A Comparison of Regression models for Cylindrical data in Psychology.

Jolien Cremers[*1], *Helena J.M. Pennings*[2,3], *Christophe Ley* [4]
[1]*Department of Methodology and Statistics, Utrecht University*
[2]*TNO*
[3]*Department of Education, Utrecht University*
[4]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University*

Abstract

Cylindrical dats is multivariate data which consists of a directional, in this paper circular, component and a linear component. Examples of cylindrical data in psychology include human navigation (direction and distance of movement), eye-tracking research (direction and length of saccades) and data from an interpersonal circumplex (type and strength of interpersonal behavior). In this paper we adapt four models for cylindrical data to include a regression of the circular and linear component onto a set of covariates. Subsequently, we illustrate how to fit these models and interpret their results on a dataset on the interpersonal behavior of teachers.

1 Introduction

Cylindrical data are data that consist of a linear variable and a directional variable. In this paper, the directional variable is circular meaning that it consists of a single angle instead of a set of angles. A circular variable is different from a linear variable in the sense that it is measured on a different scale. Figure 1 shows the difference between a circular scale (right) and a linear scale (left). The most important difference is that on a circular scale the datapoints 0° and 360° are connected and in fact represent the same number while on a linear scale the two ends, $-\infty$ and ∞ are not connected. This difference requires us to use different statistical methods for circular variables (see e.g. Fisher (1995) for an introduction to circular data and Mardia & Jupp (2000), Jammalamadaka & Sengupta (2001) and Ley & Verdebout (2017) for a more elaborate overview).

Cylindrical data occur in several fields of research, such as for instance in meteorology (García-Portugués, Crujeiras, & González-Manteiga, 2013), ecology (García-Portugués, Barros, Crujeiras, González-Manteiga, & Pereira, 2014) or marine research (Lagona, Picone, Maruotti, & Cosoli, 2015). Several types of data in psychology are also of a cylindrical nature. For example, in research on human navigation in the field of cognitive psychology both distance, a linear variable, and direction, a circular variable, of movement are of interest (Chrastil & Warren, 2017). In eye-tracking research, which can be used for investigating various cognitive processes (e.g. those involved in reading a text), we can also speak of cylindrical data. When measuring eye-movements we speak of the eye-movements themselves, the saccades, and fixations, the periods of time between movements when the eyes are looking at one point. Of the saccades both the direction, a circular variable, and the duration, a linear variable, are of interest (for a review of eye-tracking research see Rayner (2009)). Data from circumplex measurement instruments, e.g. the interpersonal circumplex as used in personality psychology, are also of a cylindrical nature (see Section 2 for a more detailed explanation).

[*]Corresponding author: j.cremers@uu.nl

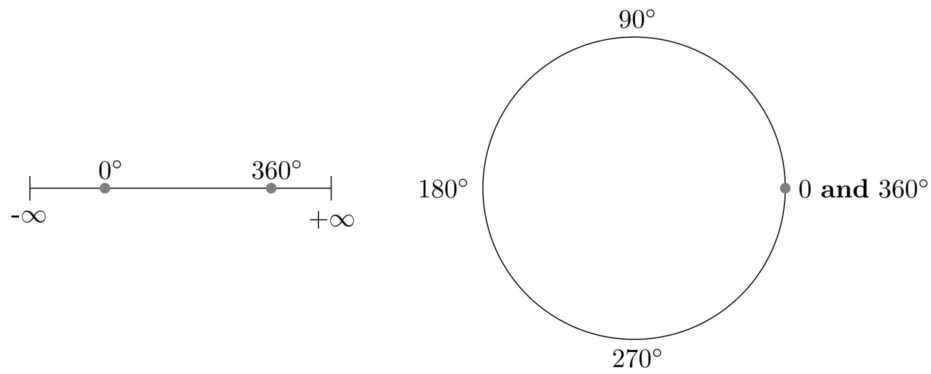

Figure 1: The difference between a linear scale (left) and a circular scale (right).

In this paper we will discuss how a correct statistical treatment of such cylindrical data can lead to new insights. In particular we will show how cylindrical models pave the way for circular-linear and linear-circular regression. We will do this for a motivating example, the teacher data, from the field of educational psychology . In this example, apart from modelling the relation between the linear and circular component of a cylindrical variable we would also like to predict the two components from a set of covariates in a regression model. The teacher data will be further introduced in Sections 2 and 4.

As is the case for circular data, the analysis of cylindrical data requires special methods. Several methods have been put forward to model the relation between the linear and circular component of a cylindrical variable. Some of these are based on regressing the linear component onto the circular component using the following type of relation:

$$y = \beta_0 + \beta_1 * \cos(\theta) + \beta_2 * \sin(\theta) + \epsilon,$$

where y is the linear component and θ the circular component (Johnson & Wehrly, 1978; Mardia & Sutton, 1978; Mastrantonio, Maruotti, & Jona-Lasinio, 2015). Others model the relation in a different way, e.g. by specifying a multivariate model for several linear and circular variables and modelling their covariance matrix (Mastrantonio, 2018) or by proposing a joint cylindrical distribution. For example, Abe & Ley (2017) introduce a cylindrical distribution based on a Weibull distribution for the linear component and a sine-skewed von Mises distribution for the circular component and link these through their shape and concentration parameters respectively. However, none of the methods that have been proposed thus far include additional covariates onto which both the circular and linear component are regressed.

Our aim in this paper is to fill this gap in the literature by adapting four existing cylindrical methods in such a way that they include a regression of both the linear and circular component of a cylindrical variable onto a set of covariates. First however, we will introduce the teacher data in Section 2. In Section 3 we introduce the four modified models for cylindrical data that we use to analyze the data from the motivating example. We also choose a model selection criterion to compare the models. The teacher data will be analysed in Section 4. The paper will be concluded with a discussion in Section 5. The Appendix contains technical details.

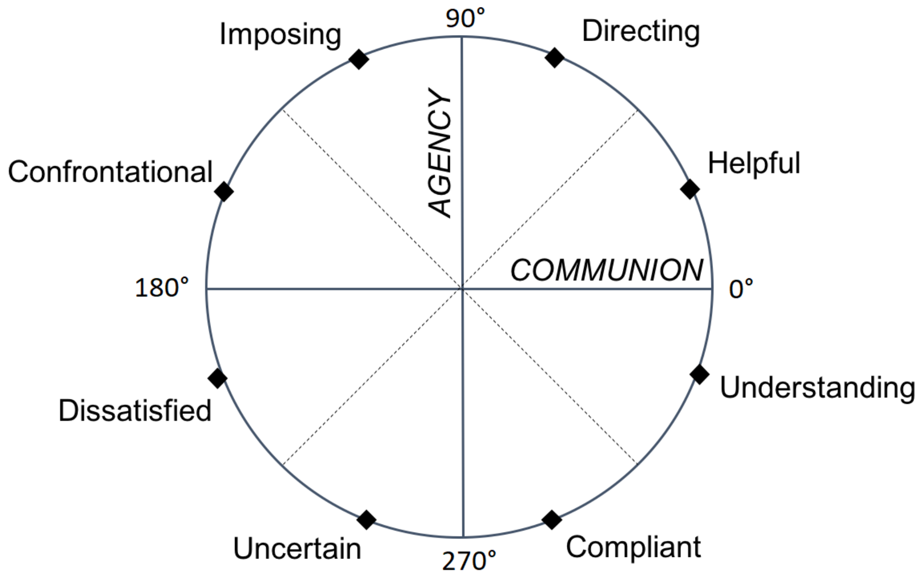

Figure 2: The interpersonal circle for teachers (IPC-T). The words presented in the circumference of the circle are anchor words to describe the type of behavior located in each part of the IPC.

2 Teacher data

The motivating example for this article comes from the field of educational psychology and was collected for the studies on classroom climate of Pennings et al. (2017), Claessens (2016) and Want (2015). An indicator of classroom climate is the students' perceptions of their teachers' interpersonal behavior. These interpersonal perceptions, both in educational psychology as well as in other areas of psychology, can be measured using circumplex measurement instruments (see Horowitz & Strack (2011) for an overview of many such instruments).

The Questionnaire on Teacher Interaction (QTI) (Wubbels, Brekelmans, Brok, & Tartwijk, 2006) is one such circumplex measurement instrument. It is used to study student perceptions of their teachers' interpersonal behavior and contains items that load on two dimensions: Agency and Communion. Agency refers to the degree of power or control a teacher exerts in interaction with his/her students. Communion refers to the degree of friendliness or affiliation a teacher conveys in interaction with his/her students. The loadings on the two dimensions of the QTI can be placed in a two-dimensional space formed by Agency (vertical) and Communion (horizontal), see Figure 2. Different parts of this space are characterized by different teacher behavior, e.g. 'helpful' or 'uncertain'. We call the two-dimensional space the interpersonal circle (IPC). The idea is that the IPC is "a continuous order with no beginning or end" (Gurtman, 2009, p. 2). We call such ordering a circumplex ordering and the IPC is therefore often called the interpersonal circumplex. The ordering also implies that scores on the IPC could be viewed as a circular variable.

Cremers, Mainhard, & Klugkist (2018a) show how data from the IPC can be considered circular data and analyzed as such using a regression model. The two-dimension scores Agency and Communion can be converted to a circular score using the two-argument arctangent function in (1), where A represents a score on the Agency dimension and C represents a score on the Communion dimension. The resulting circular variable θ can then be modeled. However, when two-dimensional data are converted to the circle we lose some information, the length of the two-dimensional vector (A, C), *i.e.*, its euclidean norm $|| \, (A, C) \, ||$. This length represents the strength of the type of interpersonal behavior a teacher shows towards his/her students. In a cylindrical model we are able to incorporate this information, and model a circular variable θ together with a linear variable corresponding to $|| \, (A, C) \, ||$. This leads to an improved analysis of data from the IPC. In the next section we introduce several models that can be used for more accurate and informative regression analysis on the teacher data. Descriptives for the teacher data are given in Section 4.

$$\theta = \operatorname{atan2}(A,\ C) = \left| \begin{array}{ll} \arctan\left(\frac{A}{C}\right) & \text{if} \quad C > 0 \\ \arctan\left(\frac{A}{C}\right) + \pi & \text{if} \quad C < 0 \ \ \& \ \ A \geq 0 \\ \arctan\left(\frac{A}{C}\right) - \pi & \text{if} \quad C < 0 \ \ \& \ \ A < 0 \\ +\frac{\pi}{2} & \text{if} \quad C = 0 \ \ \& \ \ A > 0 \\ -\frac{\pi}{2} & \text{if} \quad C = 0 \ \ \& \ \ A < 0 \\ \text{undefined} & \text{if} \quad C = 0 \ \ \& \ \ A = 0. \end{array} \right. \tag{1}$$

3 Four cylindrical regression models

In this section we introduce four cylindrical regression models to contain predictors for the linear and circular outcomes, Θ and Y. We extend the models from Mastrantonio (2018) and Abe & Ley (2017). Additionally, we introduce two models where the relation between Θ and Y is modelled as follows (following Mastrantonio et al. (2015)):

$$y = \gamma_0 + \gamma_{cos} * \cos(\theta) * r + \gamma_{sin} * \sin(\theta) * r + \gamma_1 * x_1 + \cdots + \gamma_q * x_q + \epsilon, \tag{2}$$

where r will be introduced in Section 3.1, the error term $\epsilon \sim N(0, \sigma)$, $\gamma_0, \gamma_{cos}, \gamma_{sin}, \gamma_1, \ldots, \gamma_q$ are the intercept and regression coefficients and $x_1, \ldots, x_q$ are the q covariate values. In this model Y follows a normal distribution and Θ a projected normal (PN) or general projected normal (GPN) distribution on the circle.

3.1 The modified CL-PN and modified CL-GPN models

In both of these models the relation between $\Theta \in [0, 2\pi)$ and $Y \in (-\infty, +\infty)$ is specified as in (2) with the following conditional distribution:

$$f(y \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{c^2 + (y - (\gamma_0 + \gamma_1 x_1 + \cdots + \gamma_q x_q))^2 - 2c(y - (\gamma_0 + \gamma_1 x_1 + \cdots + \gamma_q x_q))}{2\sigma^2}\right], \tag{3}$$

where $c = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}^t \begin{bmatrix} \gamma_{cos} \\ \gamma_{sin} \end{bmatrix}$, $r \geq 0$, $\gamma_0, \gamma_{cos}, \gamma_{sin}, \gamma_1, \ldots, \gamma_q$ are the intercept and regression coefficients and $\sigma^2 \geq 0$ is the error variance. The linear outcome thus has a normal distribution, conditional on Θ.

For the circular outcome we assume either a projected normal (PN) or a general projected normal (GPN) distribution. These distributions arise from the radial projection of a distribution defined on the plane onto the circle. The relation between a bivariate variate $\boldsymbol{S}$ in the plane and the circular outcome Θ is defined as follows:

$$\boldsymbol{S} = \begin{bmatrix} S^I \\ S^{II} \end{bmatrix} = R\boldsymbol{u} = \begin{bmatrix} R\cos\Theta \\ R\sin\Theta \end{bmatrix}, \tag{4}$$

where $R =\mid\mid S \mid\mid$, the euclidean norm of the bivariate vector $\boldsymbol{S}$. In the PN distribution we assume $\boldsymbol{S} \sim N_2(\boldsymbol{\mu}, \boldsymbol{I})$ and in the GPN we assume $\boldsymbol{S} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \rho^2 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho \in (-\infty, +\infty)$ and $\tau^2 \geq 0$ (as in Hernandez-Stumpfhauser, Breidt, & Woerd (2016)).

Following Nuñez-Antonio, Gutiérrez-Peña, & Escarela (2011), the joint density of Θ and R for the PN distribution in a regression set-up equals:

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{I}) = [2\pi]^{-1} \exp\left[\frac{-r^2 - \boldsymbol{\mu}^2\boldsymbol{\mu} + 2r\boldsymbol{u}^t\boldsymbol{\mu}}{2}\right], \tag{5}$$

In a regression setup the outcome θ_i for each individual $i = 1, \ldots, n$, where n is the sample size, is generated independently from (5). The mean vector $\boldsymbol{\mu}_i \in \mathbb{R}^2$ is defined as $\boldsymbol{\mu}_i = \boldsymbol{z}_i\boldsymbol{B}$. The vector $\boldsymbol{z}_i$ is a a vector of p covariate values and $\boldsymbol{B} = (\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II})$ contain the regression coefficients. Note however that the dimensions of $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^I$ need not necessarily be the same and we are thus allowed to have a different set of predictor variables and vectors $\boldsymbol{z}_i^I$ and $\boldsymbol{z}_i^{II}$ for the two components of $\boldsymbol{\mu}_i$.

Following Wang & Gelfand (2013) and Hernandez-Stumpfhauser et al. (2016) the joint density of r and θ for the GPN distribution equals:

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = r(2\pi\tau)^{-1} \exp\left[-0.5\sigma^2(r\boldsymbol{u} - \boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(r\boldsymbol{u} - \boldsymbol{\mu})\right], \tag{6}$$

where $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \rho^2 & \rho \\ \rho & 1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$. In a regression setup the outcome θ_i for each individual is generated independently from (6). The mean vector $\boldsymbol{\mu}_i \in \mathbb{R}^2$ is defined as $\boldsymbol{\mu}_i = \boldsymbol{z}_i(\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II})$. The vector $\boldsymbol{z}_i$ is a a vector of p covariate values and each $\boldsymbol{\beta}^k$ is a vector with two regression coefficients, one for each of the two components of $\boldsymbol{\mu}_i$. Note that for the CL-GPN model we do need to have the same predictors for both components of $\boldsymbol{\mu}_i$.

Both cylindrical models introduced here are estimated using MCMC methods based on Nuñez-Antonio et al. (2011), Wang & Gelfand (2013) and Hernandez-Stumpfhauser et al. (2016) for the regression of the circular outcome. A detailed description of the Bayesian estimation and MCMC samplers can be found in Appendices A.1 and A.2.

3.2 The modified Abe-Ley model

This model is an extension of the cylindrical model introduced by Abe & Ley (2017) to the regression context. The joint density of Θ and Y, in this model defined on the positive real half-line $[0, +\infty)$, is:

$$f(\theta, y) = \frac{\alpha(\beta)^\alpha}{2\pi\cosh(\kappa)}(1 + \lambda\sin(\theta - \mu))y^{\alpha-1}\exp[-((\beta y)^\alpha(1 - \tanh(\kappa)\cos(\theta - \mu)))], \quad (7)$$

In a regression setup the outcome vector $(\theta_i, y_i)^t$ for each individual is generated independently from (7). The parameters $\alpha > 0$ and $\beta_i = \exp(\boldsymbol{x}_i^t\boldsymbol{\nu}) > 0$ are linear shape and scale parameters, $\mu_i = \eta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t\boldsymbol{\eta}) \in [0, 2\pi)$, $\kappa > 0$ and $\lambda \in [-1, 1]$ are circular location, concentration and skewness parameters. The parameter $\boldsymbol{\nu}$ is a vector of q regression coefficients $\nu_j \in (-\infty, +\infty)$ for the prediction of y where $j = 0, \ldots, q$ and ν_0 is the intercept. The parameter $\eta_0 \in [0, 2\pi)$ is the intercept and $\boldsymbol{\eta}$ is a vector of p regression coefficients $\eta_j \in (-\infty, +\infty)$ for the prediction of θ where $j = 1, \ldots, p$. The vector $\boldsymbol{x}_i$ is a vector of predictor values for the prediction of y and $\boldsymbol{z}_i$ is a vector of predictor values for the prediction of θ.

As in Abe & Ley (2017), the conditional distribution of y given θ is a Weibull distribution with scale α and shape $\beta(1-\tanh(\kappa)\cos(\theta-\mu))^{1-\alpha}$ and the conditional distribution of θ given y is a sine skewed von Mises distribution with location parameter μ and concentration parameter $(\beta y)^\alpha \tanh(\kappa)$.

The log-likelihood for this model equals:

$$\begin{aligned} l(\alpha, \boldsymbol{\nu}, \lambda, \kappa, \boldsymbol{\eta}) = n[\ln(\alpha) - \ln(2\pi\cosh(\kappa))] + \alpha\sum_{i=1}^{n}\ln(\exp(\boldsymbol{x}_i^t\boldsymbol{\nu})) \\ + \sum_{i=1}^{n}\ln(1 + \lambda\sin(\theta_i - \eta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t\boldsymbol{\eta}))) + (\alpha - 1)\sum_{i=1}^{n}\ln(y_i) \\ - \sum_{i=1}^{n}(\exp(\boldsymbol{x}_i^t\boldsymbol{\nu})y_i)^\alpha(1 - \tanh(\kappa)\cos(\theta_i - \eta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t\boldsymbol{\eta}))) \end{aligned} \quad (8)$$

We can use numerical optimization (Nelder-Mead) to find solutions for the maximum likelihood (ML) estimates for the parameters of the model.

3.3 Modified joint projected and skew normal (GPN-SSN)

This model is an extension of the cylindrical model introduced by Mastrantonio (2018) to the regression context. It contains p circular outcomes and q linear outcomes. The circular outcomes $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_p)$ are modeled together by a multivariate GPN distribution and the linear outcomes $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_q)$ are modeled together by a multivariate skew normal distribution (Sahu, Dey, & Branco, 2003). Because the GPN distribution is modelled using a so-called augmented representation (as in (4) and (6)) it is convenient to use a similar tactic for modelling the multivariate skew normal distribution. Following Mastrantonio (2018) the linear outcomes are represented as:

$$\boldsymbol{Y} = \boldsymbol{\mu}_y + \boldsymbol{\Lambda}\boldsymbol{D} + \boldsymbol{H},$$

where where $\boldsymbol{\mu}_y$ is a mean vector for the linear outcome $\boldsymbol{Y}$, $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ is a $q \times q$ diagonal matrix with diagonal elements $\boldsymbol{\lambda} = \lambda_1, \ldots, \lambda_q$ (skewness parameters), $\boldsymbol{D} \sim HN_q(\boldsymbol{0}_q, \boldsymbol{I}_q)$, a q-dimensional half normal distribution (Olmos, Varela, Gómez, & Bolfarine, 2012), and $\boldsymbol{H} \sim N_q(\boldsymbol{0}_q, \boldsymbol{\Sigma}_y)$. This means that conditional on the auxiliary data $\boldsymbol{D}$, $\boldsymbol{Y}$ is normally distributed with mean $\boldsymbol{\mu}_y + \boldsymbol{\Lambda}\boldsymbol{D}$ and covariance matrix $\boldsymbol{\Sigma}_y$. The joint density for $(\boldsymbol{Y}, \boldsymbol{D})^t$ is defined as follows:

$$f(\boldsymbol{y}, \boldsymbol{d}) = 2^q \phi_q(\boldsymbol{y} \mid \boldsymbol{\mu}_y + \boldsymbol{\Lambda d}, \boldsymbol{\Sigma}_y)\phi_1(\boldsymbol{d} \mid \boldsymbol{0}_q, \boldsymbol{I}_q). \tag{9}$$

As in Mastrantonio (2018) dependence between the linear and circular outcome is created by modelling the augmented representations of $\boldsymbol{\Theta}$, and $\boldsymbol{Y}$ together in a $2 \times p + q$ dimensional normal distribution. The joint density of the model is then represented by:

$$f(\boldsymbol{\theta}, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{d}) = 2^q \phi_{2p+q}((\boldsymbol{s}, \boldsymbol{y})^t \mid \boldsymbol{\mu} + (\boldsymbol{0}_{2p}, diag(\boldsymbol{\lambda})\boldsymbol{d})^t, \boldsymbol{\Sigma})\phi_q(\boldsymbol{d} \mid \boldsymbol{0}_q, \boldsymbol{I}_q) \prod_{j=1}^{p} r_j, \tag{10}$$

where the mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_s, \boldsymbol{\mu}_y)^t$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_{sy} \\ \boldsymbol{\Sigma}_{sy^t} & \boldsymbol{\Sigma}_y \end{pmatrix}$. The matrix $\boldsymbol{\Sigma}_s$ is the covariance matrix for the variances of and covariances between the augmented representations of the circular outcome and the matrix $\boldsymbol{\Sigma}_{sy}$ contains covariances between the augmented representations of the circular outcome and the linear outcome.

In our regression extension we have $i = 1, \ldots, n$ observations of p circular outcomes, q linear outcomes and k covariates. The mean in the density in (10) then becomes $\boldsymbol{\mu}_i = \boldsymbol{x}_i\boldsymbol{B}$ where $\boldsymbol{B}$ is a $k \times (2 \times p + q)$ matrix with regression coeffients. We estimate the model using MCMC methods. A detailed description of these methods is given in Appendix A.3.

3.4 Model fit criterion

For the four cylindrical models we focus on their out-of-sample predictive performance to determine the fit of the model. To do so we split our data in a training and holdout set (10 % of the sample). A proper criterion to compare out-of-sample predictive performance is the the Predictive Log Scoring Loss (PLSL) (Gneiting & Raftery, 2007). The lower the value of this criterion, the better the predictive performance of the model. Using ML estimates this criterion can be computed as follows:

$$PLSL = -2 \sum_{i=1}^{M} \log l(x_i \mid \hat{\boldsymbol{\vartheta}}), \tag{11}$$

where l is the model likelihood, M is the sample size of the holdout set, x_i is the i^{th} datapoint from the holdout set and $\hat{\boldsymbol{\vartheta}}$ are the ML estimates of the model parameters. Using posterior samples the criterion is similar to the log pointwise predictive density (lppd) as outlined in Gelman et al. (2014) and can be computed as:

$$PLSL = -2 \frac{1}{B} \sum_{j=1}^{B} \sum_{i=1}^{M} \log l(x_i \mid \boldsymbol{\vartheta}^{(j)}), \tag{12}$$

where B is the amount of posterior samples and $\boldsymbol{\vartheta}^{(j)}$ are the posterior estimates of the model parameters for the j^{th} iteration. Note that although we fit the CL-PN, CL-GPN and joint GPN-SSN models using Bayesian statistics, we do not take prior information into account when assessing model fit. According to Gelman et al. (2014) (p. ???) this is not necessary since we are assessing the fit of a model to data, the holdout set, only. They argue that the prior in such case is only of

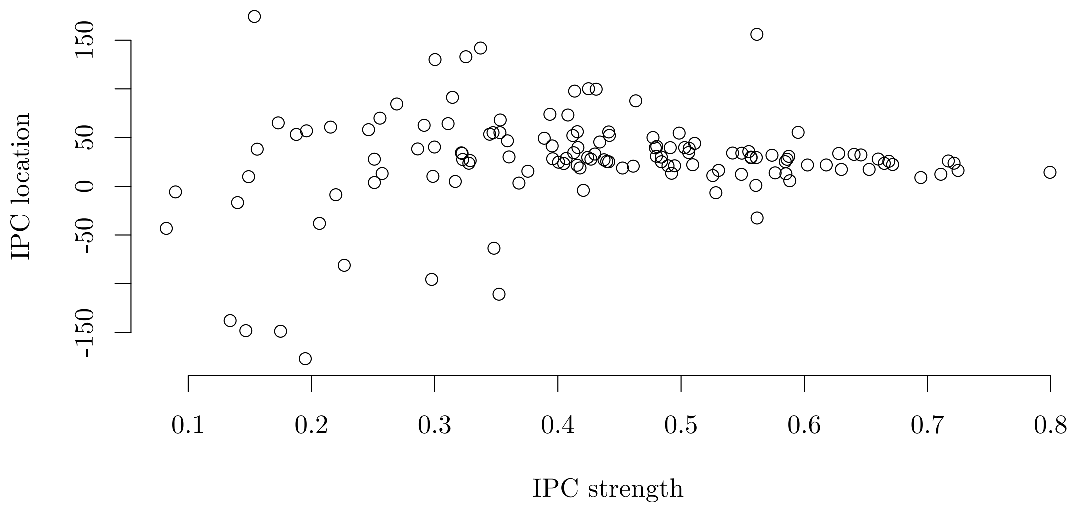

Figure 3: Plot showing the relation between the linear and circular component of the teacher data.

interest for estimating the parameters of the model but not for determining the predictive accuracy. For each of the cylindrical models we can then compute a PLSL for the circular and linear outcome by using the conditional log-likelihoods of the respective outcome.

4 Data Analysis

In this section we analyze the teacher data with the help of the four cylindrical models from Section 3. First however, we give a more detailed description of the dataset.

4.1 Data description

The teacher data was collected between 2010 and 2015 and contains several repeated measures on the IPC of 161 teachers. Measurements were obtained using the QTI and taken in different years and classes. For this paper we only consider one measurement, the first occasion (2010) and largest class if data for multiple classes were available. In addition to the score on the IPC, the circular outcome, and the strength of the score on the IPC, the linear outcome, a teachers' self-efficacy (`SE`) concerning classroom management ia used as covariate in the analysis. After the removal of missings we have a sample of 148 teachers of which 10% (n = 15) is used as a holdout set for the computation of the PLSL. Table 1 shows descriptives for the dataset. Figure 3 is a scatterplot showing the relation between the linear and circular outcome of the teacher data.

Table 1: Descriptives for the teacher dataset

Variable	mean/$\bar{\theta}$	sd/$\hat{\rho}$	Range	Type
IPC	33.22°	0.76	-	Circular
strength IPC	0.43	0.15	0.08 - 0.80	Linear
SE	5.04	1.00	1.5 - 7.0	Linear

Note that $\hat{\rho}$ is an sample estimate for the circular concentration where a value of 0 means that the data is not concentrated at all, i.e. spread over the entire circle, and a value of 1 means that all data is concentrated at a single point on the circle.

4.2 Models

The regression equations for the linear and the circular outcome of the four cylindrical models fit to the teacher dataset are as follows:

- For the modified CL-PN and CL-GPN models:

  $\hat{\boldsymbol{\mu}}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} \beta_0^I + \beta_1^I \text{SE}_i \\ \beta_0^{II} + \beta_1^{II} \text{SE}_i \end{pmatrix},$

  $\hat{y}_i = \gamma_0 + \gamma_{cos} \cos \theta_i r_i + \gamma_{sin} \sin \theta_i r_i + \gamma_1 \text{SE}_i.$

- For the modified Abe-Ley model:

  $\hat{\mu}_i = \eta_0 + 2 * \tan^{-1}(\eta_1 \text{SE}_i),$

  $\hat{\beta}_i = \exp(\nu_0 + \nu_1 \text{SE}_i).$

- For the modified joint projected and skew normal model:

  $\hat{\boldsymbol{\mu}}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{SE}_i$, where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{s_i}, \boldsymbol{\mu}_{y_i})^t$, $\boldsymbol{\beta}_0 = (\beta_{0_s^I}, \beta_{0_s^{II}}, \beta_{0_y})$ and $\boldsymbol{\beta}_1 = (\beta_{1_s^I}, \beta_{1_s^{II}}, \beta_{1_y})$.

We use the loglikelihoods of the following conditional densities for the computation of the PLSL:

- for the modified CL-PN model:

  $y_i \mid \mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$, where $\mu_i = \hat{y}_i$ and for θ_i we use (5).

- for the modified CL-GPN model:

  $y_i \mid \mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$, where $\mu_i = \hat{y}_i$ and for θ_i we use (6).

- for the modified Abe-Ley model:

  $y_i \mid \boldsymbol{\theta}_i, \beta_i, \mu_i, \kappa, \alpha \sim W\left(\beta_i(1 - \tanh(\kappa)\cos(\theta_i - \mu_i))^{1/\alpha}, \alpha\right)$, a Weibull distribution.

  $\theta_i \mid y_i, \beta_i, \mu_i, \kappa, \alpha\lambda \sim SSVM\left(\mu_i, (\beta_i y_i)^\alpha (\tanh \kappa)\right)$, a sine-skewed von Mises distribution.

- for the modified joint projected and skew normal model:

  $y_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\theta}_i, r_i \sim SSN(\mu_{i_y} + \lambda d_i + \boldsymbol{\Sigma}_{sy}^t \boldsymbol{\Sigma}_w^{-1}(\boldsymbol{s}_i - \boldsymbol{\mu}_{i_s}), \boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_{sy}^t \boldsymbol{\Sigma}_s \boldsymbol{\Sigma}_{sy}),$

  $\theta_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, y_i, d_i \sim GPN(\boldsymbol{\mu}_{i_s} + \boldsymbol{\Sigma}_{sy} \boldsymbol{\Sigma}_y^{-1}(y_i - \mu_{i_y} - \lambda d_i), \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_{sy} \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\Sigma}_{sy}^t)$

  where SSN is the skew normal distribution.

Table 2: Results for the modified CL-PN and CL-GPN model

Parameter	CL-PN			CL-GPN		
	Mode	HPD LB	HPD UB	Mode	HPD LB	HPD UB
β_0^I	1.75	1.46	2.00	2.39	1.87	3.00
β_1^I	0.62	0.37	0.84	0.74	0.36	1.19
β_0^{II}	1.12	0.89	1.33	1.40	1.12	1.73
β_1^{II}	0.54	0.36	0.76	0.65	0.43	0.90
γ_0	0.37	0.31	0.43	0.36	0.30	0.41
γ_{cos}	0.04	0.02	0.07	0.03	0.01	0.05
γ_{sin}	-0.02	-0.05	0.02	-0.01	-0.04	0.02
γ_1	0.03	0.00	0.07	0.03	0.00	0.06
σ	0.14	0.12	0.16	0.13	0.12	0.16
$\sum_{1,1}$	NA	NA	NA	3.02	1.87	4.97
$\sum_{1,2}$	NA	NA	NA	0.59	0.20	0.88
$\sum_{2,2}$	NA	NA	NA	1.00	1.00	1.00

Table 3: Results for the modified Abe-Ley model

Parameter	ML-estimate
η_0	0.36
η_1	-0.02
ν_0	1.19
ν_1	0.01
α	3.82
κ	1.58
λ	0.68

4.3 Results & Analysis

Before analysis there were a couple of settings that we had to specify for the cylindrical models. Starting values for the Abe-Ley model were the following $\eta_0 = 0.9, \eta_1 = 0.9, \nu_0 = 0.9, \nu_1 = 0.9, \kappa = 0.9, \alpha = 0.9, \lambda = 0$. The initial amount of iterations for the three MCMC samplers was set to 2000. After we checked convergence via traceplots we concluded that some of the parameters of the joint GPN-SSN model did not converge. Therefore we set the amount of iterations of the MCMC models to 20,000 and subtracted a burn-in of 5000. With this specification the MCMC chains for all parameters converge. Note that we choose the same amount of iterations for all three Bayesian cylindrical models to make their comparison via the PLSL as fair as possible. Lastly, the predictor `SE` was centered before inclusion in the analysis.

Tables 2, 3 and 4 show results from the four models that were fit to the teacher dataset. For the models in which Bayesian estimation was used we show both the estimated posterior mode and the 95% highest posterior density (HPD) interval for each parameter. Before we turn to an evaluation of the fit of the models, we discuss results of and the interpretation of parameters from the four models. We will cover parameters concerning the circular outcome, linear outcome and the

Table 4: Results for the modified joint projected and skew normal model

Parameter	Unconstrained			Constrained		
	Mode	HPD LB	HPD UB	Mode	HPD LB	HPD UB
$\beta_{0_s^I}$	0.31	0.26	0.34	2.07	1.77	2.50
$\beta_{0_s^{II}}$	0.19	0.16	0.21	1.28	1.06	1.53
β_{0_y}	0.33	0.30	0.36	0.33	0.30	0.36
$\beta_{1_s^I}$	0.08	0.04	0.12	0.54	0.28	0.82
$\beta_{1_s^{II}}$	0.07	0.04	0.09	0.46	0.29	0.65
β_{1_y}	0.08	0.05	0.12	0.08	0.05	0.12
$\sum_{s1,1}$	0.05	0.04	0.06	2.30	1.66	3.34
$\sum_{s2,2}$	0.02	0.02	0.03	1.00	1.00	1.00
$\sum_{y3,3}$	0.03	0.02	0.04	0.03	0.02	0.04
$\sum_{s1,2}$	0.00	0.00	0.01	0.09	-0.20	0.35
$\sum_{sy1,3}$	0.03	0.03	0.05	0.23	0.18	0.32
$\sum_{sy2,3}$	0.01	0.01	0.02	0.09	0.05	0.12
λ	0.15	0.13	0.17	0.15	0.13	0.17

association between them separately.

4.3.1 Circular component

Firstly, we can compare the estimated mean of the circular outcome between the four models. In Table 5 we see the means of the posterior predictive distributions for the circular outcome of the CL-PN, CL-GPN and joint GPN-SSN model. For the CL-PN model we can actually also compute this mean from the estimates in Table 2 as follows: $\mu_{circ} = atan2(\beta_0^{II}, \beta_0^I) = 32.62°$. The estimates for the CL-PN, CL-GPN and joint GPN-SSN model are about equal and correspond quite well to the actual data mean of 33.22° in Table 1. The estimate from the Abe-Ley model which is 0.36 radians or 20.62° is different. This difference could be caused by the fact that the densities for the circular ouctome, projected normal or sine-skewed von Mises, differ between the models.

We can also compare the effect of self-efficacy (`SE`) on the circular outcome. For the CL-PN model and Abe-Ley models we can get estimates of a circular regression coefficient of this effect. For the Abe-Ley model this estimate is the parameter $\eta_1 = -0.02$. This means that for each unit increase in self-efficacy, at the inflection point of the circular regression line, the score of the teacher on the IPC decreases with $0.02 * (180/\pi) = 1.15°$. CHRISTOPHE: KLOPT DIT, OF MOET IK λ HIER NOG MEENEMEN; NEGEER IK NU HET SINE-SKEWED GEDEELTE The inflection point is the point at which the regression line starts flattening off, i.e. the steepness of the slope decreases. For the CL-PN model we can use methods from Cremers, Mulder, & Klugkist (2018b). The estimated posterior mode of b_c, a parameter that is comparable to η_1 in the Abe-Ley model, equals 1.53 and its 95% HPD interval is $(-17.37; 21.26)$. This means that for each unit increase in self-efficacy, at the inflection point of the circular regression line, the score of the teacher on the IPC increases with $1.53 * (180/\pi) = 87.66°$. The values of b_c and η_1 are thus quite different even though they should represent the same effect. However, the slope at the inflection point that b_c and η_1 describe is not necessarily representative of the effect of `SE` in the data range. In Figure 4 we see that the inflection

Table 5: Posterior estimates for the circular mean in the CL-PN, CL-GPN and joint GPN-SSN models

	Mean (degrees)
Cl-PN	31.79
CL-GPN	33.38
GPN-SSN	34.47

[a] Note that these means are based on the posterior predictive distribution for the intercepts following (Wang & Gelfand, 2013).

point (square) of the predicted circular regression line for the CL-PN model lies just ouside of the data range and that for the Abe-Ley model even further (outside the x-axis range). The slope of the regression lines in the data range is much more alike even though one is positive and the other negative. The slope at the mean self-efficicay, which has a value of 0 because it was centered, SAM is estimated at $0.07(-0.03; 0.17)$. The SAM can be interpreted as the amount of change on the circle (in radians) for a unit increase at the mean of the predictor variable. However, for the CL-PN model the 95% HPD interval of SAM for the effect of `SE` includes 0 meaning that the value is not different from 0. THis explains in part how the slopes of the two regression lines are predicted to be different. For the Abe-Ley model standard errors of the parameters are not known so we cannot formally test whether η_1 differs from 0. Neither is there a known way of computing the effect at predictor values other than the one of the inflection point.

For the CL-GPN and GPN-SSN we cannot compute circular regression coefficients. Instead, we will compute posterior predictive distributions for the predicted circular outcome of individuals scoring the minimum, maximum and average self-efficacy. The modes and 95% HPD intervals of these posterior predictive distributions in the CL-GPN model are $\hat{\theta}_{SE_{min}} = 211.31°(144.41°\ 43.43°)$, $\hat{\theta}_{SE_{mean}} = 25.82°(335.55°\ 145.11°)$ and $\hat{\theta}_{SE_{max}} = 30.96°(6.72°\ 74.35°)$ respectively. For the GPN-SSN model the modes and 95% HPD intervals of the posterior predictive distributions are $\hat{\theta}_{SE_{min}} = 352.06°(121.82°\ 77.75°)$, $\hat{\theta}_{SE_{mean}} = 22.74°(336.71°\ 132.25°)$ and $\hat{\theta}_{SE_{max}} = 30.87°(358.80°\ 82.13°)$ respectively. For both models we can thus conclude that as the self-efficacy increases, the score of the teacher on the IPC moves counterlockwise. We can however not reach any conclusions about the size of this effect or whether it is different from zero.

4.3.2 Linear component

For the linear outcome, we compare the regression coefficients of `SE` for the linear outcome, the strength of the interpersonal behavior of a teacher. For the CL-PN, CL-GPN and joint GPN-SSN model this are the parameters γ_1 and β_{1_y} respectively. Note that in the CL-PN and CL-GPN model the effect of `SE` is the effect controlled for $\cos\theta$ and $\sin\theta$ similar to a 'normal' multiple regression model. For the GPN-SSN model the effect of `SE` is also controlled for the circular outcome. However in this model the associaton is modelled in a way that is similar to modelling the association

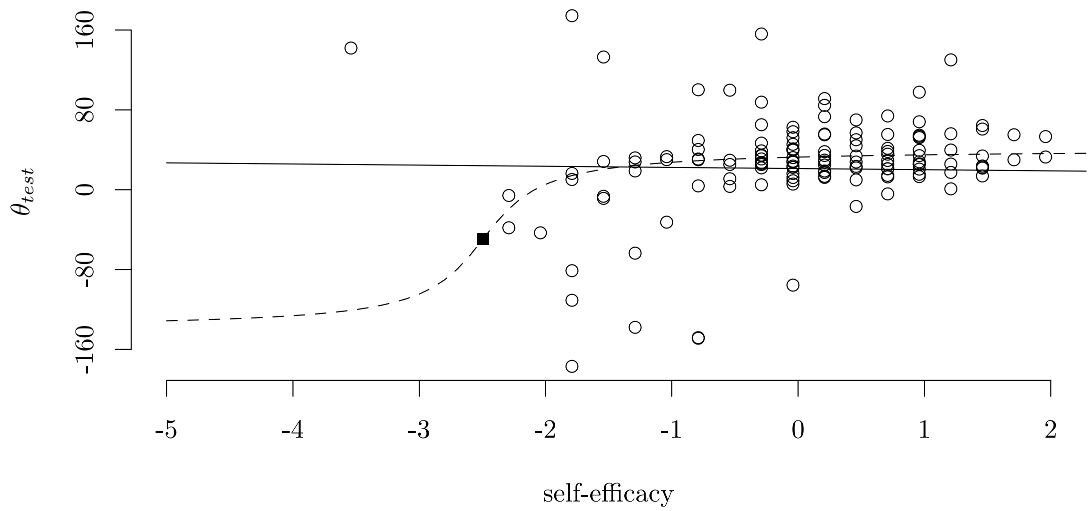

Figure 4: Plot showing circular regresion lines for the effect of self-efficacy as predicted by the Abe-Ley (solid line) and CL-PN model (dashed line). The black square indicated the inflection point of the circular regression line for the CL-PN model.

between several outcome variables in a multivariate regression or ANOVA model. Additionally, the distribution we assume for the linear outcome is different. For the CL-PN and CL-GPN models we assume a normal distribution while for the GPN-SSN model we assume a skewed normal distribution. These differences will lead to differences in the estimated effect of `SE` on the linear outcome. The interpretation of the regression coefficients is however similar for all three models and is like the interpretation of a regression coefficient in any other standard linear regression model.

For all three models there is an effect of `SE` on the strength of the interpersonal behavior of a teacher, the HPD interval does not include 0: $\gamma_1 = 0.03\,(0.00; 0.07)$ (CL-PN), $\gamma_1 = 0.03\,(0.00; 0.06)$ (CL-GPN) and $\beta_{1_y} = 0.08\;(0.05; 0.12)$ (GPN-SSN). This means that teachers with a higher self-efficacy for classroom management are stronger in their interpersonal behavior than teachers with a lower self-efficacy. In particular, for each unit increase in self-efficacy the strength of their interpersonal behavior is 0.03 or 0.08 (depending on the model) higher. This is an increase of 4.17 or 11.11 % considering the range of the strength on the IPC in the data.

For the Abe-Ley model we make use of the fact that the conditional distribution for the linear outcome is Weibull. This means that we can use methods from survival analsis in medical statistics to interpret the effect of self-efficacy. In survival analysis they make use of a 'survival' function in which the time is plotted against the probability of survival of subjects suffering from a specific medical condition. In our data however we plot the strength on the IPC (instead of time) against the probability of a teacher having such a strength. This probability can be computed using the so-called 'survival-function' which in our case is equal to $\exp(-\alpha y_i^{\beta})$ where $\beta = \exp(\nu_0 + \nu_1 SE_i)$. In Figure 5 we plot the survival function for the minimum, mean and maximum value of self-efficacy. From this Figure we conclude that the stronger interpersonal behaviors are less probable. We also see that the self-efficacy positively influences the strength on the IPC, the probability of having a

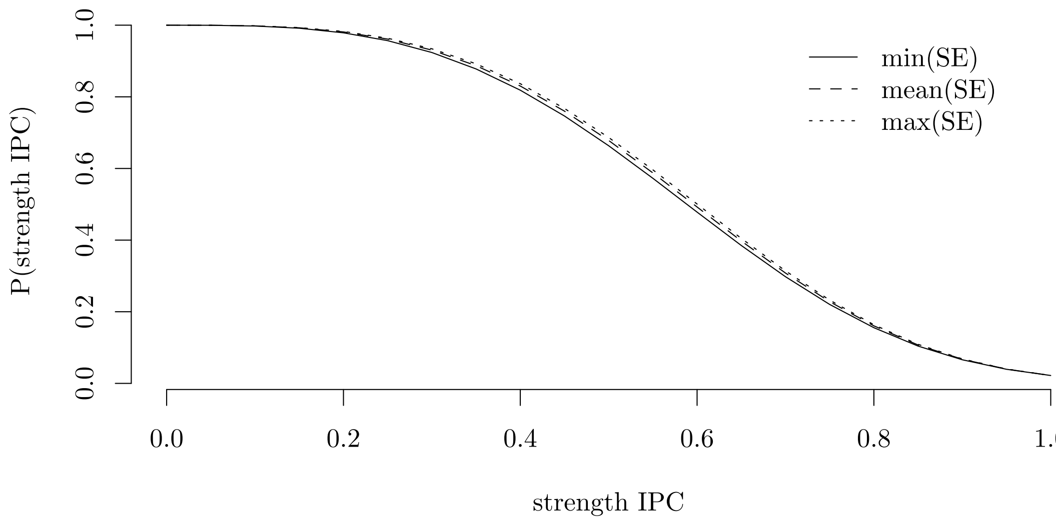

Figure 5: Plot showing the probability of having a particular strength of interersonal behavior (survival plot) for the minimum, mean and maximum self-efficacy in the data.

stronger interpersonal behavior increases with increasing self-efficacy. This result is equivalent to the effect of self-efficacy found using the CL-PN, CL-GPN and GPN-SSN models.

4.3.3 Association between the linear and circular component

For all four models we can also investigate the association between the linear and circular component. In the CL-PN and CL-GPN model we do this using the parameters $\gamma_{\cos}$ and γ_{sin}, the coefficients of the regression of the linear outcome onto the sine and cosine of the circular outcome. In the joint SSN-GPN model we look at the covariances between the linear outcome and the sine and cosine of the circular outcome $\sum_{sy_{2,3}}$ and $\sum_{sy_{1,3}}$. For the Abe-Ley model we can use the model parameters in Table 3 to compute a circular-linear correlation (Abe & Ley, 2017).

In both the CL-PN and CL-GPN model there is an effect of the cosine of the circular outcome onto the linear outcome (HPD interval does not include 0). In the teacher data the sine and cosine component have a substantive meaning. In this case the Communion component of the IPC positively effects the strength of a teachers' type of interpersonal behavior. Thus the teachers showing interpersonal behavior types with higher communion scores (e.g. 'helpful' and 'understanding' in Figure 2) are stronger in their behavior. In the joint SSN-GPN model we reach a sligtly different conclusion. In this model both covariances, $\sum_{sy_{2,3}} = 0.09$ and $\sum_{sy_{1,3}} = 0.23$, are different from 0, but the one between the cosine of the circular outcome and the linear outcome is larger. This means that both the Communion and Agency component of the IPC positively effect the strength of a teachers' type of interpersonal behavior but the effect of the Communon component is larger. Thus, teachers scoring high on Agency but slightly higher on Communion (the 'helpful' category) are strongest in their behavior. Following Abe & Ley (2017) the estimated correlation between the circular and linear outcome is

Table 6: PLSL criteria for the circular and linear outcome in the four cylindrical models

	CL-PN	CL-GPN	Abe-Ley	Joint GPN-SSN
circular	86.49	46.83	59.34	111.63
linear	-20.85	-20.94	51.46	15.48

equal to CHRISTOPHE: KAN JIJ DEZE BEREKENEN, IN R KAN IK GEEN MANIER VINDEN OM LEGENDRE FUNCTIES MET NIET INTEGER DEGREE TE EVALUEREN EN ALS IK http://functions.wolfram.com/webMathematica/FunctionEvaluation.jsp?name=LegendreP2General GEBRUIKT KRIJG IK IMAGINAIRE GETALLEN (voor m = 1 en m = 2), IK WEET NIET OF DAT DE BEDOELING IS... This correlation is not affected by μ or β meaning that the covariate, self-efficacy, can not influence the correlation between the type and strength of interpersonal behavior.

4.3.4 Model fit

Instead of interpreting specific model parameters we can also look at the overall fit of the models to the data. In Section 3.4 we introduced the criterion, PLSL, that we use to assess model fit in this paper. This criterion is focused on the out-of-sample predictive performance. Table 6 shows the values of this criterion for the linear and circular outcome of the four different models fit to the teacher data.

The CL-PN and CL-GPN model have the best out-of-sample predictive performance for the linear outcome. Note that they have roughly the same performance since they model the linear outcome in the same way apart from the value of r in (2) which is a parameter computed in the estimation of the parameters for the circular variable. This similarity in fit is also shown in Figure 6 which shows histograms of the test and holdout data together with the (posterior) predicitive distributions for all four cylindrical models. The posterior predictive distributions for the CL-PN and CL-GPN model are almost the same for the linear outcome.

The CL-GPN model has the best out-of-sample predicitve performance for the circular outcome. The Abe-Ley model has the second best performance. This means that for the circular variable a slightly skewed distribution (see Figure 6) fits best. The GPN-SSN model fits worst even though it assumes the same distribution for the circular outcome as the CL-GPN model. This could be due to the fact that the relation between the circular and linear outcome is modeled differently in these two models. In the CL-GPN model the linear outcome depends on the circular one following the relation in (2) but the circular outcome is not dependent on the linear one. In the GPN-SSN model both outcomes are modelled with a joint variance-covariance matrix governing their dependence. In Figure 6 we see however that the actual predictive densities do not differ that much between the four models. It is possible that the large difference in PLSL are caused by the fact that they are computed for a relatively small holdout dataset (n = 15) CHRISTOPHE: IS DIT LOGISCH?.

5 Discussion

In this paper we have modified four models for cylindrical data such that they include a regression of both the linear and circular components onto a set of covariates. Subsequently we have used these

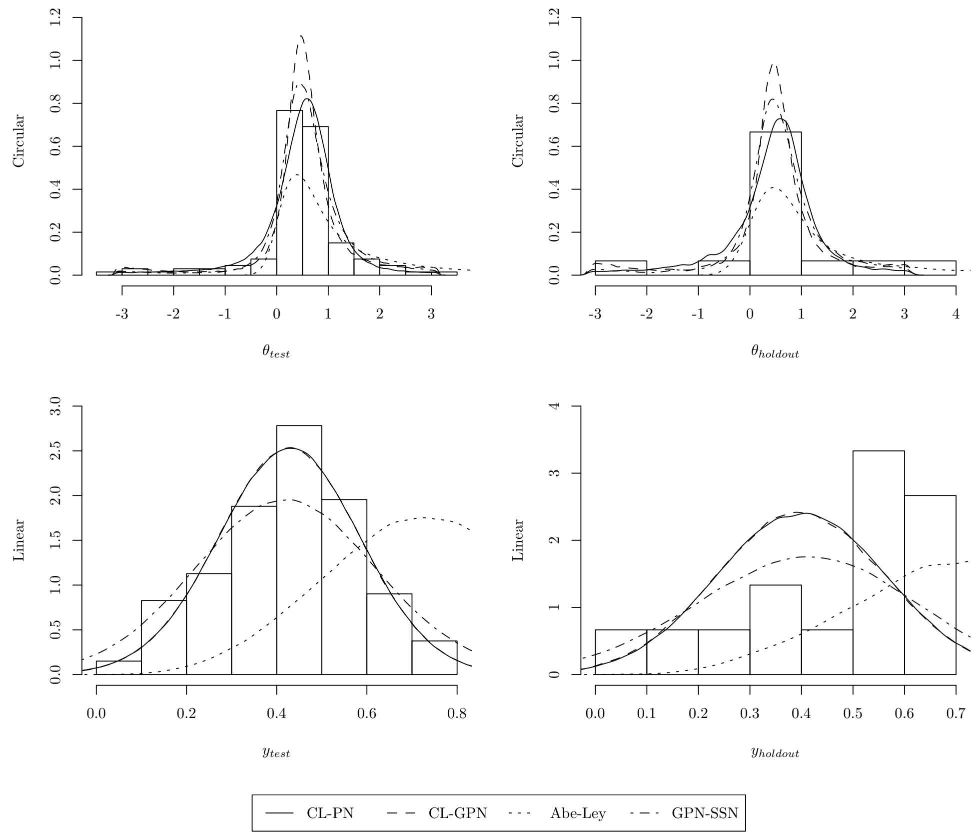

Figure 6: Histograms of linear and circular values of the test and holdout set of the teacher data plotted together with (posterior) predictive density estimates for the modified CL-PN, CL-GPN, Abe-Ley and joint GPN-SSN models based on $n \times 1000$ samples from the (posterior) predictive distribution.

four methods to analyze a dataset on the interpersonal behavior of teachers. Here we will comment on the differences between these models, the results from the analysis of the teacher data and how cylindrical models improve the analysis of this type of data; data from an interpersonal circumplex.

In terms of interpretability, the CL-PN and Abe-Ley model score best. As shown in the previous section it is relatively easy to interpret the parameters of the linear and the circular component for these models. In the CL-GPN and joint GPN-SSN model the interpretation of the parameters of the circular component it not easy, if at all possible. This is caused by the fact that in addition to the mean vector the covariance matrix of the GPN distribution affects the location of the circular data, making it difficult to compute regression coefficients on the circle. Wang & Gelfand (2013) state that we can use Monte Carlo integration to compute a circular mean and variance for the GPN distribution. This might also be a solution to computing coefficients on the circle and could be applied to the methods of Cremers et al. (2018b) such that they can be used for circular regression coefficients in GPN models as well. HIER OOK NOG EEN OPMERKING OVER HOE DE ASSOCIATION TUSSEN CIRCULAR AND LINEAR COMPONENT WORDT GEMODELLEERD.

In terms of flexibility the joint GPN-SSN scores best. In this model multiple linear and circular outcomes can be included and we can thus apply it to multivariate cylindrical data. In addition for both the linear anc circular components this model and the CL-GPN and CL-PN models are extendable to a mixed-effects model and can thus also be fit to longitudinal data. Extension to mixed-effects, longitudinal (see Nuñez-Antonio & Gutiérrez-Peña (2014) and Hernandez-Stumpfhauser et al. (2016) for hierarchical/mixed-effects models for the PN and GPN distributions respectively). For the Abe-Ley model this may also be possible and has been done for the conditional distribution of its linear component, the Weibull distribution, but has not been done in previous literature for the conditional distribution of its circular component (sine-skewed von Mises)CHRISTOPHE: KLOPT DIT?. In addition, the joint GPN-SSN model allows for non-symmetrical shapes of the distributions of both the linear (sine-skewed normal) and the circular (general projected normal) components. The Abe-Ley model also allows for this and the CL-GPN model allows for a non-symmetrical shape of the circular component.

To investigate model fit for the teacher data we asessed the fit to the linear and circular component separately using the PLSL, a criterion that is designed to assess the out-of-sample preditive performance. The CL-PN and CL-GPN model have the best fit for the linear component and the CL-GPN model has the best fit for the circular component. Differences in fit, in addition to being a result of the different distributions that were used to model the linear and circular component of the data, may also be caused by the way in which the relation between the linear and circular component is modeled. Whereas in both the Abe-Ley and joint GPN-SSN model the distribution of the linear component is conditional on the circular component and vice versa, the distribution of the circular component in the CL-PN and CL-GPN model is independent of the linear component. In these models the linear component are related through a regression structure where the circular component serves as a predictor and the linear component as the outcome. We also investigated the (posterior) predictive distributions of each model and its fit to the training data, the part of the teacher data the model was fit on, and the holdout data, the part of the teacher data that we used to assess the out-of-sample predictive performance. From these plots we saw that whereas the PLSL criteria seem to indicate a substantive difference in fit, the (posterior) predicitve distributions seem to be quite similar (especially for the circular component). This discrepancy might be caused by the fact that the holdout dataset, for which the PLSL criteria were computed, is quite small and thus that small deviations in fit could lead to large deviations in the criterion value.

The four cylindrical models that were modified to the regression context in this paper are not the only cylindrical distributions available from the literature. Several other cylindrical distributions, amongst which are Kato & Shimizu (2008), Sugasawa (2015) and Fernández-Durán (2007) have been introduced (for more references we refer to chapter 2 of Ley & Verdebout (2017)). In the present research we have decided not to include these models for reasons of space, complexity of the models and ease of implementing a regression structure. In further research however it would be interesting to investigate other types of cylindrical distributions as well.

Even though some of the models have downsides regarding interpretation, e.g. the Abe-Ley model does not provide standard errors and parameters from the models with a GPN distribution for the circular component are hard to interpret, cylindrical models in general offer new insights into data of a cylindrical nature in psychology. Concerning the particular example in this paper, the teacher data that were measured using an interpersonal circumplex, we were able to analyze all information in the data simultaneously. Before, the two components from the interpersonal circumplex were analyzed separately. With such an approach we get information about the strength of a teachers' score on the two axes of interpersonal behavior, in our case Agency and Communion. This however does not allow us to distinguish between different types of behavior that are specific combinations of Agency and Communion. A solution to that would be to treat circumplex data as circular (Cremers et al., 2018a). This solution however, does not retain the information about the strength of the specific type of interpersonal behavior a teacher displays. This is possible using the cylindrical models in the present paper, we can simultaneosly model the information about the type (a circular variable) and strength (a linear variable) of interpersonal behavior that a teacher displays and the association between them.

References

Abe, T., & Ley, C. (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics*, *4*, 91–104. doi:10.1016/j.ecosta.2016.04.001

Chrastil, E. R., & Warren, W. H. (2017). Rotational error in path integration: Encoding and execution errors in angle reproduction. *Experimental Brain Research*, *235*(6), 1885–1897. doi:10.1007/s00221-017-4910-y

Claessens, L. C. (2016). *Be on my side i'll be on your side : Teachers' perceptions of teacher–student relationships* (PhD thesis).

Cremers, J., Mainhard, M. T., & Klugkist, I. (2018a). Assessing a bayesian embedding approach to circular regression models. *Methodology*, *14*(2), 69–81.

Cremers, J., Mulder, K. T., & Klugkist, I. (2018b). Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology*, *71*(1), 75–95. doi:10.1111/bmsp.12108

Fernández-Durán, J. (2007). Models for circular–linear and circular–circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, *63*(2), 579–585. doi:10.1111/j.1541-0420.2006.00716.x

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

García-Portugués, E., Barros, A. M., Crujeiras, R. M., González-Manteiga, W., & Pereira, J. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic Environmental Research and Risk Assessment*, *28*(5), 1261–1275. doi:10.1007/s00477-013-0819-6

García-Portugués, E., Crujeiras, R. M., & González-Manteiga, W. (2013). Exploring wind direction and so2 concentration by circular–linear density estimation. *Stochastic Environmental Research and Risk Assessment*, *27*(5), 1055–1067. doi:10.1007/s00477-012-0642-5

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. doi:10.1198/016214506000001437

Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass*, *3*(4), 601–619. doi:10.1111/j.1751-9004.2009.00172.x

Hernandez-Stumpfhauser, D., Breidt, F. J., & Woerd, M. J. van der. (2016). The general projected normal distribution of arbitrary dimension: Modeling and bayesian inference. *Bayesian Analysis*, *12*(1), 113–133. doi:10.1214/15-BA989

Horowitz, L. M., & Strack, S. (2011). *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*. Hoboken, NJ: John Wiley & Sons.

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Vol. 5). World Scientific.

Johnson, R. A., & Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, *73*(363), 602–606.

Kato, S., & Shimizu, K. (2008). Dependent models for observations which include angular ones. *Journal of Statistical Planning and Inference*, *138*(11), 3538–3549. doi:10.1016/j.jspi.2006.12.009

Lagona, F., Picone, M., Maruotti, A., & Cosoli, S. (2015). A hidden markov approach to the analysis of space–time environmental data with linear and circular components. *Stochastic Environmental Research and Risk Assessment*, *29*(2), 397–409. doi:10.1007/s00477-014-0919-y

Ley, C., & Verdebout, T. (2017). *Modern directional statistics*. CRC Press.

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics* (Vol. 494). Chichester, England: Wiley.

Mardia, K. V., & Sutton, T. W. (1978). A model for cylindrical variables with applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, *40*(2), 229–233.

Mastrantonio, G. (2018). The joint projected normal and skew-normal: A distribution for poly-cylindrical data. *Journal of Multivariate Analysis*, *165*, 14–26. doi:10.1016/j.jmva.2017.11.006

Mastrantonio, G., Maruotti, A., & Jona-Lasinio, G. (2015). Bayesian hidden markov modelling using circular-linear general projected normal distribution. *Environmetrics*, *26*(2), 145–158. doi:10.1002/env.2326

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. doi: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*(3), 185–201. doi:10.1177/1471082X1001100301

Olmos, N. M., Varela, H., Gómez, H. W., & Bolfarine, H. (2012). An extension of the half-normal distribution. *Statistical Papers*, *53*(4), 875–886. doi:10.1007/s00362-011-0391-4

Pennings, H. J. M., Brekelmans, M., Sadler, P., Claessens, L. C., Want, A. C. van der, & Tartwijk, J. van. (2017). Interpersonal adaptation in teacher-student interaction. *Learning and Instruction*, *55*, 41–57. doi:10.1016/j.learninstruc.2017.09.005

Rayner, K. (2009). The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics*, *31*(2), 129–150. doi:10.2307/3316064

Sugasawa, S., S. (2015). *A flexible family of distributions on the cylinder*. Retrieved from arXiv: 1501.06332v2

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

Want, A. C. van der. (2015). *Teachers' interpersonal role identity*. (PhD thesis).

Wubbels, T., Brekelmans, M., Brok, P. den, & Tartwijk, J. van. (2006). An interpersonal perspective on classroom management in secondary classrooms in the netherlands. In C. Evertson & C. S.

Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1161–1191). Malwah, NJ: Lawrence Erlbaum Associates.

A Appendix

In this appendix we outline the MCMC procedures to fit the cylindrical regression models from Section 3. R-code for the MCMC sampler and the analysis of the teacher data can be found here: https://github.com/joliencremers/CylindricalComparisonCircumplex.

A.1 Bayesian Model and MCMC procedure for the modified CL-PN model

We use the following algorithm to obtain posterior estimates from the model:

1. Split the data, with the circular outcome $\boldsymbol{\theta} = \theta_i, \ldots, \theta_n$, the linear outcome $\boldsymbol{y} = y_i, \ldots, y_n$ where n is the sample size, and the design matrices $\boldsymbol{Z}^k$ and $\boldsymbol{X}$ for the two components of the circular and the linear outcome respectively in a training (90%) and holdout (10%) set.
2. Define the prior parameters for the training set. In this paper we use:
   - Prior for $\boldsymbol{\gamma}$: $N_q(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, with $\boldsymbol{\mu}_0 = (0,0,0,0)^t$ and $\boldsymbol{\Lambda}_0 = 10^{-4}\boldsymbol{I}_4$.
   - Prior for σ^2: $IG(\alpha_0, \beta_0)$, an inverse-gamma prior with $\alpha_0 = 0.001$ and $\beta_0 = 0.001$.
   - Prior for $\boldsymbol{\beta^k}$: $N_2(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, with $\boldsymbol{\mu}_0 = (0,0)^t$ and $\boldsymbol{\Lambda}_0 = 10^{-4}\boldsymbol{I}_2$ for $k \in I, II$.
3. Set starting values $\boldsymbol{\gamma} = (0,0,0,0)^t$, $\sigma^2 = 1$ and $\boldsymbol{\beta^k} = (0,0)$ for $k \in I, II$. Also set starting values $r_i = 1$ in the training and holdout set.
4. Compute the latent bivariate outcome $\boldsymbol{s}_i = (s_i^I, s_i^{II})^t$ underlying the circular outcome for the holdout and training dataset as follows:

$$\begin{bmatrix} s_i^I \\ s_i^{II} \end{bmatrix} = \begin{bmatrix} r_i \cos\theta_i \\ r_i \sin\theta_i \end{bmatrix}$$

5. Sample $\boldsymbol{\gamma}$, σ^2 and $\boldsymbol{\beta^k}$ for $k \in I, II$ for the training dataset from their conditional posteriors:
   - Posterior for $\boldsymbol{\gamma}$: $N_q(\boldsymbol{\mu}_n, \sigma^2\boldsymbol{\Lambda}_n^{-1})$, with $\boldsymbol{\mu}_n = (\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{\Lambda}_0)^{-1}(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{X}^t\boldsymbol{y})$ and $\boldsymbol{\Lambda}_n = (\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{\Lambda}_0)$.
   - Posterior for σ^2: $IG(\alpha_n, \beta_n)$, an inverse-gamma posterior with $\alpha_n = \alpha_0 + n/2$ and $\beta_n = \beta_0 + \frac{1}{2}(\boldsymbol{y}^t\boldsymbol{y} + \boldsymbol{\mu}_0^t\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{\mu}_n^t\boldsymbol{\Lambda}_n\boldsymbol{\mu}_n)$.
   - Posterior for $\boldsymbol{\beta^k}$: $N_2(\boldsymbol{\mu}_n^k, \boldsymbol{\Lambda}_n^k)$, with $\boldsymbol{\mu}_n^k = ((\boldsymbol{Z}^k)^t\boldsymbol{Z}^k + \boldsymbol{\Lambda}_0^k)^{-1}(\boldsymbol{\Lambda}_0^k\boldsymbol{\mu}_0^k + (\boldsymbol{Z}^k)^t\boldsymbol{s}^k)$ and $\boldsymbol{\Lambda}_n^k = ((\boldsymbol{Z}^k)^t\boldsymbol{Z}^k + \boldsymbol{\Lambda}_0^k)$.
6. Sample new r_i for the training and holdout dataset from the following posterior:

$$f(r_i \mid \theta_i, \boldsymbol{\mu}_i) \propto r_i \exp\left(-\frac{1}{2}(r_i)^2 + b_i r_i\right)$$

   where $b_i = \begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix}^t \boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i = \boldsymbol{z_i}\boldsymbol{B}$ and $\boldsymbol{B} = (\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II})$.

   We can sample from this posterior using a slice sampling technique (Cremers et al., 2018):

- In a slice sampler the joint density for an auxiliary variable v_i with r_i is:

$$p(r_i, v_i \mid \theta_i, \boldsymbol{\mu}_i = \boldsymbol{z}_i\boldsymbol{B}) \propto r_i\mathbf{I}\left(0 < v_i < \exp\left\{-\frac{1}{2}(r_i - b_i)^2\right\}\right)\mathbf{I}(r_i > 0).$$

  The full conditional for v_i, $p(v_i \mid r_i, \boldsymbol{\mu}_i, \theta_i)$, is:

$$U\left(0, \exp\left\{-\frac{1}{2}(r_i - b_i)^2\right\}\right)$$

  and the full conditional for r_i, $p(r_i \mid v_i, \boldsymbol{\mu}_i, \theta_i)$, is proportional to:

$$r_i\mathbf{I}\left(b_i + \max\left\{-b_i, -\sqrt{-2\ln v_i}\right\} < r_i < b_i + \sqrt{-2\ln v_i}\right).$$

  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 $r_{i_2} = b_i + \sqrt{-2\ln v_i}$.

7. Compute the PLSL for the circular and linear outcome on the holdout set using the estimates of $\boldsymbol{\gamma}$, σ^2 and $\boldsymbol{\beta^k}$ for $k \in I, II$ for the training dataset.

8. Repeat steps 4 to 8 until the sampled parameter estimates have converged.

A.2 Bayesian Model and MCMC procedure for the modified CL-GPN mode

We use the following algorithm to obtain posterior estimates from the model:

1. Split the data, with the circular outcome $\boldsymbol{\theta} = \theta_i, \ldots, \theta_n$, the linear outcome $\boldsymbol{y} = y_i, \ldots, y_n$ where n is the sample size, and the design matrices $\boldsymbol{Z}^k$ and $\boldsymbol{X}$ for the two components of the circular and the linear outcome respectively in a training (90%) and holdout (10%) set.

2. Define the prior parameters for the training set. In this paper we use:
   - Prior for $\boldsymbol{\gamma}$: $N_q(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, with $\boldsymbol{\mu}_0 = (0,0,0,0)^t$ and $\boldsymbol{\Lambda}_0 = 10^{-4}\boldsymbol{I}_4$.
   - Prior for σ^2: $IG(\alpha_0, \beta_0)$, an inverse-gamma prior with $\alpha_0 = 0.001$ and $\beta_0 = 0.001$.
   - Prior for $\boldsymbol{\beta}_j$: $N_2(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, with $\boldsymbol{\mu}_0 = (0,0)^t$ and $\boldsymbol{\Sigma}_0 = 10^5\boldsymbol{I}_2$ for $j \in 1, \ldots, p$ where p is the number of covariates in $\boldsymbol{Z}$.
   - Prior for ρ: $N(\mu_0, \sigma^2)$, with $\mu_0 = 0$ and $\sigma^2 = 10^4$.
   - Prior for τ: $IG(\alpha_0, \beta_0)$, an inverse gamma prior with $\alpha_0 = 0.01$ and $\beta_0 = 0.01$.

3. Set starting values $\boldsymbol{\gamma} = (0,0,0,0)^t$, $\sigma^2 = 1$, $\boldsymbol{\beta}_j = (0,0)^t$, $\rho = 0$, $\tau = 1$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \rho^2 & \rho \\ \rho & 1 \end{bmatrix}$.

   Also set starting values $r_i = 1$ in the training and holdout set.

4. Compute the latent bivariate outcome $\boldsymbol{s}_i = (s_i^I, s_i^{II})^t$ underlying the circular outcome for the holdout and training dataset as follows:

$$\begin{bmatrix} s_i^I \\ s_i^{II} \end{bmatrix} = \begin{bmatrix} r_i \cos\theta_i \\ r_i \sin\theta_i \end{bmatrix}$$

5. Sample $\boldsymbol{\gamma}$, σ^2, $\boldsymbol{\beta}_j$, ρ and τ for the training dataset from their conditional posteriors:
   - Posterior for $\boldsymbol{\gamma}$: $N_q(\boldsymbol{\mu}_n, \sigma^2\boldsymbol{\Lambda}_n^{-1})$, with $\boldsymbol{\mu}_n = (\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{\Lambda}_0)^{-1}(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{X}^t\boldsymbol{y})$ and $\boldsymbol{\Lambda}_n = (\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{\Lambda}_0)$.
   - Posterior for σ^2: $IG(\alpha_n, \beta_n)$, an inverse-gamma posterior where $\alpha_n = \alpha_0 + n/2$ and $\beta_n = \beta_0 + 0.5(\boldsymbol{y}^t\boldsymbol{y} + \boldsymbol{\mu}_0^t\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{\mu}_n^t\boldsymbol{\Lambda}_n\boldsymbol{\mu}_n)$.
   - Posterior for $\boldsymbol{\beta}_j$: $N_2(\boldsymbol{\mu}_{j_n}, \boldsymbol{\Sigma}_{j_n})$, with $\boldsymbol{\mu}_{j_n} = \boldsymbol{\Sigma}_{j_n}\boldsymbol{\Sigma}^{-1}\Bigg(-\sum_{i=1}^n z_{i,j-1}\sum_{l\neq j} z_{i,l-1}\boldsymbol{\beta}_l + \sum_{i=1}^n z_{i,j-1}r_i\begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}\Bigg)$ and $\boldsymbol{\Sigma}_{j_n} = \left(\sum_{i=1}^n z_{i,j-1}^2\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0\right)^{-1}$ for $j \in 1, \ldots, p$ where p is the number of covariates in $\boldsymbol{Z}$.
   - Posterior for ρ: $N(\mu_n, \sigma_n^2)$, with $\mu_n = \frac{\tau^{-2}\sum_{i=1}^n (s_i^I - \mu_i^I)(s_i^{II} - \mu_i^{II}) + \mu_0\sigma_0^{-2}}{\tau^{-2}\sum_{i=1}^n (s_i^{II} - \mu^{II})^2 + \sigma_0^{-2}}$ and $\sigma_n^2 = \frac{1}{\tau^{-2}\sum_{i=1}^n (s_i^{II} - \mu^{II})^2 + \sigma_0^{-2}}$ where $\mu_i^I = \boldsymbol{z}_i\boldsymbol{\beta}^I$ and $\mu_i^{II} = \boldsymbol{z}_i\boldsymbol{\beta}^{II}$.
   - Posterior for τ: $IG(\alpha_n, \beta_n)$, an inverse-gamma posterior with $\alpha_n = \frac{n}{2} + \alpha_0$ and $\beta_n = \sum_{i=1}^n (s_i^I - \{\mu_i^{II} + \rho(s_i^{II} - \mu^{II})\})^2 + \beta_0$

6. Sample new r_i for the training and holdout dataset from the following posterior:

$$f(r_i \mid \theta_i, \boldsymbol{\mu}_i) \propto r_i \exp\left\{-0.5A_i\left(r_i - \frac{B_i}{A_i}\right)^2\right\}$$

where $B_i = \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}^t \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i = \boldsymbol{z_i B}$, $\boldsymbol{B} = (\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II})$ and $A_i = \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}^t \boldsymbol{\Sigma}^{-1} \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}$.

We can sample from this posterior using a slice sampling technique (Hernandez-Stumpfhauser et.al. 2018):

- In a slice sampler the joint density for an auxiliary variable v_i with r_i is:

$$p(r_i, v_i \mid \theta_i, \boldsymbol{\mu}_i = \boldsymbol{z}_i\boldsymbol{B}^t) \propto r_i\mathbf{I}\left(0 < v_i < \exp\left\{-.5A_i(r_i - \frac{B_i}{A_i})^2\right\}\right)\mathbf{I}(r_i > 0)$$

- The full conditional for v_i, $p(v_i \mid r_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \theta_i)$, is:

$$U\left(0, \exp\left\{-.5A_i\left(r_i - \frac{B_i}{A_i}\right)^2\right\}\right)$$

  and the full conditional for r_i, $p(r_i \mid v_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \theta_i)$, is proportional to:

$$r_i\mathbf{I}\left(\frac{B_i}{A_i} + \max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln v_i}{A_i}}\right\} < r_i < \frac{B_i}{A_i} + \sqrt{\frac{-2\ln v_i}{A_i}}\right)$$

- 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 $r_{i_1} = \frac{B_i}{A_i} + \max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln v_i}{A_i}}\right\}$ and $r_{i_2} = \frac{B_i}{A_i} + \sqrt{\frac{-2\ln v_i}{A_i}}$.

7. Compute the PLSL for the circular and linear outcome on the holdoutset using the estimates of $\boldsymbol{\gamma}$, σ^2, $\boldsymbol{\beta^k}$, ρ and τ for the training dataset. Use the density $f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the circular outcome.

8. Repeat steps 4 to 8 until the sampled parameter estimates have converged.

A.3 Bayesian Model and MCMC procedure multivariate GPN model

1. Split the data, with the circular outcome $\boldsymbol{\theta} = \theta_i, \ldots, \theta_n$, the linear outcome $\boldsymbol{y} = y_i, \ldots, y_n$ where n is the sample size, and the design matrix $\boldsymbol{X}$ in a training (90%) and holdout (10%) set. Note that in this paper we have only one circular outcome and one linear outcome and the MCMC procedure outlined here is specified for this situation. It can however be generalized to a situation with multiple circular and linear outcomes without too much effort.
2. Define the prior parameters for the training set. Note that in this paper we have only one circular outcome and one linear outcome, so $p = 1$ and $q = 1$. In this paper we use the following priors:
   - Prior for $\boldsymbol{\Sigma}$: $IW(\boldsymbol{\Psi}_0, \nu_0)$, an inverse-Wishart with $\boldsymbol{\Psi}_0 = 10^{-4}\boldsymbol{I}_{2p+q}$ and $\nu_0 = 1$.
   - Prior for $\boldsymbol{\beta}$: $MN(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0 \otimes \boldsymbol{\kappa}_0)$, where $\boldsymbol{\beta}$ is a vectorized $\boldsymbol{B}$, the matrix with regression coefficients, $\boldsymbol{\beta}_0 = \boldsymbol{0}_{k(2p+q)}$, $\boldsymbol{B}_0 = \boldsymbol{0}_{k\times(2p+q)}$ and $\boldsymbol{\kappa}_0 = 10^{-4}\boldsymbol{I}_k$.
   - Prior for λ: $N(\gamma_0, \omega_0)$, with $\gamma_0 = 0$ and $\omega_0 = 10000$.
3. Set starting values $\boldsymbol{\beta} = (0, 0, 0, 0, 0, 0)^t$, $\boldsymbol{\Sigma} = \boldsymbol{I}_3$ and $\lambda = 0$. Also set starting values $r_i = 1$ and $d_i = 1$ in the training and holdout set.
4. Compute the latent bivariate outcome $\boldsymbol{s}_i = (s_i^I, s_i^{II})^t$ underlying the circular outcome for the holdout and training dataset as follows:

$$\begin{bmatrix} s_i^I \\ s_i^{II} \end{bmatrix} = \begin{bmatrix} r_i \cos\theta_i \\ r_i \sin\theta_i \end{bmatrix}$$

5. Compute the latent outcomes $\tilde{y}_i$ underlying the linear outcome for the holdout and training dataset as follows:

$$\tilde{y}_i = \lambda d_i$$

6. Compute $\boldsymbol{\eta}_i$ defined as follows for each individual i:

$$\boldsymbol{\eta}_i = (\boldsymbol{s}_i, y_i)^t - (\boldsymbol{0}_{2p}, \lambda d_i)$$

7. Sample $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and λ for the training dataset from their conditional posteriors:
   - Posterior for $\boldsymbol{B}$: $MN(\boldsymbol{B}_n, \boldsymbol{\kappa}_n, \boldsymbol{\Sigma}_n)$, with $\boldsymbol{B}_n = \boldsymbol{\kappa}_n^{-1}\boldsymbol{X}^t\boldsymbol{\eta} + \boldsymbol{\kappa}_0\boldsymbol{B}_0$ and $\boldsymbol{\kappa}_n = \boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{\kappa}_0$.
   - Posterior for Σ: $IW(\boldsymbol{\Psi}_n, \nu_n)$, an inverse-Wishart with $\boldsymbol{\Psi}_n = \boldsymbol{\Psi}_0 + (\boldsymbol{\eta} - \boldsymbol{X}^t\boldsymbol{B})^t(\boldsymbol{\eta} - \boldsymbol{X}^t\boldsymbol{B}) + (\boldsymbol{B} - \boldsymbol{B}_0)^t\boldsymbol{\kappa}_0(\boldsymbol{B} - \boldsymbol{B}_0)$ and $\nu_n = \nu_0 + n$
   - Posterior for λ: $N(\gamma_n, \omega_n)$, with $\omega_n = \left(\sum_{i=1}^n d_i^2\boldsymbol{\Sigma}_{y|s}^{-1} + \omega_0^{-1}\right)^{-1}$ and $\gamma_n = \omega_n\left(\sum_{i=1}^n d_i\boldsymbol{\Sigma}_{y|s}^{-1}(y_i - \boldsymbol{\mu}_{y_i|s_i}) + \omega_0^{-1}\gamma_0\right)$
8. Sample new d_i for the training and holdout dataset from the following posterior:

$$f(d_i) \propto \phi_q(y_i|\boldsymbol{\mu}_{y_i|s_i} + \lambda d_i, \boldsymbol{\Sigma}_{y|s})\phi_q(d_i|0, 1),$$

where $\boldsymbol{\mu}_{y_i|s_i} = \boldsymbol{x}_i \boldsymbol{B}_{y_i|s_i}$. We can see d_i as a positive regressor with λ as covariate and $\phi_q(d_i|0,1)$ as prior (Mastrantonio, 2018). The full conditional is then a q-dimensional truncated normal with support $\mathbb{R}^+$ as follows:

$$N_q(\boldsymbol{M}_{d_i}, \boldsymbol{V}_q),$$

where $\boldsymbol{V}_q = (\lambda^2 \boldsymbol{\Sigma}_{y|s}^{-1} + 1)$ and $\boldsymbol{M}_{d_i} = \boldsymbol{V}_d \lambda \boldsymbol{\Sigma}_{y|s}^{-1}(y_i - \boldsymbol{\mu}_{y_i,s_i})$.

9. Sample new r_i for the training and holdout dataset from the following posterior:

$$f(r_i \mid \theta_i, \boldsymbol{\mu}_i) \propto r_i \exp\left\{-0.5 A_i \left(r_i - \frac{B_i}{A_i}\right)^2\right\}$$

where $B_i = \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}^t \boldsymbol{\Sigma}_{s_i|y_i}^{-1} \boldsymbol{\mu}_{s_i|y_i}$, $\boldsymbol{\mu}_{s_i|y_i} = \boldsymbol{x}_i \boldsymbol{B}_{s_i|y_i}$ and $A_i = \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}^t \boldsymbol{\Sigma}_{s_i|y_i}^{-1} \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}$. The parameters $\boldsymbol{\mu}_{s_i|y_i}$ and $\boldsymbol{\Sigma}_{s_i|y_i}$ are the conditional mean and covariance matrix of $\boldsymbol{s}_i$ assuming that $(\boldsymbol{s}_i, y_i)^t \sim N_{2p+q}(\boldsymbol{\mu} + (\boldsymbol{0}_{2p}, \lambda d_i)^t, \boldsymbol{\Sigma})$.

Because in this paper $\boldsymbol{\theta}$ originates from a bivariate variable that is known we can simply define the r_i as the euclidean norm of the bivariate datapoints. However, for didactic purposes continue with the explanation of the sampling procedure. We can sample from posterior for r_i using a slice sampling technique (Hernandez-Stumpfhauser et.al. 2018):

- In a slice sampler the joint density for an auxiliary variable v_i with r_i is:

$$p(r_i, v_i \mid \theta_i, \boldsymbol{\mu}_i = \boldsymbol{x}_i \boldsymbol{B}^t) \propto r_i \mathbf{I}\left(0 < v_i < \exp\left\{-.5 A_i (r_i - \frac{B_i}{A_i})^2\right\}\right) \mathbf{I}(r_i > 0)$$

- The full conditionals for v_i, $p(v_i \mid r_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \theta_i)$, is:

$$U\left(0, \exp\left\{-.5 A_i \left(r_i - \frac{B_i}{A_i}\right)^2\right\}\right)$$

  and the full conditional for r_i, $p(r_i \mid v_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \theta_i)$, is proportional to :

$$r_i \mathbf{I}\left(\frac{B_i}{A_i} + \max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln v_i}{A_i}}\right\} < r_i < \frac{B_i}{A_i} + \sqrt{\frac{-2\ln v_i}{A_i}}\right)$$

- 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 $r_{i_1} = \frac{B_i}{A_i} + \max\left\{-\frac{B_i}{A_i}, -\sqrt{\frac{-2\ln v_i}{A_i}}\right\}$ and $r_{i_2} = \frac{B_i}{A_i} + \sqrt{\frac{-2\ln v_i}{A_i}}$.

10. Compute the PLSL for the circular and linear outcome on the holdoutset using the estimates of $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and λ for the training dataset.

11. Repeat steps 4 to 10 until the sampled parameter estimates have converged.

12. In the MCMC sampler we have estimated an unconstrained $\mathbf{\Sigma}$. However, for identification of the model we need to apply the following constraint to both $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_w & \mathbf{0}_{2p \times q} \\ \mathbf{0}^t_{2p \times q} & \boldsymbol{I}_q \end{bmatrix}$$

where $\boldsymbol{C}_w$ is a $2p \times 2p$ diagonal matrix with every $(2(j-1)+k)^{th}$ entry > 0 where $k \in 1, 2$ and $k = 1, \ldots, p$ (Mastrantonio, 2018). The estimates $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ can then be related to their constrained versions, $\tilde{\mathbf{\Sigma}}$ and $\tilde{\boldsymbol{\mu}}$ as follows:

$$\boldsymbol{\mu} = \boldsymbol{C}\tilde{\boldsymbol{\mu}}$$

$$\mathbf{\Sigma} = \boldsymbol{C}\tilde{\mathbf{\Sigma}}\boldsymbol{C}.$$