Working title: Regression models for Cylindrical data.

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

This is the abstract

1 Introduction

Cylindrical data are data that consist of a linear component and a directional component (a set of angles). In this paper, the directional component is circular meaning that it consists of a single angle instead of a set of angles. Figure **??** NOG MAKEN shows the difference between a circular variable (left) and a linear variable (right). The most obviuous 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. 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 psychological research 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 is used for investigating cognitive processes 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 still and 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). 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 from the field of educational psychology. We will refer to this example data as the teacher data. 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. For an overview of methods for circular data see Mardia & Jupp (2000) and Jammalamadaka & Sengupta (2001) and for modern treatments Ley & Verdebout (2017). In the literature 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

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

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 on 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 of the existing methods for the analysis of cylindrical data 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.

2 Teacher data

The motivating example for this article comes from the field of educational psychology and was collected for the studies of Pennings et al. (2017), Claessens (2016) and Want (2015). These studies concern research on classroom climate. 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 that can be used to study student perceptions of their teachers' interpersonal behavior. This instrument 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 1. 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 et al. (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 Equation (1), where A represents a score on the Agency dimension and C represents a score on the Communion dimension. The

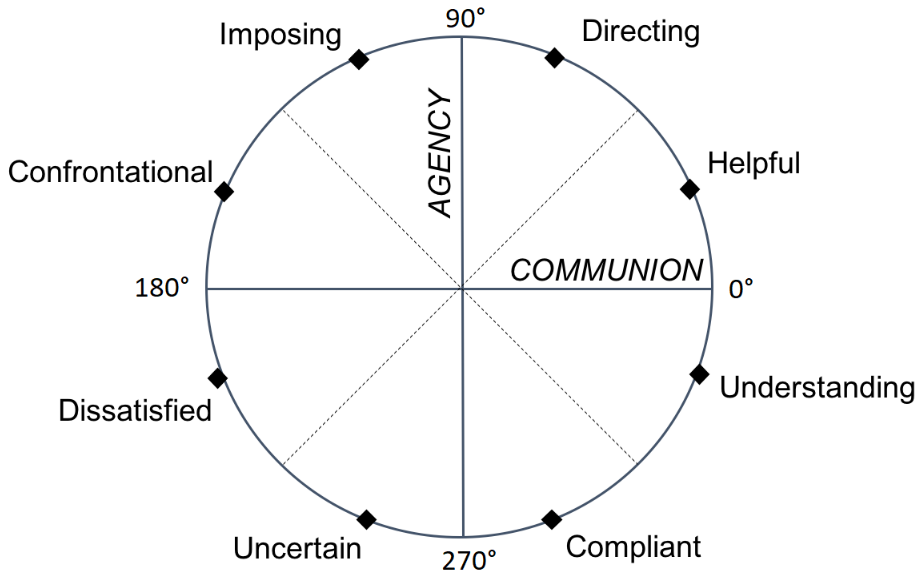

Figure 1: The interpersonal circle for teachers (IPC-T). The words presented in the circumference of the circle are anchor words to describe the type of behavior located in each part of the IPC.

resulting circular variable θ can then be modeled and for example be regressed onto covariates in a circular regression model. However, when two-dimensional data are converted to the circle we lose some information, namely 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 will lead to an improved analysis of data from the IPC. In the next section we will introduce several models that we can use for a more accurate and more informative regression analysis on the teacher data. More information on the teacher data will be given in Section 4.

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

3 Four cylindrical regression models

In this section we will introduce four cylindrical regression models. We will extend the models from Mastrantonio (2018) and Abe & Ley (2017) to contain predictors for the linear and circular outcomes in the cylindrical model. Additionally we introduce two models where both variates Θ and Y are predicted by covariates and that model the relation between the linear and circular outcome as follows (following Wang, Gelfand, & Jona-Lasinio (2015) and 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 the circular variable $\Theta \in [0, 2\pi)$ and linear variable $Y \in (-\infty, +\infty)$ is specified as in Equation 2. The distribution of Y conditional on Θ is then given as:

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

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 $R = \| S \|$, 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 Equation (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 $i = 1, \ldots, n$, where n is the sample size, is generated independently from Equation (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 in this section are estimated using MCMC methods. These methods are 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 density of the model for the pair of the circular variable $\theta \in [0, 2\pi)$ and linear variable y on the positive real half-line $[0, +\infty)$ is:

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

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

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

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. The model 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. As in Mastrantonio (2018) we use the following representation of the linear outcomes:

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

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 $\boldsymbol{\lambda} = \lambda_1, \ldots, \lambda_q$ (the parameters that indicate the skewness of the linear outcomes), $\boldsymbol{D} \sim HN_q(\boldsymbol{0}_q, \boldsymbol{I}_q)$, 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:

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

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 $\mathbf{\Sigma}_{sy}$ contains covariances between the augmented representations of the circular outcome and the linear outcome.

In our regression extension we have $i = 1, \dots, n$ 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 (only the likelihood) into account when assessing model fit. According to Gelman et al. (2014) this is not necessary since we are assessing the fit of a model to data only AFMAKEN!. 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 will analyze the teacher dataset with the help of the four cylindrical models introduced in Section 3. We will however first 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 gathered for the studies of Want (2015), Claessens (2016) and Pennings

Table 1: Descriptives for the teacher dataset

Variable	mean/$\bar{\theta}$	sd/$\hat{\rho}$	Range	Type
IPC	33.23°	0.76	-	Circular
length IPC	0.43	0.15	0.08 - 0.80	Linear
SE	4.72	0.87	2 - 6.5	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.

et al. (2017). The measurements were obtained using the QTI and taken in different years and in different classes. For this paper we only consider one measurement, namely the one for the first measurement occasion (2010) and the one for the largest class if data for multiple classes were available. In addition to a variable `IPC` containing the circular outcome and the 'length' of the IPC, the linear outcome, a teachers' self-efficacy (`SE`) will be used as covariate in the analysis. After the removal of missing cases we end up with a sample of 148 teachers Table 1 shows descriptives for the dataset.

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:

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

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

- For the modified Abe-Ley model:

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

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

- For the modified joint projected and skew normal model:

  $\hat{\boldsymbol{\mu}}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathrm{SE}_i$, 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 $\boldsymbol{\beta}_1 = (\beta_{1_s^I}, \beta_{1_s^{II}}, \beta_{1_y})$.

The variable `SE` is thus used as a predictor for the linear and circular outcomes in all four cylindrical models. We will use the loglikelihoods of the following conditional densities CHRISTOPHE, MAG HIER WEL $\sim$ GEBRUIKT WORDEN?for the computation of the PLSL criterion to evaluate the model fit:

- 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 Equation (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 Equation (6).

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	1.63	1.36	1.87	2.15	1.64	2.59
β_0^I	0.52	0.23	0.80	0.65	0.14	1.06
β_0^{II}	1.00	0.80	1.23	1.24	0.98	1.51
β_1^{II}	0.39	0.16	0.64	0.46	0.18	0.72
γ_0	0.35	0.29	0.41	0.35	0.30	0.40
γ_{cos}	0.05	0.02	0.07	0.04	0.02	0.06
γ_{sin}	-0.01	-0.04	0.03	-0.01	-0.04	0.03
γ_1	0.03	-0.01	0.06	0.03	-0.01	0.06
σ	0.14	0.13	0.16	0.14	0.12	0.16
$\sum_{1,1}$	NA	NA	NA	2.67	1.77	4.12
$\sum_{1,2}$	NA	NA	NA	0.81	0.53	1.07
$\sum_{2,2}$	NA	NA	NA	1.00	1.00	1.00

- 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.

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 WeiSSVM model were the following $\eta_0 = 0.5, \eta_1 = 0.5, \nu_0 = 0.5, \nu_1 = 0.5, \kappa = 0.5, \alpha = 0.5, \lambda = 0$. 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 multivariate GPN 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. There are several parameters that we can compare between these results. Firstly, we have the estimated mean of the circular outcome. In Table 5 we see the posterior (predictive) distributions for the estimated circular means of the CL-PN, CL-GPN and joint GPN-SSN model. For the CL-PN model we can actually also compute this means from the estimates in Table 2 as follows: $\mu_{circ} = atan2(\beta_0^{II}, \beta_0^{I})$.

Table 3: Results for the modified Abe-Ley model

Parameter	ML-estimate
η_0	0.37
η_1	-0.01
ν_0	1.19
ν_1	0.00
α	3.82
κ	1.57
λ	0.66

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
$\beta_{0_s^I}$	0.30	0.26	0.34	1.94	1.63	2.33
$\beta_{0_s^{II}}$	0.19	0.16	0.21	1.20	0.98	1.43
β_{0_y}	0.33	0.30	0.37	0.33	0.30	0.37
$\beta_{1_s^I}$	0.07	0.03	0.12	0.50	0.16	0.79
$\beta_{0_s^{II}}$	0.04	0.01	0.08	0.27	0.09	0.50
β_{1_y}	0.07	0.04	0.12	0.07	0.04	0.12
$\sum_{s_{1,1}}$	0.05	0.04	0.07	2.16	1.51	2.99
$\sum_{s_{2,2}}$	0.02	0.02	0.03	1.00	1.00	1.00
$\sum_{y_{3,3}}$	0.03	0.02	0.05	0.03	0.02	0.05
$\sum_{s_{1,2}}$	0.00	0.00	0.01	0.19	-0.06	0.46
$\sum_{sy_{1,3}}$	0.04	0.03	0.05	0.24	0.18	0.33
$\sum_{sy_{2,3}}$	0.01	0.01	0.02	0.10	0.06	0.13
λ	0.14	0.13	0.17	0.14	0.13	0.17

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

	Mean (degrees)
Cl-PN	31.98
CL-GPN	32.98
GPN-SSN	34.50

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

This leads to an estimate of 31.53°. 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 33.23° in Table 1. The estimate from the Abe-Ley model which is 0.37 radians or 21.20° 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 `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 $\nu_1 = -0.01$. This means that for each unit increate in self-efficacy, at the inflection point of the circular regression line, the score of the teacher on the IPC decreases with $0.01 * (180/\pi) = 0.57°$. For the CL-PN model we can use methods from Cremers et al. (2018b). The estimated posterior mode of b_c, a parameter that is comparable to ν_1 in the Abe-Ley model, equals 1.13 and its 95% HPD interval is $(-26.55; 25.46)$. This means that for each unit increate in self-efficacy, at the inflection point of the circular regression line, the score of the teacher on the IPC increases with $1.13 * (180/\pi) = 74.74°$. Values of b_c and ν_1 are quite different even though they should represent the same effect. However, the slope at the inflection point, which b_c and ν_1 describe is not necessarily representative of the effect of `SE` in the data range. In Figure **??** we see that the inflection point (square) 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 both 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 $0.02(-0.09; 0.15)$ The SAM can be interpreted as the amount of change on the circle 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. For the Abe-Ley model standard errors of the parameters are not known so we cannot formally test whether ν_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 multivariate CL-GPN models we cannot compute circular regression coefficients. Instead, we will compute posterior predictive distributions for the predicted circular outcome of individuals scoring 1 unit above and 1 unit below the average self-efficacy.

Next, we compare the intercepts and regression coefficients for the linear outcome.

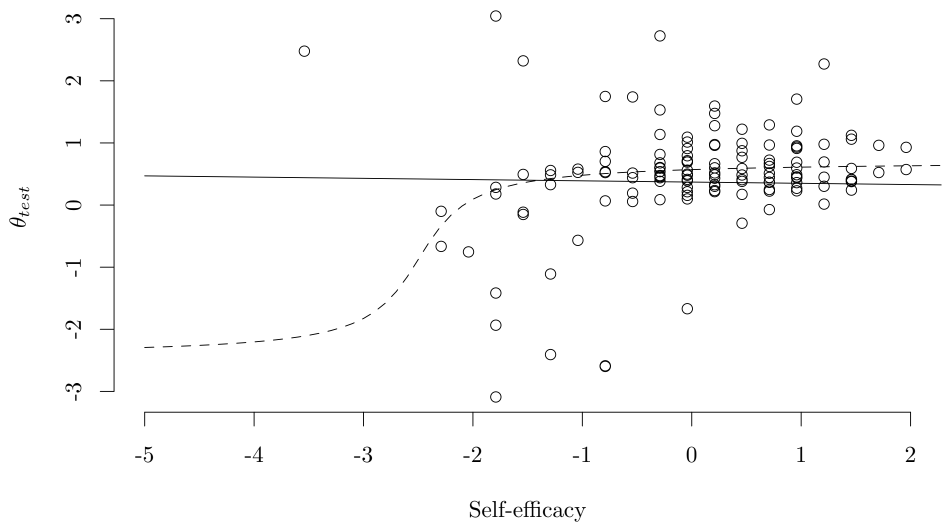

Figure 2: Plot showing circular regresion lines as predicted by the Abe-Ley (solid line) and CL-PN model (dashed line).

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	86.15	52.88	58.53	103.83
linear	-20.17	-20.32	51.83	11.98

- How to interpret weibull regression for Abe-Ley model?
- Comment on different distributions for linear and circular outcome (λ in multivariate CL-GPN), integrate with model fit does this correspond.
- Comment on different estimation methods, in CL-PN and CL-GPN the regression of θ is independent of y but not vice-versa. In the Abe-Ley and Multivariate method dependence between θ and y is modelled both ways.

4.3.1 Model fit

5 Discussion

- Comment on difficulty of interpreting parameters in GPN model.

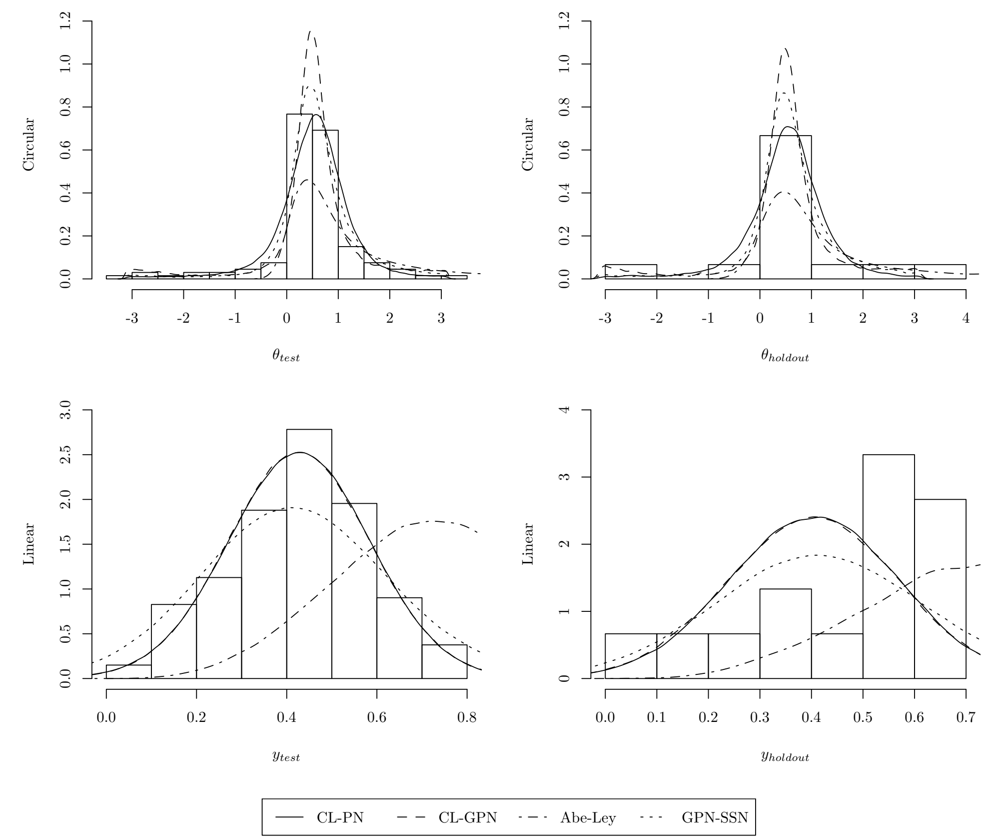

Figure 3: Histograms of $\boldsymbol{y}$ and $\boldsymbol{\theta}$ values of the test and holdout set of the teacher data plotted together with (posterior predictive) density estimates for the modified CL-PN, CL-GPN, Abe-Ley and joint GPN-SSN models.

References

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

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.

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.*

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

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.

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.

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.

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). 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.

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.

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., & Sutton, T. (1978). A model for cylindrical variables with applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, 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.

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

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.

Pennings, H. J., 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*.

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.

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.

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

Wang, F., Gelfand, A. E., & Jona-Lasinio, G. (2015). Joint spatio-temporal analysis of a linear and a directional variable: Space-time modeling of wave heights and wave directions in the adriatic sea. *Statistica Sinica*, 25–39.

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 $\boldsymbol{\beta}^{\boldsymbol{k}}$ 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 $\beta_n = \sum\limits_{i=1}^n (s_i^I - \{\mu_i^{II} + \rho(s_i^{II} - \mu^{II})\})^2 + \beta_0$

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

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

where $B_i = \begin{bmatrix}\cos\theta_i \\ \sin\theta_i\end{bmatrix}^t \boldsymbol{\Sigma}_{s_i|y_i}^{-1}\boldsymbol{\mu}_{s_i|y_i}$, $\boldsymbol{\mu}_{s_i|y_i} = \boldsymbol{x}_i \boldsymbol{B}_{s_i|y_i}$ 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:

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

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

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

  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 $r_{i_2} = \frac{B_i}{A_i} + \sqrt{\frac{-2\ln v_i}{A_i}}$.

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 $\boldsymbol{\Sigma}$. However, for identification of the model we need to apply the following constraint to both $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$:

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

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 $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ can then be related to their constrained versions, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\mu}}$ as follows:

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