Supplementary Material

In this Supplementary Material we give a more detailed description of the cylindrical regression models and outline the MCMC procedures to fit them. R-code for the MCMC sampler and the analysis of the teacher data can be found here: https://github.com/joliencremers/CylindricalComparisonCircumplex. Note that the dimensions of the objects (design matrices, mean vectors, etc.) are those that were used in the analysis of the teacher data where we have 1 circular component, 1 linear component and estimate an intercept and regression coefficient for the covariate self-efficacy. Note that for the regression of the linear component in the CL-PN and CL-GPN models we also have the sine and cosine of the circular component in the regression equation, this makes the vector with regression coefficients, $\boldsymbol{\gamma}$, four-dimensional.

Four cylindrical regression models

The modified CL-PN and modified CL-GPN models

Following Mastrantonio, Maruotti, & Jona-Lasinio (2015) we consider in this section two models where the prediction equation for the linear component is specified as

$$\hat{y}_i = \gamma_0 + \gamma_{cos} * \cos(\theta_i) * r_i + \gamma_{sin} * \sin(\theta_i) * r_i + \gamma_1 * x_1 + \cdots + \gamma_q * x_q, \tag{1}$$

where r_i is a realization of the unobserved the random variable $R \geq 0$ that will be introduced below, $\gamma_0, \gamma_{cos}, \gamma_{sin}, \gamma_1, \ldots, \gamma_q$ are the intercept and regression coefficients and $x_1, \ldots, x_q$ are the q covariates. In both of these models the conditional distribution of Y given $\Theta = \theta$ and $R = r$ is given by

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

where $c = \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix}^t \begin{bmatrix} \gamma_{cos} \\ \gamma_{sin} \end{bmatrix}$, $r \geq 0$. The linear component thus has a normal distribution conditional on Θ and R and contains already linear covariates $x_1, \ldots, x_q$ in its location part.

For the circular component 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 vector $\boldsymbol{S}$ in the plane and the circular component Θ is defined as follows

$$\boldsymbol{S} = \begin{bmatrix} S^I \\ S^{II} \end{bmatrix} = R\boldsymbol{u} = \begin{bmatrix} R\cos(\Theta) \\ R\sin(\Theta) \end{bmatrix}, \tag{2}$$

where $R =|| \boldsymbol{S} ||$, 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 $\boldsymbol{\mu} \in \mathbb{R}^2$, $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \xi^2 & \xi \\ \xi & 1 \end{bmatrix}$, and $\xi, \tau \in (-\infty, +\infty)$ (as in Hernandez-Stumpfhauser, Breidt, & Van der Woerd (2016)). This leads to the circular-linear PN (CL-PN) and circular-linear GPN (CL-GPN) distributions. We will now detail how we modify both cylindrical distributions to also incorporate covariates for the circular part.

The modified CL-PN distribution

Following Nuñez-Antonio, Gutiérrez-Peña, & Escarela (2011), the joint density of Θ and R for the PN distribution equals

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} \exp\{-0.5 \mid\mid \boldsymbol{\mu}^2 \mid\mid\} \exp\{-0.5 \left[r^2 - 2r(\boldsymbol{u}^t\boldsymbol{\mu})\right]\}, \tag{3}$$

where $\boldsymbol{u} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$ and r is defined in (2). In a regression setup the outcome components θ_i, r_i for each individual $i = 1, \ldots, n$, where n is the sample size, are generated independently from the distribution with density (3). The mean vector $\boldsymbol{\mu}_i \in \mathbb{R}^2$ is then defined as $\boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{z}_i$

where the vector $\boldsymbol{z}_i$ is a vector of dimension $p+1$ that contains the covariate values and the value 1 to estimate an intercept and $\boldsymbol{B} = (\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II})$ contains the regression coefficients and intercepts.

The modified CL-GPN distribution

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

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{r}{2\pi\tau} \exp\left[-\frac{(r\boldsymbol{u} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(r\boldsymbol{u} - \boldsymbol{\mu})}{2\tau^2}\right], \tag{4}$$

where we recall that $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \xi^2 & \xi \\ \xi & 1 \end{bmatrix}$. In a regression setup the outcome components θ_i and r_i for each individual are generated independently from (4). The mean vector $\boldsymbol{\mu}_i \in \mathbb{R}^2$ is defined in the same way via covariates as for the modified CL-PN distribution.

Parameter estimation

Both cylindrical models introduced here are estimated using Markov Chain Monte Carlo (MCMC) methods based on Nuñez-Antonio et al. (2011), Wang & Gelfand (2013) and Hernandez-Stumpfhauser et al. (2016) for the regression of the circular component.

The modified Abe-Ley model

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

$$f(\theta, y) = \frac{\alpha\nu^\alpha}{2\pi\cosh(\kappa)}(1 + \lambda\sin(\theta - \mu_c))y^{\alpha-1}\exp[-(\nu y)^\alpha(1 - \tanh(\kappa)\cos(\theta - \mu_c))], \tag{5}$$

where $\alpha > 0$ is a linear shape parameter, $\kappa > 0$ and $\lambda \in [-1, 1]$ are circular concentration and skewness parameters with κ also regulating the circular-linear dependence. Our modification occurs at the level of the linear scale parameter $\nu > 0$ and circular location parameter $\mu_c \in [0, 2\pi)$, both of which we express in terms of covariates: $\nu_i = \exp(\boldsymbol{x}_i^t \boldsymbol{\gamma}) > 0$ and $\mu_{c,i} = \beta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t \boldsymbol{\beta})$. The parameter $\boldsymbol{\gamma}$ is a vector of q regression coefficients $\gamma_j \in (-\infty, +\infty)$ for the prediction of y where $j = 0, \ldots, q$ and ν_0 is the intercept. The parameter $\beta_0 \in [0, 2\pi)$ is the intercept and $\boldsymbol{\beta}$ is a vector of p regression coefficients $\beta_j \in (-\infty, +\infty)$ 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. In a regression setup the outcome component vector $(\theta_i, y_i)^t$ for each individual is generated independently from the modified density (5).

As in Abe & Ley (2017), the conditional distribution of Y given $\Theta = \theta$ is a Weibull distribution with shape α and scale $\nu(1 - \tanh(\kappa)\cos(\theta - \mu_c))^{1/\alpha}$ and the conditional distribution of Θ given $Y = y$ is a sine skewed von Mises distribution with location parameter μ_c and concentration parameter $(\nu y)^\alpha \tanh(\kappa)$. The log-likelihood for this model equals

$$\begin{aligned} l(\alpha, \boldsymbol{\gamma}, \lambda, \kappa, \boldsymbol{\beta}) = {} & n[\ln(\alpha) - \ln(2\pi\cosh(\kappa))] + \alpha \sum_{i=1}^{n} \boldsymbol{x}_i^t \boldsymbol{\gamma} \\ & + \sum_{i=1}^{n} \ln(1 + \lambda \sin(\theta_i - (\beta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t \boldsymbol{\beta})))) + (\alpha - 1) \sum_{i=1}^{n} \ln(y_i) \\ & - \sum_{i=1}^{n} (\exp(\boldsymbol{x}_i^t \boldsymbol{\gamma}) y_i)^\alpha (1 - \tanh(\kappa)\cos(\theta_i - (\beta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t \boldsymbol{\beta})))). \end{aligned}$$

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

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. Both models contain m independent circular components and w

independent linear components. The circular components $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_m)$ are modelled together by a multivariate GPN distribution. The joint distribution of $\boldsymbol{\Theta}$ and $\boldsymbol{R}$ can thus be modeled as the product of (4) for each of the m circular components. The linear components $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_w)$ are modelled 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 (2) and (4)) it is convenient to use a similar tactic for modelling the multivariate skew normal distribution. Following Mastrantonio (2018) the linear components are represented as

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

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

$$f(\boldsymbol{y}, \boldsymbol{d}) = 2^w \phi_w(\boldsymbol{y} \mid \boldsymbol{M}_y + \boldsymbol{\Lambda}\boldsymbol{d}, \boldsymbol{\Sigma}_y)\phi_w(\boldsymbol{d} \mid \boldsymbol{0}_w, \boldsymbol{I}_w),$$

where $\phi_\ell(\cdot|\boldsymbol{M}_\ell, \boldsymbol{\Sigma}_\ell)$ stands for the ℓ-dimensional normal density with mean vector $\boldsymbol{M}_\ell$ and covariance $\boldsymbol{\Sigma}_\ell$. As in Mastrantonio (2018) dependence between the linear and circular component is created by modelling the augmented representations of $\boldsymbol{\Theta}$ and $\boldsymbol{Y}$ together in a $2m + w$ dimensional normal distribution. The joint density of the model is then represented by:

$$f(\boldsymbol{\theta}, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{d}) = 2^w \phi_{2m+w}((\boldsymbol{s}^t, \boldsymbol{y}^t)^t \mid \boldsymbol{M} + (\boldsymbol{0}_{2m}^t, (\text{diag}(\boldsymbol{\lambda})\boldsymbol{d})^t)^t, \boldsymbol{\Sigma})\phi_w(\boldsymbol{d} \mid \boldsymbol{0}_w, \boldsymbol{I}_w) \prod_{j=1}^{m} r_j, \tag{6}$$

where $\boldsymbol{s} = (r_1(\cos(\theta_1), \sin(\theta_1)), \ldots, r_m(\cos(\theta_m), \sin(\theta_m)))^t$, the mean vector $\boldsymbol{M} = (\boldsymbol{M}_s^t, \boldsymbol{M}_y^t)^t$ 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 component and the matrix $\mathbf{\Sigma}_{sy}$ contains covariances between the augmented representations of the circular component and the linear component.

In our regression extension we have $i = 1, \ldots, n$ observations of m circular components, w linear components and g covariates. The mean in the density in (6) then becomes $\boldsymbol{M}_i = \boldsymbol{B}^t \boldsymbol{x}_i$ where $\boldsymbol{B}$ is a $(g+1) \times (2m+w)$ matrix with regression coefficients and intercepts and $\boldsymbol{x}_i$ is a $g+1$ dimensional vector containing the value 1 to estimate an intercept and the g covariate values.

Model fit

We use the following (conditional) loglikelihoods for the computation of the PLSL in the teacher data:

- For the modified CL-PN model:

$$l(y \mid \theta, r) = \log(1) - \log(\sqrt{2\pi\sigma^2}) + \sum(\hat{y}_i - (\gamma_0 + \gamma_{cos}\cos(\theta_i)r_i + \gamma_{sin}\sin(\theta_i)r_i + \gamma_1 \mathrm{SE}_i))^2/2\sigma^2$$

$$l(\theta, r) = \log(1) - \log(2\pi) + \sum -0.5\hat{\boldsymbol{\mu}}_i^2 - 0.5(r_i^2 - 2r_i u_i^t \hat{\boldsymbol{\mu}}_i)$$

where $u_i = (cos\theta_i, \sin\theta_i)$ and $\hat{\boldsymbol{\mu}}_i = (\beta_0^I + \beta_0^I \mathrm{SE}_i, \beta_0^{II} + \beta_0^{II} \mathrm{SE}_i)^t$.

- For the modified CL-GPN model:

$$l(y \mid \theta, r) = \log(1) - \log(\sqrt{2\pi\sigma^2}) + \sum(\hat{y}_i - (\gamma_0 + \gamma_{cos}\cos(\theta_i)r_i + \gamma_{sin}\sin(\theta_i)r_i + \gamma_1 \mathrm{SE}_i))^2/2\sigma^2$$

$$l(\theta, r) = \log(1) - \log(2\pi + \tau) - \sum \log(r_i) + (u_i^t \hat{\boldsymbol{\mu}}_i \Sigma^{-1} (u_i^t \hat{\boldsymbol{\mu}}_i)^t)/2\tau^2$$

where $u_i = (cos\theta_i, \sin\theta_i)$ and $\hat{\boldsymbol{\mu}}_i = (\beta_0^I + \beta_0^I \mathrm{SE}_i, \beta_0^{II} + \beta_0^{II} \mathrm{SE}_i)^t$.

- For the modified Abe-Ley model:

$$l(y \mid \theta) = \log \alpha + \sum \log h_i^\alpha + \sum \log y_i^{\alpha - 1} - \sum (h_i y_i)^\alpha$$

where $h_i = \exp(\hat{y}_i)\{1 - \tanh(\kappa)\cos(\theta_i - \hat{\theta}_i)\}^{1/\alpha}$, $\hat{y}_i = \gamma_0 + \gamma_1 \mathrm{SE}_i$ and $\hat{\theta}_i = \beta_0 + 2\tan^{-1}(\beta_1 \mathrm{SE}_i))$.

$$l(\theta \mid y) = \log(1) - \sum \log 2\pi I_0(c_i) + \sum \log\{1 + \lambda \sin(\theta_i - \hat{\theta}_i\} + \sum c_i \cos(\theta_i - \hat{\theta}_i)$$

where $c_i = y_i^\alpha \exp(\hat{y}_i)^\alpha \tanh \kappa$, and I_0 is a modified Bessel function of order 0.

- For the modified joint projected and skew normal model we take the loglikelihoods of the following distributions:

$y_i \mid \boldsymbol{M}_i, \boldsymbol{\Sigma}, \theta_i, r_i \sim SSN(M_{i_y} + \lambda d_i + \boldsymbol{\Sigma}_{sy}^t \boldsymbol{\Sigma}_s^{-1}(\boldsymbol{s}_i - \boldsymbol{M}_{i_s}), \sigma_y^2 + \boldsymbol{\Sigma}_{sy}^t \boldsymbol{\Sigma}_s^{-1} \boldsymbol{\Sigma}_{sy}),$

$\theta_i \mid \boldsymbol{M}_i, \boldsymbol{\Sigma}, y_i, d_i \sim GPN(\boldsymbol{M}_{i_s} + \boldsymbol{\Sigma}_{sy} \sigma_y^{-2}(y_i - M_{i_y} - \lambda d_i), \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_{sy} \sigma_y^{-2} \boldsymbol{\Sigma}_{sy}^t)$

where SSN is the skew normal distribution. Computationally this comes down to taking the log of the density values for a univariate and multivariate normal distribution (with mean and variance specified as above) for the linear and circular component respectively.

MCMC procedures

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 component $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear component $\boldsymbol{y} = y_1, \ldots, y_n$ where n is the sample size, and the design matrices $\boldsymbol{Z}^k_{n\times 2}$ (for $k \in \{I, II\}$) and $\boldsymbol{X}_{n\times 4}$ of the circular and the linear component 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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, 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 $\boldsymbol{\beta^k} = (0, 0)^t$ for $k \in \{I, II\}$. Also set starting values $r_i = 1$ in the training and holdout set.

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

$$\begin{bmatrix} s^I_i \\ s^{II}_i \end{bmatrix} = \begin{bmatrix} r_i \cos(\theta_i) \\ r_i \sin(\theta_i) \end{bmatrix}.$$

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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_n, \sigma^2\boldsymbol{\Lambda}_n^{-1})$, 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 $\boldsymbol{\beta^k}$: $N_2(\boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n)$, with $\boldsymbol{\mu}_n = ((\boldsymbol{Z}^k)^t\boldsymbol{Z}^k + \boldsymbol{\Lambda}_0)^{-1}(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + (\boldsymbol{Z}^k)^t\boldsymbol{s}^k)$ and $\boldsymbol{\Lambda}_n = ((\boldsymbol{Z}^k)^t\boldsymbol{Z}^k + \boldsymbol{\Lambda}_0)$.

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

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

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

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

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

7. Compute the PLSL for the circular and linear component 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 7 until the sampled parameter estimates have converged. We assess convergence visually using traceplots.

Bayesian Model and MCMC procedure for the modified CL-GPN model

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

1. Split the data, with the circular component $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear component $\boldsymbol{y} = y_1, \ldots, y_n$ where n is the sample size, and the design matrices $\boldsymbol{Z}_{n\times 2}$ and $\boldsymbol{X}_{n\times 4}$ of the circular and the linear component 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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, 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 $j \in \{0, \ldots, p\}$ where p is the number of covariates, 1, in $\boldsymbol{Z}$.

   - Prior for ξ: $N(\mu_0, \sigma^2)$, 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 $\boldsymbol{\gamma} = (0,0,0,0)^t$, $\sigma^2 = 1$, $\boldsymbol{\beta}_j = (0,0)^t$ for $j \in \{0,1\}$, $\xi = 0$, $\tau = 1$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \xi^2 & \xi \\ \xi & 1 \end{bmatrix}$. Also set starting values $r_i = 1$ in the training and holdout set.

4. Compute the latent bivariate scores $\boldsymbol{s}_i = (s_i^I, s_i^{II})^t$ underlying the circular component 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, $\boldsymbol{\beta}_j$ for $j \in \{0,1\}$, ξ and τ for the training dataset from their conditional posteriors:

- Posterior for $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_n, \sigma^2\boldsymbol{\Lambda}_n^{-1})$, 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 $\beta_n = \beta_0 + \frac{1}{2}(\boldsymbol{y}^t\boldsymbol{y} + \boldsymbol{\mu}_0^t\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \boldsymbol{\mu}_n^t\boldsymbol{\Lambda}_n\boldsymbol{\mu}_n)$.

- Posterior for $\boldsymbol{\beta}_j$: $N_2(\boldsymbol{\mu}_{j_n}, \boldsymbol{\Sigma}_{j_n})$, 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 $j \in \{0, \ldots, p\}$ where p is the number of covariates, 1, in $\boldsymbol{Z}$.

- Posterior for ξ: $N(\mu_n, \sigma_n^2)$, with $\mu_n = \frac{\tau^{-2}\sum_{i=1}^n (s_i^I - \mu_i^I)(s_i^{II} - \mu_i^{II}) + \mu_0\sigma_0^{-2}}{\tau^{-2}\sum_{i=1}^n (s_i^{II} - \mu_i^{II})^2 + \sigma_0^{-2}}$ and $\sigma_n^2 = \frac{1}{\tau^{-2}\sum_{i=1}^n (s_i^{II} - \mu_i^{II})^2 + \sigma_0^{-2}}$ where $\mu_i^I = (\boldsymbol{\beta}^I)^t\boldsymbol{z}_i$ and $\mu_i^{II} = (\boldsymbol{\beta}^{II})^t\boldsymbol{z}_i$.

- Posterior for τ: $IG(\alpha_n, \beta_n)$, an inverse gamma posterior with $\alpha_n = \frac{n}{2} + \alpha_0$ and $\beta_n = \sum_{i=1}^n (s_i^I - \{\mu_i^I + \xi(s_i^{II} - \mu_i^{II})\})^2 + \beta_0$

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

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

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

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

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

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

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

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

8. Repeat steps 4 to 7 until the sampled parameter estimates have converged. We visually assess convergence using traceplots.

Bayesian Model and MCMC procedure for the modified GPN-SSN model

1. Split the data, with the circular component $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear component $\boldsymbol{y} = y_1, \ldots, y_n$ where n is the sample size, and the design matrix $\boldsymbol{X}_{n\times 2}$ in a training (90%) and holdout (10%) set. Note that in this paper we have only one circular component and one linear component and the MCMC procedure outlined here is specified for this situation. It can however be generalized to a situation with multiple circular and linear components without too much effort.

2. Define the prior parameters for the training set. Since we have only one circular component, one linear component and one covariate, we have $m = 1$, $w = 1$ and $g = 1$. In this paper we use the following priors:

   - Prior for $\boldsymbol{\Sigma}$: $IW(\boldsymbol{\Psi}_0, \nu_0)$, an inverse Wishart with $\boldsymbol{\Psi}_0 = 10^{-4}\boldsymbol{I}_{2m+w}$ and $\nu_0 = 1$.
   - Prior for $\boldsymbol{B}$ in vectorized form: $N_{(g+1)(2m+w)}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma} \otimes \boldsymbol{\kappa}_0)$, where $\otimes$ stands for the Kronecker product, $\boldsymbol{\beta}_0 = \text{vec}(\boldsymbol{B}_0)$, the matrix with prior values for the regression coefficients. We choose $\boldsymbol{\beta}_0 = \boldsymbol{0}_{(g+1)(2m+w)}$, $\boldsymbol{B}_0 = \boldsymbol{0}_{(g+1)\times(2m+w)}$ and $\boldsymbol{\kappa}_0 = 10^{-4}\boldsymbol{I}_{g+1}$.
   - 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 scores $\boldsymbol{s}_i = (s_i^I, s_i^{II})^t$ underlying the circular component 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 scores $\tilde{y}_i$ underlying the linear component 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:

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

7. Sample $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and λ for the training dataset from their conditional posteriors:

   - Posterior for $\boldsymbol{\Sigma}$: $IW(\boldsymbol{\Psi}_n, \nu_n)$, 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 $\nu_n = \nu_0 + n$ where n is the sample size.
   - Posterior for $\boldsymbol{B}$ in matrix form: $MN(\boldsymbol{B}_n, \boldsymbol{\kappa}_n, \boldsymbol{\Sigma})$, 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 λ: $N(\gamma_n, \omega_n)$, with $\omega_n = \Big(\sum_{i=1}^n d_i^2\sigma_{y|s}^{-2} + \omega_0^{-1}\Big)^{-1}$ and $\gamma_n = \omega_n\Big(\sum_{i=1}^n d_i\sigma_{y|s}^{-2}(y_i - \mu_{y_i|s_i}) + \omega_0^{-1}\gamma_0\Big)$ where $\mu_{y_i|s_i} = \mu_y + \boldsymbol{\Sigma}_{sy}^t\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{s_i} - \boldsymbol{\mu_s})$ and $\sigma_{y|s}^2 = \sigma_y^2 - \boldsymbol{\Sigma}_{sy}^t\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{sy}$.

8. Sample new d_i for the training and holdout dataset from the following posterior:

$$f(d_i) \propto \phi(y_i|\mu_{y_i|s_i} + \lambda d_i, \sigma_{y|s}^2)\phi(d_i|0,1),$$

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

$$N(m_{d_i}, v),$$

where $v = \Big(\lambda^2\sigma_{y|s}^{-2} + 1\Big)$ and $m_{d_i} = v\lambda\sigma_{y|s}^{-2}\Big(y_i - \mu_{y_i|s_i}\Big)$.

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 $B_i = \begin{bmatrix}\cos(\theta_i)\\ \sin(\theta_i)\end{bmatrix}^t \boldsymbol{\Sigma}_{s_i|y_i}^{-1}\boldsymbol{\mu}_{s_i|y_i}$, $\boldsymbol{\mu}_{s_i|y_i} = \boldsymbol{B}_{s_i|y_i}^t\boldsymbol{x}_i$ 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 $(\boldsymbol{s}_i^t, y_i)^t \sim N_{2m+w}(\boldsymbol{\mu} + (\mathbf{0}_{2m}^t, \lambda d_i)^t, \boldsymbol{\Sigma})$. Because in this paper $\boldsymbol{\theta}$ originates from a bivariate variable that is known we can in this model (where the variance-covariance matrix of the circular component is not constrained in the estimation procedure) simply define the r_i as the Euclidean norm of the bivariate datapoints. However, for didactic purposes we continue with the explanation of the sampling procedure. We can sample from the 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

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

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

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

  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 component on the holdout set 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 constraints to both $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. Therefore we need the matrix

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_s & \boldsymbol{0}_{2m\times w} \\ \boldsymbol{0}^t_{2m\times w} & \boldsymbol{I}_w \end{bmatrix}$$

where $\boldsymbol{C}_s$ is a $2m \times 2m$ diagonal matrix with every $(2(j-1)+k)^{th}$ entry > 0 where $k \in \{1,2\}$ and $j = 1, \ldots, m$ (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:

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

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

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

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

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., & 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

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

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