



Contents lists available at SciVerse ScienceDirect

## Statistical Methodology

journal homepage: [www.elsevier.com/locate/stamet](http://www.elsevier.com/locate/stamet)

# Directional data analysis under the general projected normal distribution

Fangpo Wang<sup>\*</sup>, Alan E. Gelfand

Department of Statistical Science, Duke University, Durham, NC 27708-0251, United States

## ARTICLE INFO

### Article history:

Received 17 December 2011

Received in revised form

20 July 2012

Accepted 24 July 2012

### Keywords:

Bivariate normal distribution

Circular data

Concentration

Latent variables

Markov chain Monte Carlo

Mean direction

## ABSTRACT

The projected normal distribution is an under-utilized model for explaining directional data. In particular, the general version provides flexibility, e.g., asymmetry and possible bimodality along with convenient regression specification. Here, we clarify the properties of this general class. We also develop fully Bayesian hierarchical models for analyzing circular data using this class. We show how they can be fit using MCMC methods with suitable latent variables. We show how posterior inference for distributional features such as the angular mean direction and concentration can be implemented as well as how prediction within the regression setting can be handled. With regard to model comparison, we argue for an out-of-sample approach using both a predictive likelihood scoring loss criterion and a cumulative rank probability score criterion.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Directional data arise naturally in many scientific fields where observations are recorded as directions or angles relative to a system with a fixed orientation. Examples of circular data include wind directions, wave directions or directions of animal movement. A different type of circular data arises from converting periodic time data to angular measurements, for example, time data measured on a 24 h clock can be converted with 0:00 corresponding to 0 and 24:00 to  $2\pi$ . The circular nature of such data creates difficulties in applying ordinary statistical methods used for inline data. For example, with angular observations  $1^\circ$ ,  $0^\circ$  and  $359^\circ$ , it is inappropriate to use the arithmetic mean, which is  $120^\circ$ . (Of course, the variance and other *linear* features suffer from similar problems.) A more sensible measurement is the angular mean direction (formally defined in Section 2) which is  $0^\circ$ . Over the last fifty years, substantial techniques and methods have been developed paralleling those for linear data [12,14,7,1]. Lee [11] gives a recent review of statistical methods for analyzing circular data.

<sup>\*</sup> Corresponding author.

E-mail addresses: [fw19@stat.duke.edu](mailto:fw19@stat.duke.edu) (F. Wang), [alan@stat.duke.edu](mailto:alan@stat.duke.edu) (A.E. Gelfand).

Corresponding to the usual parametric distributions for linear variables, there are several parametric distributions for circular variables. The most commonly used circular distribution is the von Mises distribution (also known as circular normal distribution). The von Mises distribution is the maximum entropy distribution on the circle for a fixed mean direction and dispersion and can be considered as an analogue of the normal distribution on the real line. The von Mises distribution  $M(\mu, \kappa)$  has two parameters: the mean direction  $\mu$  and the concentration parameter  $\kappa$ . It is unimodal and symmetric about its mode which is the same as the mean direction  $\mu$ . The mean resultant length has the expression  $I_1(\kappa)/I_0(\kappa)$ , where  $I_v$  is the modified Bessel function of the first kind of order  $v$ .

Circular distributions can also be obtained by radial projection of bivariate distributions on the plane. For any random vector  $\mathbf{Y} \in \mathbb{R}^2$  with  $\Pr(\mathbf{Y} = \mathbf{0}) = 0$ ,  $\mathbf{U} = \mathbf{Y}/\|\mathbf{Y}\|$  is a random point on the unit circle. The angle between the unit vector to this point and, say the  $x$ -axis becomes a random angle. An important distribution within this family is the projected normal distribution, where  $\mathbf{Y}$  follows a bivariate normal distribution  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . The projected normal distribution, denoted as  $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , is also referred to as an *offset normal* distribution [12,7] or an *angular Gaussian* distribution [22]. The projected normal distribution with identity covariance matrix, also known as the *displaced normal* [10], has attracted some work in the literature; see references below. However, like the von Mises, it suffers the limitation of symmetry and unimodality. For general  $\boldsymbol{\Sigma}$ , the projected normal distribution will be asymmetric and can be bimodal, and is thus a more flexible family compared to the von Mises distribution. However, due to the unwieldy expression for the density function (see (1) below), the general version has received little attention in the literature.

The contribution of this paper is to demonstrate how to handle circular data using the general projected normal distribution  $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . We show how to implement flexible circular regression modeling with linear covariates under the assumption of a general projected normal distribution. We work in a fully Bayesian framework and provide full inference including mean direction, concentration, regression coefficients, and predictive distributions with highest posterior density (HPD) arcs to produce credible sets. We also argue for the use of out-of-sample model comparison, employing a log likelihood scoring rule, a continuous rank probability score rule, and the mean length of HPD intervals of the predictive distribution.

The regression model with circular response and linear covariates is referred to as Linear–Circular regression [7]. Previously proposed circular regression models are mostly based on the assumption of the von Mises distribution [4,8,2]. Gould [4] sets  $\mu_i = \mu_0 + \mathbf{x}_i' \boldsymbol{\beta}$  and  $\kappa_i = \kappa$  for  $i = 1, \dots, n$ , allowing the mean of von Mises distribution to vary by levels of covariates. Johnson and Wehrly [8] consider only one covariate and propose the location model and concentration model, where a link function is used to convert the covariate to a value between 0 and 1. The location model takes  $\mu_i = \mu_0 + 2\pi F(x_i)$  and  $\kappa_i = \kappa$ , where  $F$  is chosen to be a known cumulative distribution function. Fisher and Lee [2] further generalize the Johnson–Wehrly models by a monotone link function  $g(\cdot)$  mapping the linear predictor  $\mathbf{x}_i' \boldsymbol{\beta}$  to  $(-\pi, \pi)$ , with  $g(0) = 0$ . (A common choice for  $g(\cdot)$  is the arctan function.) Their location model takes  $\mu_i = \mu_0 + g(\mathbf{x}_i' \boldsymbol{\beta})$  and  $\kappa_i = \kappa$ . However, computational difficulty arises when fitting a Fisher–Lee model with more than one covariate due to the multimodality of the likelihood.

Returning to the projected normal model with identity covariance matrix, Presnell et al. [17] introduce the spherically projected multivariate linear model based on the assumption of a projected normal distribution. Their paper demonstrates that the projected normal distribution has convenient regression specification and obtains the maximum likelihood estimates for the circular regression model using the EM algorithm. Extending Presnell et al. [17], Nuñez-Antonio and Gutiérrez-Peña [15] and Nuñez-Antonio et al. [16] show how to fit the projected normal distribution in a Bayesian framework using a Gibbs sampler, without and with covariates, respectively. Hernandez-Stumpfhauser et al. [6] build a mixed effect model based on the projected normal distribution by including the factors into the mean vector  $\boldsymbol{\mu}$ . Again, the above papers set the covariance matrix of the projected normal  $\boldsymbol{\Sigma}$  to be  $\mathbf{I}$ , arguing that this avoids an identifiability issue, that it is a rich enough model, and that it provides computational convenience. Again, our goal is to demonstrate the benefits of working with the more general  $\boldsymbol{\Sigma}$ .

The format of the paper is as follows. Section 2 illustrates the general projected normal distribution and its properties. Section 3 demonstrates how to implement model fitting, inference and prediction

and proposes the criteria for model comparison. Section 4 provides simulation examples and real data examples on directions of animal movements. Section 5 offers a summary and future directions.

## 2. The general projected normal model

In the sequel, we shall use  $\phi_r$  and  $\Phi_r$  to denote the pdf and cdf of a  $r$ -dimensional standard normal distribution. Assume a random vector  $\mathbf{Y} = (Y_1, \dots, Y_r)'$  follows an  $r \geq 2$ -dimensional multivariate normal distribution, with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The corresponding random unit vector  $\mathbf{U} = \mathbf{Y}/\|\mathbf{Y}\|$  is said to follow a *projected normal* distribution [18,14] with the same parameters, denoted as  $PN_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Here, we confine ourselves to the circular projected normal distribution, the special case when  $r = 2$ . Let  $\phi_2(y_1, y_2; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  be the density function of the bivariate normal distribution with mean vector  $\boldsymbol{\mu} = (\mu_1, \mu_2)'$  and covariance matrix  $\boldsymbol{\Sigma}$ . Let  $\sigma_1^2$  and  $\sigma_2^2$  be the variances of  $Y_1$  and  $Y_2$  and  $\rho$  denote the correlation between the two variables. The corresponding covariance matrix has the expression,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

In this case, we convert  $\mathbf{Y}$ , equivalently  $\mathbf{U}$  to a random angular variable  $\Theta$  through  $U_1 = \cos \Theta$  and  $U_2 = \sin \Theta$  whence  $\tan \Theta = Y_2/Y_1 = U_2/U_1$  which exists with probability 1. Inversion for  $\Theta$  requires care since the period for the tan function is only  $\pi$ . The usual choice is the  $\arctan^*$  function. Following [7, pp. 13],  $\arctan^*(S/C)$  is formally defined as  $\arctan(S/C)$  if  $C > 0, S \geq 0$ ;  $\pi/2$  if  $C = 0, S > 0$ ;  $\arctan(S/C) + \pi$  if  $C < 0$ ;  $\arctan(S/C) + 2\pi$  if  $C \geq 0, S < 0$ ; undefined if  $C = 0, S = 0$ . The resulting density function,  $f(\theta; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ , is [12, pp. 52],

$$f(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\phi_2(\mu_1, \mu_2; \mathbf{0}, \boldsymbol{\Sigma})}{C(\theta)} + \frac{aD(\theta)\Phi_1\{D(\theta)\}\phi_1\left[a\{C(\theta)\}^{-\frac{1}{2}}(\mu_1 \sin \theta - \mu_2 \cos \theta)\right]}{C(\theta)}, \quad (1)$$

where  $\phi$  and  $\Phi$  are defined at the beginning of this section, and

$$a = \left\{ \sigma_1\sigma_2\sqrt{1-\rho^2} \right\}^{-1},$$

$$C(\theta) = a^2(\sigma_2^2 \cos^2 \theta - \rho\sigma_1\sigma_2 \sin 2\theta + \sigma_1^2 \sin^2 \theta),$$

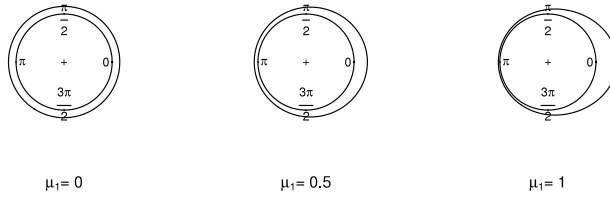
$$D(\theta) = a^2\{C(\theta)\}^{-\frac{1}{2}}\{\mu_1\sigma_2(\sigma_2 \cos \theta - \rho\sigma_1 \sin \theta) + \mu_2\sigma_1(\sigma_1 \sin \theta - \rho\sigma_2 \cos \theta)\}.$$

Because of the complicated form of this density function, the general version of the projected normal has only been considered theoretically. Data analysis and inference in the literature has been confined to the case of the projected normal  $PN_2(\boldsymbol{\mu}, \mathbf{I})$  [17,15,6,16]. In this case,  $a = 1$ ,  $C(\theta) = 1$  and  $D(\theta) = \mu_1 \cos \theta + \mu_2 \sin \theta$ .

With the identity covariance matrix, the corresponding density function is symmetric and unimodal. Thus  $PN_2(\boldsymbol{\mu}, \mathbf{I})$  and the von Mises distribution  $M(\mu, \kappa)$  each have two parameters and they can be aligned by matching the mean direction and the mean resultant length. When  $\mu_1 = \mu_2 = 0$ , the distribution reduces to the uniform distribution on the circle.

The mean direction for the projected normal is the unit vector,  $\mathbf{E}(\mathbf{U})/\|\mathbf{E}(\mathbf{U})\|$  and the associated mean resultant length,  $\gamma = \|\mathbf{E}(\mathbf{U})\| \leq 1$ , provides a concentration. The angular mean direction,  $\omega$ , arises from  $\mathbf{E}(\mathbf{U}) = (E \cos \Theta, E \sin \Theta)'$ . That is,  $\cos \omega = E \cos \Theta / \gamma$ ,  $\sin \omega = E \sin \Theta / \gamma$  and  $\tan \omega = E \sin \Theta / E \cos \Theta = E(U_2)/E(U_1)$ . Thus, since  $Y_1 = \|\mathbf{Y}\| \cos \Theta$  and  $Y_2 = \|\mathbf{Y}\| \sin \Theta$ , we can see that, in general,

$$\tan \omega = \frac{E(Y_2/\|\mathbf{Y}\|)}{E(Y_1/\|\mathbf{Y}\|)}.$$



**Fig. 1.** Density of  $PN_2$  with  $\mu_2 = 0$ ,  $\Sigma = \mathbf{I}$  for different values of  $\mu_1$ .

In [Appendix A](#) we show that when  $\Sigma = \mathbf{I}$ ,  $\tan \omega = \mu_2/\mu_1$  but that this is not true for general  $\Sigma$ . For instance, with  $\Sigma = \mathbf{I}$ , for  $\mu_1 = 0$  and  $\mu_2 \neq 0$ , the angular mean direction is  $\pi/2$  if  $\mu_2$  is positive and is  $3\pi/2$  if  $\mu_2$  is negative. When  $\mu_1 \neq 0$  and  $\mu_2 = 0$ , the angular mean direction is 0 if  $\mu_1$  is positive and is  $\pi$  if  $\mu_1$  is negative. To our knowledge, a closed form expression for the angular mean direction in the general case  $\mu_1 \neq 0$  and  $\mu_2 \neq 0$  is not available. When  $\Sigma = \mathbf{I}$ , Kendall [10] provides a formula for the mean resultant length (or concentration),  $\gamma = \|\mu\|$ , i.e.,  $\gamma = (\pi\zeta/2)^{1/2} \exp(-\zeta)(I_0(\zeta) + I_1(\zeta))$  where  $\zeta = \gamma^2/4$ , and  $I_v$  is the modified Bessel function of the first kind of order  $v$ . Again, a closed form expression in the general case is not available. However, both mean direction and mean resultant length can be computed using numerical integration or Monte Carlo methods. Below, we obtain these in a regression context using Monte Carlo methods.

Special cases of  $PN_2(\mu, \mathbf{I})$  with varying parameter  $\mu_1$  are shown in [Fig. 1](#). In general, with increasing  $\gamma$ , the density becomes increasingly peaked.

Another special case of a general projected normal distribution is the angular central Gaussian distribution  $PN_2(\mathbf{0}, \Sigma)$ . The angular central Gaussian is symmetric and antipodal, with two equal modes in exactly opposite directions. When  $\sigma_1 = \sigma_2 = \sigma$ , the density reduces to

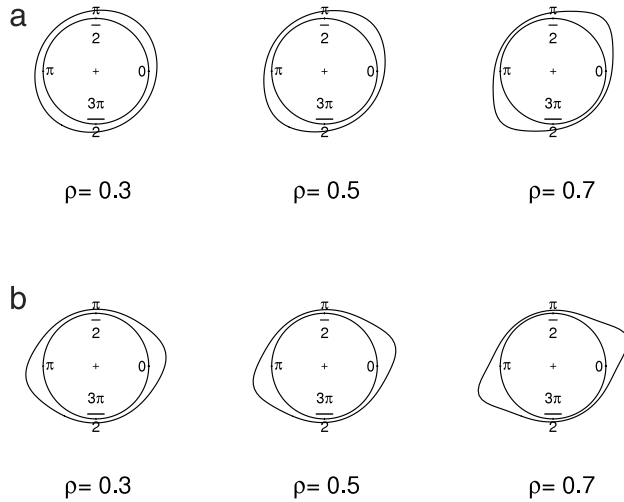
$$f(\theta|0, 0, \sigma, \rho) = \frac{\sqrt{1 - \rho^2}}{2\pi(1 - \rho \sin 2\theta)},$$

revealing symmetry relative to the increasing  $45^\circ$ -line when  $\rho > 0$ , relative to the decreasing  $45^\circ$ -line when  $\rho < 0$ . Equivalently,  $f(\theta|0, 0, \sigma, -\rho) = f(\theta + \pi/2|0, 0, \sigma, \rho)$ .

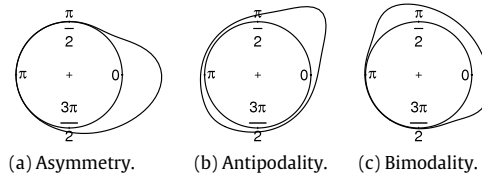
Let  $\Sigma = V\Lambda V'$  be an eigen-decomposition of  $\Sigma$  with  $V$  the matrix of unit eigenvectors and  $\Lambda$  the diagonal matrix of eigenvalues. Thus the bivariate random vector  $\mathbf{Y} \sim V\Lambda^{1/2}N(\mathbf{0}, \mathbf{I})$ , where  $V$  is the rotation matrix. For any  $\rho > 0$  and  $\sigma_1 = \sigma_2$ , the rotation matrix is  $V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$  and the modes occur at  $\pi/4$  and  $5\pi/4$ ; for any  $\rho < 0$  and  $\sigma_1 = \sigma_2$ , the rotation matrix is  $V = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$  and the modes occur at  $3\pi/4$  and  $7\pi/4$ . In the case of  $\sigma_1 = \sigma_2$ , the parameter  $\rho$  changes the concentration of the circular density, but not the modal axis (see [Fig. 2\(a\)](#)). When  $\sigma_1 \neq \sigma_2$ , no simple rotations are available (see [Fig. 2\(b\)](#)). The angular central Gaussian distribution always produces a symmetric and antipodal density, except in the uniform case where  $\sigma_1 = \sigma_2$  and  $\rho = 0$ .

In general, the shape of the  $PN_2(\mu, \Sigma)$  density with  $\mu \neq \mathbf{0}$  and  $\Sigma \neq \mathbf{I}$  can be asymmetric or possibly bimodal (antipodality with different amplitudes or more general bimodality but no more than two modes) (see [Fig. 3](#)). The five parameters  $\mu_1, \mu_2, \sigma_1, \sigma_2$  and  $\rho$  together determine the shape of the density. However, unlike in the previous special cases, it is not clear how changing a single parameter affects the shape of the density. This suggests that, in the general case, interpretation of the parameters can be difficult. However, if the goal is flexibility and prediction, this specification can be attractive.

Since the distribution of  $\mathbf{U}$ , hence,  $\Theta$  does not change if the random vector  $\mathbf{Y}$  is scaled by any  $a > 0$ , a further constraint on the parameters is needed for identifiability. To fix the scale of  $\Sigma$ , without loss of generality, we set  $\sigma_2 = 1$ , so that the covariance matrix  $\Sigma$  becomes  $\Sigma = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}$ , and thus, we have a four parameter ( $\mu_1, \mu_2, \tau$  and  $\rho$ ) distribution.



**Fig. 2.** Density of  $PN_2$  with  $\mu_1 = \mu_2 = 0$ , (a)  $\sigma_1 = \sigma_2$ , (b)  $\sigma_1 = 2\sigma_2$  for different values of  $\rho$ .



**Fig. 3.** Shapes for the general projected normal distribution.

In many applications, one encounters asymmetric circular data so that fitting them with a unimodal and symmetric model may result in misleading conclusions. In Section 4 below, we revisit the classical turtle movement dataset [14]; this dataset, shown in Fig. 6, provides an example of circular data where an assumption of symmetry and unimodality does not seem appropriate. Asymmetry and bimodality can also be obtained by mixing two von Mises distributions, with directly interpretable parameters. However, such mixtures need to be identified and parameter estimation can be difficult if the component modes are not well-separated; see [9,19] for parameter estimation based on numerical optimization. Moreover, if the goal is to capture asymmetry, it is not clear that a two-component mixture model is an appropriate specification. The general projected normal distribution provides an alternative which is straightforward to fit (within a Bayesian framework) and allows assessment of departure from symmetry and unimodality through the departure of  $\Sigma$  from  $\mathbf{I}$ .

The regression specification under the general projected model is  $\Theta_i \stackrel{\text{ind}}{\sim} PN_2(B^T \mathbf{X}_i, \Sigma)$ , with regression coefficient matrix  $B_{2 \times p} = (\beta_1, \dots, \beta_p)$  and covariates  $\mathbf{X}_i = (1, X_{i,1}, \dots, X_{i,p-1})'$ . In this model, changing the levels of the covariates is equivalent to changing the mean vector  $\mu_i$ . With a fixed covariance matrix, different levels of the covariates can result in very different predictive distributions. However, smoothness of  $\mathbf{X}$  in  $\mu_1(\mathbf{X})$  and  $\mu_2(\mathbf{X})$  induces smoothness in  $PN_2(B^T \mathbf{X}, \Sigma)$  over  $\mathbf{X}$  and thus smoothness in the mean directions and concentrations, as we illustrate in Section 4.3 below.

### 3. Bayesian inference for the general PN

With conditionally independent observations, fitting a general projected normal model in the Bayesian framework is straightforward; we illustrate both without and with covariates.

### 3.1. Model fitting

As discussed in the previous section, the covariance matrix is set to be  $\Sigma = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}$ , so that, with no covariates, we have four parameters to estimate:  $\mu_1, \mu_2, \tau$  and  $\rho$ . For data  $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ , the likelihood is  $\prod_{i=1}^n f(\theta_i)$ , where  $f(\theta_i)$  has the expression shown in (1). A prior is added for  $\mu_1, \mu_2, \tau$  and  $\rho$ . In implementing MCMC to fit the model, it is easiest to introduce a latent realization  $r_i$  associated with  $\theta_i$ , whence a vector  $\mathbf{r} = (r_1, r_2, \dots, r_n)$ . The joint density of  $(\Theta, R)$  can be obtained by changing variables from  $\mathbf{Y} = (Y_1, Y_2)'$  to  $(R, \Theta)'$ . The likelihood under this data augmentation is:

$$\prod_{i=1}^n f(r_i, \theta_i | \boldsymbol{\mu}, \Sigma) = (2\pi)^{-n} |\Sigma|^{-\frac{n}{2}} \times \exp \left( -\frac{\sum_{i=1}^n [(r_i \mathbf{u}_i - \boldsymbol{\mu})' \Sigma^{-1} (r_i \mathbf{u}_i - \boldsymbol{\mu})]}{2} \right) \prod_{i=1}^n r_i \quad (2)$$

where  $\mathbf{u}_i = (\cos \theta_i, \sin \theta_i)'$ . Conjugacy for  $\boldsymbol{\mu}$  arises under a bivariate normal prior, e.g.,  $\boldsymbol{\mu} \sim N(\mathbf{0}, \lambda_0 \mathbf{I})$ . However, there is still no closed form for the full conditional posteriors of  $\tau^2, \rho$  and the augmented parameter  $\mathbf{r}$ . For  $\tau^2$ , we choose the prior as an inverse Gamma  $IG(a_\tau, b_\tau)$  with mean  $b_\tau/(a_\tau - 1) = 1$ ; for  $\rho$ , a uniform distribution between  $-1$  and  $1$ . For the MCMC implementation, the full conditional distributions are given in Appendix B.

For the PN regression model with  $p - 1$  covariates, the parameters to infer about are the regression coefficients  $B_{2 \times p}^T = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$ , along with  $\tau^2$  and  $\rho$ . The likelihood under the data augmentation becomes:

$$\prod_{i=1}^n f(r_i, \theta_i | B, \Sigma) = (2\pi)^{-n} |\Sigma|^{-\frac{n}{2}} \prod_{i=1}^n r_i \times \exp \left( -\frac{\sum_{i=1}^n [(r_i \mathbf{u}_i - B^T \mathbf{x}_i)' \Sigma^{-1} (r_i \mathbf{u}_i - B^T \mathbf{x}_i)]}{2} \right) \quad (3)$$

where  $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,p-1})'$ . For convenience, we partition the coefficient matrix  $B_{2 \times p}^T$  as  $B^T = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$ , with  $\boldsymbol{\beta}_j \in \mathbb{R}^2, j = 1, \dots, p$ . Again, a Gaussian prior is chosen for  $\boldsymbol{\beta}_j$ , e.g.,  $\boldsymbol{\beta}_j \sim N(\mathbf{0}, \lambda_0 \mathbf{I})$ . The full conditionals are again provided in Appendix B.

### 3.2. Inference

Using MCMC model fitting, we obtain posterior samples for the parameters  $\boldsymbol{\mu}$  and  $\Sigma$ , or  $B$  and  $\Sigma$  for the density estimation or regression model, respectively.

In the “no covariates” case, the posterior density estimate is  $f(\theta | \text{data}) = \int f(\theta | \Psi) f(\Psi | \text{data}) d\Psi$ , where  $\Psi = \{\boldsymbol{\mu}, \Sigma\}$ . As a mixture of general PN densities, its form is very flexible. Posterior samples of the parameters provide the usual Rao-Blackwellized estimator. That is, given a posterior sample of the parameters  $\mu_1^{(g)}, \mu_2^{(g)}, (\tau^2)^{(g)}$  and  $\rho^{(g)}, g = 1, \dots, G$ , we evaluate  $f(\theta | \boldsymbol{\mu}^{(g)}, \Sigma^{(g)})$  (expression (1)) on a uniform grid of points on  $[0, 2\pi)$ . Then, averaging over  $g$  yields an estimate of the posterior at the resolution of the grid points.

As previously indicated, closed form expressions for the mean direction and the mean resultant length of the projected normal distribution are only available for the special case  $PN_2(\boldsymbol{\mu}, \mathbf{I})$ . MCMC Bayesian model fitting enables inference for the mean direction and mean resultant length without requiring analytical forms; in fact, we obtain posterior samples. According to the definition, the mean direction is  $\omega = \arctan^* \left( \frac{\beta}{\alpha} \right)$  and the resultant length is  $\gamma = \sqrt{\alpha^2 + \beta^2}$ , where  $\alpha =$

$E \cos \Theta$  and  $\beta = E \sin \Theta$ . Thus, for each posterior draw,  $\mu_1^{(g)}, \mu_2^{(g)}, (\tau^2)^{(g)}$  and  $\rho^{(g)}$ , we obtain  $\alpha^{(g)}$  and  $\beta^{(g)}$  by Monte Carlo integration. Therefore, we obtain the posterior samples  $\omega^{(g)}$  and  $\gamma^{(g)}$  as  $\arctan^*(\beta^{(g)}/\alpha^{(g)})$  and  $\sqrt{(\alpha^{(g)})^2 + (\beta^{(g)})^2}$ , respectively. Illustrative posterior distributions are provided below.

It is worth noting that  $\omega$  and  $\gamma$  are parametric functionals associated with  $f(\theta|\mu, \Sigma)$  and we are obtaining the posterior distribution associated with these functionals. If we wanted  $\omega$  and  $\gamma$  associated with  $f(\theta|\text{data})$ , we are seeking constants and it would be easiest to obtain sample estimates by taking draws from this distribution. This would be done by drawing  $\theta^{(g)}$  from  $f(\theta|\mu^{(g)}, \Sigma^{(g)})$  created as above. In fact, we would draw  $\mathbf{Y}^{(g)}$  from  $N_2(\mu^{(g)}, \Sigma^{(g)})$  and convert to  $\theta^{(g)}$ . We recall the usual method-of-moments estimates, i.e., for each  $\theta^{(g)}$ , obtain  $(\cos \theta^{(g)}, \sin \theta^{(g)})$ . Compute  $\bar{C} =$

$$\Sigma_g \cos \theta^{(g)} / G \text{ and } \bar{S} = \Sigma_g \sin \theta^{(g)} / G. \text{ Then, } \hat{\omega} = \arctan^* \frac{\bar{S}}{\bar{C}} \text{ and } \hat{\gamma} = \sqrt{\bar{C}^2 + \bar{S}^2}.$$

For the PN regression model, besides estimation of the regression coefficients, we are also interested in prediction at new levels of the covariates say,  $\mathbf{x}^*$ . However, given the posterior sample of parameters and the covariates, the values of the density function  $f(\theta|\mathbf{x}^*, B^{(g)}, \Sigma^{(g)})$  are calculated on a uniform grid over  $[0, 2\pi)$ . By taking the average of the density values on each grid over the samples, we obtain the predictive distribution given the covariates  $\mathbf{x}^*$ . Again, as a mixture of general PN densities, its shape is very flexible. In particular, equal tail credible intervals may be inappropriate; HPD credible intervals are more sensible. In general, the HPD arc is not trivial to calculate but it can be obtained numerically.

### 3.3. Model comparison using proper scoring rules

For the PN regression model, we focus on the the performance of out-of-sample prediction. We fit models using the training data  $\theta = (\theta_1, \dots, \theta_n)'$ , and evaluate models using a holdout set  $\theta^* = (\theta_1^*, \dots, \theta_m^*)'$ . As previously discussed, given the covariate levels,  $\mathbf{x}^*$ , we can calculate the length of the HPD credible interval and also see whether the holdout value  $\theta^*$  falls in this HPD region.

However, if predictive distributions are a goal, then model performance should be judged with regard to the entire predictive distribution. We need to compare the predictive distribution at  $\mathbf{x}^*$  with the held out  $\theta^*$ . Two proper scoring rules for doing this are the *predictive log scoring loss* (PLSL) [3] and the *continuous ranked probability score* (CRPS) as in [5]. Using posterior samples under (3), for the former, we calculate the quantity

$$\text{PLSL} = -\frac{2}{G} \sum_{g=1}^G \sum_{j=1}^m \log f(\theta_j^* | \mathbf{x}_j^*, B^{(g)}, \Sigma^{(g)}).$$

As a model comparison criterion, the model with smaller PLSL is favored.

For the latter, Gritti et al. [5] developed the continuous ranked probability score (CRPS) for circular variables,

$$\text{CRPS}_{\text{circ}}(F, \theta^*) = E\{\alpha(\Theta, \theta^*)\} - \frac{1}{2} E\{\alpha(\Theta, \Theta')\},$$

where  $F$  is the circular predictive distribution,  $\theta^*$  is the holdout value,  $\alpha$  is the angular distance and  $\Theta$  and  $\Theta'$  are independent copies of a circular random variable with distribution  $F$ . The angular distance  $\alpha(\theta, \theta^*)$  employed here is formally defined as,  $|\theta - \theta^*|$  if  $|\theta - \theta^*| \leq \pi$ ;  $2\pi - |\theta - \theta^*|$  if  $\pi \leq |\theta - \theta^*| < 2\pi$  (see [5, pp. 2939]). Although we do not have an explicit expression for the predictive distribution,  $F$ , the CRPS is readily approximated in the Bayesian framework. With posterior samples  $\theta_i^{(g)}$  from  $PN_2((B^{(g)})^T \mathbf{x}_i^*, \Sigma^{(g)})$ , the CRPS associated with the  $i$ -th holdout value can be approximated by

$$\text{CRPS}_i = \frac{1}{G} \sum_{g=1}^G \alpha(\theta_i^{(g)}, \theta_i^*) - \frac{1}{2G^2} \sum_{g=1}^G \sum_{k=1}^G \alpha(\theta_i^{(g)}, \theta_i^{(k)}).$$

Finally, we compare the models by taking average CRPS over holdout samples. Again, the model with the smaller CRPS is favored. Both criteria are utilized in Section 4.



## 4. Examples

Having shown how to fit a general PN model in the Bayesian framework using MCMC, we now present several illustrations. In Section 4.1 we consider simulated examples while in Sections 4.2 and 4.3 we consider two real datasets. The first is the classic turtle movement dataset (mentioned in Section 2), which we use to illustrate density estimation; the second dataset concerns butterfly orientations [21,17], which we use to illustrate circular regression with linear covariates.

### 4.1. Simulation: Density estimation and PN regression

First, we illustrate the density estimation using samples generated from a skewed PN distribution. The parameter values used in this example are  $\mu_1 = -0.19$ ,  $\mu_2 = 1.25$ ,  $\sigma_1 = 1.58$ ,  $\sigma_2 = 1.4$  and  $\rho = -0.84$ ; the distribution is shown as the solid line in Fig. 4. Four different sample sizes were used,  $n = 200, 500, 1000$  and  $2000$ . We fit the general projected normal and displaced normal ( $\Sigma = \mathbf{I}$ ) as described in Section 3 and display the posterior estimates, using a set of grid points, in Fig. 4 (only shown for  $n = 200$  and  $2000$ ). In addition, we fit a two component mixture of von Mises by utilizing the *movMF* function in the R package “*movMF*”. The result is also shown in Fig. 4. We see that the projected normal recovers the true distribution quite well, even for the smaller sample size. The displaced normal and the von Mises mixture are not adequate. Posterior samples of the mean direction and mean resultant length for all four sample sizes are obtained, following Section 3.2, and the resultant posterior density plots are shown in Fig. 5. The figure provides results from fitting an illustrative sample for each size  $n$ .

Next, we turn to the regression problem, employing simulation examples with focus on prediction and model comparison. We simulate  $\theta_i$ ,  $i = 1, \dots, n+m$ , from a general projected normal distribution  $PN_2(B^T \mathbf{x}_i, \Sigma)$ , where  $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,p-1})'$ . From many simulation examples we have run, to illustrate here, we choose two  $\tau^2$  values, 1 and 2.56, keeping  $\rho = .4$  with  $n = 60$  and  $200$ . We take  $p = 3$  where the covariates  $x_{i,1}$  and  $x_{i,2}$  are generated uniformly on  $(-1, 1)$  and the regression coefficients are set to be  $\beta_1 = (2, -2)'$ ,  $\beta_2 = (1, -1)'$  and  $\beta_3 = (1, 1)'$ . From Section 3, the prior for  $\tau^2$  is  $IG(5, 4)$  and for the  $\beta$ 's,  $\lambda_0 = 0.5$ . In the absence of useful prior information, the priors were made vague; sensitivity analyses (not presented) showed the inference to be essentially unchanged. The performance of prediction is evaluated on a validation set of size  $m = 100$ . We run Metropolis–Hastings within the Gibbs sampler for 25 000 iterations with a burn-in of 5000, thinning the remainder by collecting every 5th sample.

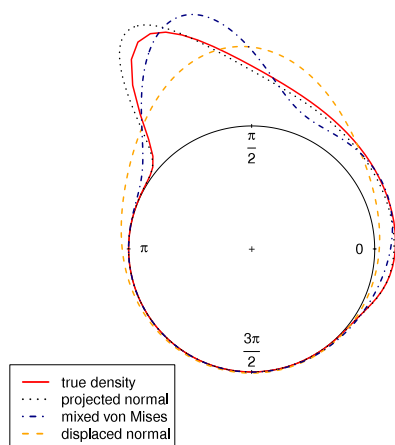
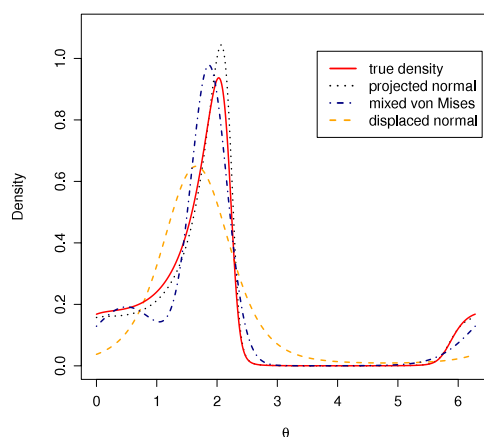
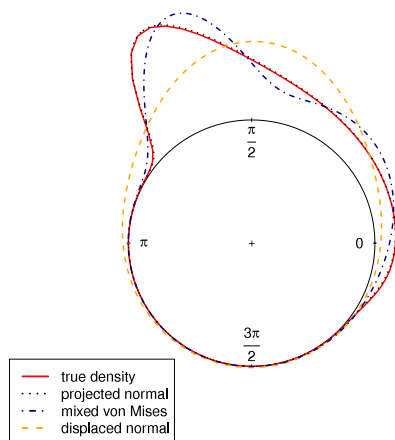
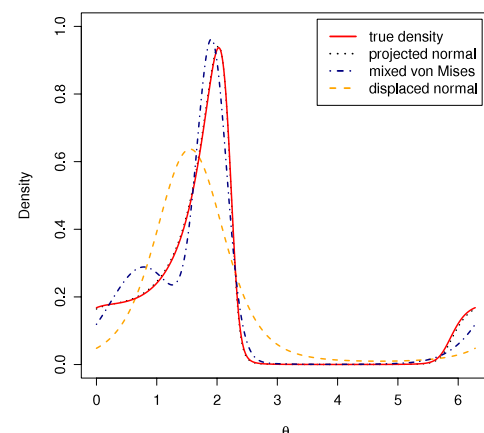
For each of the four cases, both the general projected normal regression  $PN(B^T \mathbf{x}_i, \Sigma)$  and the displaced normal regression  $PN(B^T \mathbf{x}_i, \mathbf{I})$  are fitted for comparison. The posterior summaries for the parameters are shown in Table 1. As we would expect, the general PN outperforms the displaced PN. The latter gives point estimates that are less well-centered with credible intervals that are too tight. It performs relatively worse for the larger  $\tau^2$  and for the larger training sample size. We see the benefit of using  $\Sigma \neq \mathbf{I}$  when this is the case.

To compare predictive performance, we look at the average length of predictive 95% credible intervals over all holdout samples and also the percentage of holdout values falling into their corresponding 95% credible intervals. We also compute the predictive log scoring loss (PLSL) and the average predictive CRPS. The results are summarized in Table 2. For the smaller training set, the results are a bit mixed, perhaps not surprising. However, for the larger sample size, PLSL and average CRPS (which is known to be conservative) favor the general version; so does the average length of the predictive interval. We also see that the PLSL and average CRPS decrease with sample size and increase as the concentration decreases.

### 4.2. Turtle movement

We consider a turtle dataset due to Dr. E. Gould of the Johns Hopkins School of Hygiene. The data are the directions of turtle movements after a certain treatment and can be found in [20, pp. 27] or [14, pp. 9]. The circular plot of the raw data is given in Fig. 6. We can see that, while most of the turtles moved in one main direction, a small portion of them moved in another, almost opposite direction.



(a.1) on the circle  $n = 200$ .(b.1) inline  $n = 200$ .(a.2) on the circle  $n = 2000$ .(b.2) inline  $n = 2000$ .**Fig. 4.** Simulation results of density estimation.

We fit the general PN and displaced normal and report the posterior mean of the density in Fig. 6(a). The prior used here follows Section 3.1:  $IG(5, 4)$  for  $\tau^2$  and  $N(\mathbf{0}, 0.5\mathbf{I})$  for  $\mu$ . We also fit a single von Mises using the *mle.vonmises* function of R package “circular”, which is also shown in Fig. 6(a). It is clear that the data can not be adequately described by a single symmetric and unimodal distribution.

Stephens [20] fitted a special case of a two component von Mises, with the same concentration and modes  $\pi$  radians apart. Mardia [13] later fitted a general two component mixture with different concentrations and modes. In Fig. 6(b), we plot the posterior mean density estimate from our PN model together with the density based on the parameter estimates of [13]. The general PN distribution provides a density estimation comparable with the mixture of two von Mises distributions. The latter does have five parameters, compared with four for the general PN, and also introduces potential

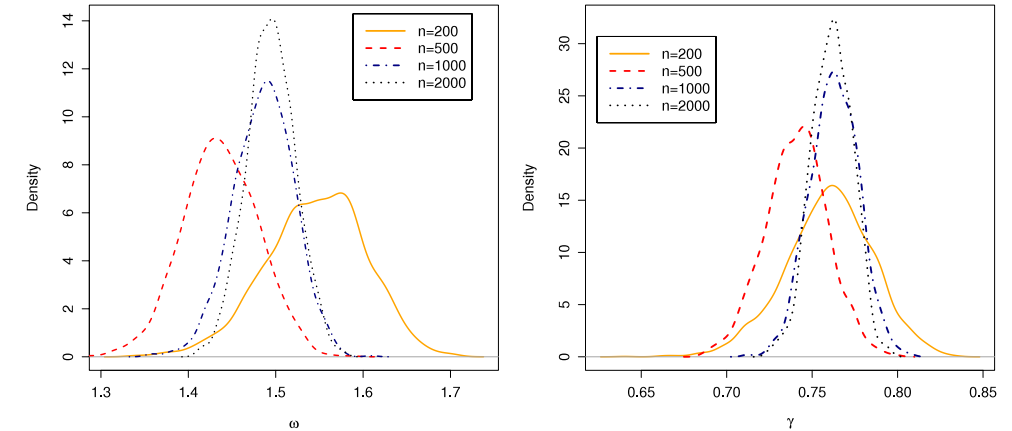


Fig. 5. Posterior distributions of mean direction  $\omega$  and mean resultant length  $\gamma$ .

**Table 1**  
Simulation results of PN regression ( $n, \tau^2, \rho$ ): posterior mean, lower and upper bounds of 95% credible interval for parameters  $\tau^2, \rho$  and  $\beta_j = (\beta_{j1}, \beta_{j2})' (j = 1, 2, 3)$ .

		$n = 60, \tau^2 = 1, \rho = 0.4$							$n = 200, \tau^2 = 1, \rho = 0.4$						
		$\Sigma \neq I$			$\Sigma = I$				$\Sigma \neq I$			$\Sigma = I$			
par.	true	mean	2.5%	97.5%	mean	2.5%	97.5%		mean	2.5%	97.5%	mean	2.5%	97.5%	
$\beta_{11}$	2	1.66	1.27	2.22	1.81	1.41	2.23		1.69	1.51	1.95	1.50	1.32	1.70	
$\beta_{21}$	1	1.14	0.36	1.86	1.32	0.64	2.09		0.58	0.28	0.86	0.34	0.05	0.61	
$\beta_{31}$	1	0.88	0.27	1.56	0.85	0.22	1.42		1.21	0.91	1.49	0.84	0.56	1.14	
$\beta_{12}$	-2	-1.78	-2.28	-1.39	-1.90	-2.31	-1.49		-1.78	-2.01	-1.54	-1.58	-1.76	-1.40	
$\beta_{22}$	-1	-0.55	-1.26	0.19	-0.70	-1.45	-0.05		-0.61	-0.92	-0.31	-0.38	-0.65	-0.10	
$\beta_{32}$	1	0.89	0.26	1.58	0.96	0.38	1.56		0.59	0.28	0.90	0.58	0.29	0.84	
$\tau^2$	1	0.91	0.49	1.55	-	-	-		0.76	0.56	1.04	-	-	-	
$\rho$	0.4	0.23	-0.15	0.62	-	-	-		0.42	0.12	0.68	-	-	-	

		$n = 60, \tau^2 = 2.56, \rho = 0.4$							$n = 200, \tau^2 = 2.56, \rho = 0.4$						
		$\Sigma \neq I$			$\Sigma = I$				$\Sigma \neq I$			$\Sigma = I$			
par.	true	mean	2.5%	97.5%	mean	2.5%	97.5%		mean	2.5%	97.5%	mean	2.5%	97.5%	
$\beta_{11}$	2	1.41	0.96	1.98	1.18	0.86	1.49		1.61	1.27	1.91	1.22	1.03	1.40	
$\beta_{21}$	1	0.58	-0.37	1.46	0.41	-0.20	1.00		0.81	0.43	1.20	0.41	0.12	0.71	
$\beta_{31}$	1	1.03	0.25	1.90	0.74	0.15	1.36		0.65	0.26	1.07	0.49	0.22	0.76	
$\beta_{12}$	-2	-1.79	-2.27	-1.06	-1.52	-1.22	-0.65		-1.67	-1.88	-1.43	-1.25	-1.43	-1.08	
$\beta_{22}$	-1	-0.44	-1.20	0.56	-0.39	-1.00	0.21		-0.64	-0.94	-0.27	-0.28	-0.54	-0.02	
$\beta_{32}$	1	0.93	0.26	1.65	0.66	0.04	1.30		0.60	0.29	0.91	0.35	0.12	0.60	
$\tau^2$	2.56	2.03	1.10	3.29	-	-	-		1.83	1.15	2.53	-	-	-	
$\rho$	0.4	0.58	0.15	0.83	-	-	-		0.62	0.47	0.75	-	-	-	

identifiability problems. As noted in Section 2, interpretation of the parameters in the general projected normal is difficult, hence, less clear than in the case of the mixture of two von Mises distributions. But again, with the goal of capturing departure from symmetry, it is not apparent that a two-component mixture model is an appropriate specification.

4.3. Butterfly migration

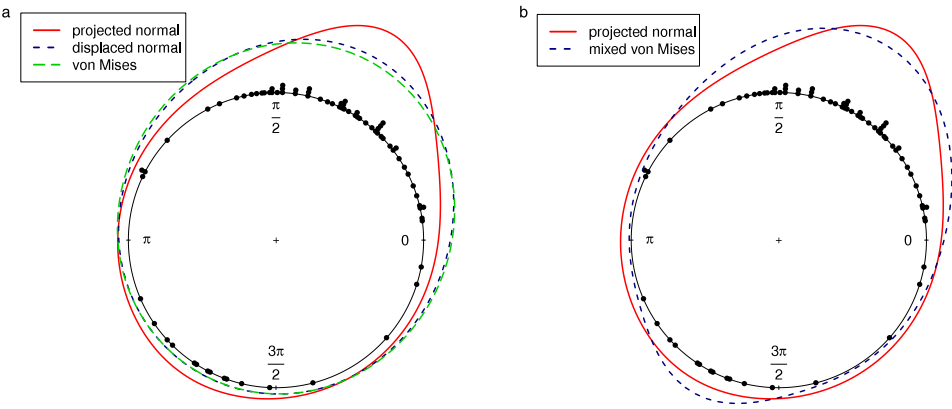
Finally, we illustrate the PN regression model using data from a study by Walker and Whitesell of migratory butterflies entering northern Florida. More detail regarding the biological background and data collection methods can be found in [21]. For illustration, we only use the data for the species *cloudless sulphur*. One purpose of this study is to examine the effect of distance from the coast on

**Table 2**  
Model comparison of PN regression under each combination of  $(n, \tau^2, \rho)$ : general PN  $\Sigma \neq \mathbf{I}$  (left column) and displaced normal  $\Sigma = \mathbf{I}$  (right column).

$(n, \tau^2, \rho)$	(60, 1, 0.4)		(60, 2.56, 0.4)	
PLSL	121.89	122.70	183.65	171.45
Average CRPS	0.2511	0.2531	0.3124	0.3129
Average length of predictive CI	1.88	1.92	2.20	3.74
Percentage (%)	98	94	95	98

$(n, \tau^2, \rho)$	(200, 1, 0.4)		(200, 2.56, 0.4)	
PLSL	106.64	117.09	148.35	164.50
Average CRPS	0.2421	0.2444	0.2997	0.3072
Average length of predictive CI	1.68	2.15	2.08	2.94
Percentage (%)	98	100	96	100



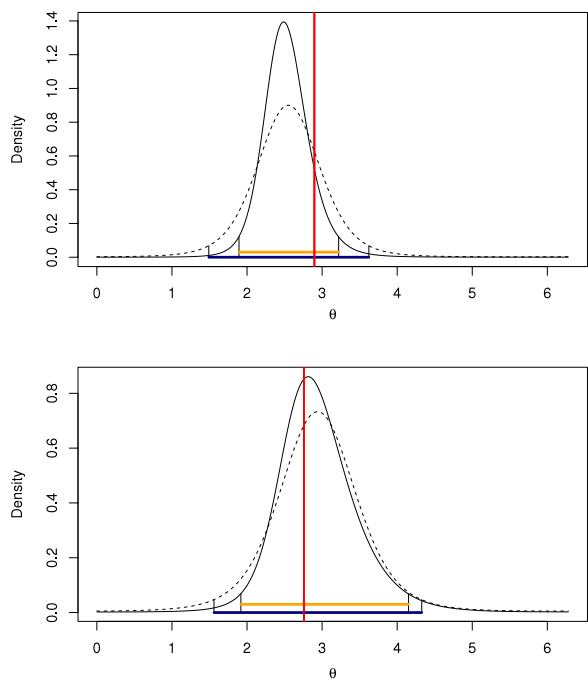
**Fig. 6.** Turtle data example.

the direction of butterfly migration. Therefore, we include the standardized distance to the coast (in kilometers) as a covariate in the regression model. We also include the standardized temperature. So, for the  $i$ -th observation,  $\mathbf{x}_i = (1, \text{distcoast}_i, \text{temp}_i)'$ . The total number of observations is 233 and we hold out 40, chosen at random, as a validation set. Again, we fit the general PN and the displaced PN, adding to the priors of the previous subsection for  $\tau^2$  and  $\rho$  a normal prior with  $\lambda_0 = 0.5$  for the  $\beta$ 's. The posterior estimates and uncertainties are reported in Table 3. The percentage of hold out values falling into the 95% HPD predictive intervals is 100% for both; the average lengths of the predictive intervals for the holdout set are 2.09 and 2.56, respectively. The average CRPS for the general version is 0.1987; for the identity version it is 0.2044. The mean PLSLs are 34.55 and 40.10, respectively. All these criteria suggest that the general PN regression is a preferable model for this dataset. Also, we provide predictive density plots for two illustrative holdout observations in Fig. 7. We see shorter HPD credible intervals for the general PN, revealed by highlighting the lengths at the bottom.

Finally, in order to see how the mean direction  $\omega$  and the concentration  $\gamma$  respond to changes in the  $\mathbf{X}$ 's, we present Fig. 8. Here, we show the change in mean direction and mean resultant length with respect to change in the temperature, while fixing the value of the standardized distance to the coast ( $\text{distcoast}_i = 1$ ). We allow the standardized temperature  $X_2$  to range from  $-2$  to  $2$ . Using posterior samples, we obtain estimates of  $\omega(x_2)$  and  $\gamma(x_2)$  associated with  $f(\theta|x_2, \text{data})$  as in Section 3.2. In Fig. 8, we can observe that change in temperature leads to smooth change in the mean direction of the butterfly migration, while also resulting in smooth change in the mean resultant length. As the temperature increases, the butterfly migration direction shifts and is relatively less concentrated in the mean direction.

**Table 3**  
Butterfly data example: posterior summaries of parameters.

par.	$\Sigma \neq \mathbf{I}$			$\Sigma = \mathbf{I}$		
	mean	2.5%	97.5%	mean	2.5%	97.5%
$\beta_{10}$	−1.95	−2.10	−1.80	−1.81	−1.97	−1.63
$\beta_{11}$	0.10	−0.08	0.27	−0.04	−0.20	0.11
$\beta_{12}$	0.22	0.01	0.40	0.11	−0.11	0.30
$\beta_{20}$	0.53	0.38	0.68	0.43	0.24	0.59
$\beta_{21}$	0.22	0.05	0.40	0.33	0.17	0.48
$\beta_{22}$	−0.94	−1.11	−0.76	−0.82	−0.98	−0.66
$\tau^2$	0.82	0.66	1.02	–	–	–
$\rho$	−0.21	−0.37	−0.04	–	–	–

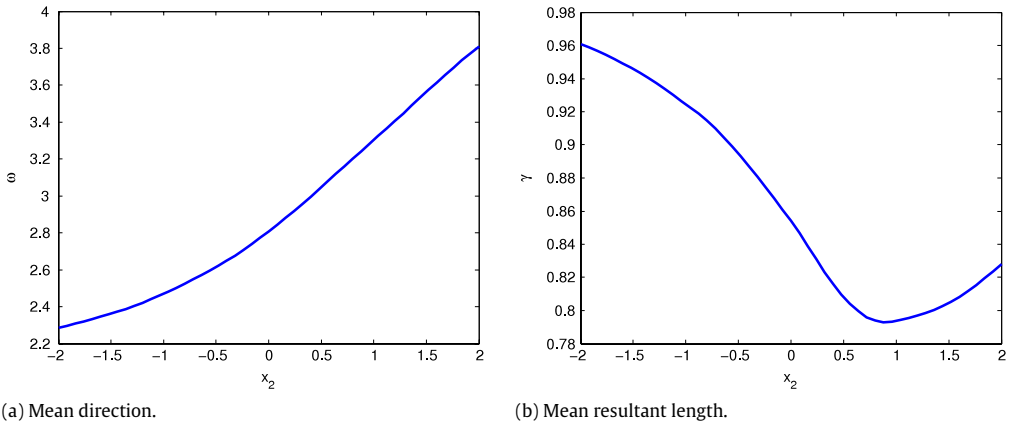


**Fig. 7.** Predictive density and its 95% HPD credible intervals for two hold out observations: the solid line is the predictive density for the general PN regression and the dotted line is for the identity version. The vertical line shows the true hold out value  $\theta^*$ .

**5. Summary and future work**

We have argued that the flexibility of the general PN model as well as the convenient regression linkage to linear explanatory variables makes it an attractive model for angular data. In particular, we have shown how to straightforwardly fit these models within a Bayesian framework, obtaining full posterior inference. Using both PLSL and CRPS, we have demonstrated that the general PN outperforms the displaced PN ( $\Sigma = \mathbf{I}$ ) using simulated data when the former is true and with two real datasets. Finally, in the Bayesian context, we have shown how we can see the response of the mean direction and concentration to change in covariate level.

As in the turtle data example, it may be asserted that, for independent observations on a sphere, mixtures of von Mises distributions can achieve comparable flexibility to the general PN. In this regard, we note that we are currently investigating mixing with PN's, both discrete and continuous mixtures. The former leads to a dense class theorem, the latter leads to projected  $t$ -distributions.



**Fig. 8.** Change in mean direction  $\omega$  and mean resultant length  $\gamma$  with changing temperature  $x_2$  under the general PN regression for the butterfly data.

PN's also offer convenient multivariate extension to dependent observations, each on a circle or, more generally, on a sphere, through suitable multivariate normal distributions; the von Mises does not easily extend beyond two or three dimensions in this case. Of particular interest for space and space-time directional data is a spatial and spatio-temporal projected Gaussian process which we will report on in a forthcoming manuscript.

## Acknowledgments

The authors thank Giovanna Jona Lasinio for useful conversations and Ramon Littell for help in obtaining the butterfly data in Section 4.3. The work of the authors was supported in part by NSF CMG 0934595 and NSF CDI 0940671.

## Appendix A. Mean direction calculation for $PN_2(\mu, I)$ from Section 2

Assume  $X_1$  and  $X_2$  are two independent variables,  $X_1 \sim N(\mu_1, 1)$  and  $X_2 \sim N(\mu_2, 1)$ . The joint density of  $X_1$  and  $X_2$  is

$$f(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}[(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2]\right).$$

Change variables  $(X_1, X_2)$  to  $(R, \Theta)$ , i.e.,  $X_1 = R \cos \Theta$ ,  $X_2 = R \sin \Theta$  and  $|J| = R$ . Let  $\mu_0^2 = \mu_1^2 + \mu_2^2$ , where  $\mu_1 = \mu_0 \cos \alpha$  and  $\mu_2 = \mu_0 \sin \alpha$ ,  $\forall \alpha$ .

$$\begin{aligned} f(\theta, r) &= \frac{1}{2\pi} r \exp\left(-\frac{1}{2}(r^2 - 2r(\mu_1 \cos \theta + \mu_2 \sin \theta) + \mu_1^2 + \mu_2^2)\right) \\ &= \frac{1}{2\pi} r \exp\left(-\frac{1}{2}(r^2 - 2r\mu_0(\cos \alpha \cos \theta + \sin \alpha \sin \theta) + \mu_0^2)\right) \\ &= \frac{1}{2\pi} r \exp\left(-\frac{1}{2}[r - \mu_0 \cos(\theta - \alpha)]^2\right) \exp\left(-\frac{1}{2}\mu_0^2 \sin^2(\theta - \alpha)\right) \\ f(\theta) &= \int_0^\infty f(\theta, r) dr \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mu_0^2 \sin^2(\theta - \alpha)\right) \frac{1}{\sqrt{2\pi}} \int_0^\infty r \exp\left(-\frac{1}{2}[r - \mu_0 \cos(\theta - \alpha)]^2\right) dr. \end{aligned}$$

Use the result,

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty r \exp\left(-\frac{1}{2}(r-a)^2\right) dr = \phi_1(a) + a\Phi_1(a)$$

where  $\phi_1(\cdot)$  and  $\Phi_1(\cdot)$  are the standard univariate normal pdf and cdf, respectively. Let  $g(x) = \phi_1(x) + x\Phi_1(x)$ , then

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty r \exp\left(-\frac{1}{2}[r - \mu_0 \cos(\theta - \alpha)]^2\right) dr = g(\mu_0 \cos(\theta - \alpha)).$$

Under careful calculation,

$$\begin{aligned} E \sin(\theta - \alpha) &= \int_0^{2\pi} \sin(\theta - \alpha) f(\theta) d\theta \\ &= \int_0^{2\pi} \sin(\theta - \alpha) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mu_0^2 \sin^2(\theta - \alpha)\right) g(\mu_0 \cos(\theta - \alpha)) d\theta \\ &= 0 \end{aligned}$$

$$E(\sin \theta \cos \alpha - \sin \alpha \cos \theta) = 0 \Rightarrow \frac{\mu_1}{\mu_0} E \sin \theta = \frac{\mu_2}{\mu_0} E \cos \theta \Rightarrow \mu_1 E \sin \theta = \mu_2 E \cos \theta.$$

According to the definition of mean direction,  $\arctan\left(\frac{E \sin \theta}{E \cos \theta}\right) = \arctan\left(\frac{\mu_2}{\mu_1}\right)$  when  $\mu_1 \neq 0$ .

## Appendix B. Posterior computation

The full conditionals for the parameters of the projected normal model without covariates are:

- $f(\boldsymbol{\mu}|\boldsymbol{\theta}, \mathbf{r}, \tau^2, \rho) \sim N(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$ , where  $\tilde{\boldsymbol{\mu}} = \tilde{\Sigma}^{-1} \sum_{i=1}^n r_i \mathbf{u}_i$  and  $\tilde{\Sigma} = (n\Sigma^{-1} + I/\lambda_0)^{-1}$ .
- $f(r_i|\theta_i, \boldsymbol{\mu}, \tau^2, \rho) \propto r_i \exp\left(-\frac{\mathbf{u}_i^T \Sigma^{-1} \mathbf{u}_i r_i^2 - 2\mathbf{u}_i^T \Sigma^{-1} \boldsymbol{\mu} r_i}{2}\right) I_{(0,\infty)}(r_i)$ .
- $f(\tau^2, \rho|\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\mu}, \rho) \propto \prod_{i=1}^n f(\theta_i, r_i|\boldsymbol{\mu}, \Sigma)(\tau^2)^{-a_\tau-1} e^{-\frac{b_\tau}{\tau^2}}$ . Use the truncated normal kernel  $(\tau^2)^* \sim N_+(\tau^2, c_2)$  and  $\rho^* \sim TN(\rho, c_3, -1, 1)$ .

The full conditionals for the parameters of the PN regression model are:

- $\boldsymbol{\beta}_j \sim BN(\tilde{\boldsymbol{\mu}}_j, \tilde{\Sigma}_j)$ , where  $\tilde{\boldsymbol{\mu}}_j = \tilde{\Sigma}_j^{-1} \left(-\sum_{i=1}^n x_{i,j-1} \sum_{k \neq j} x_{i,k-1} \boldsymbol{\beta}_k + \sum_{i=1}^n x_{i,j-1} r_i \mathbf{u}_i\right)$  and  $\tilde{\Sigma} = \left(\sum_{i=1}^n x_{i,j-1}^2 \Sigma^{-1} + \lambda_0^{-1} \mathbf{I}\right)^{-1}$ .
- $f(r_i|\theta_i, \boldsymbol{\mu}, \tau^2, \rho) \propto r_i \exp\left(-\frac{\mathbf{u}_i^T \Sigma^{-1} \mathbf{u}_i r_i^2 - 2\mathbf{u}_i^T \Sigma^{-1} \boldsymbol{\mu} r_i}{2}\right) I_{(0,\infty)}(r_i)$ , where  $\boldsymbol{\mu}_i = B_{2 \times p}^T \mathbf{x}_i$ .
- $f(\tau^2, \rho|\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\mu}, \rho) \propto \prod_{i=1}^n f(\theta_i, r_i|\boldsymbol{\mu}, \Sigma)(\tau^2)^{-a_\tau-1} e^{-\frac{b_\tau}{\tau^2}}$ . Use the truncated normal kernel  $(\tau^2)^* \sim N_+(\tau^2, c_2)$  and  $\rho^* \sim TN(\rho, c_3, -1, 1)$ .

## References

- [1] N.I. Fisher, Statistical Analysis of Circular Data, Cambridge University Press, 1993.
- [2] N.I. Fisher, A.J. Lee, Regression models for an angular response, Biometrics 48 (1992) 665–677.
- [3] T. Gneiting, A.E. Raftery, Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association 102 (2007) 359–378.
- [4] A.L. Gould, A regression technique for angular variates, Biometrics 25 (1969) 683–700.
- [5] E. Gneiting, T. Gneiting, V. Berrocal, N. Johnson, The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, Quarterly Journal Royal Meteorological Society 132 (2006) 2925–2942.
- [6] D. Hernandez-Stumpfhauser, J.F. Breidt, J.D. Opsomer, Hierarchical Bayesian Small Area Estimation for Circular Data, Technical Report, Colorado State University, 2010.
- [7] S.R. Jammalamadaka, A. SenGupta, Topics in Circular Statistics, World Scientific, 2001.
- [8] R.A. Johnson, T. Wehrly, Some angular-linear distributions and related regression models, Journal of the American Statistical Association 73 (1978) 602–606.

- [9] T.A. Jones, W.R. James, Analysis of bimodal orientation, *Mathematical Geology* 1 (1969) 129–135.
- [10] D.G. Kendall, Pole-seeking Brownian motion and bird navigation, *Journal of the Royal Statistical Science. Series B (Methodological)* 36 (1974) 365–417.
- [11] A. Lee, Circular data, *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010) 477–486.
- [12] K.V. Mardia, *Statistics of Directional Data*, Academic Press, London and New York, 1972.
- [13] K.V. Mardia, Statistics of directional data, *Journal of the Royal Statistical Society. Series B (Methodological)* 37 (1975) 349–393.
- [14] K.V. Mardia, P.E. Jupp, *Directional Statistics*, John Wiley & Sons, 2000.
- [15] G. Nuñez-Antonio, E. Gutiérrez-Peña, A Bayesian analysis of directional data using the projected normal distribution, *Journal of Applied Statistics* 32 (2005) 995–1001.
- [16] G. Nuñez-Antonio, E. Gutiérrez-Peña, G. Escarela, A Bayesian regression model for circular data based on the projected normal distribution, *Statistical Modeling* 11 (2011) 185–201.
- [17] B. Presnell, S.P. Morrison, R.C. Littell, Projected multivariate linear models for directional data, *Journal of the American Statistical Association* 93 (1998) 1068–1077.
- [18] C.G. Small, *The Statistical Theory of Shape*, Springer, 1996.
- [19] B.D. Spurr, M.A. Koutbeiy, A comparison of various methods for estimating the parameters in mixtures of von Mises distributions, *Communications in Statistics. Simulation and Computation* 20 (1991) 725–741.
- [20] M.A. Stephens, *Techniques for Directional Data*, Technical Report, Stanford University, 1969.
- [21] T.J. Walker, R.C. Littell, Orientation of fall-migrating butterflies in north peninsular Florida and source areas, *Ethology* 98 (1994) 60–84.
- [22] G.S. Watson, *Statistics on Spheres*, John Wiley & Sons, 1983.