Supplementary Material

In this Supplementary Material we outline the MCMC procedures to fit the cylindrical regression models. 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 outcome, 1 linear outcome 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 outcome in the regression equation, this makes the vector with regression coefficients, $\boldsymbol{\gamma}$, four-dimensional.

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 $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear outcome $\boldsymbol{y} = y_1, \ldots, y_n$ where n is the sample size, and the design matrices $\boldsymbol{Z}_{n \times 2}^k$ (for $k \in \{I, II\}$) and $\boldsymbol{X}_{n \times 4}$ 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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, 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 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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_n, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$, 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 outcome on the holdout set using the estimates of $\boldsymbol{\gamma}$, σ^2 and $\boldsymbol{\beta^k}$ for $k \in \{I, II\}$ for the training dataset.

8. Repeat steps 4 to 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 outcome $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear outcome $\boldsymbol{y} = y_1, \ldots, y_n$ 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 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 $\boldsymbol{\gamma}$: $N_4(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$, 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 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 $\boldsymbol{\gamma} = (0, 0, 0, 0)^t$, $\sigma^2 = 1$, $\boldsymbol{\beta}_j = (0, 0)^t$ for $j \in \{0, 1\}$, $\rho = 0$, $\tau = 1$ 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, $\boldsymbol{\beta}_j$ for $j \in \{0, 1\}$, ρ 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 $\boldsymbol{\mu}_{j_n} = \boldsymbol{\Sigma}_{j_n} \boldsymbol{\Sigma}^{-1}\left(- \sum_{i=1}^n z_{i,j-1} \sum_{l \neq j} z_{i,l-1} \boldsymbol{\beta}_l + \right.$

$$\left. \sum_{i=1}^n z_{i,j-1} r_i \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix} \right) \text{ and } \boldsymbol{\Sigma}_{j_n} = \left(\sum_{i=1}^n z_{i,j-1}^2 \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0 \right)^{-1} \text{ for } j \in \{0, \ldots, p\}$$

where p is the number of covariates, 1, in $\boldsymbol{Z}$.

- Posterior for ρ: $N(\mu_n, \sigma_n^2)$, with $\mu_n = \frac{\tau^{-2} \sum_{i=1}^n (s_i^I - \mu_i^I)(s_i^{II} - \mu_i^{II}) + \mu_0 \sigma_0^{-2}}{\tau^{-2} \sum_{i=1}^n (s_i^{II} - \mu_i^{II})^2 + \sigma_0^{-2}}$ 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 $\beta_n = \sum_{i=1}^n (s_i^I - \{\mu_i^I + \rho(s_i^{II} - \mu_i^{II})\})^2 + \beta_0$

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 outcome on the holdout set using the estimates of $\boldsymbol{\gamma}$, σ^2, $\boldsymbol{\beta^k}$ for $k \in \{I, II\}$, ρ 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 outcome $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and the linear outcome $\boldsymbol{y} = y_1, \ldots, y_n$ 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 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. Since we have only one circular outcome, one linear outcome 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 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:

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

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 $\omega_n = \left(\sum_{i=1}^n d_i^2 \sigma_{y|s}^{-2} + \omega_0^{-1} \right)^{-1}$ and $\gamma_n = \omega_n \left(\sum_{i=1}^n d_i \sigma_{y|s}^{-2} (y_i - \mu_{y_i|s_i}) + \omega_0^{-1} \gamma_0 \right)$ 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 $v = \left(\lambda^2 \sigma_{y|s}^{-2} + 1 \right)$ and $m_{d_i} = v \lambda \sigma_{y|s}^{-2} \left(y_i - \mu_{y_i|s_i} \right).$

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

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

where $B_i = \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix}^t \boldsymbol{\Sigma}_{s_i \mid y_i}^{-1} \boldsymbol{\mu}_{s_i \mid y_i}$, $\boldsymbol{\mu}_{s_i \mid y_i} = \boldsymbol{B}_{s_i \mid y_i}^t \boldsymbol{x}_i$ and $A_i = \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix}^t \boldsymbol{\Sigma}_{s_i \mid y_i}^{-1} \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix}$.

The parameters $\boldsymbol{\mu}_{s_i \mid y_i}$ and $\boldsymbol{\Sigma}_{s_i \mid y_i}$ are the conditional mean and covariance matrix of $\boldsymbol{s}_i$ assuming that $(\boldsymbol{s}_i^t, y_i)^t \sim N_{2m+w}(\boldsymbol{\mu} + (\boldsymbol{0}_{2m}^t, \lambda d_i)^t, \boldsymbol{\Sigma})$. 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 outcome 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 outcome 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

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

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:

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

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