The joint projected normal and skew-normal: A distribution for poly-cylindrical data

The joint projected normal and skew-normal: A distribution for poly-cylindrical data

Gianluca Mastrantonio

Department of Mathematics, Polytechnic of Turin, Turin, Italy

Abstract

This paper introduces a multivariate circular-linear (or poly-cylindrical) distribution obtained by combining the projected and the skew-normal. We show the flexibility of our proposal, its closure under marginalization, and how to quantify multivariate dependence. Due to a non-identifiability issue that our proposal inherits from the projected normal, a computational problem arises. We overcome it in a Bayesian framework, adding suitable latent variables and showing that posterior samples can be obtained with a post-processing of the estimation algorithm output. Under specific prior choices, this approach enables us to implement a Markov chain Monte Carlo algorithm relying only on Gibbs steps, where the updates of the parameters are done as if we were working with a multivariate normal likelihood. The proposed approach can also be used with the projected normal. As a proof of concept, on simulated examples we show the ability of our algorithm in recovering the parameter values and to solve the identification problem. Then the proposal is used in a real data example, where the turning-angles (circular variables) and the logarithm of the step-lengths (linear variables) of four zebras are modeled jointly.

Keywords: Circular data, circular-linear distribution, multivariate distribution, projected Normal, skew-Normal

1. Introduction

The analysis of circular data, i.e., observations on the unit circle, requires specific statistical tools since the circular domain is intrinsically different from the real line, and this inhibits the use of standard statistics that, if not properly modified, lead to uninterpretable results; for a general review, see [16, 22, 34]. A similar type of problem holds for circular densities which, beyond being non-negative and integrating to 1, should possess the property of "invariance" [26], i.e., they must be a location model under the group of rotations and reflections of the circle. This property, which expresses the need of densities that can represent equivalently the same phenomena under different reference systems, is specific to circular densities and it is sometimes overlooked [26].

Circular data are often observed along with "linear" variables, i.e., variables taking values on the real line; they are then called cylindrical in the bivariate case and poly-cylindrical in higher dimension. For example in marine research, wind and wave directions are modeled with wind speed and wave height [8, 28, 39] and, in ecology, animal behavior is described using measures of speed and direction, e.g., step-length and turning-angle [9, 19, 30, 33]. In most applications, cylindrical data are modeled assuming independence between the circular and linear components; see, e.g., [8, 20, 31]. Ignoring dependence can lead to misleading inference since we are not considering a component of the data that can help in understanding the phenomenon under study; see, e.g., [27]. In the literature, to date, no poly-cylindrical distributions have been proposed and there are only few distributions for cylindrical data; the best known examples are those in [2, 17, 23], and the new density of [1]. The aim of this work is to introduce what is, to the best of our knowledge, the first poly-cylindrical distribution.

Circular and linear variables live in very different spaces and the definition of a mixed-domain distribution is not easy. The issue is even more complicated if we require flexibility, interpretable parameters and the possibility to define an efficient estimation algorithm that is easy to implement. We choose to adopt a Bayesian framework because, as we show in Section 3.1 that using standard Markov chain Monte Carlo (MCMC) methods, we are able to propose an algorithm with the required characteristics while Monte Carlo (MC) procedures [7, 36] allow us to obtain posterior distributions for all the statistics we may need to describe the results.

Since circular observations often exhibit bimodality [38, 40], our aim is to propose a distribution with circular marginals that can model such data. In the literature, the best known bimodal circular distributions are the projected normal ($\mathcal{PN}$) [40] and the generalized von Mises [10]. The former can be easily generalized to the multivariate setting and it has an interesting augmented density representation, based on a normal probability density function (pdf), that can be used to define circular-linear dependence. The $\mathcal{PN}$ is very flexible [25, 41] with shapes that range from unimodal and symmetric to bimodal and antipodal; it is closed under marginalization and, as we show in the Appendix, it possesses the invariance property. However, multivariate extensions of the generalized von Mises are not easy to handle and, in our opinion, it is not straightforward to use it as a component of a poly-cylindrical distribution.

We define our proposal constructively, starting from the $\mathcal{PN}$ and choosing a distribution for the linear component which, taking advantage of the $\mathcal{PN}$ augmented density representation, allows us to define a poly-cylindrical distribution that is flexible enough to model real data and whose parameters can be easily estimated with MCMC algorithms.

For the linear component we use a skew-normal, i.e., a generalization of the Gaussian distribution which allows more flexibility introducing asymmetry in the normal density. Its first univariate version was proposed by Azzalini [4]. Subsequent works introduced multivariate extensions and different formalizations; see, e.g., [5, 13, 18, 37]. Among these, we found that of Sahu et al. [37] (hereafter $\mathcal{SSN}$) interesting: it can be closed under marginalization and it has an augmented density representation that, as the $\mathcal{PN}$, is based on a normal pdf.

Using this particular form of the skew-normal distribution, due to the properties listed above, we are able to define the *joint projected normal and skew-normal* ($\mathcal{JPNSN}$) poly-cylindrical distribution by introducing dependence in the normal pdfs of the augmented representations. The distribution retains the $\mathcal{PN}$ and $\mathcal{SSN}$ as marginal distributions and is closed under marginalization, i.e., any subset of circular and linear variables is $\mathcal{JPNSN}$ distributed. The MCMC algorithm we propose can be based only on Gibbs steps, updating parameters as if we were working with a multivariate normal likelihood. The density cannot be expressed in closed form but we do not consider this an issue from the point of view of model fitting, since we are able to estimate its parameters easily.

The $\mathcal{JPNSN}$ has the same identification problem as the $\mathcal{PN}$ [40], but we show that posterior values can be obtained by a post-processing of the MCMC algorithm based on the non-identifiable likelihood. The proposed algorithm can also be used with the univariate and multivariate $\mathcal{PN}$ and the spherical $\mathcal{PN}$ distribution of [14], solving their identification problem in a new way.

The algorithm, tested on simulated datasets, shows its ability to retrieve the parameters used to simulate the data, and posterior samples do not suffer from an identification issue. We used the $\mathcal{JPNSN}$ to jointly model the logarithm of step-lengths and turning-angles of four zebras observed in Botswana. Through a comparison based on the continuous rank probability scores (CRPSs) [11, 12] of our proposal, the cylindrical distribution of [1] and a cylindrical version of the $\mathcal{JPNSN}$, i.e., assuming independence between zebras, we show that ignoring multivariate dependence can lead to loss of predictive ability.

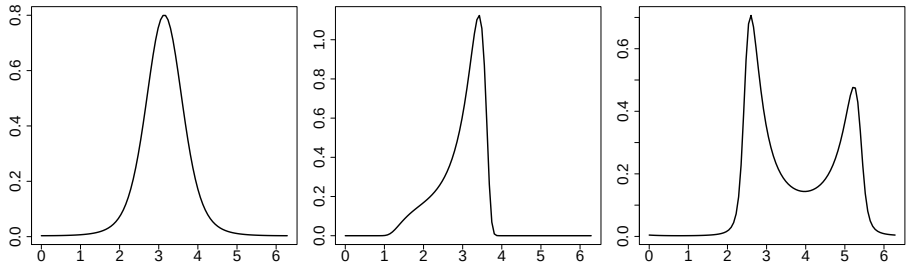
The paper is organized as follows. Section 2 is devoted to the constructive definition of the distribution. In Section 3 we introduce the identification problem and how to estimate the $\mathcal{JPNSN}$ parameters. The proposal is applied to simulated examples in Section 4.1 and the real data application is discussed in Section 4.2. The paper ends with concluding remarks in Section 5. In the Appendix we prove the invariance property of the $\mathcal{PN}$ and we show MCMC implementation details.

2. The joint projected normal and skew-normal distribution

In this section we build the poly-cylindrical density by first introducing the circular and linear marginals and then showing how to induce dependence.

2.1. The projected normal distribution

The $\mathcal{PN}$ is a distribution for a p-dimensional vector $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ of circular variables, i.e., $\Theta_i \in [0, 2\pi)$ is an angle expressed in radiant, obtained starting from a $2p$-dimensional vector $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_p)$, where $\mathbf{W}_i = (W_{i1}, W_{i2})^\top \in \mathbb{R}^2$, distributed as a $2p$-variate normal with mean vector $\boldsymbol{\mu}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$. For each $i \in \{1, \ldots, p\}$, $\mathbf{W}_i$ is normally distributed with parameters $\boldsymbol{\mu}_{w_i}$, and $\boldsymbol{\Sigma}_{w_i}$ is a point in the 2-dimensional space expressed



(a) (b) (c)

Figure 1: Univariate projected normal densities with mean $\boldsymbol{\mu}_{w_i}$ and covariance matrix $\boldsymbol{\Sigma}_{w_i}$ with diagonal entries equal to 1 and off-diagonal entries ρ_{w_i}: (a) $\boldsymbol{\mu}_{w_i} = (2, 0)^\top$, $\rho_{w_i} = 0$; (b) $\boldsymbol{\mu}_{w_i} = (2, 0)^\top$, $\rho_{w_i} = 0.9$; (c) $\boldsymbol{\mu}_{w_i} = (-0.1, -0.2)^\top$, $\rho_{w_i} = -0.9$.

using the Cartesian system. The same point can be also represented in polar coordinates with the angle Θ_i and the distance vector $R_i \in \mathbb{R}^+$. Between $\mathbf{W}_i$, Θ_i and R_i the following relations exist:

$$\Theta_i = \text{ atan}^* (W_{i2}/W_{i1}) \tag{1}$$

and

$$\mathbf{W}_i = R_i \begin{pmatrix} \cos\Theta_i \\ \sin\Theta_i \end{pmatrix}, \quad R_i = \|\mathbf{W}_i\|,$$

where

$$\text{atan}^* (S/C) = \begin{cases} \text{atan}\,(S/C) & \text{if } C > 0,\ S \geq 0, \\ \pi/2 & \text{if } C = 0,\ S > 0, \\ \text{atan}\,(S/C) + \pi & \text{if } C < 0, \\ \text{atan}\,(S/C) + 2\pi & \text{if } C \geq 0,\ S < 0, \\ \text{undefined} & \text{if } C = 0,\ S = 0, \end{cases}$$

is a modified arctangent function used to define a quadrant-specific inverse of the tangent.

If we transform each $\mathbf{W}_i$ in $(\Theta_i, R_i)^\top$ the Jacobian of the transformation is $R_1 \times \cdots \times R_p$ and then the joint density of $(\boldsymbol{\Theta}, \mathbf{R})^\top$, where $\mathbf{R} = (R_1, \ldots, R_p)$, is given by

$$f(\boldsymbol{\theta}, \mathbf{r}) = \prod_{i=1}^{p} r_i \phi_{2p}(\mathbf{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w), \tag{2}$$

where f stands for the density of its arguments, $\mathbf{r}$ is a realization of $\mathbf{R}$, $\phi_{2p}(\mathbf{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$ is the pdf evaluated at $\mathbf{w}$ of a $2p$-variate normal distribution with mean $\boldsymbol{\mu}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$; here $\mathbf{w}$ must be seen as a function of $(\boldsymbol{\theta}, \mathbf{r})^\top$.

The marginal density of $\boldsymbol{\Theta}$, obtained by integrating out $\mathbf{R}$ in (2), is a p-variate projected normal with parameters $\boldsymbol{\mu}_w$ and $\boldsymbol{\Sigma}_w$, i.e., $\boldsymbol{\Theta} \sim \mathcal{PN}_p(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$. As shown in [40], the $\mathcal{PN}$ can be symmetric, asymmetric and bimodal; univariate shapes are depicted in Figure 1.

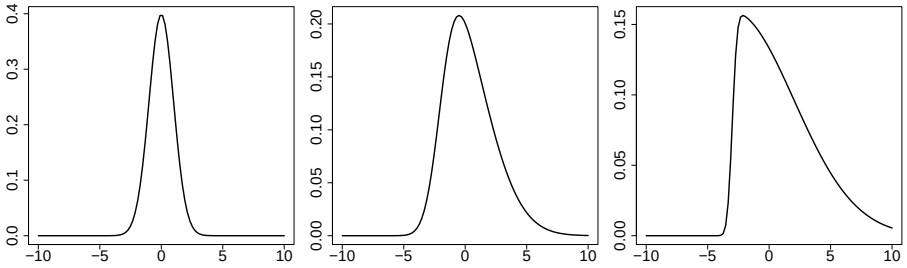
A closed form expression for the $\mathcal{PN}_p$ density is only available in the univariate case ($p = 1$) and it is

$$f(\theta_i) = \frac{\phi_2(\boldsymbol{\mu}_{w_i}|\mathbf{0}_2, \boldsymbol{\Sigma}_{w_i}) + |\boldsymbol{\Sigma}_{w_i}|^{-1} D(\theta_i)\Phi_1\{D(\theta_i)|0, 1\}\phi_1\{|\boldsymbol{\Sigma}_{w_i}|^{-1} C(\theta_i)^{-1/2}\,(\mu_{w_{i1}} \sin\theta_i - \mu_{w_{i2}} \cos\theta_i)\}}{C(\theta_i)},$$

where

$$C(\theta_i) = |\boldsymbol{\Sigma}_{w_i}|^{-2}(\sigma^2_{w_{i2}} \cos^2\theta_i - \rho_{w_i}\sigma_{w_{i1}}\sigma_{w_{i2}} \sin 2\theta_i + \sigma^2_{w_{i1}} \sin^2\theta_i),$$

$$D(\theta_i) = \frac{|\boldsymbol{\Sigma}_{w_i}|^{-2}\{\mu_{w_{i1}}\sigma_{w_{i2}}\,(\sigma_{w_{i2}} \cos\theta_i - \rho_{w_i}\sigma_{w_{i1}} \sin\theta_i) + \mu_{w_{i2}}\sigma_{w_{i1}}\,(\sigma_{w_{i1}} \sin\theta_i - \rho_{w_i}\sigma_{w_{i2}} \cos\theta_i)\}}{\sqrt{C(\theta_i)}},$$



(a) (b) (c)

Figure 2: Univariate skew-normal densities under three sets of parameters: (a) $\boldsymbol{\mu}_y = 0$, $\boldsymbol{\Sigma}_y = 1$, $\boldsymbol{\lambda} = 0$; (b) $\boldsymbol{\mu}_y = -2$, $\boldsymbol{\Sigma}_y = 1$, $\boldsymbol{\lambda} = 3$; (c) $\boldsymbol{\mu}_y = -3$, $\boldsymbol{\Sigma}_y = 0.1$, $\boldsymbol{\lambda} = 5$.

$\Phi_\ell(\cdot|\cdot,\cdot)$ denotes the normal ℓ-variate cumulative distribution function with given mean vector and covariance matrix, $\mathbf{0}_\ell$ is a vector of 0s of length ℓ, $\mu_{w_{ij}}$ and $\sigma^2_{w_{ij}}$ are the mean and variance of W_{ij}, respectively, and ρ_{w_i} is the correlation between W_{i1} and W_{i2}. In practical applications, it is generally preferable to work with the pair $(\boldsymbol{\Theta}, \mathbf{R})^\top$ that has the nice closed form density given in Eq. (2), treating $\mathbf{R}$ as a vector of latent variables; see, e.g., [24, 25, 41].

The multivariate $\mathcal{PN}$ is closed under marginalization since $\boldsymbol{\Theta}_A \sim \mathcal{PN}_{n_a}(\boldsymbol{\mu}_{w,A}, \boldsymbol{\Sigma}_{w,A})$, where $A \subset \{1, \ldots, p\}$, n_a denotes the cardinality of the set A and $\{\boldsymbol{\mu}_{w,A}, \boldsymbol{\Sigma}_{w,A}\}$ are mean and covariance matrix of $\mathbf{W}_A$. Moreover, as we show in Appendix A, the univariate $\mathcal{PN}$ possesses the invariance property and then inference does not depend on the reference system chosen for the circular variables [26].

2.2. The skew-normal distribution

We now introduce the skew-normal distribution of [37] as the distribution of a vector $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$, with $Y_j \in \mathbb{R}$ for each $j \in \{1, \ldots, q\}$. Let $\boldsymbol{\mu}_y$ be a vector of length q, $\boldsymbol{\Sigma}_y$ be a $q \times q$ non-negative definite (nnd) matrix and $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ be a $q \times q$ diagonal matrix with main diagonal $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q) \in \mathbb{R}^q$. We say that $\boldsymbol{Y}$ is distributed according to a q-variate skew-normal with parameters $\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\lambda}$, denoted $\boldsymbol{Y} \sim \mathcal{SSN}_q(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y, \boldsymbol{\lambda})$, if its pdf is given by

$$f(\mathbf{y}) = 2^q \phi_q(\mathbf{y}|\boldsymbol{\mu}_y, \boldsymbol{\Upsilon}) \Phi_q\{\boldsymbol{\Lambda}^\top \boldsymbol{\Upsilon}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)|\mathbf{0}_q, \boldsymbol{\Gamma}\}, \tag{3}$$

where $\boldsymbol{\Upsilon} = \boldsymbol{\Sigma}_y + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$, $\boldsymbol{\Gamma} = \mathbf{I}_q - \boldsymbol{\Lambda}^\top \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Lambda}$ and $\mathbf{I}_q$ is the identity matrix of dimension q. Although in [37] $\boldsymbol{\Lambda}$ is defined as a full matrix, here we constrain it to be diagonal to have a $\mathcal{SSN}$ closed under marginalization; the same property will be inherited by our poly-cylindrical distribution; see Section 2.3. From Eq. (3) we clearly see that $\mathbf{Y} \sim \mathcal{N}_q(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$ if $\boldsymbol{\Lambda}$ is a null matrix and for this reason it is called the *skewness parameter*. Examples of univariate $\mathcal{SSN}$ densities are shown in Figure 2.

The $\mathcal{SSN}$ has a nice stochastic representation [3] that is useful for the definition of the poly-cylindrical distribution. Let $\mathbf{D} \sim \mathcal{HN}_q(\mathbf{0}_q, \mathbf{I}_q)$, where $\mathcal{HN}_q(\cdot,\cdot)$ indicates the q-dimensional half normal [32], and $\mathbf{H} \sim \mathcal{N}_q(\mathbf{0}_q, \boldsymbol{\Sigma}_y)$, then $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \boldsymbol{\mu}_y + \boldsymbol{\Lambda}\mathbf{D} + \mathbf{H}. \tag{4}$$

From (4) we can see that $\mathbf{Y}|\mathbf{D} = \mathbf{d}$ is normally distributed with mean $\boldsymbol{\mu}_y + \boldsymbol{\Lambda}\mathbf{d}$ and covariance matrix $\boldsymbol{\Sigma}_y$. Consequently, the joint density of $(\mathbf{Y}, \mathbf{D})^\top$ expressed as the product of the ones of $\mathbf{Y}|\mathbf{D}$ and $\mathbf{D}$, is given by

$$f(\mathbf{y}, \mathbf{d}) = 2^q \phi_q(\mathbf{y}|\boldsymbol{\mu}_y + \boldsymbol{\Lambda}\mathbf{d}, \boldsymbol{\Sigma}_y) \phi_q(\mathbf{d}|\mathbf{0}_q, \mathbf{I}_q). \tag{5}$$

2.3. The joint linear-circular distribution

In this section, we define the poly-cylindrical distribution starting from the augmented circular and linear marginals shown in Eqs. (2) and (5). As in the previous sections, we denote by p and q the dimensions of the vectors of circular and linear variables, respectively.

It is natural to introduce dependence between $\boldsymbol{\Theta}$ and $\mathbf{Y}$ by substituting the two normal pdfs of the augmented representation, i.e., $\phi_{2p}(\mathbf{w}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$ and $\phi_q\{\mathbf{y}|\boldsymbol{\mu}_y + \text{diag}(\boldsymbol{\lambda})\mathbf{d}, \boldsymbol{\Sigma}_y\}$, with a $(2p+q)$-dimensional normal pdf that has the two pdfs as marginals; after marginalization we obtain the density of $(\boldsymbol{\Theta}, \mathbf{Y})^\top$. More precisely, we define the joint density of $(\boldsymbol{\Theta}, \mathbf{R}, \mathbf{Y}, \mathbf{D})^\top$ as

$$f(\boldsymbol{\theta}, \mathbf{r}, \mathbf{y}, \mathbf{d}) = 2^q \phi_{2p+q}\{(\mathbf{w}, \mathbf{y})^\top | \boldsymbol{\mu} + (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \boldsymbol{\Sigma}\} \phi_q(\mathbf{d}|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} r_i, \tag{6}$$

with $\boldsymbol{\mu} = (\boldsymbol{\mu}_w, \boldsymbol{\mu}_y)^\top$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_w & \boldsymbol{\Sigma}_{wy} \\ \boldsymbol{\Sigma}_{wy}^\top & \boldsymbol{\Sigma}_y \end{pmatrix}.$$

We then say that $(\boldsymbol{\Theta}, \mathbf{Y})^\top$ is marginally distributed as a (p, q)-variate *joint projected normal and skew-normal* with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$, i.e., $(\boldsymbol{\Theta}, \mathbf{Y})^\top \sim \mathcal{JPNSN}_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Since $(\mathbf{W}, \mathbf{Y})^\top | \mathbf{d} \sim \mathcal{N}_{2p+q}[\boldsymbol{\mu} + (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \boldsymbol{\Sigma}]$, transformation of $\mathbf{W}$ into $\boldsymbol{\Theta}$ using (1) implies that $(\boldsymbol{\Theta}, \mathbf{Y})^\top$ is $\mathcal{JPNSN}$ distributed; this last remark can be used to easily simulate random samples from the $\mathcal{JPNSN}$.

Closure under marginalization of the $\mathcal{PN}$ and the $\mathcal{SSN}$ shows that any subset of $(\boldsymbol{\Theta}, \mathbf{Y})^\top$ is still $\mathcal{JPNSN}$ distributed (see Eq. (6)) and, as limit cases, $\boldsymbol{\Theta} \sim \mathcal{PN}_p(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$ and $\mathbf{Y} \sim \mathcal{SSN}_q[\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y, \text{diag}(\boldsymbol{\lambda})]$. The flexibility of the $\mathcal{PN}$ and $\mathcal{SSN}$ are then inherited by the marginal distributions of our proposal that allows also multivariate dependence between its components. Conditional densities are not standard, but from $(\mathbf{W}, \mathbf{Y})^\top | \mathbf{d} \sim \mathcal{N}_{2p+q}[\boldsymbol{\mu} + (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \boldsymbol{\Sigma}]$ we can easily see that

$$\boldsymbol{\Theta}|\mathbf{y}, \mathbf{d} \sim \mathcal{PN}_p[\boldsymbol{\mu}_w + \boldsymbol{\Sigma}_{wy}\boldsymbol{\Sigma}_y^{-1}\{\mathbf{y} - \boldsymbol{\mu}_y - \text{diag}(\boldsymbol{\lambda})\mathbf{d}\}, \boldsymbol{\Sigma}_w + \boldsymbol{\Sigma}_{wy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{wy}^\top]$$

and

$$\mathbf{Y}|\boldsymbol{\theta}, \mathbf{r} \sim \mathcal{SSN}_q[\boldsymbol{\mu}_y + \text{diag}(\boldsymbol{\lambda})\mathbf{d} + \boldsymbol{\Sigma}_{wy}^\top\boldsymbol{\Sigma}_w^{-1}(\mathbf{w} - \boldsymbol{\mu}_w), \boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_{wy}^\top\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}_{wy}].$$

$\mathcal{JPNSN}$ shapes are depicted in Figure 3.

Note that $\Theta_i \perp \Theta_j$, where $\perp$ indicates independence, if and only if $\mathbf{W}_i \perp \mathbf{W}_j$ and, by construction, matrix $\boldsymbol{\Sigma}_{wy}$ rules the circular-linear dependence since if $\mathbf{W}_i \perp Y_j$, then $\Theta_i \perp Y_j$. Parameters $\boldsymbol{\mu}_y$ and $\boldsymbol{\Sigma}_y$ are easily interpretable since $\text{E}(\mathbf{Y}) = \boldsymbol{\mu}_y + \boldsymbol{\lambda}\sqrt{2/\pi}$, $\text{var}(\mathbf{Y}) = \boldsymbol{\Sigma}_y + (1 - 2/\pi)\,\text{diag}(\boldsymbol{\lambda})\text{diag}(\boldsymbol{\lambda})$ and if $[\boldsymbol{\Sigma}_y]_{j,k} = 0$, where $[\boldsymbol{\Sigma}_y]_{j,k}$ represents the element positioned in the jth row and kth column, then Y_j and Y_k are independent. The skewness of the linear component is controlled by $\boldsymbol{\lambda}$ and Y_i is normally distributed if $\lambda_i = 0$. Parameters $\boldsymbol{\mu}_{w_i}$ and $\boldsymbol{\Sigma}_{w_i}$ determine the shape of the density of Θ_i, which is always $\mathcal{PN}$. It is not clear how changing one of the elements of $\boldsymbol{\mu}_{w_i}$ or $\boldsymbol{\Sigma}_{w_i}$ affects the density, but special cases exist: a circular uniform distribution is obtained with $\boldsymbol{\mu}_{w_i} = \mathbf{0}_2$ and $\boldsymbol{\Sigma}_{w_i} = d\mathbf{I}_2$, $\boldsymbol{\mu}_{w_i} = \mathbf{0}_2$ produces an antipodal density and $\mathcal{PN}(\boldsymbol{\mu}_{w_i}, d\mathbf{I}_2)$ is unimodal and symmetric. All the other statistics of the distribution that cannot be computed directly from the parameters, e.g., the circular mean, can be approximated with MC procedures.

3. Identifiability and Bayesian inference

Let $\mathbf{C}_w$ be a $2p \times 2p$ diagonal matrix with $\{2(i-1)+j\}$th entry equal to $c_i > 0$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, 2\}$. Then, since

$$\Theta_i = \text{atan}^*(W_{i2}/W_{i1}) = \text{atan}^*(c_i W_{i2}/c_i W_{i1}), \tag{7}$$

the two random vectors $\mathbf{W}_i \sim \mathcal{N}_2(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w)$ and $\mathbf{C}_w\mathbf{W}_i \sim \mathcal{N}_{2p}(\mathbf{C}_w\boldsymbol{\mu}_w, \mathbf{C}_w\boldsymbol{\Sigma}_w\mathbf{C}_w)$ produce the same $\boldsymbol{\Theta}$, i.e., the c_is cancel out in Eq. (7). It follows that $\{\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w\}$ and $\{\mathbf{C}_w\boldsymbol{\mu}_w, \mathbf{C}_w\boldsymbol{\Sigma}_w\mathbf{C}_w\}$ represent the same $\mathcal{PN}$ density which is then not identifiable.

The $\mathcal{JPNSN}$ is based on the $\mathcal{PN}$, which is also its circular marginal distribution, and it has the same identification issue; identifiability constraints on the parameter space are needed. Following and extending [40], we set to 1 the variance of each W_{i2} and from now on, to avoid confusion, we write $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{W}, \mathbf{R}\}$ as $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{W}}, \tilde{\mathbf{R}}\}$ when such constraints are imposed; $\boldsymbol{\lambda}$, $\mathbf{D}$, $\boldsymbol{\mu}_y$ and $\boldsymbol{\Sigma}_y$ are always identified since they are related only to the linear component. Let

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_\mathbf{w} & \mathbf{0}_{2p,q} \\ \mathbf{0}_{2p,q}^\top & \mathbf{I}_q \end{pmatrix},$$

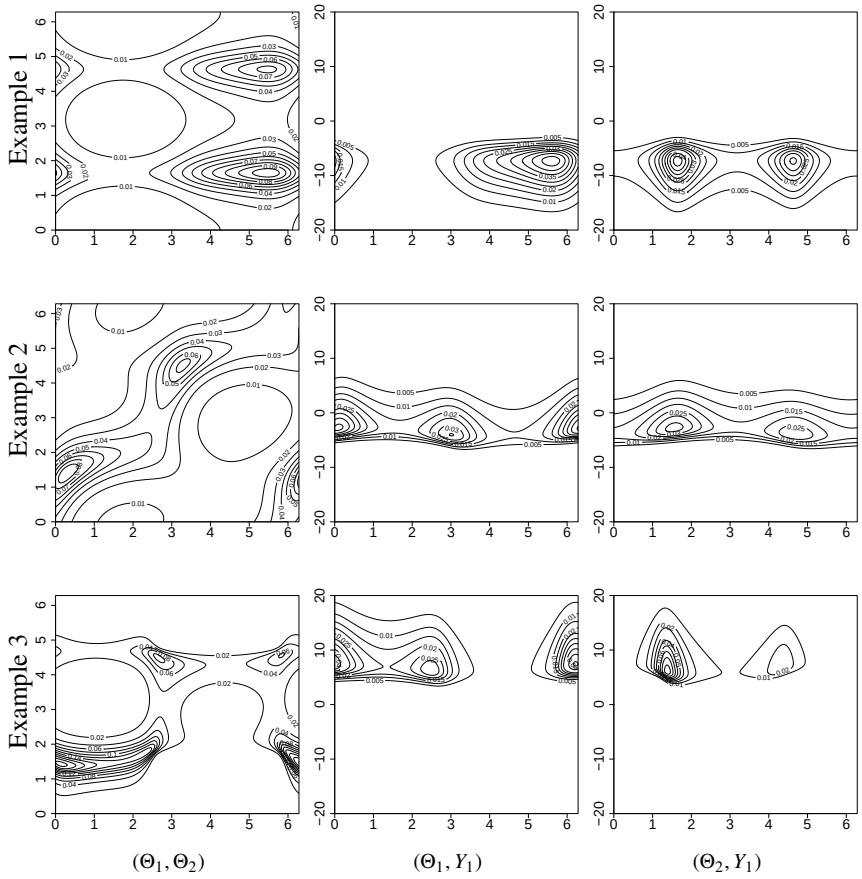


Figure 3: Bivariate marginal distributions of a $\mathcal{JPNSN}_{2,1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ under three sets of parameters (row) reported in Section 4.1. Displayed in the first column are the marginal distributions of (Θ_1, Θ_2); displayed in the second those of (Θ_1, Y_1), and in the third the marginals of (Θ_2, Y_1).

where $\mathbf{0}_{2p,q}^\top$ is a $2p \times q$ zero matrix, then the sets $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\lambda}\}$ and $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\}$ with $\boldsymbol{\mu} = \mathbf{C}\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma} = \mathbf{C}\tilde{\boldsymbol{\Sigma}}\mathbf{C}$ produce the same $\mathcal{JPNSN}$ density and the following relation holds:

$$\begin{aligned} f(\boldsymbol{\theta}, \tilde{\mathbf{r}}, \mathbf{y}, \mathbf{d}) &= 2^q \phi_{2p+q}[(\tilde{\mathbf{w}}, \mathbf{y})^\top | \tilde{\boldsymbol{\mu}} + (\mathbf{0}_{2p}, \operatorname{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \tilde{\boldsymbol{\Sigma}}] \phi_q(\mathbf{d}|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} \tilde{r}_i \\ &= 2^q \phi_{2p+q}[\mathbf{C}(\tilde{\mathbf{w}}, \mathbf{y})^\top | \mathbf{C}\tilde{\boldsymbol{\mu}} + (\mathbf{0}_{2p}, \operatorname{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \mathbf{C}\tilde{\boldsymbol{\Sigma}}\mathbf{C}] \phi_q(\mathbf{d}|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} c_i \tilde{r}_i. \end{aligned} \tag{8}$$

Note that there is a one-to-one relation between sets $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ and $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \mathbf{C}\}$ since $c_i = \sqrt{[\boldsymbol{\Sigma}]_{2i,2i}}$.

Due to the unavailability of MCMC algorithms for a constrained covariance matrix estimate, a computational problem arises and we show how to overcome it in the next section.

3.1. The MCMC algorithm

Suppose that T observations are drawn from a (p, q)-variate $\mathcal{JPNSN}$, i.e., $(\boldsymbol{\Theta}_t, \mathbf{Y}_t)^\top \sim \mathcal{JPNSN}_{p,q}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\lambda})$ for all $t \in \{1, \ldots, T\}$. As the $\mathcal{JPNSN}$ does not have a closed form density, we introduce $\tilde{\mathbf{R}}_t = (\tilde{R}_{t1}, \ldots, \tilde{R}_{tp})$ and $\mathbf{D}_1, \ldots, \mathbf{D}_1$ as latent variables for each $t \in \{1, \ldots, p\}$. Letting $g_1(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda}) g_2(\boldsymbol{\lambda})$ be the prior distribution, we want to evaluate the posterior of $\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_T$ and $\mathbf{D}_1, \ldots, \mathbf{D}_T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\lambda}$ given by

$$\frac{\prod_{t=1}^{T} 2^q \phi_{2p+q}\{(\tilde{\mathbf{w}}_t, \mathbf{y}_t)^\top | \tilde{\boldsymbol{\mu}} + (\mathbf{0}_{2p}, \operatorname{diag}(\boldsymbol{\lambda})\mathbf{d}_t)^\top, \tilde{\boldsymbol{\Sigma}}\} \phi_q(\mathbf{d}_t|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} \tilde{r}_{ti} g_1(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda}) g_2(\boldsymbol{\lambda})}{Z\{(\boldsymbol{\theta}_1, \mathbf{y}_1), \ldots, (\boldsymbol{\theta}_T, \mathbf{y}_T)\}}, \tag{9}$$

where $Z\{(\boldsymbol{\theta}_1, \mathbf{y}_1), \ldots, (\boldsymbol{\theta}_T, \mathbf{y}_T)\}$ is the normalization constant. Some difficulties arise in the definition of g_1 since its domain must contain the space of constrained nnd matrices and, to the best of our knowledge, no priors with such domain are available.

Our proposed MCMC algorithm starts by defining a prior $f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{\lambda})$ over $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. We indicate with $f^*(\mathbf{C}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda})$ the distribution over $\{\mathbf{C}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\}$ induced by $f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{\lambda})$ and we define $g_1(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda})$ as

$$g_1(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda}) = \int_{\mathbb{R}^+} \cdots \int_{\mathbb{R}^+} f^*(\mathbf{C}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda}) dc_1 \cdots dc_p. \tag{10}$$

Then, using (8) and (10) we can write (9) as

$$\begin{aligned} \int_{\mathbb{R}^+} \cdots \int_{\mathbb{R}^+} \prod_{t=1}^{T} 2^q \phi_{2p+q}\{\mathbf{C}(\tilde{\mathbf{w}}_t, \mathbf{y}_t)^\top | \mathbf{C}\tilde{\boldsymbol{\mu}} + (\mathbf{0}_{2p}, \mathrm{diag}(\boldsymbol{\lambda})\mathbf{d})^\top, \mathbf{C}\tilde{\boldsymbol{\Sigma}}\mathbf{C}\} \\ \times \frac{\phi_q(\mathbf{d}_t|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} c_i \tilde{r}_{ti} f^*(\mathbf{C}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}|\boldsymbol{\lambda}) g_2(\boldsymbol{\lambda})}{Z\{(\boldsymbol{\theta}_1, \mathbf{y}_1), \ldots, (\boldsymbol{\theta}_T, \mathbf{y}_T)\}} dc_1 \cdots dc_p, \end{aligned} \tag{11}$$

and if we transform $\{\mathbf{C}, \tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\}$ into $\{\mathbf{R}_1, \ldots, \mathbf{R}_T, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, the integrand of Eq. (11) becomes

$$\frac{\prod_{t=1}^{T} \phi_{2p+q}\{(\mathbf{w}_t, \mathbf{y}_t)^\top | \boldsymbol{\mu} + (\mathbf{0}_{2p}, \mathrm{diag}(\boldsymbol{\lambda})\mathbf{d}_t)^\top, \boldsymbol{\Sigma}\} \phi_q(\mathbf{d}_t|\mathbf{0}_q, \mathbf{I}_q) \prod_{i=1}^{p} r_{ti} f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{\lambda}) g_2(\boldsymbol{\lambda})}{Z\{(\boldsymbol{\theta}_1, \mathbf{y}_1), \ldots, (\boldsymbol{\theta}_T, \mathbf{y}_T)\}}. \tag{12}$$

Then, relying on standard MC integration rules (see, e.g., [7, 36]), a set of B draws from (9) is obtained by taking B samples of $\{\mathbf{R}_1, \ldots, \mathbf{R}_T, \mathbf{D}_1, \ldots, \mathbf{D}_T, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\}$ from (12) and transforming them to $\{\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_T, \mathbf{D}_1, \ldots, \mathbf{D}_T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\lambda}\}$.

In a schematic way, our proposal is

- to define a prior over $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\}$ that induces a prior g_1 (see Eq. (10));
- to obtain a set of samples of $\{\mathbf{R}_1, \ldots, \mathbf{R}_T, \mathbf{D}_1, \ldots, \mathbf{D}_T, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\}$ from distribution (12);
- to transform the posterior samples of $\{\mathbf{R}_1, \ldots, \mathbf{R}_T, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ into $\{\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\}$ after the model fitting.

The resulting posterior samples are from the distribution of interest, Eq. (9). The proposed MCMC algorithm can be used with the $\mathcal{JPNSN}$, the univariate projected normal ($q = 0$ and $p = 1$), the multivariate projected normal ($q = 0$) and also with the proposal of [14], i.e., a distribution defined over the K-dimensional sphere, since all of them share the same identification problem.

There are no restrictions on the choice of g_1 and g_2 but, as shown in Appendix B, if ease of implementation and conjugate priors are required, a normal inverse-Wishart ($\mathcal{NIW}$) can be used for $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ and a normal for λ; these are the ones we use in the examples of Section 4. Regardless of the priors chosen, the updates of $\mathbf{D}_t$ and R_{ti} can be done using Gibbs steps.

4. Examples

4.1. Synthetic data

The aim of these simulated examples is to show that the proposed MCMC algorithm is able to retrieve the parameters used to simulate the data and to solve the identification problem. We simulate three datasets with $p = 2$, $q = 1$, i.e., two circular variables and 1 linear, $T = 1000$ and parameters

$$\tilde{\boldsymbol{\mu}}_1 = \begin{bmatrix} 0.5 \\ -1.0 \\ -0.1 \\ 0.1 \\ -5.0 \end{bmatrix}, \quad \tilde{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.0 \\ 0.1 \\ -5.0 \end{bmatrix}, \quad \tilde{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.0 \\ 0.5 \\ 5.0 \end{bmatrix},$$

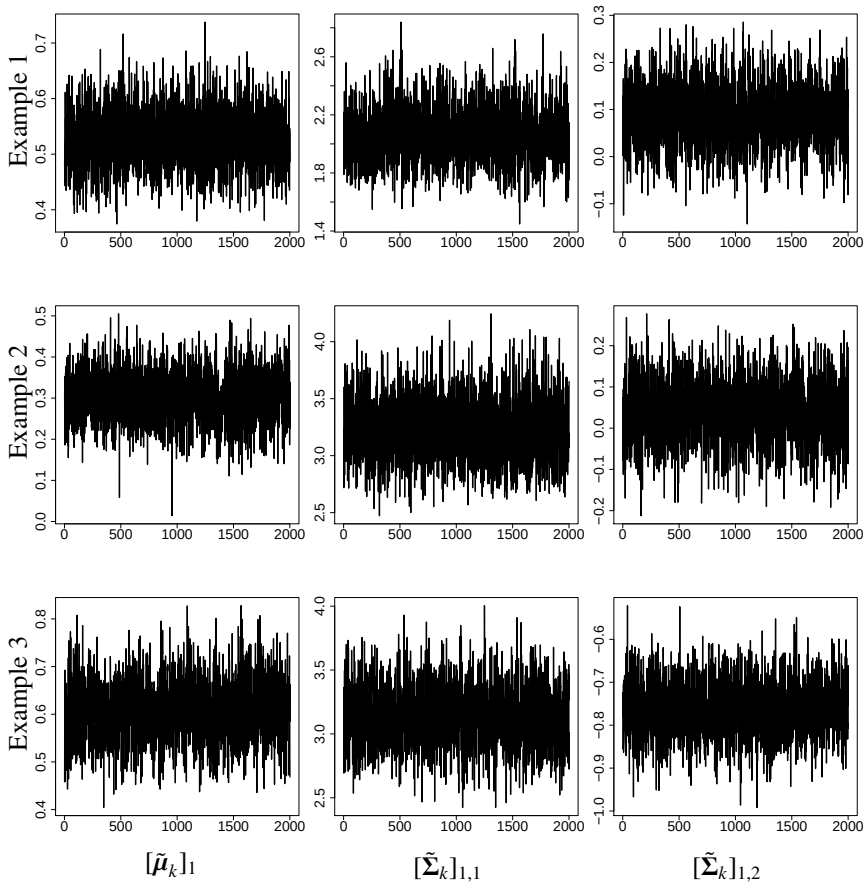


Figure 4: Simulated examples — trace plots of parameters $[\tilde{\mu}_k]_1$, $[\tilde{\Sigma}_k]_{1,1}$ and $[\tilde{\Sigma}_k]_{1,2}$ (columns) in the three examples (rows).

$\lambda_1 = -5, \lambda_2 = 5, \lambda_3 = 6,$

$$\tilde{\Sigma}_1 = \begin{bmatrix} 2 & 0 & 0.0 & 0 & 0 \\ 0 & 1 & 0.0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.0 & 1 & 0 \\ 0 & 0 & 0.0 & 0 & 2 \end{bmatrix}, \quad \tilde{\Sigma}_2 = \begin{bmatrix} 3.000 & 0.000 & 0.551 & 0.779 & 0.857 \\ 0.000 & 1.000 & -0.318 & 0.450 & 0.495 \\ 0.551 & -0.318 & 0.500 & 0.000 & -0.318 \\ 0.779 & 0.450 & 0.000 & 1.000 & 0.450 \\ 0.857 & 0.495 & -0.318 & 0.450 & 1.000 \end{bmatrix},$$

Table 1: Simulated examples — posterior mean estimates ($\hat{}$) and 95% credible intervals (CI) of $\tilde{\mu}_k$ and λ_k.

	Example		
	$k = 1$	$k = 2$	$k = 3$
$[\hat{\tilde{\mu}}_k]_1$	0.529	0.301	0.608
CI	$(0.428, 0.638)$	$(0.179, 0.427)$	$(0.480, 0.738)$
$[\hat{\tilde{\mu}}_k]_2$	-0.943	0.207	0.5
CI	$(-1.029, -0.858)$	$(0.137, 0.280)$	$(0.428, 0.576)$
$[\hat{\tilde{\mu}}_k]_3$	-0.112	-0.02	0.007
CI	$(-0.145, -0.080)$	$(-0.068, 0.029)$	$(-0.023, 0.036)$
$[\hat{\tilde{\mu}}_k]_4$	0.095	0.111	0.503
CI	$(0.020, 0.163)$	$(0.041, 0.178)$	$(0.429, 0.576)$
$[\hat{\tilde{\mu}}_k]_5$	-5.095	-4.765	4.705
CI	$(-5.415, -4.780)$	$(-4.925, -4.596)$	$(4.489, 5.105)$
$\hat{\lambda}_k$	-4.941	4.872	6.237
CI	$(-5.320, -4.561)$	$(4.630, 5.135)$	$(5.920, 6.556)$

Table 2: Simulated examples — posterior mean estimates (ˆ) and 95% credible intervals (CI) of $\tilde{\mathbf{\Sigma}}_k$.

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$[\hat{\tilde{\mathbf{\Sigma}}}_1]_{1,j}$	2.062	0.085	−0.031	0.011	−0.001
CI	(1.726, 2.462)	(−0.040, 0.214)	(−0.083, 0.021)	(−0.0980.121)	(−0.287, 0.278)
$[\hat{\tilde{\mathbf{\Sigma}}}_1]_{2,j}$	·	1	−0.008	−0.009	0.149
CI	(· ·)	(1 1)	(−0.046, 0.028)	(−0.094, 0.075)	(−0.060, 0.368)
$[\hat{\tilde{\mathbf{\Sigma}}}_1]_{3,j}$	·	·	0.201	0.002	0.007
CI	(· ·)	(· ·)	(0.168, 0.237)	(−0.039, 0.041)	(−0.079, 0.097)
$[\hat{\tilde{\mathbf{\Sigma}}}_1]_{4,j}$	·	·	·	1	0.047
CI	(· ·)	(· ·)	(· ·)	(1 1)	(−0.144, 0.248)
$[\hat{\tilde{\mathbf{\Sigma}}}_1]_{5,j}$	·	·	·	·	2.151
CI	(· ·)	(· ·)	(· ·)	(· ·)	(1.532, 2.908)
$[\hat{\tilde{\mathbf{\Sigma}}}_2]_{1,j}$	3.23	0.04	0.581	0.862	0.882
CI	(2.719, 3.793)	(−0.112, 0.191)	(0.467, 0.713)	(0.722, 1.015)	(0.619, 1.151)
$[\hat{\tilde{\mathbf{\Sigma}}}_2]_{2,j}$	·	1	−0.293	0.429	0.419
CI	(· ·)	(1 1)	(−0.354, −0.235)	(0.361, 0.492)	(0.284, 0.557)
$[\hat{\tilde{\mathbf{\Sigma}}}_2]_{3,j}$	·	·	0.521	0.034	−0.223
CI	(· ·)	(· ·)	(0.438, 0.617)	(−0.024, 0.095)	(−0.333, −0.120)
$[\hat{\tilde{\mathbf{\Sigma}}}_2]_{4,j}$	·	·	·	1	0.496
CI	(· ·)	(· ·)	(· ·)	(1 1)	(0.354, 0.645)
$[\hat{\tilde{\mathbf{\Sigma}}}_2]_{5,j}$	·	·	·	·	1.001
CI	(· ·)	(· ·)	(· ·)	(· ·)	(0.737, 1.316)
$[\hat{\tilde{\mathbf{\Sigma}}}_3]_{1,j}$	3.129	−0.762	0.38	0.623	0.834
CI	(2.684, 3.620)	(−0.889, −0.633)	(0.309, 0.455)	(0.469, 0.775)	(0.566, 1.132)
$[\hat{\tilde{\mathbf{\Sigma}}}_3]_{2,j}$	·	1	0.207	0.389	−0.176
CI	(· ·)	(1 1)	(0.177, 0.238)	(0.319, 0.458)	(−0.325, −0.021)
$[\hat{\tilde{\mathbf{\Sigma}}}_3]_{3,j}$	·	·	0.189	0.223	0.17
CI	(· ·)	(· ·)	(0.164, 0.216)	(0.193, 0.255)	(0.108, 0.240)
$[\hat{\tilde{\mathbf{\Sigma}}}_3]_{4,j}$	·	·	·	1	−0.337
CI	(· ·)	(· ·)	(· ·)	(1 1)	(−0.493, −0.188)
$[\hat{\tilde{\mathbf{\Sigma}}}_3]_{5,j}$	·	·	·	·	0.875
CI	(· ·)	(· ·)	(· ·)	(· ·)	(0.620, 1.165)

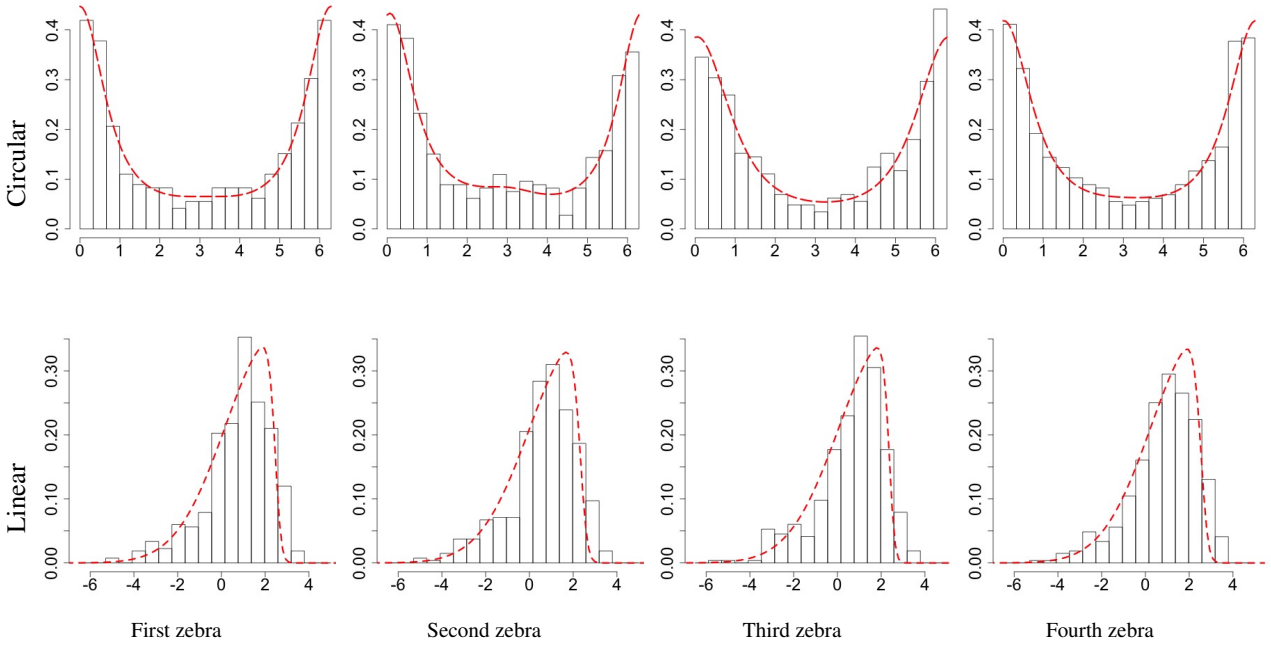


Figure 5: Zebras movement example — histograms of the observed data and posterior marginal densities of turning-angles (first row) and the logarithm of step-lengths (second row).

$$\tilde{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} 3.000 & -0.783 & 0.377 & 0.684 & 0.781 \\ -0.783 & 1.000 & 0.214 & 0.335 & -0.092 \\ 0.377 & 0.214 & 0.200 & 0.231 & 0.209 \\ 0.684 & 0.335 & 0.231 & 1.000 & -0.382 \\ 0.781 & -0.092 & 0.209 & -0.382 & 1.000 \end{bmatrix}.$$

The marginal bivariate densities are plotted in Figure 3. We chose the parameters so to have independent (first example) and dependent variables (second and third), highly skew linear densities and, at least, one bimodal circular marginal for each example.

In the three examples inference is carried out considering 40,000 iterations, burn-in 30,000, thin 5 and by taking 2000 posterior samples. As prior distributions we choose $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \sim \mathcal{NIW}(\mathbf{0}_5, 0.001, 15, \mathbf{I}_5)$ and $\lambda_k \sim \mathcal{N}_1(0, 100)$, that are standard weak informative priors. From Tables 1–2 we see that, with the exception of $[\hat{\tilde{\boldsymbol{\mu}}}_2]_5$, all true values are inside the associated 95% credible intervals (CIs), proving that our algorithm is able to estimate the $\mathcal{JPNSN}$ parameters. To further corroborate the validity of the proposed MCMC scheme in solving the identification problem, in Figure 4 we show, as examples, the trace plots of parameters $[\tilde{\boldsymbol{\mu}}_k]_1$, $[\tilde{\boldsymbol{\Sigma}}_k]_{1,1}$ and $[\tilde{\boldsymbol{\Sigma}}_k]_{1,2}$. These chains have reached their stationary distributions (we also checked it by using the R package `coda` [35]) with weak informative priors; this shows that the identification problem is no more relevant.

4.2. Zebras movements example

In this section we estimate the $\mathcal{JPNSN}$ parameters on an animal movement dataset taken from the movebank repository (`www.movebank.org`). Our aim is to show that the $\mathcal{JPNSN}$ can give information on the dependence of poly-cylindrical observations. Seven zebras are jointly observed in Botswana between the Okavango Delta and the Makgadikgadi Pans, and their hourly positions are recorded with GPS devices during the years 2007–09 [6]. In the observational period, the zebras migrate from the dry season habitat, i.e., the Okavango Delta, to the rainy season habitat, i.e., the Makgadikgadi Pans. We select data from four zebras, observed between November 18, 2008 and

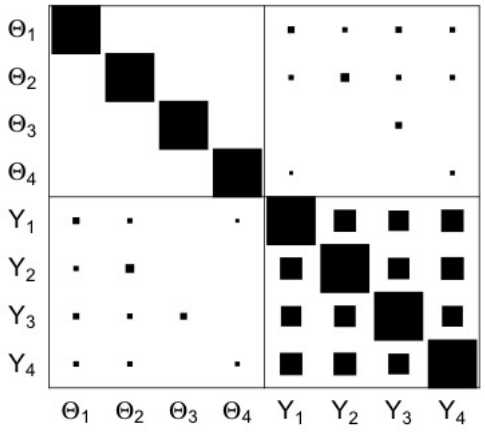


Figure 6: Zebras movement example — dependence matrix computed using Eqs. (13) (circular-circular), (14) (circular-linear) and the Pearson correlation coefficient (linear-linear). The size of the square is proportional to the posterior mean value. All values are positive.

February 18, 2009, when they have ended the migration. For each animal we compute the turning-angles and logarithm of step-lengths, having then poly-cylindrical observations composed of four circular and four linear variables. It is beyond the scope of this work to introduce complex models based on the $\mathcal{JPNSN}$, that are left to future developments, and we assume that observations are independent and identically distributed. For this reason, to mitigate temporal dependence we use data five hours apart, having then 442 observations for parameters estimate. The Pearson coefficient and the circular-circular correlation of [15], i.e.,

$$\rho_{(\Theta_i,\Theta_{i'})} = \frac{\mathrm{E}\{\sin(\Theta_i - \Theta_i^*)\sin(\Theta_{i'} - \Theta_{i'}^*)\}}{\sqrt{\mathrm{E}\{\sin^2(\Theta_i - \Theta_i^*)\}\mathrm{E}\{\sin^2(\Theta_{i'} - \Theta_{i'}^*)\}}} \in [-1, 1], \tag{13}$$

where Θ_i^* and $\Theta_{i'}^*$ are two circular variables distributed, respectively, as Θ_i and $\Theta_{i'}$, show that all the autocorrelations have values lower than 0.05 for the subset of data used. The histograms of the data can be seen in Figure 5.

The MCMC algorithm is implemented using the same number of iterations, thin and burn-in used in the previous section while $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\} \sim \mathcal{NIW}(\mathbf{0}_{12}, 0.001, 15, \mathbf{I}_{12})$ and $\boldsymbol{\lambda} \sim \mathcal{N}_4(0, 100\mathbf{I}_4)$. In Figures 5–6 we depicted, respectively, the marginal posterior $\mathcal{JPNSN}$ densities and the dependence matrix. The latter shows the MC estimates of the posterior mean circular-circular correlation of [15], the circular-linear dependence of [21], viz.

$$\rho^2_{(\Theta_i,Y_j)} = \frac{\mathrm{cor}(\cos\Theta_i, Y_j)^2 + \mathrm{cor}(\sin\Theta_i, Y_j)^2}{1 - \mathrm{cor}(\cos\Theta_i, \sin\Theta_i)} + \frac{-2\mathrm{cor}(\cos\Theta_i, Y_j)\mathrm{cor}(\sin\Theta_i, Y_j)\mathrm{cor}(\cos\Theta_i, \sin\Theta_i)}{1 - \mathrm{cor}(\cos\Theta_i, \sin\Theta_i)} \in [0, 1], \tag{14}$$

and linear-linear correlation, evaluated with the Pearson coefficient. A circular-circular and linear-linear correlation is plotted only if the associated CI does not contain zero while since the CI of $\rho^2(\Theta_i, Y_j)$ has zero probability to contain 0, we plot the value of $\rho^2(\Theta_i, Y_j)$ using a different rationality. More precisely, since $\boldsymbol{W}_i \perp Y_j$ iff $\Theta_i \perp Y_j$, in Figure 6 we plot the posterior mean value of $\rho^2(\Theta_i, Y_j)$ only if at least one of the CIs of $\boldsymbol{\Sigma}_{wy}$ that measure the correlation between $\boldsymbol{W}_i$ and Y_j does not contain 0. From Figure 5 we appreciate that the $\mathcal{JPNSN}$ is able to fit satisfactorily the data and to find significant circular-linear and linear-linear correlations (Figure 6).

4.3. Comparison with cylindrical distributions

With this section we want to show that ignoring multivariate dependence leads to loss of predictive ability. Then we compare our proposal with the cylindrical distribution of Abe–Ley [1] and a cylindrical version of the $\mathcal{JPNSN}$, i.e., for both we assume dependence between the variables belonging to the same animal and independence between

zebras. Since the $\mathcal{JPNSN}$ is not available in closed form, a comparison based on informational criteria, such as AIC or BIC, is not possible. We decided to make the comparison in terms of predictive ability measured used the CRPS, that is a proper scoring rule defined for both circular [12] and linear [11] variables that measure the distance between cumulative distribution functions [29]; lower values are then preferable. The Abe–Ley distribution is defined only for a positive linear variable and then, to make a fair comparison, we use the distribution that arises by taking the logarithm of its linear component, viz.

$$f(\theta_i, y_j) = \frac{\alpha^{AL}(\beta^{AL})^{\alpha^{AL}}}{2\pi\cosh\kappa^{AL}}\{1 + \lambda^{AL}\sin(\theta_i - \mu^{AL})\}e^{y_j(\alpha^{AL}-1)}e^{-(\beta^{AL}e^{y_j})^{\alpha^{AL}}\{1-\tanh\kappa^{AL}\cos(\theta_i-\mu^{AL})\}}e^{y_j}.$$

$\alpha^{AL} \in \mathbb{R}^+$ and $\beta^{AL} \in \mathbb{R}^+$ are linear scale and shape parameters, $\mu^{AL} \in [0, 2\pi)$ and $\lambda^{AL} \in [-1, 1]$ endorse the role of circular location and skewness parameters and $\kappa^{AL} \in \mathbb{R}^+$ plays the role of circular concentration and circular-linear dependence parameter. For the Abe–Ley parameters we use standard weak informative priors, i.e., an inverse gamma with parameters (1,1) for α^{AL}, β^{AL} and κ^{AL} while uniform distributions on the respective domains are used for μ^{AL} and λ^{AL}. Under the cylindrical $\mathcal{JPNSN}$, a $\mathcal{N}_1(0, 100)$ is used for the skewness parameters and a $\mathcal{NIW}(\mathbf{0}_3, 0.001, 6, \mathbf{I}_3)$ for the others that are the marginal priors deriving from the ones of the poly-cylindrical $\mathcal{JPNSN}$.

We select 10% of the circular and linear observations to be set aside and not used to estimate the posterior distributions. We predict their values based on the posterior samples and we measure how the models perform in terms of posterior estimates. Then, let $C_i \subset \{1, 2, \ldots, T\}$ and $\mathcal{L}_j \subset \{1, \ldots, T\}$ be sets of indices, where $t \in C_i$ if θ_{ti} is missing and $t \in \mathcal{L}_j$ if y_{tj} is missing, and let θ^b_{ti}, $t \in C_i$, and y^b_{tj}, $t \in \mathcal{L}_j$, be, respectively, the bth posterior sample of θ_{ti} and y_{tj}. An MC approximation of the CRPS for the circular variable based on B posterior samples is computed as

$$CRPS\, c_i \approx \frac{1}{B}\sum_{b=1}^{B} d(\theta_{ti}, \theta^b_{ti}) - \frac{1}{2B^2}\sum_{b=1}^{B}\sum_{b'=1}^{B} d(\theta^b_{ti}, \theta^{b'}_{ti}), \quad t \in C_i,$$

where $d(\cdot, \cdot)$ is the angular distance, while the CRPS for the linear variable is approximated by

$$CRPS\, \ell_j \approx \frac{1}{B}\sum_{b=1}^{B} |y_{tj} - y^b_{tj}| - \frac{1}{2B^2}\sum_{b=1}^{B}\sum_{b'=1}^{B} |y^b_{tj} - y^{b'}_{tj}|, \quad t \in \mathcal{L}_j.$$

We then compute the overall mean CRPSs for the sets of circular and linear variables and we use these indices to measure the goodness of fit.

Circular and linear CRPSs have values 0.383 and 0.693 for the $\mathcal{JPNSN}$, 0.385 and 0.762 for the cylindrical $\mathcal{JPNSN}$, and 0.412 and 0.753 for the Abe–Ley density, showing that the $\mathcal{JPNSN}$ performs better and, moreover, it is also able to give a measure of dependence between all the circular and linear components (Figure 6) that is not possible with cylindrical distributions.

5. Concluding remarks

In this work we introduced a poly-cylindrical distribution. The proposal is highly flexible, it is closed under marginalization and it allows to have dependent components, bimodal marginal circular distributions and asymmetric linear ones. We showed how the MCMC algorithm, used to obtain posterior samples, can be easily implemented using only Gibbs steps. The proposal suffers from an identification problem and we showed how to overcome it with a post-processing of posterior samples that can also be used with the $\mathcal{PN}$ distribution. With the aim to prove the validity of our sampling scheme, the algorithm was applied to simulated examples. Then the proposed distribution was used to model real data taken from the movebank data repository. The predictive ability of our proposal was compared with the ones of cylindrical distributions, showing that the $\mathcal{JPNSN}$ performs better.

Future work will lead us to use the $\mathcal{JPNSN}$ as emission distribution in a hidden Markov model and to incorporate covariates to model mean and covariance of the circular-linear observations.

Acknowledgments

The author wishes to thank Antonello Maruotti, Giovanna Jona Lasinio, Alessio Pollice, the Editor-in-Chief, the Associate Editor and the two anonymous reviewers for comments that greatly improved the manuscript. This work is partially developed under the PRIN2015 supported-project Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT) funded by MIUR (Italian Ministry of Education, University and Scientific Research).

Appendix

A. The invariance property of the $\mathcal{PN}$

Here we prove that the univariate marginal density of the circular variables is invariant. Let $\Theta_i^* = \delta(\Theta_i + \xi)$, where $\delta \in \{-1, 1\}$ and $\xi \in [0, 2\pi)$, following Theorem 1 of [26], the density of $\Theta_i \sim \mathcal{PN}_1(\boldsymbol{\mu}_{w_i}, \boldsymbol{\Sigma}_{w_i})$, i.e., f_{Θ_i}, has the invariant property if $f_{\Theta_i^*}$, i.e., the density of Θ_i^*, belongs to the same parametric family of f_{Θ_i}.

The random variables Θ_i^* can be written as

$$\Theta_i^* = \text{atan}^* \frac{\sin \Theta_i^*}{\cos \Theta_i^*} = \text{atan}^* \frac{\sin\{\delta(\Theta_i + \xi)\}}{\cos\{\delta(\Theta_i + \xi)\}} = \text{atan}^* \frac{\delta \sin(\Theta_i + \xi)}{\cos(\Theta_i + \xi)}, \tag{A.1}$$

and using relations $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ and $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$, Eq. (A.1) can be stated equivalently as

$$\Theta_i^* = \text{atan}^* \frac{\delta(R_i \sin\Theta_i \cos\xi + R_i \cos\Theta_i \sin\xi)}{R_i \cos\Theta_i \cos\xi - R_i \sin\Theta_i \sin\xi} = \text{atan}^* \frac{\delta(W_{i2} \cos\xi + W_{i1} \sin\xi)}{W_{i1} \cos\xi - W_{i2} \sin\xi}.$$

To prove that Θ_i^* is $\mathcal{PN}$ distributed, let us consider the random variable $\mathbf{W}_i^* = \Delta\mathbf{T}\mathbf{W}_i$, where $\Delta = \text{diag}\{(1, \delta)^\top\}$ and

$$\mathbf{T} = \begin{pmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{pmatrix}.$$

$\mathbf{W}_i^*$ is normally distributed and Eq. (1) applied to $\mathbf{W}_i^*$ gives

$$\text{atan}^* \frac{W_{i2}^*}{W_{i1}^*} = \text{atan}^* \frac{\delta(W_{i2} \cos\xi + W_{i1} \sin\xi)}{W_{i1} \cos\xi - W_{i2} \sin\xi} = \Theta_i^*.$$

Then Θ_i^* follows a projected normal distribution; this proves the invariance of the $\mathcal{PN}$. □

B. MCMC implementation details

Sampling $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The full conditional of $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ is proportional to

$$\prod_{t=1}^{T} \phi_{2p+q}\{(\mathbf{w}_t, \mathbf{y}_t)^\top - (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d}_t)^\top | \boldsymbol{\mu}, \boldsymbol{\Sigma}\} f(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{\lambda}). \tag{B.2}$$

Eq. (B.2) is equivalent to the full conditional of the mean and covariance matrix in a model with iid normally distributed observations. If we assume $f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{\lambda}) \equiv f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_0 + 2p + q)/2 - 1} \exp\left\{ -\frac{\text{tr}(\boldsymbol{\Psi}_0 \boldsymbol{\Sigma}^{-1}) + \kappa_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)}{2} \right\}$$

i.e., $f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a $\mathcal{NIW}(\boldsymbol{\mu}_0, \kappa_0, \nu_0, \boldsymbol{\Psi}_0)$, where $\kappa_0 > 0$ and $\nu_0 > 2p + q - 1$ are real numbers, $\boldsymbol{\mu}_0 \in \mathbb{R}^{2p+q}$ and $\boldsymbol{\Psi}_0$ is a $(2p + q) \times (2p + q)$ nnd matrix, and we let $\boldsymbol{\eta}_t = (\mathbf{w}_t, \mathbf{y}_t)^\top - (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d}_t)^\top$ and

$$\bar{\boldsymbol{\eta}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\eta}_t,$$

the full conditional is $\mathcal{NIW}(\boldsymbol{\mu}_{\text{post}}, \kappa_{\text{post}}, \nu_{\text{post}}, \boldsymbol{\Psi}_{\text{post}})$ with

$$\boldsymbol{\mu}_{\text{post}} = \frac{\kappa_0 \boldsymbol{\mu}_0 + T\bar{\boldsymbol{\eta}}}{\kappa_0 + T}, \quad \kappa_{\text{post}} = \kappa_0 + T, \quad \nu_{\text{post}} = \nu_0 + T,$$
$$\boldsymbol{\Psi}_{\text{post}} = \boldsymbol{\Psi}_0 + \sum_{t=1}^{T} (\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})^\top + \frac{\kappa_0 T}{\kappa_0 + T} (\bar{\boldsymbol{\eta}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{\eta}} - \boldsymbol{\mu}_0)^\top.$$

Sampling $\boldsymbol{\lambda}$. The full conditional of $\boldsymbol{\lambda}$ is proportional to

$$\prod_{t=1}^{T} \phi_q\{\mathbf{y}_t | \boldsymbol{\mu}_{y_t|w_t} + \text{diag}(\mathbf{d}_t)\boldsymbol{\lambda}, \boldsymbol{\Sigma}_{y|w}\} g_2(\boldsymbol{\lambda}), \tag{B.3}$$

where $\boldsymbol{\mu}_{y_t|w_t} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{wy}^\top \boldsymbol{\Sigma}_w^{-1} (\mathbf{w}_t - \boldsymbol{\mu}_w)$ and $\boldsymbol{\Sigma}_{y|w} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{wy}^\top \boldsymbol{\Sigma}_w^{-1} \boldsymbol{\Sigma}_{wy}$. In (B.3) we can see $\boldsymbol{\lambda}$ as a vector of regression coefficients, where the matrix of covariates is $\text{diag}(\mathbf{d}_t)$. Then, standard results tell us that a normal $g_2(\boldsymbol{\lambda})$ induces a normal full conditional. More precisely, let $\boldsymbol{\lambda} \sim \mathcal{N}_q(\boldsymbol{\gamma}_0, \boldsymbol{\Omega}_0)$, then the full conditional is $\mathcal{N}_q(\boldsymbol{\gamma}_{\text{post}}, \boldsymbol{\Omega}_{\text{post}})$ with

$$\boldsymbol{\Omega}_{\text{post}} = \left\{ \sum_{t=1}^{T} \text{diag}(\mathbf{d}_t) \boldsymbol{\Sigma}_{y|w}^{-1} \text{diag}(\mathbf{d}_t) + \boldsymbol{\Omega}_0^{-1} \right\}^{-1}, \quad \boldsymbol{\gamma}_{\text{post}} = \boldsymbol{\Omega}_{\text{post}} \left\{ \sum_{t=1}^{T} \text{diag}(\mathbf{d}_t) \boldsymbol{\Sigma}_{y|w}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_{y_t|w_t}) + \boldsymbol{\Omega}_0^{-1} \boldsymbol{\gamma}_0 \right\}.$$

Sampling $\mathbf{D}_t$. The full conditional of the latent vector $\mathbf{D}_t$ is proportional to

$$\phi_q\{\mathbf{y}_t | \boldsymbol{\mu}_{y_t|w_t} + \text{diag}(\boldsymbol{\lambda})\mathbf{d}_t, \boldsymbol{\Sigma}_{y|w}\} \phi_q(\mathbf{d}_t | \mathbf{0}_q, \mathbf{I}_q).$$

The vector $\mathbf{d}_t$ can be seen as a vector of (positive) regressors with $\text{diag}(\boldsymbol{\lambda})$ as matrix of covariates and $\phi_q(\mathbf{d}_t | \mathbf{0}_q, \mathbf{I}_q)$ as prior. The full conditional is then $\mathcal{N}_q(\mathbf{M}_{d_t}, \mathbf{V}_q) I_{\mathbf{0_q},\infty}$, where $\mathcal{N}_q(\cdot,\cdot) I_{\mathbf{0_q},\infty}$ is a q-dimensional truncated normal distribution with components having support $\mathbb{R}^+$,

$$\mathbf{V}_d = (\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{y|w}^{-1} \boldsymbol{\Lambda} + \mathbf{I}_q)^{-1}$$

and

$$\mathbf{M}_{d_t} = \mathbf{V}_d \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{y|w}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_{y_t|w_t}).$$

Sampling R_{ti}. Let $\mathbf{u}_{ti} = (\cos\theta_{ti}, \sin\theta_{ti})^\top$, $A_{ti} = \mathbf{u}_{ti}^\top \boldsymbol{\Sigma}_{w_{ti}|w_{t-i},y}^{-1} \mathbf{u}_{ti}$ and $B_{ti} = \mathbf{u}_{ti}^\top \boldsymbol{\Sigma}_{w_{ti}|w_{t-i},y}^{-1} \boldsymbol{\mu}_{w_{ti}|w_{t-i},y_t}$, where $\boldsymbol{\mu}_{w_{ti}|w_{t-i},y_t}$ and $\boldsymbol{\Sigma}_{w_{ti}|w_{t-i},y}$ are the conditional mean and covariance matrix of $\mathbf{w}_{ti}$ assuming $(\mathbf{w}_t, \mathbf{y}_t)^\top \sim \mathcal{N}_{2p+q}[\boldsymbol{\mu} + (\mathbf{0}_{2p}, \text{diag}(\boldsymbol{\lambda})\mathbf{d}_t)^\top, \boldsymbol{\Sigma}]$. The full conditional of R_{ti} is then proportional to

$$r_{ti} \exp\left\{ -\frac{1}{2} A_{ti} \left(r_{ti} - \frac{B_{ti}}{A_{ti}} \right)^2 \right\}. \tag{B.4}$$

Eq. (B.4) is the same full conditional of the latent variable of the spherical $\mathcal{PN}$ of [14] and then we can use their *slice sampling* strategy to sample from it. In detail, if

$$v_{ti} \sim \mathcal{U}\left[0, \exp\left\{ -\frac{1}{2} A_{ti} \left(r_{ti} - \frac{B_{ti}}{A_{ti}} \right)^2 \right\} \right], \quad v_{ti}^* \sim \mathcal{U}(0,1),$$

then

$$r_{ti} = \sqrt{(\varrho_{2ti}^2 - \varrho_{1ti}^2) v_{ti}^* + \varrho_{1ti}^2},$$

with

$$\varrho_{1ti} = \frac{B_{ti}}{A_{ti}} + \max\left(-\frac{B_{ti}}{A_{ti}}, -\sqrt{\frac{-2\ln v_{ti}}{A_{ti}}} \right), \quad \varrho_{2ti} = \frac{B_{ti}}{A_{ti}} + \sqrt{\frac{-2\ln v_{ti}}{A_{ti}}},$$

is distributed accordingly to the full conditional (B.4).

References

[1] T. Abe, C. Ley, A tractable, parsimonious and flexible model for cylindrical data, with applications, Econometrics and Statistics 4 (2017) 91–104.
[2] C. Anderson-Cook, An extension to modeling cylindrical variables, Statist. Probab. Lett. 35 (1997) 215–223.
[3] R. Arellano-Valle, H. Bolfarine, V. Lachos, Bayesian inference for skew-normal linear mixed models, J. Appl. Stat. 34 (2007) 663–682.
[4] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.
[5] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[6] H. L. A. Bartlam-Brooks, P. S. A. Beck, G. Bohrer, S. Harris, In search of greener pastures: Using satellite images to predict the effects of environmental change on zebra migration, J. Geophys. Res. Biogeosciences 118 (2013) 1427–1437.
[7] S. Brooks, A. Gelman, G. Jones, X. Meng, Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, London, 2011.
[8] J. Bulla, F. Lagona, A. Maruotti, M. Picone, A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series, J. Agric. Biol. Environ. Statist. 17 (2012) 544–567.
[9] A. D'Elia, A statistical model for orientation mechanism, Statist. Methods Appl. 10 (2001) 157–174.
[10] R. Gatto, S. R. Jammalamadaka, The generalized von mises distribution, Statist. Methodol. 4 (2007) 341–353.
[11] T. Gneiting, A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, J. Am. Statist. Assoc. 102 (2007) 359–378.
[12] E. P. Grimit, T. Gneiting, V. J. Berrocal, N. A. Johnson, The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, Q. J. Royal Meteorol. Soc. 132 (2006) 2925–2942.
[13] A. K. Gupta, G. González-Farias, J. Dominguez-Molina, A multivariate skew normal distribution, J. Multivar. Anal. 89 (2004) 181–190.
[14] D. Hernandez-Stumpfhauser, F. J. Breidt, M. J. van der Woerd, The general projected normal distribution of arbitrary dimension: modeling and Bayesian inference., Bayesian Anal. 12 (2017) 113-133.
[15] S. Jammalamadaka, Y. Sarma, A correlation coefficient for angular variables, Statist. Theor. Data Anal. II (1988) 349–364.
[16] S. R. Jammalamadaka, A. SenGupta, Topics in Circular Statistics, World Scientific, Singapore, 2001.
[17] R. A. Johnson, T. E. Wehrly, Some angular-linear distributions and related regression models, J. Am. Statist. Assoc. 73 (1978) 602–606.
[18] M. C. Jones, A. Pewsey, Sinh-arcsinh distributions, Biometrika 96 (2009) 761–780.
[19] I. D. Jonsen, J. M. Flemming, R. A. Myers, Robust state-space modeling of animal movement data, Ecol. 86 (2005) 2874–2880.
[20] F. Lagona, M. Picone, A latent-class model for clustering incomplete linear and circular data in marine studies, J. Data Sci. 9 (2011) 585–605.
[21] K. V. Mardia, Linear-circular correlation coefficients and rhythmometry, Biometrika 63 (1976) 403-405.
[22] K. V. Mardia, P. E. Jupp, Directional Statistics, John Wiley and Sons, Chichester, 1999.
[23] K. V. Mardia, T. W. Sutton, A model for cylindrical variables with applications, J. Royal Stat. Soc. 40 (1978) 229–233.
[24] A. Maruotti, A. Punzo, G. Mastrantonio, F. Lagona., A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure, Stoch. Environ. Res. Risk. Assess. 30 (2016) 1725–1740.
[25] G. Mastrantonio, G. Jona Lasinio, A. E. Gelfand, Spatio-temporal circular models with non-separable covariance structure 25 (2016) 331–350.
[26] G. Mastrantonio, G. Jona Lasinio, A. Maruotti, G. Calise, Invariance properties and statistical inference for circular data, Statist. Sinica To appear.
[27] G. Mastrantonio, A. Maruotti, G. Jona Lasinio, Bayesian hidden Markov modelling using circular-linear general projected normal distribution, Environmetrics 26 (2015) 145–158.
[28] G. Mastrantonio, G. Calise, Hidden Markov model for discrete circular-linear wind data time series, J. Statist. Comput. Simul. 86 (2016) 2611–2624
[29] J. E. Matheson, R. L. Winkler, Scoring rules for continuous probability distributions, Manag. Sci. 22 (1976) 1087–1096.
[30] J. M. Morales, P. R. Moorcroft, J. Matthiopoulos, J. L. Frair, J. G. Kie, R. A. Powell, E. H. Merrill, D. T. Haydon, Building the bridge between animal movement and population dynamics, Philos. Trans. Royal Soc. B 365 (2010) 2289–2301.
[31] J. M. Morales, D. T. Haydon, J. Frair, K. E. Holsinger, J. M. Fryxell, Extracting more out of relocation data: building movement models as mixtures of random walks, Ecol. 85 (2004) 2436–2445.
[32] N. M. Olmos, H. Varela, H. W. Gómez, H. Bolfarine, An extension of the half-normal distribution, Statist. Papers 53 (2012) 875–886.
[33] T. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen, J. Matthiopoulos, State-space models of individual animal movement, Trends Ecol. Evol. 23 (2008) 87–94.
[34] A. Pewsey, M. Neuhäuser, G. D. Ruxton, Circular Statistics in R, Oxford University Press, Croydon, 2013.
[35] M. Plummer, N. Best, K. Cowles, K. Vines, Coda: Convergence diagnosis and output analysis for MCMC, R News 6 (2006) 7–11.
[36] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
[37] S. K. Sahu, D. K. Dey, M. D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, Can. J. Statist. 31 (2003) 129–150.
[38] K.-F. Storch, O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong, C. J. Weitz, Extensive and divergent circadian gene expression in liver and heart, Nature 417 (2002) 78–83.
[39] F. Wang, A. E. Gelfand, G. Jona Lasinio, Joint spatio-temporal analysis of a linear and a directional variable: space-time modeling of wave heights and wave directions in the Adriatic sea, Statist. Sinica 25 (2015) 25–39.
[40] F. Wang, A. E. Gelfand, Directional data analysis under the general projected normal distribution, Statist. Methodol. 10 (2013) 113–127.
[41] F. Wang, A. E. Gelfand, Modeling space and space-time directional data using projected Gaussian processes, J. Am. Statist. Assoc. 109 (2014) 1565–1580.