{i=1}^{n} \log(1 + \lambda\sin(\theta_i - \mu))$$

Model parameters are easily estimated via maximum likelihood technology, score functions are given below.

$$S_\beta = \frac{n}{\beta} - \sum_{i=1}^{n} x_i(1 - \tanh\kappa\cos(\theta_i - \mu))$$
$$+ \sum_{i=1}^{n} \exp[-x_i\beta(1 - \tanh\kappa\cos(\theta_i - \mu))]x_i(1 - \tanh\kappa\cos(\theta_i - \mu))I_\beta(0, \infty)$$
$$S_\mu = -\sum_{i=1}^{n} \frac{\lambda\cos(\theta_i - \mu)}{1 + \lambda\sin(\theta_i - \mu)} + \beta\sum_{i=1}^{n} x_i\tanh\kappa\sin(\theta_i - \mu)$$
$$- \sum_{i=1}^{n} \exp[-x_i\beta(1 - \tanh\kappa\cos(\theta_i - \mu))]x_i\beta\tanh\kappa\sin(\theta_i - \mu)I_\mu[-\pi, \pi]$$
$$S_\kappa = -n\tanh\kappa + \beta\sum_{i=1}^{n} x_i\cos(\theta_i - \mu)(\cosh\kappa)^{-2}$$

Table 1 Bias(MSE) of estimators. True $\kappa = 0$.

λ	n	$\hat{\beta}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$
−1	100	0.0192(0.0251)	0.0014(0.0249)	−0.0009(0.0022)	0.0567(0.0108)
	250	0.0095(0.0088)	−0.0000(0.0045)	0.0012(−0.0000)	0.0263(0.0011)
	500	0.0004(0.0045)	−0.0004(0.0022)	0.0001(0.0000)	0.0152(0.0004)
	1000	0.0007(0.0022)	0.0008(0.0011)	0.0000(0.0000)	0.0098(0.0001)
−0.5	100	0.0202(0.0245)	0.0034(0.0827)	0.0008(0.0012)	−0.0121(0.0182)
	250	0.0085(0.0089)	0.0014(0.0304)	0.0001(0.0000)	−0.0066(0.0066)
	500	0.0015(0.0045)	−0.0011(0.0153)	0.0000(0.0000)	−0.0040(0.0033)
	1000	0.0019(0.0022)	0.0009(0.0076)	0.0000(0.0000)	−0.0021(0.0016)
0	100	0.0219(0.0233)	−0.0107(0.5258)	0.0000(0.0010)	−0.0003(0.0063)
	250	0.0071(0.0089)	0.0132(0.5239)	−0.0006(0.0004)	−0.0014(0.0024)
	500	0.0017(0.0045)	−0.0012(0.5349)	−0.0002(0.0002)	0.0012(0.0012)
	1000	0.0022(0.0021)	−0.0029(0.5351)	0.0000(0.0001)	−0.0003(0.0006)
0.5	100	0.0232(0.0235)	−0.0020(0.0848)	0.0006(0.0013)	0.0084(0.0184)
	250	0.0060(0.0090)	−0.0020(0.0309)	0.0001(0.0001)	0.0066(0.0063)
	500	0.0037(0.0045)	−0.0005(0.0148)	0.0000(0.0000)	0.0025(0.0032)
	1000	0.0021(0.0022)	−0.0008(0.0076)	0.0000(0.0000)	0.0018(0.0017)
1	100	0.0220(0.0257)	−0.0036(0.0222)	−0.0018(0.0020)	−0.0545(0.0082)
	250	0.0054(0.0090)	−0.0003(0.0044)	0.0004(0.0001)	−0.0260(0.0011)
	500	0.0044(0.0044)	0.0008(0.0022)	0.0000(0.0000)	−0.0160(0.0004)
	1000	0.0016(0.0023)	−0.0004(0.0011)	0.0000(0.0000)	−0.0096(0.0002)

Table 2 Bias(MSE) of estimators. True $\kappa = 1$.

λ	n	$\hat{\beta}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$
-1	100	0.1457(0.0994)	-0.0324(0.0067)	0.0269(0.0099)	0.0752(0.0100)
	250	0.0559(0.0279)	-0.0163(0.0023)	0.0109(0.0036)	0.0365(0.0022)
	500	0.0300(0.0126)	-0.0089(0.0010)	0.0069(0.0018)	0.0217(0.0008)
	1000	0.0182(0.0061)	-0.0058(0.0005)	0.0039(0.0008)	0.0134(0.0003)
-0.5	100	0.0743(0.0617)	-0.0027(0.0081)	0.0096(0.0081)	0.0011(0.0318)
	250	0.0300(0.0215)	-0.0017(0.0029)	0.0041(0.0032)	0.0001(0.0115)
	500	0.0133(0.0100)	0.0005(0.0014)	0.0016(0.0016)	0.0004(0.0058)
	1000	0.0066(0.0049)	0.0000(0.0007)	0.0008(0.0008)	-0.0009(0.0028)
0	100	0.0659(0.0535)	0.0016(0.0078)	0.0063(0.0079)	-0.0036(0.0349)
	250	0.0227(0.0182)	-0.0007(0.0030)	0.0014(0.0029)	0.0007(0.0143)
	500	0.0112(0.0086)	0.0003(0.0014)	0.0013(0.0015)	-0.0004(0.0071)
	1000	0.0059(0.0043)	0.0003(0.0007)	0.0008(0.0007)	-0.0017(0.0036)
0.5	100	0.0691(0.0590)	0.0019(0.0080)	0.0079(0.0083)	-0.0023(0.0339)
	250	0.0283(0.0208)	0.0015(0.0029)	0.0037(0.0032)	-0.0035(0.0115)
	500	0.0138(0.0101)	0.0011(0.0014)	0.0017(0.0016)	-0.0018(0.0057)
	1000	0.0064(0.0047)	0.0001(0.0007)	0.0009(0.0008)	-0.0009(0.0029)
1	100	0.1433(0.0961)	0.0343(0.0067)	0.0268(0.0096)	-0.0740(0.0095)
	250	0.0581(0.0289)	0.0150(0.0022)	0.0115(0.0035)	-0.0364(0.0022)
	500	0.0307(0.0130)	0.0087(0.0010)	0.0065(0.0018)	-0.0217(0.0008)
	1000	0.0158(0.0060)	0.0052(0.0005)	0.0028(0.0008)	-0.0134(0.0003)

$$- \sum_{i=1}^{n} \exp[-x_i\beta(1 - \tanh\kappa\cos(\theta_i - \mu))]x_i\beta\cos(\theta_i - \mu)(\cosh\kappa)^{-2} I_\kappa(0, \infty)$$

$$S_\lambda = \left\{ \sum_{i=1}^{n} \frac{\sin(\theta_i - \mu)}{1 + \lambda\sin(\theta_i - \mu)} \right\} I_\lambda[-1, 1] \tag{10}$$

where I denotes an indicator function for the parameter space of interest. Letting $S(\beta, \mu, \kappa, \lambda) = (S_\beta, S_\mu, S_\kappa, S_\lambda)^T$, maximum likelihood estimators (MLEs) of the parameters are obtained as a solution of $S(\beta, \mu, \kappa, \lambda) = 0$. Numerical methods are required to solve the score equations as closed form solutions are not available. We used the gbnm function in Matlab performing globalized Nelder-Mead method (Luersen and Le Riche (2004)). One can also use optimx package in R that performs constrained optimization (Nash and Varadhan (2011)). Letting Ω denote the parameter vector, asymptotic standard error estimates of MLEs are computed from $\hat{I}_T^{-1}(\Omega)$ where $\hat{I}_T$ is the estimator of the total information matrix and given by $\sum_{i=1}^{n} \frac{\partial}{\partial\Omega} \log f(\theta_i, x_i) \frac{\partial}{\partial\Omega^T} \log f(\theta_i, x_i)$.

Dependence of circular and extremal random variables can easily be investigated through testing the null hypothesis $H_0 : \kappa = 0$ versus the alternative $H_1 : \kappa > 0$ thanks to the philosophy of proof by contradiction. We recommend, in this context,

Table 3 Bias(MSE) of estimators. True $\kappa = 2$.

λ	n	$\hat{\beta}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$
-1	100	0.1089(0.0793)	-0.0082(0.0007)	0.0175(0.0081)	0.1145(0.0221)
	250	0.0424(0.0263)	-0.0038(0.0002)	0.0065(0.0032)	0.0583(0.0057)
	500	0.0217(0.0122)	-0.0024(0.0001)	0.0035(0.0016)	0.0381(0.0024)
	1000	0.0118(0.0058)	-0.0015(0.0001)	0.0020(0.0008)	0.0235(0.0009)
-0.5	100	0.0809(0.0666)	-0.0014(0.0007)	0.0097(0.0077)	0.0102(0.0426)
	250	0.0290(0.0227)	0.0001(0.0003)	0.0031(0.0030)	-0.0002(0.0185)
	500	0.0153(0.0108)	-0.0002(0.0001)	0.0020(0.0014)	0.0012(0.0093)
	1000	0.0074(0.0053)	-0.0002(0.0001)	0.0005(0.0007)	-0.0006(0.0046)
0	100	0.0767(0.0671)	0.0005(0.0007)	0.0094(0.0077)	-0.0040(0.0561)
	250	0.0318(0.0223)	-0.0002(0.0003)	0.0049(0.0030)	0.0010(0.0235)
	500	0.0149(0.0105)	0.0001(0.0001)	0.0019(0.0014)	-0.0014(0.0107)
	1000	0.0063(0.0051)	0.0000(0.0001)	0.0005(0.0007)	0.0011(0.0058)
0.5	100	0.0783(0.0686)	0.0010(0.0007)	0.0102(0.0078)	-0.0029(0.0457)
	250	0.0300(0.0226)	0.0006(0.0003)	0.0034(0.0030)	-0.0032(0.0188)
	500	0.0155(0.0108)	0.0002(0.0001)	0.0020(0.0015)	-0.0007(0.0090)
	1000	0.0073(0.0053)	0.0001(0.0001)	0.0005(0.0007)	-0.0002(0.0045)
1	100	0.1023(0.0775)	0.0079(0.0007)	0.0159(0.0080)	-0.1174(0.0232)
	250	0.0453(0.0265)	0.0038(0.0003)	0.0071(0.0031)	-0.0613(0.0063)
	500	0.0232(0.0116)	0.0024(0.0001)	0.0038(0.0015)	-0.0374(0.0023)
	1000	0.0125(0.0058)	0.0014(0.0001)	0.0025(0.0007)	-0.0228(0.0009)

permutation based likelihood ratio test for sample sizes less than $n < 100$ and asymptotic chi-square test otherwise. Let T_{LR} denote the likelihood ratio test statistic and $T_{LR} = -2(\mathcal{L}(\hat{\beta}, \hat{\mu}, \hat{\kappa}, \hat{\lambda}) - \mathcal{L}(\tilde{\beta}, \tilde{\mu}, 0, \tilde{\lambda}))$ where $(\hat{\beta}, \hat{\mu}, \hat{\kappa}, \hat{\lambda})^T$ and $(\tilde{\beta}, \tilde{\mu}, \tilde{\lambda})^T$ are the vectors of MLEs over the global and H_0-restricted parameter spaces respectively. Distribution of T_{LR} can be obtained via permutation method when n is not sufficiently large and p value is calculated as the proportion of the upper tail with respect to the observed test statistic.

4 Simulation study

In this section we conduct a comprehensive simulation study to investigate the performance of the maximum likelihood estimation for GSSvM distribution. In our experiment, we create cylindrical datasets with various different symmetry. The data are generated from the GSSvM distribution as follows. First, sine-skewed circular observations $\{\theta_i\}_{i=1}^n$ are generated from the marginal pdf given in (2) as follows: $\theta_i = \theta_i^* I(U_i < v) + \theta_i^* I(U_i \geq v)$ where $\theta_i^* \sim wC(\mu, \tanh(\kappa/2))$, $U_i \sim Uniform(0, 1)$, $v_i = (1 + \lambda \sin(\theta_i^* - \mu))/2$ for $i = 1, 2, \cdots, n$, and I is indicator function, where this scheme follows from Abe and Ley (2017). Then, extremal observations $\{x_i\}_{i=1}^n$ are generated from the conditional pdf given in (6):

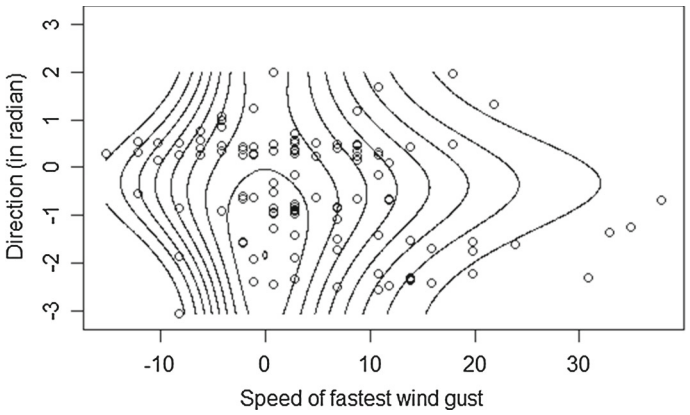

Fig. 2 The scatterplot of fastest wind speed and directions and contours of the fitted GSSvM.

$x_i \sim Gumbel - max(0, b^{-1})$ where $b = \beta(1 - \tanh\kappa\cos(\theta_i - \mu))$. We consider moderate to large samples (n=100,250,500,1000); various degrees of skewness ($\lambda = 0, 0.5, 1$); various degrees of concentration parameters ($\kappa = 0, 1, 2$) where 0 is included to represent independent circular and extremal components. In the simulations, $\beta = 2$ and $\mu = 0$ without loss of generality. We initially repeated the experiment with different values for β and observed that the results remain the same.

The bias and mean squared errors (MSE) of the estimators are given in Tables 1-3. Accordingly, the overall performance is excellent. Specifically, the bias and MSE decreases with sample size, and they are higher for λ values that are close to the boundaries. Bias and MSE of λ estimation increase with κ.

5 Real data applications

In this section we utilize GSSvM to analyze two datasets that are publicly available at data repository of Australian Bureau of Meteorology (BOM). The first dataset consists of daily fastest wind gust and its direction recorded between 1 November 2020 and 28 February 2021 in Melbourne Olympic Park while the second dataset consists of daily maximum temperature and morning wind direction during the same time period and on the same site (n=116 excluding observations with missing wind direction). This period is over the summer months in the Southern Hemisphere and includes the peak bushfire seasons, the most destructive type of wildfire in Australia. These variables play important role in bushfires and their management. Their analysis provides input for bushfire simulation and risk models (see e.g. Lopes et al. (2002); Coen (2005); Sharples et al. (2010) for wildfire simulation models).

We illustrate the fitting performance of GSSvM in relation to the independence model and the Abe-Ley model (Abe and Ley (2017)). Abe-Ley model is based on a combination of sine-skewed vM distribution and Weibull distribution. As Weibull distribution is also an extreme value distribution, Abe-Ley model can be considered as an alternative for modeling cylindrical data consisting of extremal and circular

Table 4 Estimators (ASE) of model parameters and AIC for maximum wind gust speed and its direction

Model	$\hat{\beta}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$	$\hat{\beta}_{AL}$	$\hat{\alpha}_{AL}$	AIC
GSSvM	0.15 (0.01)	−0.27 (0.27)	0.38 (0.09)	−0.10 (0.18)	–	–	1258
Abe-Ley	–	−0.15 (0.24)	0.35 (0.06)	−0.24 (0.21)	0.03 (.001)	3.71 (0.02)	1330
Independence	0.13 (0.08)	−0.37	–	0.00	–	–	1278

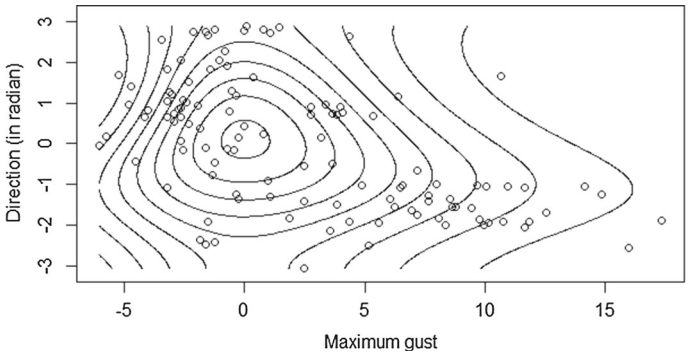

Fig. 3 The scatterplot of daily maximum temperature and wind direction at 9am and contours of the fitted GSSvM.

components. We use the Akaike Information Criterion (AIC) to compare the fitness of the models. In the following tables, the parameters of Abe-Ley distribution are indexed by AL.

5.1 Fastest wind gust and the direction

To analyse the dataset, angular direction data are transferred to radian. Wind speed data are transferred prior to applying GSSvM with $a = 0$. The scatterplot of the fastest wind gust versus their direction along with contours of the fitted GSSvM are given in Fig. 2. The scatterplot somewhat suggests an association between the two variables: concentration around a particular direction seems to increase with extrema of the gust, implying that minimum or maximum gust speed is associated with its direction.

Table 4 presents the parameter estimates and fitting capacities (AICs) of the competing models considered for this dataset, namely GSSvM and Abe-Ley. They lead to a somewhat similar location, concentration, and skewness estimates. Comparing the AICs of the models, we can see that GSSvM improves on the Abe-Ley model. The independence model is fitted to test the dependence between the two variables. The independence model is fitted by fitting GSSvM with $\kappa = 0$. Maximum for independence model occurred at $\lambda = 0$ for this dataset rendering the ASEs of μ and λ inestimable under independence assumption. Asymptotic likelihood ratio test implied strong dependence (p-value=0.00).

As conclusion, GSSvM seems to be a good fit for this dataset which in turn implies that it can be utilized for joint random generation of gust wind and direction observations in wildfire simulation models that are used to predict bushfire risks under different meteorological conditions.

5.2 Morning wind direction and maximum daily heat

For the analysis of this dataset, which consists of daily maximum temperature and wind directions at 9am, angular directions are first obtained from the original cardi-

Table 5 Estimator (ASE) and AIC of the models for daily maximum heat and morning wind direction in Melbourne.

Model	$\hat{\beta}$	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$	$\hat{\beta}_{AL}$	$\hat{\alpha}_{AL}$	AIC
GSSvM	0.24 (0.02)	− 1.46 (0.35)	0.25 (0.11)	0.36 (0.15)	–	–	1121
Abe-Ley	–	− 1.57(0.19)	0.48 (0.08)	0.40 (0.15)	0.04(0.001)	5.40 (0.03)	1119
Independence	0.26 (0.02)	0.24	–	0.00	–	–	1136

nal directions and converted to radians. The scatterplot of the observations and the fitted GSSvM contours are given in Fig. 3. The figure suggests that maximum daily temperature changes with the direction of the wind at 9am.

Table 5 presents the parameter estimates and fitting capacities (AICs) of the competing models considered for this dataset. GSSvM and Abe-Ley are comparable. As a side note, the maximum for independence model occurred at $\lambda = 0$ for this dataset rendering the ASEs of μ and λ inestimable under independence assumption. Asymptotic likelihood ratio test for independence, $H_0 : \kappa = 0$, gives p-value = 0.00 indicating highly significant dependence between daily maximum temperature and direction of the wind at 9am in Melbourne over the span of November 2020-February 2021. Contours of the fitted GSSvM given in Fig. 3 seem to show a satisfactory fit well capturing the observations. Furthermore, based on these findings, it can be inferred that GSSvM or equivalently Abe-Ley model can be used in wildfire simulations to model wind direction and extreme heat, two atmospheric conditions that affect fire behavior.

6 Future research

In this paper, a cylindrical distribution was proposed for modeling dependent extremal and circular data. The distribution is constructed based on Gumbel distribution, and a sine skewed circular distribution. The proposed distribution is quite versatile, as shown in sections 2 and 5. The model also accounts for any asymmetric feature present in the circular observations. Maximum likelihood estimation of the model parameters is straightforward, and readily available optimization algorithms with box constraints in standard statistical software can be used. MLEs are shown to maintain excellent bias and MSE properties. Also, as seen in Section 5, this model is shown to be particularly beneficial in the case of bushfire risk modeling as it provides a method to simulate dependent meteorological cylindrical variables.

In this paper, we focused on fitting the distribution only, while regression models, mixture models, prediction, and temporal or spatial dependence will be interesting future work. Mixture of GSSvM densities can be considered to analyze spatially or temporally correlated cylindrical series following Lagona and Picone (2016) and Ranalli et al. (2018) or Lagona (2019), respectively. Also, we think that our approach can be generalized on the extreme value component of the distribution by considering Kumaraswamy generalized extreme value distribution (Eljabri and Nadarajah (2017)) and this will be undertaken elsewhere.

The normalizing constant in our model is worthy of notice. Notice that it is different than the usual normalizing constant appearing in most circular densities, which is based on the Bessel function. It would be very interesting to know whether A_0 here appear to be an analytic solution of a second order differential equation, such as Bessel or Legendre functions.

References

Abe T, Ley C (2017) A tractable, parsimonous and flexible model for cylindrical data, with applications. Econom Stat 4:91–104

Abe T, Pewsey A (2011) Sine-skewed circular distributions. Stat Pap 52:683–707

Coen JL (2005) Simulation of the Big Elk Fire using coupled atmosphere-fire modeling. Int J Wildland Fire 14:49–59

Coles SG, Walsaw D (1994) Directional modelling of extreme wind speeds. Appl Stat 43:139–157

Coles SG, Walsaw D (2016) Run-up parameterization and beach vulnerability assessment on a barrier island: a downscaling approach. Nat Hazard Earth Syst Sci 16:167–180

Collet L, Beevers L, Prudhomme C (2017) Assessing the impact of climate change and extreme value uncertainty to extreme flows across Great Britain. Water 9(2):103. https://doi.org/10.3390/w9020103

Eljabri S, Nadarajah S (2017) The Kumaraswamy GEV distribution. Commun Stat Theory Methods 46:10203–10235

Johnson RA, Wehrly TE (1978) Some angular-linear distributions and related regression models. J Am Stat Assoc 73:602–606

Lagona F (2019) Copula-based segmentation of cylindrical time series. Stat Probab Lett 144:16–22

Lagona F, Picone M (2016) Model-based segmentation of spatial cylindrical data. J Stat Comput Simul 86:2598–2610

Lopes AMG, Cruz MG, Viegas DX (2002) FireStation—an integrated software system for the numerical simulation of fire spread on complex topography. Environ Model Softw 17:269–285

Luersen MA, Le Riche R (2004) Globalized Nelder-Mead method for engineering optimization. Comput Struct 82:2251–2260

Mardia KV, Sutton TW (1978) A model for cylindrical variables with applications. J R Stat Soc B 40:229–233

Nash JC, Varadhan R (2011) Unifying optimization algorithms to aid software system users: optimx for R. J Stat Softw 43(9):1–14

Ranalli M, Lagona F, Picone M (2018) Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach. JRSS Appl Stat Ser C 67:575–598

Sharples JJ, McRae RHD, Weber RO (2010) Wind characteristics over complex terrain with implications for bushfire risk management. Environ Model Softw 25:1099–1120

Yozgatligil C, Turkes M (2018) Extreme value analysis and forecasting of maximum precipitation amounts in the western Black Sea subregion of Turkey. Int J Climatol 38:5447–5458

Zeynep Kalaylioglu is Associate Professor at the Middle East Technical University (METU), Turkey. Her current research interests include circular statistics, Bayesian methods, and model selection.