British Journal of Mathematical and Statistical Psychology (2018), 71, 75–95

The British Psychological Society

www.wileyonlinelibrary.com

Circular interpretation of regression coefficients

Jolien Cremers[1]*, Kees Tim Mulder[1] and Irene Klugkist[1,2]

[1]Department of Methodology and Statistics, Utrecht University, The Netherlands
[2]Research Methodology, Measurement and Data Analysis, Behavioural Sciences, University of Twente, Enschede, The Netherlands

The interpretation of the effect of predictors in projected normal regression models is not straight-forward. The main aim of this paper is to make this interpretation easier such that these models can be employed more readily by social scientific researchers. We introduce three new measures: the slope at the inflection point (b_c), average slope (AS) and slope at mean (SAM) that help us assess the marginal effect of a predictor in a Bayesian projected normal regression model. The SAM or AS are preferably used in situations where the data for a specific predictor do not lie close to the inflection point of a circular regression curve. In this case b_c is an unstable and extrapolated effect. In addition, we outline how the projected normal regression model allows us to distinguish between an effect on the mean and spread of a circular outcome variable. We call these types of effects location and accuracy effects, respectively. The performance of the three new measures and of the methods to distinguish between location and accuracy effects is investigated in a simulation study. We conclude that the new measures and methods to distinguish between accuracy and location effects work well in situations with a clear location effect. In situations where the location effect is not clearly distinguishable from an accuracy effect not all measures work equally well and we recommend the use of the SAM.

1. Introduction

Circular models are models for data with a circular outcome variable. A circular variable measures a direction in two-dimensional space in degrees or radians and requires analysis methods that are different from standard methods for linear data. In the field of psychology circular data can be found in research on the visual perception of space (Matsushima, Vaz, Cazuza, & Ribeiro Filho, 2014), moving room experiments (Stoffregen, Bardy, Merhi, & Oullier, 2004), visual working memory experiments (Heyes, Zokaei, & Husain, 2016), movement synchronization (Kirschner & Tomasello, 2009; Ouwehand & Peper, 2015) and cognitive maps (Brunyé, Burte, Houck, & Taylor, 2015). Measurements on the interpersonal circumplex can also be regarded as circular data (König, Onnen, Karl, Rosner, & Butollo, 2016; Santos, Vandenberghe, & Tavares, 2015; Wright, Pincus, Conroy, & Hilsenroth, 2009; Zilcha-Mano *et al.*, 2015). In general, circular variables measure directions, are circumplex scales, or are a

*Correspondence should be addressed to Jolien Cremers, Sjoerd Groenman Building, room C1.20, PO box 80140, 3508 TC Utrecht, The Netherlands (email: j.cremers@uu.nl).

DOI:10.1111/bmsp.12108

measure of periodic (weekly, daily, hourly, etc.) patterns. Wright *et al.* (2009) outline how circular data are used in research using circumplex measures. They introduce how to compute a circular mean and test for differences between groups. However, circular models could be used much more effectively in psychological science. Take, for instance, a study by Locke, Sayegh, Weber, and Turecki (2017) on interpersonal characteristics of depressed outpatients. Circumplex profiles of patients were made and compared with those of normative samples. Although the authors do use circular means and variances and do compare groups, a more elaborate circular model would allow for the simultaneous evaluation of multiple predictors. It would then be possible to assess whether there still is a difference in circumplex profile between depressed and normative samples when also accounting for other variables such as gender, age or a measure of depression severity.

Analysing circular data is not straightforward and special methods are required. One of the approaches to circular data is the so-called embedding approach. In this approach projected distributions are used. Projected distributions are bivariate distributions defined in $\mathbb{R}^2$ and projected onto the circle. Other approaches to modelling circular data are the wrapping and intrinsic approaches (Mardia & Jupp, 2000). These approaches are based on respectively wrapping distributions defined in $\mathbb{R}$ onto the circle and using distributions defined on the circle itself, such as the von Mises distribution. A paper by Rivest, Duchesne, Nicosia, and Fortin (2016) provides an overview of some of the circular regression models in the literature. Although various types of models for the three approaches have been described, we will only consider regression models of the embedding approach in this paper.

Nuñez-Antonio, Gutiérrez-Peña, and Escarela (2011) and Wang and Gelfand (2013) have developed Bayesian methods for regression based on the projected normal and general projected normal distributions. These are both based on Presnell, Morrison, and Littell (1998) who first used a projected normal distribution to analyse circular regression models. In the embedding approach it is relatively easy to fit more complex models because the distributions that are used are based on distributions in $\mathbb{R}^2$. Indeed, in the literature more complex regression type models have been introduced, including random effects models (Nuñez-Antonio & Gutiérrez-Peña, 2014) and spatial and spatio-temporal models (Mastrantonio, Lasinio, & Gelfand, 2016; Wang & Gelfand, 2014). However, we will limit ourselves to the simpler multiple regression model.

Although the embedding approach is flexible with regard to model fitting, the interpretation of the effects of predictors in these models is not easy. According to Maruotti (2016) this is the major drawback of models based on projected distributions. Currently, when using the embedding approach we obtain two regression coefficients for each variable in the model. This is a result of using an underlying bivariate distribution. The two coefficients are, however, not interpretable as an effect on the circle. Additionally, they do not allow us to directly distinguish between an effect on the mean and an effect on the spread of the circular outcome. In the example of the study by Locke *et al.* (2017) we would not be able to distinguish between the differences between depressed and normative samples in average profiles and differences in the within-group homogeneity of the profiles. There is no literature yet dealing with this problem. We will therefore introduce new interpretation tools that combine the bivariate coefficients into one circular coefficient. The new tools allow us to assess whether there is an effect of a predictor on the mean of the circular outcome and how large this effect is.

In Section 2 we describe the embedding approach. We distinguish between effects on the circular mean and spread in Section 3. Next, in Section 4, we combine two bivariate

coefficients into one circular coefficient and introduce new tools for interpretation. These tools are then applied to the example data set in Section 5. Lastly, we perform a simulation study to assess the new tools in Section 6. The paper concludes with a discussion in Section 7.

2. Embedding approach

In this section we introduce the embedding approach and projected normal distribution. Both have been introduced and described previously (Nuñez-Antonio & Gutiérrez-Peña, 2005; Presnell *et al.*, 1998). Therefore, a large part of this section focuses on the interpretation of the estimates from a projected normal regression model.

2.1. Projected normal distribution

There are several representations in which one can specify circular data: angles, polar coordinates or unit vectors in $\mathbb{R}^2$. In this paper we will refer to an observation of a circular outcome variable in angles as θ_i, where $i = 1, \ldots, n$, and to its unit vector representation as $\mathbf{u}_i$. We assume that the outcome variable can be represented by an unobserved column vector $\mathbf{y}_i$ in $\mathbb{R}^2$ as follows:

$$\mathbf{u}_i = \frac{\mathbf{y}_i}{r_i}, \tag{1}$$

where r_i is the length of the vector $\mathbf{y}_i$. The angular outcome variable originates from a projection onto the circle of a vector in $\mathbb{R}^2$. If we assume that the underlying $\mathbf{y}_i$ originate from a bivariate normal distribution with mean $\boldsymbol{\mu}$ and variance–covariance matrix $\mathbf{I}$, it follows from equation (1) that θ has a projected normal (PN) distribution with density function

$$\mathrm{PN}(\theta|\boldsymbol{\mu}, \mathbf{I}) = \frac{1}{2\pi} e^{-\frac{1}{2}||\boldsymbol{\mu}||^2} \left[1 + \frac{\mathbf{u}^t\boldsymbol{\mu}\Phi(\mathbf{u}^t\boldsymbol{\mu})}{\phi(\mathbf{u}^t\boldsymbol{\mu})}\right], \tag{2}$$

where θ is the circular outcome variable measured in radians ($-\pi \leq \theta < \pi$), $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$ is the mean vector of the distribution, the variance–covariance matrix $\mathbf{I}$ is an identity matrix, and $\mathbf{u}^t = (\cos\theta, \sin\theta)$. The terms $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution, respectively. An identity variance–covariance matrix is chosen to identify the model. Due to this configuration the PN distribution is always unimodal and symmetric. Another configuration can be found in Wang and Gelfand (2013) who use different constraints on the matrix resulting in the general PN distribution that can also take a skewed and multimodal shape.

To fit a model on the mean of the projected normal distribution $\boldsymbol{\mu}$ we need r_i to obtain the unobserved bivariate normal vectors $\mathbf{y}_i$. The estimation of r_i is a missing-data problem that is solved by treating the unobserved lengths r_i as latent or auxiliary variables in the model. We can then use existing techniques such as the EM algorithm (Presnell *et al.*, 1998) or Bayesian methods (Nuñez-Antonio & Gutiérrez-Peña, 2005), to obtain inference on the $\mathbf{y}_i$.

2.2. Regression

In regression we have independent observations of a vector of linear predictors $\mathbf{x}_i$ for each individual $i = 1, \ldots, n$. The model for one of the bivariate normal vectors $\mathbf{y}_i$ has mean structure $\boldsymbol{\mu}_i = \mathbf{B}^t\mathbf{x}_i$, where $\mathbf{B} = [\boldsymbol{\beta}^{\mathrm{I}}, \boldsymbol{\beta}^{\mathrm{II}}]$ and each $\boldsymbol{\beta}$ is a vector with intercept and regression coefficients. The first component of $\mathbf{x}_i$ equals 1 so as to be able to estimate an intercept. Formally, this notation is only correct when the predictors in $\mathbf{x}_i$ are equal for both components of $\boldsymbol{\mu}_i$. The model then has the same structure as a multivariate regression model. The dimensions of $\boldsymbol{\beta}^{\mathrm{I}}$ and $\boldsymbol{\beta}^{\mathrm{II}}$ are, however, allowed to differ.

Even though our main interest lies in effects on the circular mean, the mean structure of $\mathbf{y}_i$ can influence both the mean and spread of a circular outcome. Henceforth we call an effect on the mean of a circular outcome a location effect and an effect on the spread an accuracy effect. The size of the Euclidean norm of $\boldsymbol{\mu}$ influences the circular spread. The larger it is, the smaller the spread on the circle. The consequences of this property of $\boldsymbol{\mu}$ for the interpretation of results from a PN regression model will be considered later.

To estimate PN regression models a Bayesian Markov chain Monte Carlo (MCMC) procedure is used in which r_i and $\mathbf{B}$ are sampled. The procedure is based on Nuñez-Antonio *et al.* (2011) and Hernandez-Stumpfhauser, Breidt, and van der Woerd (2017), a diffuse normal prior is used for $\mathbf{B}$ and the exact method of sampling is described in Appendix A.

3. Location and accuracy effects

There are several types of circular effects: a location effect, an accuracy effect, a mix of these or no effect. Because PN regression models are consensus models the location and spread of the circular outcome are modelled simultaneously (Rivest *et al.*, 2016). This means that we can create measures for checking whether a predictor has a location or an accuracy effect in a PN regression model. A measure to check for a location effect can be constructed by using a regression line in $\mathbb{R}^2$. Considering one predictor in a PN regression model, predicted values on the first and second bivariate component ($\hat{y}^{\mathrm{I}}$ and $\hat{y}^{\mathrm{II}}$) are determined as follows:

$$\hat{y}^{\mathrm{I}} = \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{I}} x,$$

$$\hat{y}^{\mathrm{II}} = \beta_0^{\mathrm{II}} + \beta_1^{\mathrm{II}} x,$$

where β_0^{I} and β_0^{II} are intercepts, β_1^{I} and β_1^{II} are regression coefficients of a particular predictor on the two bivariate components and x is a predictor value. Whether this regression line runs through the origin determines the type of circular effect it represents. In Figure 1 we see two regression lines in $\mathbb{R}^2$ with a unit circle. The regression line on the left passes through the origin. The regression line on the right does not pass through the origin. We also see circular predicted values, the grey dots on the unit circle. The circular predicted values lie very close together in the plot on the left, while in the plot on the right they move counterclockwise on the circle when the predictor value increases. The plot on the left represents an accuracy effect; the circular predicted values do not change with the predictor. The plot on the right represents a location effect.

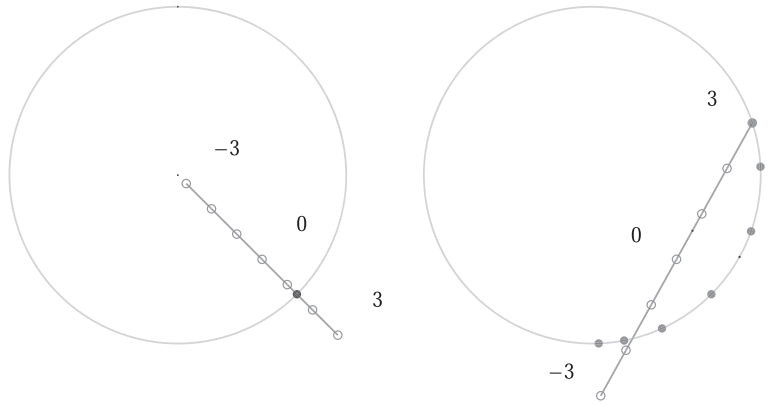

Figure 1. A bivariate regression line with linear predicted values (open dots) for $x = (-3, -2, -1, 0, 1, 2, 3)$ and a unit circle with circular predicted values (closed dots) on the circle for an accuracy (left) and location effect (right).

To assess whether the regression line runs through the origin we first compute the shortest distance (SDO) from the regression line in bivariate space to the origin. The SDO is computed as

$$\text{SDO} = \sqrt{(\beta_0^{\text{I}} + \beta_1^{\text{I}} a_x)^2 + (\beta_0^{\text{II}} + \beta_1^{\text{II}} a_x)^2}. \tag{3}$$

This is the Euclidean norm of the predictor value of the point where the line of predictions in $\mathbb{R}^2$ is closest to the origin (a_x). Because SDO ≥ 0 we give it a sign such that its posterior is defined on the entire real line. We call this new parameter the signed shortest distance to the origin (SSDO). The following equation shows how to determine whether the sign should be positive or negative:

$$\text{SSDO} = \text{sign}[\sin(a_c - \text{atan2}(\beta_1^{\text{II}}, \beta_1^{\text{I}}))]\text{SDO}. \tag{4}$$

The parameter a_c ($-\pi \leq a_c < \pi$) in this equation is the circular predicted value of a_x. The computation of a_x and a_c is outlined in Section 4.4. The function atan2() is defined in equation (5). An intuition for equation (4) is given in the Supporting Information of this paper. An example of how to use the SSDO in practice will be given in Section 5. How well it performs at distinguishing accuracy and location effects is investigated in a simulation study in Section 6.

4. Quantifying location effects for continuous predictors

In this section we show how to compute circular predicted values and make predicted circular regression curves for a marginal effect. Subsequently we outline new tools for quantifying location effects.

4.1. Circular predicted values

Using the two-argument arctangent function, atan2, we compute predicted values on a circular scale, $\hat{\theta}$, as follows:

$$
\begin{aligned}
\hat{\theta} = \text{atan2}\ (\hat{y}^{\text{II}}, \hat{y}^{\text{I}}) &= \arctan\left(\frac{\hat{y}^{\text{II}}}{\hat{y}^{\text{I}}}\right) && \text{if} \quad \hat{y}^{\text{I}} > 0 \\
&= \arctan\left(\frac{\hat{y}^{\text{II}}}{\hat{y}^{\text{I}}}\right) + \pi && \text{if} \quad \hat{y}^{\text{I}} < 0,\ \hat{y}^{\text{II}} \geq 0 \\
&= \arctan\left(\frac{\hat{y}^{\text{II}}}{\hat{y}^{\text{I}}}\right) - \pi && \text{if} \quad \hat{y}^{\text{I}} < 0,\ \hat{y}^{\text{II}} < 0 \\
&= \frac{\pi}{2} && \text{if} \quad \hat{y}^{\text{I}} = 0,\ \hat{y}^{\text{II}} > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad \hat{y}^{\text{I}} = 0,\ \hat{y}^{\text{II}} < 0 \\
&= \text{undefined} && \text{if} \quad \hat{y}^{\text{I}} = 0,\ \hat{y}^{\text{II}} = 0.
\end{aligned} \tag{5}
$$

Here $\hat{y}^{\text{I}} = \boldsymbol{\beta}^{\text{I}}\mathbf{x}$ and $\hat{y}^{\text{II}} = \boldsymbol{\beta}^{\text{II}}\mathbf{x}$ are predicted values on the two components for a vector of predictor values $\mathbf{x}$.

4.2. Predicted regression curves

To visualize the circular effect, we compute a predicted circular regression curve for a marginal effect. For the marginal effect of one linear predictor with the values of the other predictors set to zero we specify $\hat{y}^{\text{I}} = \beta_0^{\text{I}} + \beta_1^{\text{I}} x$ and $\hat{y}^{\text{II}} = \beta_0^{\text{II}} + \beta_1^{\text{II}} x$. We fill out these functions for different values of x and the intercepts and coefficients are estimated. Figure 2 shows a regression curve for one predictor together with original data points. The y-axis of this plot contains the predicted outcome, $\hat{\theta}$, in degrees and the x-axis contains values for x with a range equal to the data range. This plot illustrates the effect of the predictor on the circular outcome.

By investigating a marginal effect all predictors except one are set to a specific value. In our case they are set to zero. For continuous variables we centre the predictors and therefore zero refers to the mean value. For categorical variables zero refers to the baseline category. As in logistic regression the values to which the other predictors in the model are set influence the marginal effect we observe for the predictor of interest.

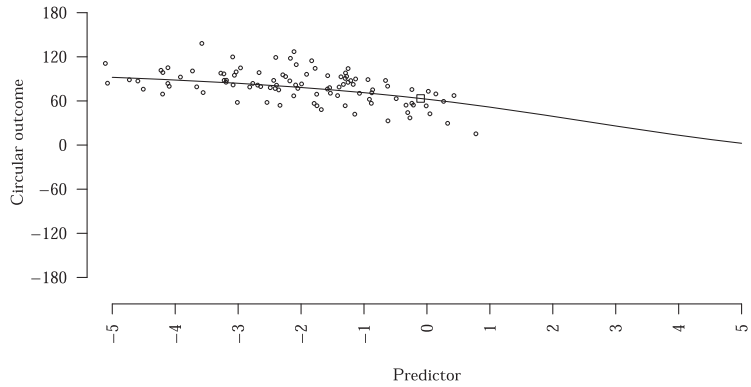

Figure 2. Predicted circular regression curve for the relation between a linear predictor and a circular outcome together with the original data points. The square indicates the inflection point of the regression curve.

4.3. A reparametrization for regression models

To quantify the slopes of circular regression curves we propose a reparametrization of equation (5). From this reparametrization we suggest three types of circular regression coefficients. The proposed reparametrization of equation (5) is as follows:

$$\begin{aligned}\hat{\theta} = \operatorname{atan2}\left(\hat{y}^{\mathrm{II}}, \hat{y}^{\mathrm{I}}\right) &= \operatorname{atan2}\left(\beta_0^{\mathrm{II}} + \beta_1^{\mathrm{II}} x, \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{I}} x\right) \\ &= a_c + \arctan[b_c(x - a_x)].\end{aligned} \tag{6}$$

Here β_0^{I} and β_0^{II} are the linear intercepts and β_1^{I} and β_1^{II} are the linear coefficients of one continuous predictor variable x. The parameters a_c, a_x and b_c describe a predicted circular regression curve, such as the one in Figure 2. The parameters a_c and a_x describe the location of the inflection point of the regression curve on the axis of the circular outcome and the axis of the predictor, respectively. The inflection point occurs at the value of the predictor for which the regression line in $\mathbb{R}^2$ is closest to the origin. Hence, a_x is both the predictor value of the point on the regression line in bivariate space that lies closest to the origin as well as the location of the inflection point of the circular regression curve on the axis of the predictor. In Figure 2, the inflection point is indicated by a square. The parameter b_c describes the slope of the tangent line at the inflection point.

4.4. Parameter derivation

4.4.1. The x-coordinate of the inflection point (a_x)

To obtain a_x, the derivative of the Euclidean norm of the point where the line of predictions in $\mathbb{R}^2$ is closest to the origin is solved for 0. We find

$$a_x = -\frac{\beta_0^{\mathrm{I}}\beta_1^{\mathrm{I}} + \beta_0^{\mathrm{II}}\beta_1^{\mathrm{II}}}{(\beta_1^{\mathrm{I}})^2 + (\beta_1^{\mathrm{II}})^2}. \tag{7}$$

This is the location of the inflection point on the axis of the predictor x.

4.4.2. The y-coordinate of the inflection point (a_c)

To obtain a_c, we insert a_x into equation (6):

$$a_c = \operatorname{atan2}\left(\beta_0^{\mathrm{II}} + \beta_1^{\mathrm{II}} a_x, \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{I}} a_x\right). \tag{8}$$

This is the location of the inflection point on the axis of the predicted circular outcome.

4.4.3. The slope at the inflection point (b_c)

We may solve the reparametrization from equation (6) for b_c to obtain

$$b_c = \frac{\tan\{\operatorname{atan2}(\beta_0^{\mathrm{II}} + \beta_1^{\mathrm{II}} x, \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{I}} x) - a_c\}}{x - a_x}, \quad x \neq a_x. \tag{9}$$

Note that for $x = a_x$ this is undefined. We can simplify the formula by plugging in $x = 0$:

$$b_c = -\frac{\tan\{\text{atan2}(\beta_0^{\text{II}}, \beta_0^{\text{I}}) - a_c\}}{a_x}, \quad a_x \neq 0. \tag{10}$$

This is the slope of the tangent line at the inflection point and a first circular regression coefficient. This coefficient is conceptually similar to the standard regression coefficient in von Mises regression models (Fisher & Lee, 1992). Like the SSDO, this regression coefficient can be used as an indicator of a location effect. In a simulation study in Section 6 we investigate how these indicators, b_c or SSDO, perform at detecting location effects.

4.4.4. Additional quantification measures

The slope b_c is not necessarily a good measure for all data sets. For some data the inflection point of the regression curve does not lie near the data. In that case, b_c can take on a large range of different values while in the asymptotes the regression curve is still a good approximation to the data. This means that in some cases b_c represents a very unstable extrapolated effect. Then it is much more interesting to investigate the slope of the regression curve near the data. We get the slope at a specific predictor value by taking the derivative of equation (6) for x and plugging in the value for which we want to know the slope:

$$\Delta\hat{\theta}(x) = \frac{\text{d}}{\text{d}x}\{a_c + \arctan[b_c(x - a_x)]\} = \frac{b_c}{1 + [b_c(x - a_x)]^2}. \tag{11}$$

An intuitive value for which we want to know the slope is the mean, $\bar{x}$. We obtain the slope at the mean (SAM) as

$$\text{SAM} = \Delta\hat{\theta}(\bar{x}) = \frac{b_c}{1 + [b_c(\bar{x} - a_x)]^2}, \tag{12}$$

where b_c is as computed in equation (10). We interpret this measure as saying that at $\bar{x}$, a one-unit increase in x results in a SAM increase in $\hat{\theta}$.

Another measure that we are interested in is the averaged slope (AS) over each data point. We compute slopes for all data values of x and average these slopes:

$$\text{AS} = \overline{\Delta}\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{b_c}{1 + [b_c(x_i - a_x)]^2}. \tag{13}$$

We interpret this measure as saying that on average, a one-unit increase in the predictor, x, results in a AS increase in $\hat{\theta}$.

5. Empirical example

To illustrate the problems that occur when interpreting output from a PN model we fit a regression model to a data set collected by Brunyé *et al.* (2015), the 'pointing north data'. In their study, 200 Tufts University students divided across 10 data collection sites were asked to point north. Pointing angles relative to the magnetic north (pointing errors) were recorded as the outcome variable and several

predictor variables were measured. The outcome variable was measured such that the real north was located at $\theta = 0°$. One of the predictor variables is self-reported spatial ability. This was measured using the Santa Barbara Sense of Direction questionnaire (SBSOD). Other predictors were age, experience with living on campus and gender (0 = male, 1 = female). Table 1 shows descriptives for these data.

Before analysis all continuous predictor variables were centred at 0. This affects the intercept but not the coefficients and will make the interpretation of individual effects more intuitive. When using the embedding approach we are predicting the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^t$ of the PN distribution for the pointing error. The prediction equations for the pointing north data are:

$$\begin{aligned} \hat{\mu}_1 &= \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{I}}\mathrm{Age} + \beta_2^{\mathrm{I}}\mathrm{Gender} + \beta_3^{\mathrm{I}}\mathrm{Experience} + \beta_4^{\mathrm{I}}\mathrm{SBSOD}, \\ \hat{\mu}_2 &= \beta_0^{\mathrm{II}} + \beta_1^{\mathrm{II}}\mathrm{Age} + \beta_2^{\mathrm{II}}\mathrm{Gender} + \beta_3^{\mathrm{II}}\mathrm{Experience} + \beta_4^{\mathrm{II}}\mathrm{SBSOD}. \end{aligned} \tag{14}$$

Table 2 shows estimates from the regression model that was fitted after careful monitoring of convergence (iterations = 3,000, burn-in = 1,000). Appendix B shows histograms of the posterior distributions of the linear intercepts and coefficients. The interpretation of the regression coefficients is the same as in linear regression. For SBSOD, if the self-reported spatial ability increases by one unit, the predicted outcome on the first bivariate component increases by 0.25 and the predicted outcome on the second bivariate component increases by 0.27. For the categorical variable, gender, we may interpret the coefficients as follows: for females the predicted outcome on the first bivariate component is 0.48 lower than for males. Because our circular outcome concerns compass data the only interpretation we can give to the bivariate components is that of a north–south and an east–west axis. The coefficients do not give us information on the change in the predicted pointing error on the circle, a circular effect, nor do they give us information on whether it is a location or accuracy effect. Next, we compute and interpret the circular regression coefficients introduced in Section 4 and the SSDO for the pointing north data.

Table 1. Descriptives for the pointing north data with linear mean and standard deviation (*SD*) for continuous variables and mean direction ($\bar{\theta}$) and mean resultant length ($\bar{R}$) for circular variables

	Mean/$\bar{\theta}$	*SD*/$\bar{R}$	Minimum	Maximum	Type
Pointing error	19.57°	0.43	–	–	Circular
Age	19.68	1.29	18.00	23.00	Continuous
Experience	1.79	1.21	0.00	4.00	Continuous
SBSOD	4.10	1.06	1.47	6.80	Continuous
Gender	–	–	–	–	Categorical
M	16.32°	0.56	–	–	
F	24.33°	0.32	–	–	

Note. Note that only absolute pointing errors are provided online with the original paper. The pointing error used here was obtained by an iterative process multiplying the absolute errors by a random vector containing 1's and −1's and computing the mean resultant length ($\bar{R}$) until a data set was found in which $\bar{R}$ was equal to 0.43, that of the original data. The mean direction was then set to 19.57, the mean direction of the original data.

Table 2. Posterior modes and 95% highest posterior density lower (LB) and upper bounds (UB) for the regression coefficients of the pointing north data. An asterisk (*) indicates that a highest posterior density interval does not contain 0

	Component I			Component II		
	Mode I	LB I	UB I	Mode II	LB II	UB II
Intercept	0.95	0.72	1.20*	0.15	−0.07	0.41
Age	−0.07	−0.29	0.14	0.00	−0.23	0.21
Gender	−0.48	−0.73	−0.10*	0.17	−0.18	0.47
Experience	0.22	−0.00	0.45	−0.02	−0.23	0.23
SBSOD	0.25	0.09	0.40*	0.27	0.07	0.38*

5.1. Circular effects

Before interpreting the circular regression coefficients we use the linear coefficients to determine whether there is any location or accuracy effect of a predictor on the pointing error. To do so we use the linear regression coefficients for a predictor (e.g., β_1^{I} and β_1^{II}). If one or both regression coefficients of the two components are different from 0 there is an accuracy and/or location effect. To check whether this holds we can use the highest posterior density (HPD) intervals of both linear coefficients. If they do not both contain zero we conclude that there is an effect of the predictor on the circle. From Table 2 we may thus conclude that only SBSOD and gender have an effect on the circle.

For the categorical variable gender we can compare the predicted angle for females, $\mathrm{atan2}(\beta_0^{\mathrm{II}} + \beta_2^{\mathrm{II}}, \beta_0^{\mathrm{I}} + \beta_2^{\mathrm{I}}) = 34.25°$, and for males, $\mathrm{atan2}(\beta_0^{\mathrm{II}}, \beta_0^{\mathrm{I}}) = 8.97°$. Females thus have a higher pointing error than males on average. If we compute such predicted angles for each iteration of the MCMC sampler we can also compute HPD intervals for this effect.

For the continuous variables of the pointing north data, Table 3 shows the posterior modes (PM) and the upper (UB) and lower bounds (LB) of the HPD intervals for the circular coefficients and the SSDO. We use the SSDO to determine whether the effect of SBSOD on the pointing error is an accuracy or a location effect. Again we use the HPD interval. If the HPD interval of the SSDO of a predictor does not include 0 we conclude that the bivariate regression line does not run through the origin and that there is a location effect. For SBSOD the HPD interval does not include zero. This means that the effect of SBSOD on the pointing error is a location effect.

Next, we investigate the circular regression coefficients for SBSOD shown in Table 3. For SBSOD, all three circular coefficients have HPD intervals that do not contain 0. This means that SBSOD has a location effect and is what we expected after concluding that the SSDO for SBSOD was different from zero. Figure 3 shows the relation between SBSOD and the pointing error. The posterior mode of the slope at the inflection point of the predicted circular regression curve, b_c, for SBSOD is equal to 0.52. This means that at the inflection point, as SBSOD increases by one unit the pointing error increases counterclockwise, by $0.52 \cdot 180/\pi \approx 29.8°$, keeping all other predictors at zero. However, the inflection point lies almost outside the data range so we would rather interpret the AS or SAM. On average an increase of one unit in SBSOD results in a counterclockwise increase in pointing error of $\mathrm{AS} \cdot 180/\pi = 0.23 \cdot 180/\pi \approx 13.2°$, keeping all other predictors at zero. At the average SBSOD an increase of one unit results in a counterclockwise increase in pointing error of $\mathrm{SAM} \cdot 180/\pi = 0.17 \cdot 180/\pi \approx 9.7°$, keeping all other predictors at zero.

Even though they do not have a circular effect, it is interesting to look at the parameter estimates of age and experience. The results show estimation issues for both predictors,

Table 3. Posterior modes (PM) and 95% HPD interval lower (LB) and upper bounds (UB) for the circular regression coefficients and SSDO for the continuous variables of the pointing north data. An asterisk (*) indicates that an HPD interval does not contain 0

	Age			Experience			SBSOD		
	PM	LB	UB	PM	LB	UB	PM	LB	UB
b_c	0.16	−2.57	2.80	−0.35	−7.04	6.50	0.52	0.09	3.48*
SAM	−0.02	−0.22	0.24	−0.03	−0.31	0.18	0.17	0.00	0.40*
AS	0.05	−0.21	0.26	−0.10	−0.36	0.27	0.23	0.01	0.42*
SSDO	−0.76	−1.06	1.00	1.30	−1.87	2.07	−1.08	−2.11	−0.59*

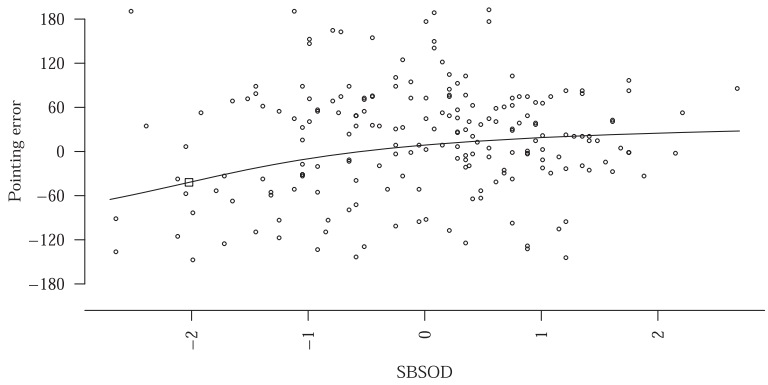

Figure 3. Predicted circular regression curve for the relation between SBSOD and the pointing error, together with the original data points. The square indicates the inflection point of the regression curve.

exemplified by the wide HPD intervals of b_c. When we look at the posterior histograms of the circular regression coefficients and SSDO for age and experience in Figure B2 in Appendix B we also see estimation issues. Whereas the histogram of b_c shows that there are some posterior estimates with extreme positive or negative values, the histograms for AS and SAM do not. These extreme values are probably the cause of the wide HPD intervals. Additionally, such estimation issues may occur when we try to estimate a location effect in a situation where there is no or just a very small location effect. Judging from the data plotted along the predicted regression curve in Figure 4, this may well be the case for the experience variable. In Section 6 we will illustrate what happens when we try to estimate a location effect in data where there is no or almost no location effect. Note, however, that experience does seem to influence the spread of the pointing error. Figure 5 shows the relation between experience and the concentration, which is the reciprocal of the spread, of the predicted values on the circle. The concentration was computed using the formula for the mean resultant length from Kendall (1974). In the Supporting Information predicted circular regression plots as well as figures showing the effect on the concentration are provided for all continuous variables in the pointing north data.

6. Simulation study

To assess the performance of the circular coefficients b_c, SAM and AS and the ability to distinguish between location and accuracy effects we conducted a simulation study with

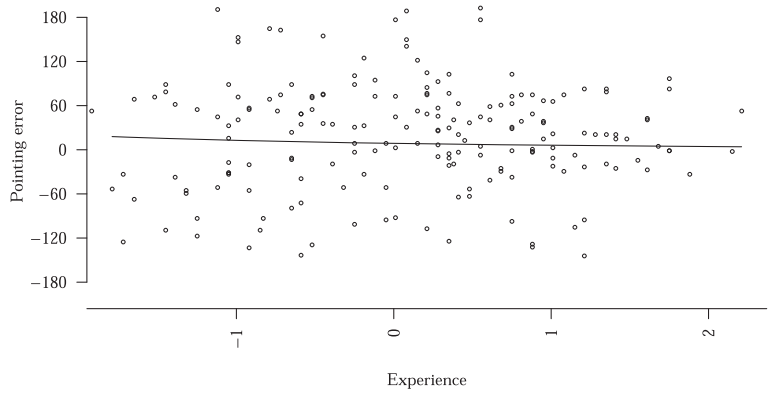

Figure 4. Predicted circular regression curve for the relation between experience and pointing error, together with the original data points. The inflection point of the regression curve is not shown as this point lies outside the range of the data.

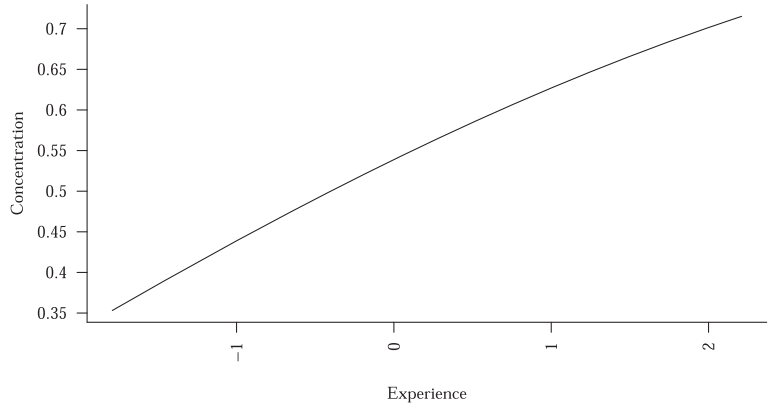

Figure 5. Relation between experience and the concentration of the predicted values on the circle.

1,225 designs with one predictor. Of these designs 1,056 were classified as location designs, 144 as accuracy and 25 as having no effect. Because the last category is so small its results are excluded from the simulation study. A description of the designs is given in Section 6.1. In Section 6.2 a summary of the simulation results is given, and in Section 6.3 we try to explain the causes of patterns observed in these results. A more detailed description of the simulation study and the results can be found in the Supporting Information for this paper.

6.1. Design

In each design different population values were chosen for the linear intercepts β_0^{I} and β_0^{II} and the regression coefficients β_1^{I} and β_1^{II}. From these values, the population values of the parameters a_x, a_c, b_c, SAM, AS, SDO and SSDO were computed. For each design 1,000 data sets were simulated, 500 with $N = 50$ and 500 with $N = 200$. Each data set contains one circular outcome θ and one linear predictor $x \sim N(0, 1)$. The relation between predictor and outcome was determined by the chosen values for the linear intercepts and coefficients. Before analysis of a data set the linear predictor was centred at 0. After

analysis the relative bias, frequentist coverage of the HPD interval and average interval width (AIW) of the estimates for b_c, SAM, AS and SSDO for each design were computed.

6.2. Results

In this section we briefly summarize the results from the simulation study with regard to how well we can detect location and accuracy effects and the performance of the MCMC sampler in estimating the circular coefficients. We will especially focus on the designs in which the measures did not perform optimally.

If there is any effect this can be found in more than 90% of the data sets of a design. This holds in all design categories, location and/or accuracy. The indicators for location effects that were tested, b_c and SSDO, work well for accuracy and location effects with SDO > 1. For location effects with SDO ≤ 1 the indicators perform worse. All indicators perform better in designs with a larger sample size.

Concerning the performance of the MCMC sampler in estimating the circular regression coefficients we can say that designs with a larger sample size and larger SSDO in most cases perform better in terms of relative bias, coverage and AIW. In accuracy designs the coefficients have smaller relative biases compared to location designs with an SSDO close to 0. In general AS and SAM have lower relative bias than b_c. In terms of coverage the AS shows slight undercoverage. The SAM has slight undercoverage in location designs and overcoverage in accuracy designs. The log AIW is largest for accuracy designs and the parameter b_c. This seems to correspond to the estimates for the example data.

6.3. Explaining patterns

To explain the patterns in relative bias, coverage and log AIW we will show what happens in the estimation of circular regression coefficients of an exemplary design. The exemplary design is a location design with an SSDO of 0.24 and a sample size of 50. The results for this design are summarized in Table 4.

Figure 6 shows histograms for the posterior modes for a_c, b_c, AS and SAM for all 500 data sets. Notice that the histograms for a_c, b_c and AS are bimodal. The bimodality is caused by the estimated a_c that switches to the other side of the circle. How often a_c switches sides is determined by the SDO. When the SDO is zero, in an accuracy design, a_c is equally likely to switch to either side of the circle and the histograms of modes will all be bimodal and symmetrical. When the SDO is large, the a_c will almost never switch to the opposite side of the circle and the histograms of modes will all be unimodal and symmetrical. In both cases the symmetry of the histograms of modes around the true value

Table 4. Simulation results of 500 simulated data sets ($N = 50$) from a population with $\beta_0^{\mathrm{I}} = 3$, $\beta_1^{\mathrm{I}} = 2$, $\beta_0^{\mathrm{II}} = 1$ and $\beta_1^{\mathrm{II}} = 0.5$

Parameter	Population value	Posterior mode	Bias	LB	UB	Coverage
a_x	−1.53	−1.46	−0.07	−1.91	−1.17	0.94
a_c	1.82	1.80	0.02	−1.28	1.89	0.96
b_c	−8.50	−2.45	−6.05	−72.62	65.01	0.95
AS	−0.41	−0.29	−0.12	−0.75	0.45	0.95
SAM	−0.05	−0.05	0.00	−0.17	0.06	0.96
SSDO	0.24	0.27	−0.02	−0.17	0.81	0.95

Note. LB and UB refer to the averaged lower and upper bounds of the 95% HPD intervals.

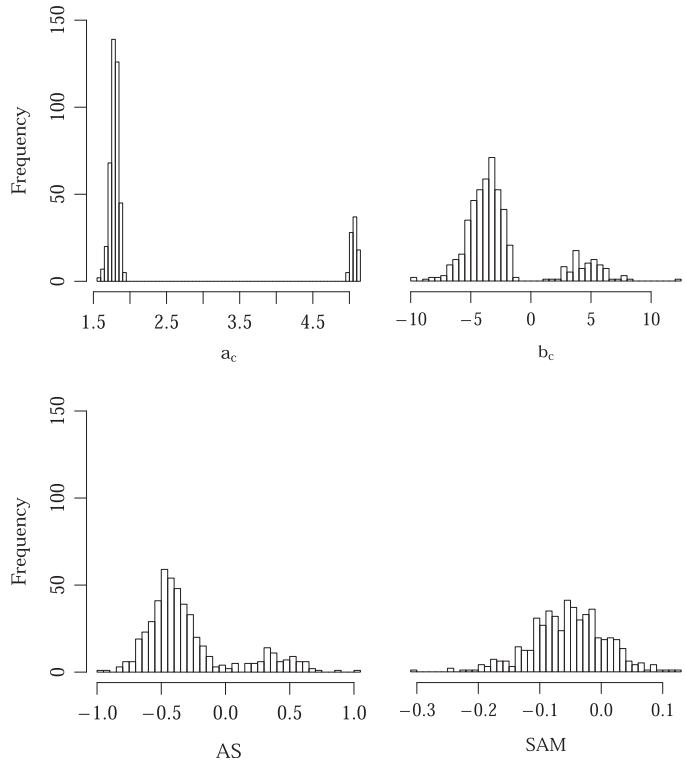

Figure 6. Histograms of the posterior modes of 500 simulated data sets ($N = 50$) from a population with $\beta_0^{\mathrm{I}} = 3$, $\beta_1^{\mathrm{I}} = 2$, $\beta_0^{\mathrm{II}} = 1$ and $\beta_1^{\mathrm{II}} = 0.5$ for the parameters a_c, b_c, AS and SAM.

results in little bias in the estimate of b_c and AS. In designs with a small SDO the histogram of modes is bimodal but not symmetrical. This causes bias in b_c and AS. The bimodality problem does not occur in the SAM, which explains the lower relative bias. Because of this property and its interpretation we prefer the SAM over b_c and AS.

In Section 5 we observed that for some iterations of the MCMC sampler the estimated b_c is either extremely negative or extremely positive. The extremes are caused by the tangent in the formula for b_c in equations (9) and (10). The tangent function has asymptotes at 0.5π radians and at -0.5π radians. If $\mathrm{atan2}(\beta_0^{\mathrm{II}} + \beta_1^{\mathrm{I}} x, \beta_0^{\mathrm{I}} + \beta_1^{\mathrm{II}} x) - a_c$ or $\mathrm{atan2}(\beta_0^{\mathrm{II}}, \beta_0^{\mathrm{I}}) - a_c$ is close to either one of these asymptotes we get extreme estimates for b_c. These extremes cause the AIWs to be large, as can be seen in Table 4. Large AIWs cause the coverage of the designs with small SDOs to be as good as or better than in designs with larger SDOs and lower the ability to detect location effects in the designs with smaller SDOs.

7. Discussion

The main contribution of this paper is to simplify the interpretation of effects in projected normal regression models. In the previous literature only the bivariate coefficients for a predictor were given, without much indication of how to properly interpret these. Therefore, we have developed methods for assessing circular effects. These methods

allow us to interpret and quantify the effect of a predictor on a circular outcome. A simulation study has shown that the performance of the methods is good in designs with an easily detectable location effect. If the location effect is harder to detect, the performance of the methods worsens. We need to increase the sample size to get more power and better performance. The slope at the mean (SAM) has an intuitive interpretation and performs best. The performance of the other circular coefficients seems to depend on the type of effect they are computed for. Therefore we recommend researchers to use the SAM and carefully investigate the circular regression plots and posterior histograms before using b_c or AS.

Additionally, we have investigated the ability of our method to detect different types of circular effects. If there is any effect it can usually be picked up by the linear coefficients. The coefficient b_c and SSDO perform equally well at detecting location effects. Assessing whether there is a location effect is harder in smaller samples, especially if it is a location effect with a small SDO. It is recommended that researchers make sure they have a large enough sample size to be able to detect the effect they are interested in. To be able to precisely say what sample size is needed more research needs to be done. From the present research we conclude that for a PN regression model with one continuous linear predictor a sample of 200 is large enough to be able to detect most location effects with small SDO.

The ability of the PN regression model to by default detect an effect, on the mean or spread, is an advantage over regression models of the intrinsic or wrapping approach to circular data because we do not need to fit two separate models. The model we use allows for investigation of the posterior distributions of the linear coefficients to check whether a location or accuracy effect is present in the data.

Although this paper focuses on regression models for the PN situation, we may consider using the tools introduced here in more complex models or in models using the general projected normal (GPN) distribution. One possibility is to use them in models where the mean of the PN distribution is partly composed of basis functions of the covariates (e.g., polynomials). Because locally polynomials look like a straight line, our tools might be applied here as well. Another example of a more complex model is the mixed-effects model that Nuñez-Antonio and Gutiérrez-Peña (2014) propose. In this model all tools introduced in this paper can be used on fixed effects. For random effects we can also consider these tools, but we have to be cautious in interpreting circular random effects. Their interpretation depends on the point relative to which the circular random effects are computed; the individual random intercepts or the average intercept. We are in this case also interested in the spread of the random effects. To assess the spread new tools will have to be developed. In GPN models the tools introduced here will be more complex to interpret and to implement. For skewed data interpretation problems could be overcome, but for bimodal data we would, for example, have to choose at what mode of the data we want to compute the SAM. In a regression model this seems redundant as we would usually try to explain possible bimodality by including predictors in the model (e.g., having different means for men and women). This can already be done in a PN model. Tools for a GPN model would also be hard to implement. There is no analytical solution for obtaining the mean direction of GPN models. This complicates the computation of circular predicted values. It is possible to get these using Monte Carlo integration (Wang & Gelfand, 2014), and we may also be able to compute the slope of a circular regression curve and the tools proposed in this paper in a similar way. However, the behaviour of these regression coefficients should then be investigated thoroughly,

especially their behaviour and use in a model where the GPN distribution is bimodal and/or skewed.

In conclusion, this paper has contributed to our knowledge about interpreting effects in projected normal regression models. We have outlined how to assess whether a predictor has an effect on either the spread or the mean of the circular outcome. But most importantly we have found a way of quantifying an effect on the mean of a circular outcome. These methods allow us to directly assess the effect of a predictor on the circle. In our opinion this has removed the major drawback of the projected normal regression model.

Acknowledgement

This work was supported by a Vidi grant awarded to I. Klugkist from the Dutch Organization for Scientific Research (NWO 452-12-010).

References

Brunyé, T. T., Burte, H., Houck, L. A., & Taylor, H. A. (2015). The map in our head is not oriented north: Evidence from a real-world environment. *PLoS One*, *10*(9), 1–12. https://doi.org/10.1371/journal.pone.0135803

Fisher, N. I., & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, *48*, 665–677. https:doi.org/10.2307/2532334

Hernandez-Stumpfhauser, D., Breidt, F. J., & van der Woerd, M. J. (2017). The general projected normal distribution of arbitrary dimension: Modeling and Bayesian inference. *Bayesian Analysis*, *12*(1), 112–133. https://doi.org/10.1214/15-BA989

Heyes, S. B., Zokaei, N., & Husain, M. (2016). Longitudinal development of visual working memory precision in childhood and early adolescence. *Cognitive Development*, *39*, 36–44. https://doi.org/10.1016/j.cogdev.2016.03.004

Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *Journal of the Royal Statistical Society, Series B*, *37*, 365–417.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *102*, 299–314. https://doi.org/10.1016/j.jecp.2008.07.005

König, J., Onnen, M., Karl, R., Rosner, R., & Butollo, W. (2016). Interpersonal subtypes and therapy response in patients treated for posttraumatic stress disorder. *Clinical Psychology & Psychotherapy*, *23*, 97–106. https://doi.org/10.1002/cpp.1946

Locke, K. D., Sayegh, L., Weber, C., & Turecki, G. (2017). Interpersonal self-efficacy, goals, and problems of persistently depressed outpatients: Prototypical circumplex profiles and distinctive subgroups. *Assessment*, Advance online publication. https://doi.org/10.1177/1073191116672330

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Chichester, UK: John Wiley & Sons.

Maruotti, A. (2016). Analyzing longitudinal circular data by projected normal models: A semi-parametric approach based on finite mixture models. *Environmental and Ecological Statistics*, *23*, 257–277. https://doi.org/10.1007/s10651-015-0338-3

Mastrantonio, G., Lasinio, G. J., & Gelfand, A. E. (2016). Spatio-temporal circular models with nonseparable covariance structure. *Test*, *25*, 331–350. https://doi.org/10.1007/s11749-015-0458-y

Matsushima, E. H., Vaz, A. M., Cazuza, R. A., & Ribeiro Filho, N. P. (2014). Independence of egocentric and exocentric direction processing in visual space. *Psychology & Neuroscience*, 7, 277–284. https://doi.org/10.3922/j.psns.2014.050

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, *31*(3), 705–741. https://doi.org/10.1214/aos/1056562461

Nuñez-Antonio, G., & Gutiérrez-Peña, E. (2005). A Bayesian analysis of directional data using the von mises–fisher distribution. *Communications in Statistics—Simulation and Computation*, *34*, 989–999. https://doi.org/10.1080/03610910500308495

Nuñez-Antonio, G., & Gutiérrez-Peña, E. (2014). A Bayesian model for longitudinal circular data based on the projected normal distribution. *Computational Statistics & Data Analysis*, *71*, 506–519. https://doi.org/10.1016/j.csda.2012.07.025

Nuñez-Antonio, G., Gutiérrez-Peña, E., & Escarela, G. (2011). A Bayesian regression model for circular data based on the projected normal distribution. *Statistical Modelling*, *11*, 185–201. https://doi.org/10.1177/1471082X1001100301

Ouwehand, P. E. W., & Peper, C. L. E. (2015). Does interpersonal movement synchronization differ from synchronization with a moving object? *Neuroscience Letters*, *606*, 177–181. https://doi.org/10.1016/j.neulet.2015.08.052

Presnell, B., Morrison, S. P., & Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, *93*, 1068–1077. https://doi.org/10.1080/01621459.1998.10473768

Rivest, L.-P., Duchesne, T., Nicosia, A., & Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Applied Statistics*, *63*, 445–463. https://doi.org/10.1111/rssc.12124

Santos, G. C., Vandenberghe, L., & Tavares, W. M. (2015). Interpersonal interactions in the marital pair and mental health: A comparative and correlational study. *Paidéia (Ribeirão Preto)*, *25*, 373–382. https://doi.org/10.1590/1982-43272562201511

Stoffregen, T. A., Bardy, B. G., Merhi, O. A., & Oullier, O. (2004). Postural responses to two technologies for generating optical flow. *Presence: Teleoperators and Virtual Environments*, *13*, 601–615. https://doi.org/10.1162/1054746042545274

Wang, F., & Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. *Statistical Methodology*, *10*(1), 113–127. https://doi.org/10.1016/j.stamet.2012.07.005

Wang, F., & Gelfand, A. E. (2014). Modeling space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association*, *109*, 1565–1580. https://doi.org/10.1080/01621459.2014.934454

Wright, A. G., Pincus, A. L., Conroy, D. E., & Hilsenroth, M. J. (2009). Integrating methods to optimize circumplex description and comparison of groups. *Journal of Personality Assessment*, *91*, 311–322. https://doi.org/10.1080/00223890902935696

Zilcha-Mano, S., McCarthy, K. S., Dinger, U., Chambless, D. L., Milrod, B. L., Kunik, L., & Barber, J. P. (2015). Are there subtypes of panic disorder? An interpersonal perspective. *Journal of Consulting and Clinical Psychology*, *83*, 938–950. https://doi.org/10.1037/a0039373

Received 29 September 2016; revised version received 19 May 2017

Supporting Information

The following supporting information may be found in the online edition of the article:

Appendix S1. Circular interpretation of projected normal regression coefficients: Supplementary material.

Appendix A: MCMC procedure for regression models

Consider the following model for a circular outcome variable:

$$\mathrm{PN}(\theta|\boldsymbol{\mu}, \mathbf{I}) = \frac{1}{2\pi} e^{-\frac{1}{2}||\boldsymbol{\mu}||^2} \left[1 + \frac{\mathbf{u}^t\boldsymbol{\mu}\Phi(\mathbf{u}^t\boldsymbol{\mu})}{\phi(\mathbf{u}^t\boldsymbol{\mu})}\right],$$

where θ is the circular outcome variable measured in radians $(-\pi \leq \theta < \pi)$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$ is the mean vector of this distribution, the variance–covariance matrix $\mathbf{I}$ is an identity matrix, $\mathbf{u}^t = (\cos\theta, \sin\theta)$, and $\Phi(\cdot)$ and $\phi(\cdot)$ respectively denote the cumulative distribution function and the probability density function of the standard normal distribution. The model for the mean vector is $\boldsymbol{\mu} = \mathbf{B}^t\mathbf{x}$, where $\mathbf{B} = [\boldsymbol{\beta}^{\mathrm{I}}, \boldsymbol{\beta}^{\mathrm{II}}]$ is a matrix of regression coefficients and $\mathbf{x}$ is a vector of predictor variables.

A method to estimate this circular regression model is presented in Nuñez-Antonio *et al.* (2011). The MCMC procedure used in this paper is the same except for the sampling of the vector of latent lengths, $\mathbf{r} = r_1, \ldots, r_n$, where n is the sample size (see equation (1)). Simulation studies have show that the method of sampling used by Nuñez-Antonio *et al.* (2011) works well; the performance in terms of bias and coverage are reasonable to good in most cases. Using a slice sampler for the latent lengths instead of a Metropolis–Hastings sampler results in improved performance and efficiency. R code for the sampler and results for the simulation can be requested from the authors.

In this appendix we only briefly mention the priors and conditional posteriors for the regression coefficients. A normal prior is specified for each of the two components of $\mathbf{B}$:

$$\boldsymbol{\beta}^j \sim N(\boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j), \quad \text{for } j = \mathrm{I}, \mathrm{II}, \tag{15}$$

where $\boldsymbol{\beta}_0^j$ is a vector with prior values for the regression coefficients and intercept and $\boldsymbol{\Lambda}_0^j$ is the prior precision matrix of component j. The full conditional density of $\boldsymbol{\beta}^j$ equals

$$\boldsymbol{\beta}^j|\boldsymbol{\theta}, \mathbf{r} \sim N(\boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j), \text{ for } j = \mathrm{I}, \mathrm{II}, \tag{16}$$

where $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$, $\boldsymbol{\mu}_F^j = (\boldsymbol{\Lambda}_F^j)^{-1}(\boldsymbol{\Lambda}_0^j\boldsymbol{\beta}_0^j + (\mathbf{X}^j)^t\mathbf{y}^j)$, and $\boldsymbol{\Lambda}_F^j = \boldsymbol{\Lambda}_0^j + (\mathbf{X}^j)^t\mathbf{X}^j$, where $\mathbf{X}^j$ is a design matrix. The latent lengths in $\mathbf{r}$ are given a prior that is uniform between 0 and ∞. The full conditional density of one latent length r_i can be found in Nuñez-Antonio *et al.* (2011) and equals

$$r_i|\theta_i, \boldsymbol{\mu}_i \propto r_i \exp(-0.5r_i^2 + b_i r_i), \tag{17}$$

where $b_i = \mathbf{u}_i^t\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_i = \mathbf{B}^t\mathbf{x}_i$. The sampler that can be used to obtain estimates for the vectors of regression coefficients $\boldsymbol{\beta}^j$ and values for the vector $\mathbf{r}$ consists of the following steps:

1. The priors for $\boldsymbol{\beta}^j$ are specified by choosing values for $\boldsymbol{\beta}_0^j$ and $\boldsymbol{\Lambda}_0^j$. In this paper we use 0 for each of the elements in $\boldsymbol{\beta}_0^j$. We specify $\boldsymbol{\Lambda}_0^j$ as a diagonal matrix with diagonal values equal to 1×10^{-4}.
2. A starting value for $\mathbf{r}$ is chosen. We choose a vector of ones.
3. Using the starting value for an r_i and $\mathbf{u}_i$ computed from the data, we may compute $\mathbf{y}_i = r_i\mathbf{u}_i$.

4. The $\boldsymbol{\beta}^j$ are sampled from their conditional posterior, equation (16).
5. Using the estimates for the $\boldsymbol{\beta}^j$, new r_i are sampled from their conditional posterior, equation (17), using slice sampling (Neal, 2003). The specifics of this slice sampler are presented by Hernandez-Stumpfhauser *et al.* (2017) and were adapted for the regression situation. The joint density for the auxiliary variable v_i with r_i for regression is

$$p(r_i, v_i|\theta_i, \boldsymbol{\mu}_i = \mathbf{B}^t\mathbf{x}_i) \propto r_i\mathbf{I}(0 < v_i < \exp\{-0.5(r_i - b_i)^2\})\mathbf{I}(r_i > 0). \tag{18}$$

The full conditionals for v_i and r_i are

$$p(v_i|r_i = r_i, \boldsymbol{\mu}_i, \theta_i) \sim U(0, \exp\{-0.5(r_i - b_i)^2\}), \tag{19}$$

$$p(r_i|v_i = v_i, \boldsymbol{\mu}_i, \theta_i) \propto r_i\mathbf{I}\left(b_i + \max\left\{-b_i, -\sqrt{-2\ln v_i}\right\} < r_i < b_i + \sqrt{-2\ln v_i}\right). \tag{20}$$

We thus sample v_i from the uniform distribution specified above. Independently we sample a value m from $U(0,1)$. We obtain a new value for r_i by computing $r_i = \sqrt{(r_{i_2}^2 - r_{i_1}^2)m + r_{i_1}^2}$, where $r_{i_1} = b_i + \max\{-b_i, -\sqrt{-2\ln v_i}\}$ and $r_{i_2} = b_i + \sqrt{-2\ln v_i}$.
6. Using the new r_i, new values for $\mathbf{y}_i$ are computed as $\mathbf{y}_i = r_i\mathbf{u}_i$
7. Steps 4, 5 and 6 are repeated for a specified number of iterations. After the iterations are completed convergence is checked. If convergence is not reached additional iterations are run.

Appendix B: Posterior histograms pointing north data

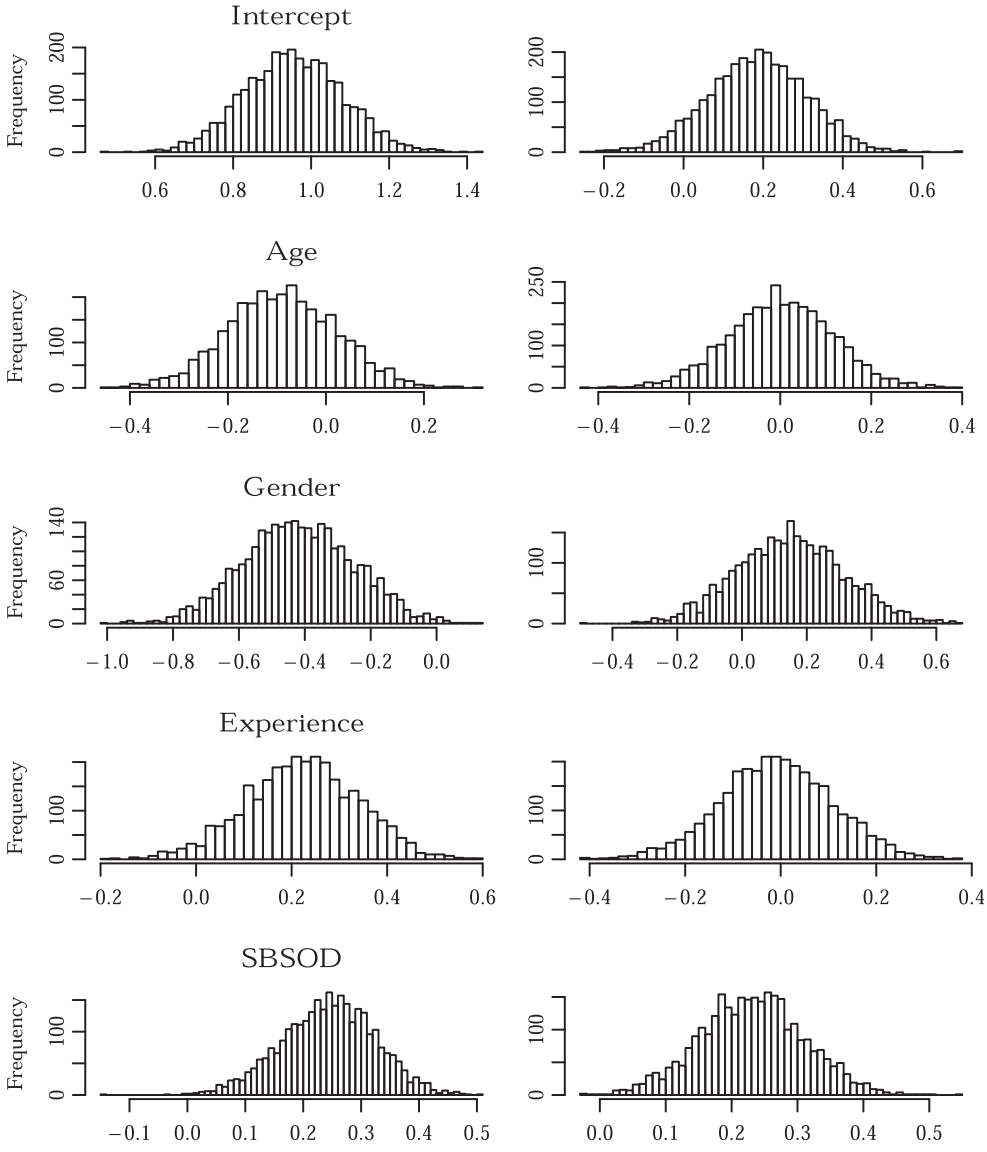

Figure B1. Posterior histograms for the linear intercepts, β_0^{I} and β_0^{I} and regression coefficients, β_1^{I}, β_2^{I}, β_3^{I}, β_4^{I}, β_1^{II}, β_2^{II}, β_3^{II}, and β_4^{II} for the pointing north data. Component I is shown on the left, component II is shown on the right.

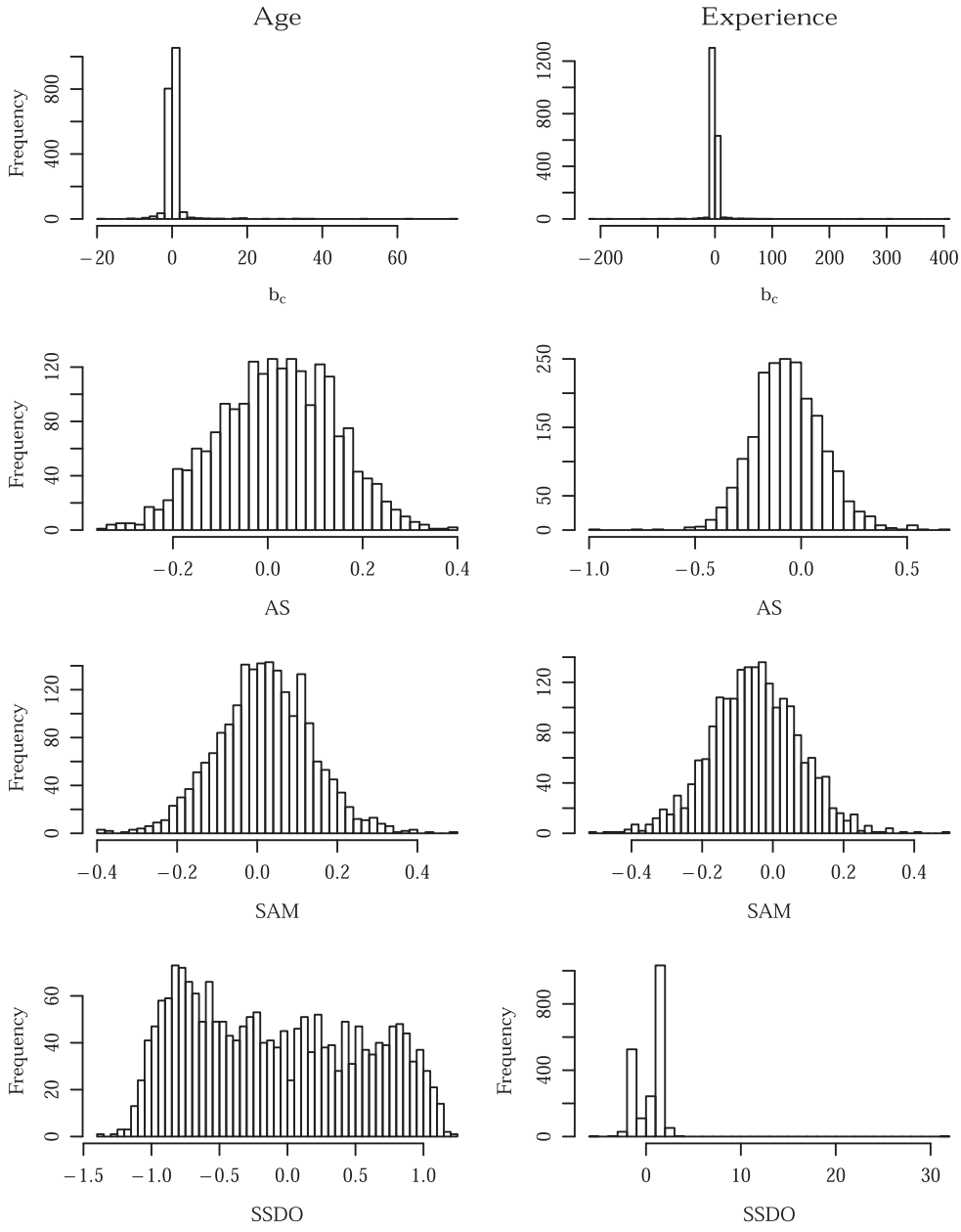

Figure B2. Posterior histograms for the circular regression coefficients and the SSDO for the age (left) and experience (right) variables of the pointing north data.