Original Article

Assessing a Bayesian Embedding Approach to Circular Regression Models

Jolien Cremers,[1] Tim Mainhard,[2] and Irene Klugkist[1,3]

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

Abstract: Circular data is different from linear data and its analysis also requires methods different from conventional methods. In this study a Bayesian embedding approach to estimating circular regression models is investigated, by means of simulation studies, in terms of performance, efficiency, and flexibility. A new Markov chain Monte Carlo (MCMC) sampling method is proposed and contrasted to an existing method. An empirical example of a regression model predicting teachers' scores on the interpersonal circumplex will be used throughout. Performance and efficiency are better for the newly proposed sampler and reasonable to good in most situations. Furthermore, the method in general is deemed very flexible. Additional research should be done that provides an overview of what circular data looks like in practice, investigates the interpretation of the circular effects and examines how we might conduct a way of hypothesis testing or model checking for the embedding approach.

Keywords: circular data, Bayesian methods, regression, interpersonal circumplex

Circular data is different from linear data in the sense that it contains information that can be converted into angles. One may come across circular data in many fields of research. Examples of circular variables include orientations of rock formations, migratory patterns of birds, eye movement patterns, and clock times. In this paper we will use data collected for the educational research of Mainhard, Brekelmans, Brok, and Wubbels (2011) as an illustration of circular data. It includes the scores of teachers from 48 classes on the interpersonal circumplex as assessed by their students in the first week of the school year. The interpersonal circumplex consists of the underlying dimensions Agency and Communion and is a measure used in personality research. Agency summarizes the aspects of status, power, dominance, and control and Communion summarizes the aspects of solidarity, friendliness, warmth, and love (Horowitz & Strack, 2011).

To consider scores on the interpersonal circumplex as circular is different from the usual treatment of such data. Before, a circumplex would be divided into several octants, for example, whether a teacher scored between 0° and 45°, 45° and 90°, and so forth. These octants or the two dimensions Agency and Communion were then analyzed separately. However, the circumplex is "[...] an interpersonal space in which the set of variables are organized theoretically as a circle – as a continuous order with no beginning or end." (Gurtman, 2009, p. 2). Wright, Pincus, Conroy, and Hilsenroth (2009) outline how different questions can be answered when we treat circumplex data as circular. They argue that the circumplex structure: "[...] is amenable to circular statistical techniques that answer more precise questions than whether groups differ on the individual scales that comprise the instruments" (p. 311). Figure 1 is a graphical representation of the interpersonal circumplex showing the two axes Agency and Communion and a unit circle on which the scores of four teachers are plotted. It shows how data from circumplex measurement instruments can be considered circular data.

Circular data requires different analysis methods. This can be illustrated by considering the linear and circular mean of the scores of four teachers on the interpersonal circumplex. The linear mean is computed by:

$$(351.40^\circ + 14.78^\circ + 133.15^\circ + 104.69^\circ)/4 = 151.01^\circ.$$

However, as can be seen from Figure 1, 151.01° is not the average direction of the scores. The correct circular mean

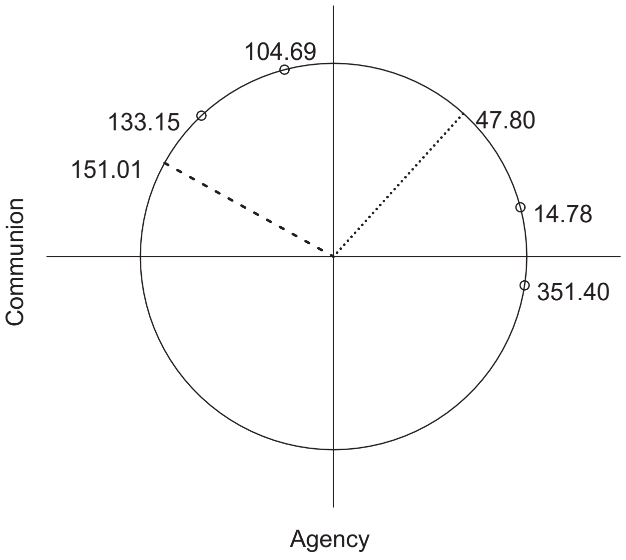

Figure 1. Interpersonal circumplex with axes Agency (vertical) and Communion (horizontal). Data points indicate the score of a teacher on the circumplex measured in degrees. The dashed line represents the direction of the linear mean and the dotted line represents the direction of the circular mean.

considers the directional nature of the data and is in this specific case computed by:

$$\tan^{-1}\left(\frac{\sin(351.40^\circ)+\sin(14.78^\circ)+\sin(133.15^\circ)+\sin(104.69^\circ)}{\cos(351.40^\circ)+\cos(14.78^\circ)+\cos(133.15^\circ)+\cos(104.69^\circ)}\right)$$
$$= 47.80^\circ$$

(Fisher, 1995).

In this paper, we will consider a Bayesian regression model for circular data. In the example data, fitting a regression model would imply predicting the score of a teacher on the interpersonal circumplex by one or more linear or circular predictors. Only a few Bayesian methods for estimating parametric circular regression models are available in the literature. Gill and Hangartner (2010), Lagona (2016), and Mulder and Klugkist (2017) provide Markov chain Monte Carlo (MCMC) methods for circular regression based on the von Mises distribution, which is directly defined on the circle. In the literature, methods using distributions directly defined on the circle are referred to as having an "intrinsic" approach. Other approaches are the "wrapping" and the "embedding" approach which make use of wrapped distributions and projected distributions, respectively. Ravindran and Ghosh (2011) provide an MCMC method for wrapped distributions. Nuñez-Antonio, Gutiérrez-Peña, and Escarela (2011) and Wang and Gelfand (2013) provide models for a circular response based on the projected normal (PN) and general projected normal (GPN) distribution. The difference between the PN and GPN distribution lies in the specification of their variance-covariance matrix. In this paper, we will focus on a regression model for PN data, since it follows the convention within models for linear data to assume the data is normally (unimodally and symmetrically) distributed and since interpreting parameters obtained from the GPN model is rather complicated. Whereas for the wrapping approach simulation studies were done by Ravindran and Ghosh (2011), no extensive simulations have been done to date for a regression model for PN data.

Therefore, we will investigate the Bayesian embedding approach as presented in Nuñez-Antonio et al. (2011) in terms of performance, efficiency, and flexibility by means of simulation studies. The approach will be investigated using two different MCMC sampling methods. These methods are a Gibbs sampler with a Metropolis-Hastings (MH) step for one of the parameters, as proposed by Nuñez-Antonio et al. (2011), and a Gibbs sampler with a slice sampling step for one of the parameters (see Electronic Supplementary Material, ESM 2 for further information). Both slice sampling and MH sampling are ways to sample from a function. In a MH sampler, we sample a candidate from an envelope function that is proportional to the function we wish to sample from. With a certain probability this candidate is then either accepted or rejected. In slice sampling we only draw samples from uniform distributions. It is based on the notion that by sampling uniformly from a region under the density function we can sample from the distribution itself.

In The Model section, the model for the Bayesian embedding approach to circular regression is introduced. The Simulations for One Predictor Models and Simulations for Multiple Predictor Models sections contain the methodology and results of the simulation studies for one and multiple predictors. These are followed by the Estimation for the Agency-Communion Data section which presents the interpretation and results from the model fit on the Agency-Communion data. The paper ends with a discussion of results.

The Model

In this section, the model as used in this paper is introduced together with a regression model to be fit on the Agency-Communion data. This is followed by an explanation of the Bayesian estimation.

The Embedding Approach

In the embedding approach we assume that the circular outcome variable Θ has a projected bivariate normal distribution $PN(\theta|\boldsymbol{\mu}, \boldsymbol{I})$, where $\boldsymbol{\mu} \in \mathbb{R}^2$ is a mean vector and $\boldsymbol{I}$ is a variance-covariance matrix equal to the identity matrix. To be able to understand how projecting a distribution

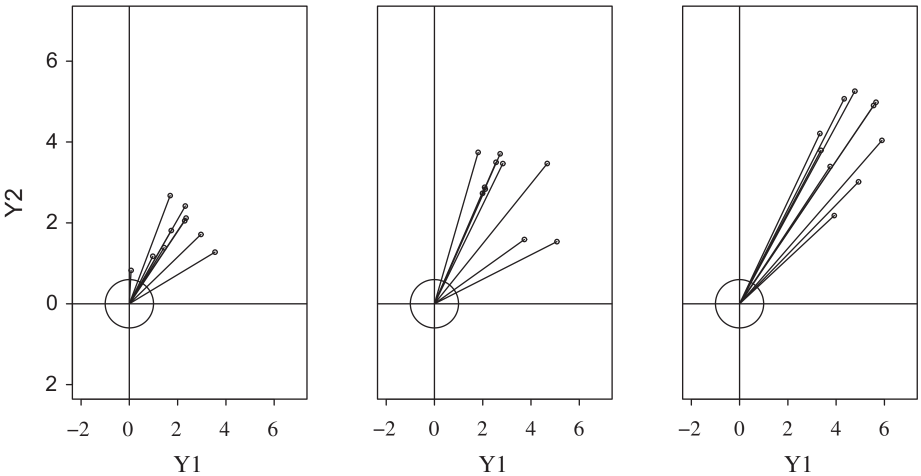

Figure 2. Three sets of bivariate normal data projected on the unit circle. From left to right mean vectors, **μ**, are: (1.5, 1.5), (3, 3), (4.5, 4.5). The lengths of the lines drawn from each datapoint to the origin represent ***r*** and their intersection with the circle corresponds to individual values of the circular outcome vector, **θ**.

works, imagine that we have one bivariate normal outcome variable, $Y \sim N_2(\boldsymbol{\mu}, \boldsymbol{I})$. Figure 2 shows the projection of this bivariate normal outcome variable on the unit circle. The three plots show datapoints from three bivariate normal distributions with different mean vectors. Lines are drawn from each datapoint to the origin (0, 0). We refer to the length of these lines as $\boldsymbol{r} = (r_1, \ldots, r_N)$ where $i = (1, \ldots, N)$ and N is the sample size. Their intersections with the unit circle can be interpreted as datapoints of the circular outcome vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. In Figure 2 we observe that the θ_i are more concentrated for PN distributions with means further from the origin.

Projecting bivariate normal data on a circle is relatively easy and produces a circular outcome vector **θ** and a vector with distances to the origin ***r***. However, when we start with circular data the process is reversed. Imagine one circular outcome variable measured in angles, **θ**. We can decompose these angles into their sine and cosine components. The distance of the decomposition of one angle $(\cos\theta_i, \sin\theta_i)$ to the origin is always 1 on a unit circle. However, the datapoints from the underlying bivariate normal outcome can theoretically be located at any distance from the origin and we cannot immediately obtain them because we have not observed ***r***. The method that treats ***r*** as a latent variable and makes inference possible is introduced in the Bayesian Estimation section.

Circular Regression

In regression models, the projected bivariate normal distribution has the following density (Nuñez-Antonio et al., 2011):

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

where $0 < \theta \leq 2\pi$ and **μ** is the mean vector of the underlying bivariate normal distribution with identity variance-covariance matrix ***I***. Furthermore, ***u*** is the vector $(\cos\theta, \sin\theta)^t$, $\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}$, where $\boldsymbol{B} = [\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II}]$, ***x*** is a matrix with predictor variables and $\boldsymbol{x}_{,1}$ is a vector of 1's to be able to estimate an intercept. The two components of ***B***, $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$ are vectors with regression coefficients and an intercept. Formally this notation is only correct when the predictors in ***x*** are equal for both components of **μ**. The structure is then like a multivariate regression model. The dimensions of $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$ are however allowed to differ. In practice that means that we have two matrices ***x***, one for each component of **μ**. Lastly, $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution.

A Regression Model for the Agency-Communion Data

Using the embedding approach, we consider the two dimensions, of the Agency-Communion data jointly in a circular regression model. As outcome variable we choose the score of the teachers on the interpersonal circumplex as assessed by their students in the first week of the school year (ACS). Regression equations for the two components of this outcome (Agency and Communion) are specified using the teachers' self-assessed score on the interpersonal circumplex (ACSSP), the variable teacher experience (TEX) and an extraversion measure (EV). Since ACSSP is a circular variable we split it up into its sine and cosine components and use these as two separate predictors in the equation (Fisher, 1995). Summary statistics for ACS and the predictor variables TEX, EV and the two components of ACSSP are shown in Table 1.

The resulting regression equations are:

$$\mu^I = \beta_0^I + \beta_1^I \cos(\text{ACSSP}) + \beta_2^I \sin(\text{ACSSP}) + \beta_3^I \text{TEX}$$

Table 1. Descriptives for the Agency-Communion data with linear mean and standard deviation (*SD*) for continuous variables and mean direction ($\bar{\theta}$) and mean resultant length ($\bar{R}$), a measure of precision, for circular variables

	$M/(\bar{\theta})$	$SD/(\bar{R})$	Minimum	Maximum	Type
ACS	79.82°	0.66	–	–	Circular
TEX	0.00	8.70	−9.80	18.69	Continuous
EV	0.00	1.16	−2.90	1.94	Continuous
sin(ACSSP)	0.34	0.62	−0.93	1.00	Continuous
cos(ACSSP)	−0.21	0.69	−0.99	0.99	Continuous

Note. TEX and EV were already centered in the original data.

$$\mu^{II} = \beta_0^{II} + \beta_1^{II}\cos(\text{ACSSP}) + \beta_2^{II}\sin(\text{ACSSP}) + \beta_3^{II}\text{TEX} + \beta_4^{II}\text{EV}$$

where μ^I and μ^{II} are the predicted values for the Communion and Agency axis. Note that in this case the two components have a meaningful interpretation because they originate from a circumplex model. In other cases, for example, clock times translated to the circle, these components might not have a meaningful interpretation on their own. The results from the analysis are reported in the Estimation for the Agency-Communion data section.

Bayesian Estimation

In our paper the modeling and notation is like that of Nuñez-Antonio et al. (2011). In Bayesian analyses, prior distributions must be specified for all model parameters. In the circular regression model, a normal prior is specified for the two components of the matrix $\boldsymbol{B}$:

$$N\left(\boldsymbol{\beta}^j \middle| \boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j\right) \forall j = I, II \tag{2}$$

where $\boldsymbol{\beta}_0^j$ are prior values for the regression coefficients and intercept and $\boldsymbol{\Lambda}_0^j$ is the prior precision matrix of component *j*. In this paper an uninformative prior was selected in which the values in $\boldsymbol{\beta}_0^j$ equal 0 and the values on the diagonal of $\boldsymbol{\Lambda}_0^j$ equal 10^{-5}. Combining this prior with a bivariate normal likelihood for $\boldsymbol{y}_i, \boldsymbol{y}_i \sim N_2(\cdot|\boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i, \boldsymbol{I})$ we obtain the following posterior:

$$f\left(\boldsymbol{\beta}^j \middle| \boldsymbol{D}\right) = N\left(\cdot \,|\, \boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j\right) \forall j = I, II \tag{3}$$

where $\boldsymbol{D} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ is a sample of independent bivariate normal observations, $\boldsymbol{\mu}_F^j = \left(\boldsymbol{\Lambda}_F^j\right)^{-1}\left(\boldsymbol{\Lambda}_0^j\boldsymbol{\beta}_0^j + \left(\boldsymbol{X}^j\right)^t\boldsymbol{y}^j\right)$, $\boldsymbol{\Lambda}_F^j = \boldsymbol{\Lambda}_0^j + \left(\boldsymbol{X}^j\right)^t\boldsymbol{X}^j$ and $\boldsymbol{X}^j$ is a design matrix. To model the underlying bivariate normal data we need to sample values for the latent vector $\boldsymbol{r}$, defined on $(0, \infty)$. Its joint distribution with $\boldsymbol{\Theta}$, the observed circular outcome, is defined by:

$$f(\theta, r|\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}) = 2\pi^{-1}\exp\{-0.5||\boldsymbol{\mu}||^2\} \times \exp\{-0.5\left[r^2 - 2r(\boldsymbol{u}^t\boldsymbol{\mu})\right]\}\,|J|, \tag{4}$$

where $|J| = r$ is the Jacobian of the transformation $\boldsymbol{y} \mapsto (\theta, r)$, $\boldsymbol{y}$ is bivariate normal, and $\boldsymbol{u} = (\cos\theta, \sin\theta)^t$. The sampler that is used to obtain estimates for $\boldsymbol{\beta}^j$ and $\boldsymbol{r}$ was developed by Nuñez-Antonio et al. (2011). It contains the following steps:

1. Starting values for the r_i in $\boldsymbol{r}$ are chosen. In this paper they are set to 1.
2. The two components of $\boldsymbol{B}$ are sampled from their conditional posterior

$$f\left(\boldsymbol{\beta}^j \middle| \theta_1, \ldots, \theta_N, \boldsymbol{r}\right) = N\left(\cdot \,|\, \boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j\right) \forall j = I, II \tag{5}$$

3. Using the estimates for the two components of $\boldsymbol{B}$ new r_i are generated from

$$f(r_i|\theta_i, \boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i) \propto r_i \exp\left(-0.5r_i^2 + b_i r_i\right) \tag{6}$$

where $0 < r_i < \infty$ and $b_i = \boldsymbol{u}_i^t\boldsymbol{\mu}_i$. This is done, either in a MH step (MCMC method 1), as in Nuñez-Antonio et al. (2011), or using a slice sampler (Neal, 2003) presented by Hernandez-Stumpfhauser, Breidt, and van der Woerd (2017) and adapted for the regression situation (MCMC method 2). The two methods for sampling r_i are outlined in ESM 2. R code can be found in ESM 3 and ESM 4.

4. Steps 2 and 3 are repeated for a specified amount of iterations.

Simulations for One Predictor Models

This section outlines the methodology and results of simulation studies for models with one predictor. For each design 500 simulated datasets were analyzed. This amount

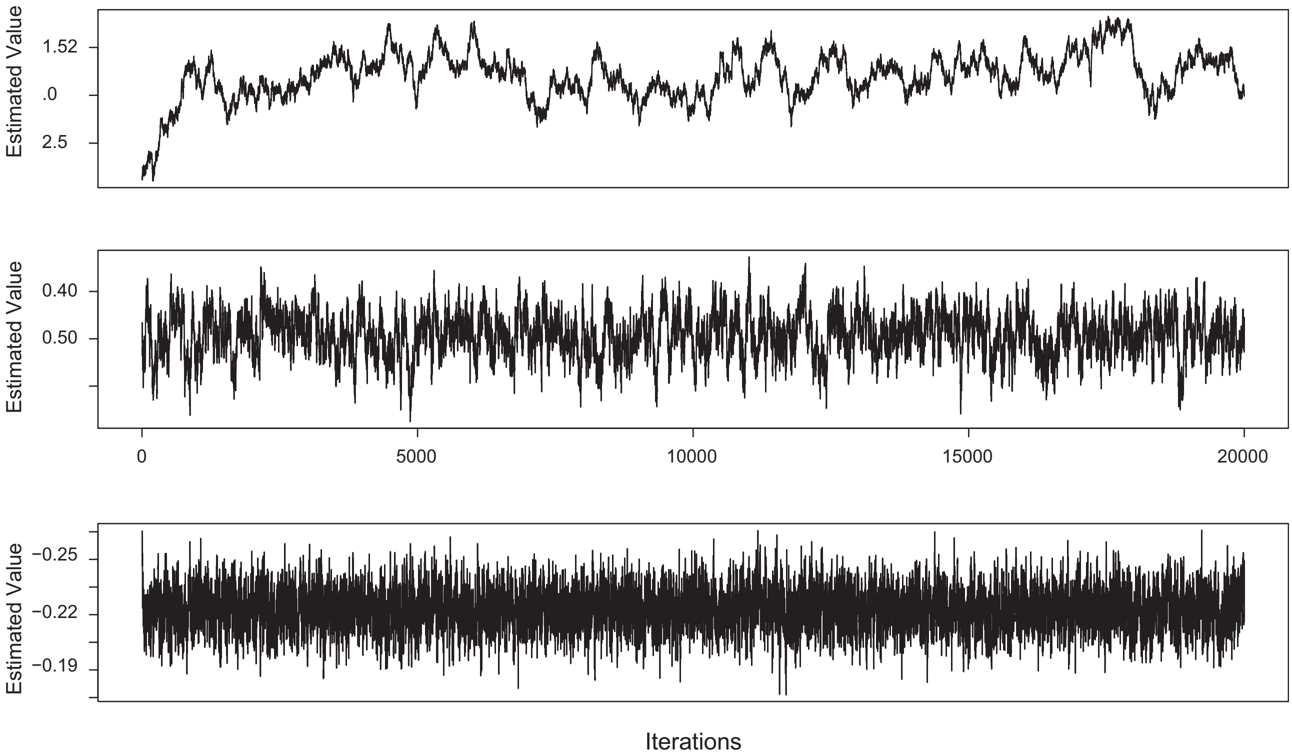

Figure 3. Example traceplots of the estimate for the regression coefficient β_2^I in the simulation studies showing no, acceptable, and good convergence.

was deemed sufficient for proper estimation of (relative) bias and coverage. All computations were performed using the programming language R version 3.1.1 (R Core Team, 2014).

Convergence of parameter estimates in separate designs (populations) was checked for both sampling methods. First, a simulation was run for 3,000 iterations of which 750 were considered burn-in. Then convergence was checked by looking at traceplots, as shown in Figure 3, for a couple of simulations of each design. When in doubt about whether the chain converged, running mean (ergodicity) plots were assessed as well. If we concluded that the chain did not converge we ran the simulation again but now with an increased number of iterations and checked convergence again. In total there were four rounds of increasing number of iterations (3,000, 20,000, 40,000, and 80,000) with initial burn-in of 750 iterations and additional burn in of 10,000, 30,000, and 70,000 for the last three rounds.

Reported results are (relative) bias, coverage, and mean computation time (MCT). MCT is defined as the time in seconds it takes to estimate the parameters for one simulated dataset, averaged over all simulated datasets. Bias is the deviation of the posterior mean of the parameters from their population value (*Population − Estimated*) averaged over the simulated datasets. Relative bias is the absolute bias divided by the absolute population value (|*Population − Estimated* |/| *Population*|). For parameters with population value zero we do not report the relative bias but the absolute bias (|*Population − Estimated*|). Coverage of the credible interval is the percentage of simulated datasets in which the population value lies within the 95% credible interval of the posterior distribution.

One Linear Predictor

Design

The circular outcome vector ($\boldsymbol{\theta}$) was generated by sampling N bivariate normal outcomes $\boldsymbol{y}_i \sim N_2(\boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i, \boldsymbol{I})$ and subsequently projecting them on the circle. The projection of a bivariate normal vector $\boldsymbol{y}_i$ with values y^I and y^{II} to obtain one circular value θ is shown in (7). The predictor vector, $\boldsymbol{x}_{,2}$, was sampled from a normal distribution $N(\mu_x, \sigma_x)$ with mean μ_x and standard deviation σ_x. This vector is sampled for each of the 500 datasets of a design.

Table 2. (Relative) Bias and Coverage of the intercepts and coefficients for various simulation designs for the study with one linear predictor using a MH step for sampling the r_i

Population values				(Relative) Bias				Coverage				MCT
β_1^I	β_1^{II}	μ_x	N	β_0^I	β_1^I	β_0^{II}	β_1^{II}	β_0^I	β_1^I	β_0^{II}	β_1^{II}	
			10	0.02	0.46	0.01	0.44	88.2	87.4	91.6	87.0	4.61
		0	50	0.00	0.04	0.01	0.06	95.8	92.6	94.8	93.2	17.76
			100	0.00	0.04	0.00	0.04	95.4	95.6	93.4	93.8	39.34*
			10	0.15	0.32	0.14	0.36	88.4	88.0	88.8	87.4	43.57**
0.5	0.5	−4	50	0.01	0.06	0.02	0.06	92.6	92.2	91.8	90.0	20.68**
			100	0.02	0.04	0.01	0.04	94.8	94.6	93.6	94.0	39.45**
			10	2.35	0.06	2.27	0.04	94.0	93.4	95.0	95.0	87.10
		10	50	0.74	0.08	0.77	0.08	94.4	94.6	94.4	95.0	345.67
			100	0.45	0.06	0.47	0.06	95.2	94.8	95.4	94.4	664.70
			10	0.02	0.65	0.01	0.55	88.0	84.0	92.8	87.4	5.43
		0	50	0.00	0.15	0.01	0.12	95.0	92.6	95.0	91.0	20.67
			100	0.00	0.00	0.00	0.02	94.6	94.2	93.6	94.6	39.37
			10	0.17	0.49	0.10	0.42	88.2	87.2	88.6	89.2	5.46
−0.2	−0.2	−4	50	0.07	0.14	0.07	0.14	93.8	92.4	92.0	91.0	20.72
			100	0.00	0.03	0.01	0.05	94.8	95.0	93.6	93.8	39.53
			10	0.83	0.05	1.03	0.10	90.0	89.6	89.2	90.0	43.70*
		10	50	0.01	0.30	0.01	0.30	91.8	92.4	93.6	93.6	173.27*
			100	0.10	0.15	0.07	0.15	94.6	94.0	93.2	94.2	330.70*
			10	0.02	0.43	0.02	0.42	90.0	82.2	88.6	81.6	43.33*
		0	50	0.00	0.06	0.01	0.06	93.4	92.2	94.4	91.2	171.56*
			100	0.00	0.04	0.02	0.04	93.2	93.2	94.6	91.8	328.36*
			10	2.53	0.09	2.44	0.10	92.0	93.0	92.6	92.6	79.49**
2.0	2.0	−4	50	0.00	0.03	0.00	0.03	91.8	92.4	93.8	92.8	331.29**
			100	0.39	0.02	0.40	0.02	95.2	95.8	95.0	94.0	600.91**
			10	Nonconvergence > 90% of datasets*								
		10	50	Nonconvergence > 90% of datasets*								
			100	Nonconvergence > 90% of datasets*								

Notes. "*" or "**" indicate that the amount of iterations was set to 20,000 and 40,000, respectively.

Parameters that were varied are: the sample size (N), μ_x, and the slopes β_1^I and β_1^{II} in the regression equation predicting the two components of the outcome. Tables 2 and 3 show the chosen values. Each row in these tables represents one design. Relatively low N were chosen to reflect common sizes in the social sciences. The lowest, $N = 10$, was chosen to investigate what happens in extremely small samples. The parameters μ_x, β_1^I, and β_1^{II} were varied to be able to investigate the influence of the variance of the circular outcome on the performance of the sampler. The mean of the linear predictor and the slopes both affect the mean vector of the bivariate normal outcome. This vector influences the variance of the projected circular outcome (see Figure 2). Positive and negative values of different sizes were chosen for μ_x and the slopes such that the circular outcome would lie on both sides of the circle. To keep the number of simulation designs manageable, the values for the intercepts β_0^I and β_0^{II} and σ_x were not varied and set to 0 and 1, respectively.

Results

Results are shown in Table 2 for the sampler with MH step and Table 3 for the slice sampler. The (relative) bias was rounded to two decimals. In Table 2 we see that a part of the estimates for the intercepts and the regression coefficients are biased. Results show that N, μ_x, β_1^I, and β_1^{II} influence the (relative) bias. Firstly, it is highest for smaller N. When changing from $N = 10$ to $N = 50$, the (relative) bias decreases a lot while it mostly does not do so when changing from 50 to 100. The (relative) bias also decreases with linear outcome means, $\mu_x(\beta_0^I, \beta_0^{II})$, closer to zero. In terms of coverage, we see that it does not reach the desired 95% level in all designs. Coverages lie between 81.6% and 95.8% and in general the coverage is best for larger N.

Table 3. (Relative) Bias and Coverage of the intercepts and coefficients for various simulation designs for the study with one linear predictor using a slice sampler for sampling the r_i

Population values				(Relative) Bias				Coverage				MCT
β_1^I	β_1^{II}	μ_x	N	β_0^I	β_1^I	β_0^{II}	β_1^{II}	β_0^I	β_1^I	β_0^{II}	β_1^{II}	
			10	0.02	0.44	0.00	0.41	89.2	88.4	92.2	87.4	2.03
		0	50	0.00	0.04	0.01	0.04	95.8	92.2	95.4	94.0	5.45
			100	0.00	0.04	0.00	0.04	96.0	96.0	94.2	94.8	9.80*
			10	0.18	0.30	0.09	0.36	88.2	88.8	89.8	88.2	15.90***
0.5	0.5	−4	50	0.01	0.06	0.01	0.06	93.0	92.6	92.8	90.8	5.60***
			100	0.03	0.04	0.02	0.04	95.4	94.8	95.0	94.4	10.68***
			10	2.10	0.00	2.04	0.00	95.0	94.8	95.6	95.4	60.60
		10	50	0.75	0.08	0.79	0.10	94.2	94.2	94.8	95.0	166.01
			100	0.44	0.06	0.46	0.06	94.6	94.6	95.2	94.4	297.56
			10	0.02	0.65	0.01	0.55	87.6	84.4	91.4	87.0	2.10
		0	50	0.00	0.10	0.01	0.10	95.0	92.8	95.2	91.6	5.49
			100	0.00	0.00	0.00	0.00	95.0	94.2	94.4	94.6	9.97
			10	0.17	0.50	0.10	0.40	88.6	88.0	89.0	89.6	2.16
−0.2	−0.2	−4	50	0.07	0.15	0.06	0.15	93.2	93.0	92.4	91.6	6.03
			100	0.00	0.00	0.01	0.05	94.4	95.2	94.0	94.4	10.89
			10	0.82	0.00	1.00	0.10	89.6	90.2	88.6	89.0	15.92*
		10	50	0.01	0.05	0.02	0.05	93.2	93.6	93.4	94.6	43.38*
			100	0.08	0.00	0.05	0.00	95.2	95.2	93.6	93.8	77.78*
			10	0.02	0.42	0.03	0.42	90.2	82.6	89.4	82.6	16.73*
		0	50	0.01	0.06	0.01	0.06	94.0	93.2	94.2	91.6	45.08*
			100	0.00	0.04	0.00	0.04	94.2	94.0	94.4	93.8	79.92*
			10	2.63	0.08	2.55	0.09	91.8	93.0	92.8	91.6	61.49***
2.0	2.0	−4	50	0.75	0.03	0.75	0.04	94.4	92.2	93.4	93.6	164.04***
			100	0.39	0.02	0.40	0.02	95.4	94.8	94.8	93.2	277.82***
			10	Nonconvergence > 90% of datasets***								
		10	50	Nonconvergence > 90% of datasets***								
			100	Nonconvergence > 90% of datasets***								

Notes. "*", "**", or "***" indicate that the amount of iterations was set to 20,000, 40,000, or 80,000, respectively.

We do see that, especially in designs with μ_x further from zero, the relation between N and coverage is less strong.

$$
\begin{aligned}
\theta = \operatorname{atan2}(y^{\mathrm{II}}, y^{\mathrm{I}}) &= \arctan\left(\frac{y^{\mathrm{II}}}{y^{\mathrm{I}}}\right) && \text{if } y^{\mathrm{I}} > 0 \\
&= \arctan\left(\frac{y^{\mathrm{II}}}{y^{\mathrm{I}}}\right) + \pi && \text{if } y^{\mathrm{I}} < 0 \;\; y^{\mathrm{II}} \geq 0 \\
&= \arctan\left(\frac{y^{\mathrm{II}}}{y^{\mathrm{I}}}\right) - \pi && \text{if } y^{\mathrm{I}} < 0 \;\; y^{\mathrm{II}} < 0 \\
&= \frac{\pi}{2} && \text{if } y^{\mathrm{I}} = 0 \;\; y^{\mathrm{II}} > 0 \\
&= -\frac{\pi}{2} && \text{if } y^{\mathrm{I}} = 0 \;\; y^{\mathrm{II}} < 0 \\
&= \text{undefined} && \text{if } y^{\mathrm{I}} = 0 \;\; y^{\mathrm{II}} = 0
\end{aligned} \tag{7}
$$

In Table 3 we see that the previous patterns also occur in the results for the slice sampler. Performance in terms of (relative) bias and coverage is comparable between the slice and MH-sampler. Although the slice sampler needed more iterations to converge in some designs, the computation time was faster in all cases ranging from a half or a fourth of the time needed for the sampler with MH-step. As can be seen in the tables, the mean computation time increases proportional to the sample size and the amount of iterations.

One Circular Predictor

Design

The circular outcome vector ($\boldsymbol{\theta}$) was generated by sampling N bivariate normal outcomes $\boldsymbol{y}_i \sim N_2(\boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i, \boldsymbol{I})$ and subsequently projecting these bivariate outcomes on the circle by using (7). The two vectors $\boldsymbol{x}_{,2}$ and $\boldsymbol{x}_{,3}$ are the cosine and sine components of a vector sampled from $VM(\mu_{\mathrm{circ},x}, \kappa_x)$; a von Mises distribution with circular mean $\mu_{\mathrm{circ},x}$, and concentration parameter κ_x. This vector is sampled for each of the 500 datasets of a design. The parameter, κ_x, is analogous to a precision, the larger it is the more homogenous the data (Fisher, 1995). The parameters that were varied

Table 4. (relative) Bias and Coverage of the intercepts and coefficients for various simulation designs for the study with one circular predictor using a MH step for sampling the r_i

Population values						(Relative) Bias						Coverage						MCT
β_1^I	β_1^I	β_1^{II}	β_2^{II}	κ_x	N	β_0^I	β_1^I	β_1^I	β_0^{II}	β_1^{II}	β_2^{II}	β_0^I	β_1^I	β_1^I	β_0^{II}	β_1^{II}	β_2^{II}	
					10	0.06	0.92	0.78	0.08	0.93	0.66	83.2	80.6	83.0	86.0	81.4	83.0	42.83*
				1	50	0.02	0.11	0.07	0.01	0.06	0.07	92.8	94.4	90.0	94.2	93.0	93.6	21.19
					100	0.01	0.03	0.02	0.00	0.02	0.04	93.0	92.8	95.6	93.8	94.4	95.8	40.36
					10	0.19	3.76	1.12	0.11	1.33	0.93	84.6	72.6	83.0	83.8	77.2	82.2	87.83**
0.5	0.5	0.5	0.5	2	50	0.01	0.19	0.10	0.01	0.18	0.09	92.0	91.8	92.2	94.4	92.2	93.4	21.12
					100	0.01	0.07	0.03	0.00	0.09	0.03	96.2	92.2	94.0	93.0	91.4	93.4	40.32
					10	0.03	3.62	0.47	0.05	2.05	0.92	79.6	71.4	80.4	82.0	74.0	78.2	43.69*
				10	50	0.00	0.17	0.14	0.01	0.40	0.18	90.8	91.4	93.4	91.2	90.0	92.0	21.17
					100	0.01	0.46	0.07	0.00	0.34	0.04	93.8	92.2	92.6	94.8	94.0	92.4	39.98
					10	0.04	0.81	0.68	0.03	1.16	0.60	82.8	76.2	83.6	83.0	77.8	84.0	42.89*
				1	50	0.01	0.04	0.05	0.00	0.17	0.08	95.4	93.4	91.0	93.6	92.4	93.8	21.22
					100	0.01	0.07	0.08	0.01	0.09	0.02	93.2	94.4	94.2	93.2	94.4	96.6	40.35
					10	0.02	2.25	0.16	0.05	1.96	1.19	85.2	73.4	82.4	84.8	78.6	81.8	83.74**
−0.2	−0.2	−0.2	−0.2	2	50	0.00	0.07	0.08	0.01	0.15	0.13	92.0	90.8	91.8	94.2	91.4	92.6	21.17
					100	0.00	0.13	0.08	0.00	0.03	0.10	95.4	93.0	93.6	93.0	91.6	92.8	40.11
					10	0.00	2.48	1.58	0.01	2.60	1.15	81.2	71.2	80.4	79.0	72.6	80.6	43.74*
				10	50	0.00	0.35	0.11	0.01	0.40	0.04	91.6	91.4	93.6	93.0	90.0	91.4	21.40
					100	0.00	0.59	0.09	0.00	0.14	0.17	94.4	91.2	93.6	93.8	92.8	93.4	40.71
					10	0.13	0.67	0.69	0.17	0.69	0.67	84.0	80.2	76.6	84.4	78.2	77.6	40.57*
				1	50	0.03	0.10	0.09	0.02	0.09	0.09	94.2	91.4	90.2	93.8	90.0	89.4	155.89*
					100	0.02	0.04	0.04	0.01	0.04	0.05	95.2	92.6	94.6	94.6	92.8	92.8	301.36*
					10	0.23	1.11	0.77	0.18	0.82	0.80	86.4	79.2	81.2	83.8	78.8	77.4	90.12**
2.0	2.0	2.0	2.0	2	50	0.02	0.10	0.08	0.01	0.10	0.08	91.4	91.2	92.4	95.8	92.0	92.0	176.99*
					100	0.02	0.06	0.04	0.01	0.07	0.04	94.4	93.0	95.4	93.2	93.2	92.0	338.77*
					10	0.03	1.39	0.82	0.02	0.81	1.06	80.0	76.4	79.4	82.2	80.4	79.2	80.22**
				10	50	0.01	0.14	0.12	0.00	0.13	0.13	91.6	91.8	91.0	92.8	93.6	90.4	149.20*
					100	0.01	0.08	0.06	0.00	0.10	0.06	94.0	94.8	94.4	94.8	95.0	92.8	289.04*

Note. "*" or "**" indicate that the amount of iterations was set to 20,000 and 40,000, respectively.

are: the sample size, N, the slopes of the cosine β_1^I and sine β_2^Icomponent of the circular predictor in the regression equation predicting the first component of the outcome, the slopes β_1^{II} and β_2^{II}in the regression equation predicting the second component of the outcome, and κ_x. Tables 4 and 5 show the chosen values. Chosen N, β_0^I, and β_0^{II} are the same as in the One Linear Predictor section and $\mu_{\text{circ},x} = 0$ to keep the amount of designs manageable. The concentration parameter of the circular predictor was varied to be able to investigate the influence of the variance of the circular outcome on the performance of the sampler. One of these was chosen to be quite high ($\kappa_x = 10$) such that we could investigate the effect of an extremely concentrated predictor, and thus also circular outcome.

Results

Results are shown in Table 4 for the sampler with MH step and Table 5 for the slice sampler. The (relative) bias was rounded to two decimals. In Table 4 we observe that part of the estimates for the intercepts and the regression coefficients is biased. Results show that N and κ_x influence the (relative) bias. Firstly, it is highest for smaller N. When changing from $N = 10$ to $N = 50$, the (relative) bias decreases a lot while it mostly does not do so when changing from 50 to 100. Furthermore, the (relative) bias decreases with lower κ_x. The relative bias of β_1^I and β_1^{II} is bigger than the relative bias of β_2^I and β_2^{II} in designs with higher κ_x and lower N. In terms of coverage we see that it is generally closer to 95% for designs with larger samples and lower κ_x. The MCT for this study shows values like those from the study with one linear predictor, it varies with N and number of iterations and it is lower for the slice sampler.

In Table 5 we see that the patterns in performance are similar for both samplers. The (relative) bias and coverage is comparable between the slice and MH-sampler for most designs. However, in the designs with relatively

Table 5. (Relative) Bias and Coverage of the intercepts and coefficients for various simulation designs for the study with one circular predictor using a slice sampler for sampling the r_i

Population values						(Relative) Bias						Coverage						MCT
β_1^I	β_1^I	β_1^{II}	β_2^{II}	κx	N	β_0^I	β_1^I	β_1^I	β_0^{II}	β_1^{II}	β_2^{II}	β_0^I	β_1^I	β_1^I	β_0^{II}	β_1^{II}	β_2^{II}	
					10	0.03	0.52	0.70	0.11	0.50	0.62	77.0	78.0	81.6	80.2	82.2	83.0	16.68*
				1	50	0.03	0.10	0.06	0.00	0.04	0.06	94.6	94.2	91.8	94.4	93.4	93.4	43.89*
					100	0.01	0.02	0.01	0.01	0.00	0.03	93.4	94.2	94.6	93.2	94.8	94.4	77.76*
					10	0.33	1.18	0.84	0.46	0.60	0.72	70.0	71.8	80.4	75.4	75.6	79.6	70.14***
0.5	0.5	0.5	0.5	2	50	0.03	0.10	0.10	0.01	0.06	0.08	93.8	93.4	93.0	91.8	91.4	92.8	6.53
					100	0.01	0.02	0.02	0.01	0.04	0.02	93.2	92.2	94.2	93.4	91.6	92.8	11.28
					10	5.20	10.06	0.48	6.70	13.16	1.00	72.2	72.4	79.4	72.8	73.2	80.2	17.56*
				10	50	0.33	0.64	0.16	0.25	0.44	0.18	90.4	91.2	92.4	89.6	90.4	90.6	6.42
					100	0.06	0.12	0.06	0.07	0.12	0.02	90.4	90.2	94.6	93.8	94.2	93.4	11.24
					10	0.01	0.40	0.60	0.04	0.70	0.55	76.6	76.4	83.6	77.4	79.2	84.8	18.50*
				1	50	0.02	0.00	0.05	0.00	0.15	0.05	95.8	92.8	91.4	93.0	91.8	94.4	5.72
					100	0.00	0.05	0.05	0.01	0.10	0.00	94.0	95.6	94.0	94.2	94.8	96.4	10.14
					10	0.93	4.75	0.30	0.43	1.75	1.10	72.8	72.2	82.6	76.0	75.2	80.8	70.04***
−0.2	−0.2	−0.2	−0.2	2	50	0.01	0.05	0.05	0.01	0.05	0.10	93.2	91.2	92.0	90.8	91.8	93.8	6.51
					100	0.00	0.05	0.10	0.01	0.05	0.10	93.6	93.8	94.0	94.2	92.6	92.8	11.58
					10	3.41	17.1	1.85	1.84	8.85	1.35	70.8	71.8	81.0	75.6	76.0	80.4	17.70*
				10	50	0.27	1.35	0.15	0.27	1.40	0.05	90.8	90.8	93.4	89.6	90.0	91.8	6.53
					100	0.21	1.05	0.05	0.13	0.70	0.10	90.8	90.6	93.0	94.6	94.4	92.6	10.57
					10	0.10	0.63	0.72	0.13	0.59	0.69	81.2	76.8	76.2	88.2	78.0	77.2	17.50*
				1	50	0.05	0.09	0.09	0.03	0.08	0.09	94.6	92.0	91.0	96.0	93.8	91.8	45.39*
					100	0.02	0.04	0.04	0.00	0.04	0.05	94.0	92.2	94.4	94.4	94.2	93.6	10.08
					10	1.18	0.04	0.81	1.78	0.36	0.78	76.4	74.0	73.0	79.0	78.2	73.8	61.55***
2.0	2.0	2.0	2.0	2	50	0.04	0.09	0.10	0.03	0.09	0.10	92.2	93.8	92.6	94.0	93.2	92.4	41.00*
					100	0.01	0.04	0.05	0.01	0.05	0.05	94.0	94.2	91.6	93.4	92.8	91.6	72.97*
					10	19.63	9.41	0.65	20.5	9.86	0.87	73.4	74.2	80.8	75.8	75.6	84.2	61.41***
				10	50	1.56	0.74	0.19	1.48	0.68	0.19	89.8	90.4	90.2	89.4	89.8	92.8	45.37*
					100	0.39	0.17	0.09	0.34	0.14	0.07	92.4	92.8	96.0	94.4	94.8	94.8	81.07*

Notes. "*", "**", or "***" indicate that the amount of iterations was set to 20,000, 40,000, or 80,000, respectively.

worse performance, small N and large κ_x, the MH-sampler outperforms the slice sampler. Patterns for MCT are like those described in the Results section.

Conclusions

Results from the two simulation studies showed that there is (relative) bias in some parameter estimates. It is of comparable size both samplers. For the regression coefficients it is lowest and in most designs its size is acceptable. There is however under coverage in designs with small N. In theory however, and as shown in our simulations, as the sample size goes to infinity the (frequentist) coverage of the posterior will reach 95% in line with Bayesian central limit theory (Gelman et al., 2014, p. 92). We have seen that (relative) bias increases and coverage is worse in designs with high κ_x and μ_x further away from zero. Since these parameters influence the concentration of the outcome variable this means that in data with a highly concentrated outcome the estimates produced by the embedding approach and their credible intervals may deviate from the truth. If we compute $\bar{R}$ for the designs using formulae from Fisher (1995) and Kendall (1974) we see that concentration starts affecting (relative) bias and coverage from $\bar{R} = 0.95$ onward. This is logical considering that when the spread of the circular outcome is smaller the effects on the circle must also be smaller. These smaller effects are harder to estimate and thus result in lower performance of the MCMC samplers.

Simulations for Multiple Predictor Models

Three simulation studies with multiple predictors were conducted: a study with two linear predictors, a study with one

Table 6. Mean and the lower and upper bounds for the 95% credible interval (CI) of the posterior distributions of the intercept and coefficients for the Agency-Communion data

Parameter	Posterior mean	Lower bound CI	Upper bound CI
Communion			
Intercept	0.39	−0.01	0.79
Communion Self Perception	0.54	0.07	1.02
Agency Self Perception	−0.13	−0.68	0.43
Experience	−0.02	−0.06	0.02
Agency			
Intercept	1.67	1.15	2.22
Communion Self Perception	0.29	−0.31	0.90
Agency Self Perception	0.50	−0.20	1.19
Experience	0.10	0.04	0.15
Extraversion	0.34	−0.01	0.71

circular and linear predictor, and a study with different regression equations for the two components of the outcome. Methods used for the simulations and convergence checks are like those for the previous studies.

No notable difference in performance from models with one predictor was detected (interested readers may contact the authors for results tables). This is useful to know as the models investigated here show more resemblance to the models estimated to answer empirical research questions.

Estimation for the Agency-Communion Data

In this section, results of the analysis of the Agency-Communion data are presented and interpreted. Teachers with missing values on any of the variables were removed resulting in a sample size of 43. Convergence was reached within 750 iterations. After subtracting a burn-in of 750 from a total of 5,000 iterations the results in Table 6 were obtained (the dataset including files for analysis can be found in ESM 5–7).

The posterior means of the coefficients from the third column of Table 6 inform us about the linear relations between predictors and the components Agency and Communion as shown in Figures 4a–4d. Predicted scores are plotted against various values of one of the predictor variables. The other predictors are kept constant at their data means. The coefficients for the two linear components can be interpreted as usual, for example, for Extraversion: "An increase of 1 unit on Extraversion leads to a 0.34 increase in predicted score on the Agency component."

The last two columns of Table 6 show the lower and upper bounds of the 95% credible intervals. Only for the coefficients Communion Self Perception for the Communion component and Experience for the Agency component credible intervals do not include zero and indicate that there is an effect. Whether this also indicates that a circular effect exists remains to be seen.

To interpret the coefficients in a circular context and combine the coefficients of both components is less straightforward. To visualize what happens if we do so, Figures 4e–4h were constructed. Here, the predicted circular outcome is plotted against one of the predictor variables. The other predictors are kept constant at their data means. To obtain circular outcomes (in radians) we use the two-argument arctangent function (7) on the two linear outcomes. In the circular relation plots of Figure 4 the slope is not constant indicating that the relation between outcome and predictor is not linear. For example, in Figure 4g we see that teachers with average experience and average values on the other predictor variables score about 1.4 rad or 80.2° on the interpersonal circumplex. These teachers thus score relatively higher on the Agency than on the Communion component. Teachers with above average experience tend to score even higher on the Agency component compared to the Communion component, they move to an asymptote of the circular regression line of about 1.6 rad or 91.67°. Teachers with below average experience tend to score relatively higher on the Communion component. The least experienced teacher with average scores on the other predictors has a score on the interpersonal circumplex of about 1 rad or 57.20°.

We may compare these results to earlier research on the effect of experience on the score on the interpersonal circumplex. An example is found in Brekelmans, Wubbels, and Van Tartwijk (2005). Even though Brekelmans et al. (2005) subdivide the circumplex and therefore do not use circular statistics to analyze their data we can infer from their eight profile types that teachers with less experience tend to score higher on the Communion component

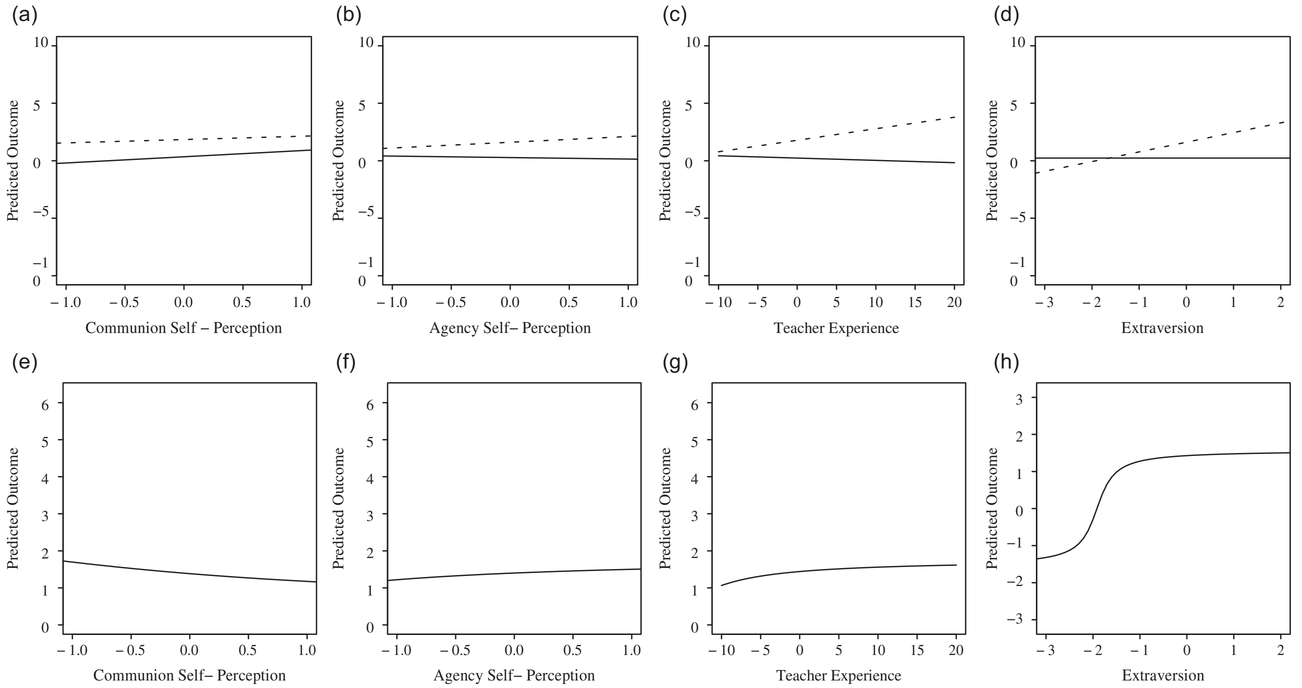

Figure 4. Predicted outcome for changes in 1 predictor. Upper subfigures are linear outcomes for the separate components (dashed = Agency, solid = Communion). Lower subfigures are circular outcomes (measured in radians).

relative to the Agency component compared to teachers with more experience. We can also infer that the teachers with the most experience show equal scores on Agency and Communion. This second result differs from the results from our data where teachers with the most experience had relatively higher scores on Agency compared to Communion. There are several possible causes of this difference. First, the data comes from different samples. Second, the analyses are based on different statistics. And third, in our analyses the effect of teacher experience is controlled for by several other predictors while this is not the case in Brekelmans et al. (2005). Modeling circumplex data in a circular setting thus allows for an interpretation of the effect on Agency and Communion relative to each other instead of separate effects on both components. This is more in line with the circular structure of the circumplex and an interpersonal variable being "[...] a particular blend of Agency and Communion, depending on that variable's location on the circle." (Gurtman, 2009, p. 2).

Discussion

The Agency-Communion data used throughout the current paper concerns the scores of teachers on the interpersonal circumplex. Even though theoretically circumplex models imply that their two dimensions should be considered jointly (Gurtman, 2009), the original research on the Agency-Communion data does not (Mainhard, Brekelmans, Brok, et al., 2011), as does other research on circumplex data. Instead, it considers statistical models for both dimensions separately (Mainhard, Brekelmans, & Wubbels, 2011; Wubbels, Brekelmans, den Brok, & van Tartwijk, 2006; Zeigler-Hill, Clark, & Beckman, 2011). Research employing circumplex data mentions the limitations of this approach and the benefits of using methods for circular data instead (Pennings et al., 2014; Wright et al., 2009). Another example where using a circular and more complex model could be beneficial can be found in a study by Locke, Sayegh, Weber, and Turecki (2016). In this study profiles on an interpersonal circumplex were made for depressed patients and normative samples. Although the authors do compare the circular means of these groups, a more complex model could, for example, allow for simultaneous evaluation of multiple predictors that influence the difference in scores of depressed and normative samples. A reason for not using different models may be that the methods are more complex and not as developed as the conventional methods for linear data and therefore not known to empirical researchers. The present paper has therefore investigated and assessed a method with which the two dimensions of circumplex data can be modeled jointly in a circular regression model by means of simulation studies and an empirical example.

The method used in this paper is very flexible. Both circular and linear predictors may be included in the model.

Although not discussed here, categorical predictors can be included by means of creating dummy variables. The effects are then interpreted by comparing predicted outcomes for persons with and without a score of 1 on these dummies. Furthermore, the two components of the outcome may be predicted by different combinations of variables, expanding possibilities to test theories of applied researchers. Additionally, both Nuñez-Antonio and Gutiérrez-Peña (2014) and Hernandez-Stumpfhauser et al. (2017) have developed a mixed effects model for a circular outcome meaning that longitudinal data or cross-sectional hierarchical data can also be modeled. The interpretation of effects in these models is however more complex. It is straightforward to estimate regression coefficients for two components separately and it is possible to interpret a circular effect qualitatively based on circular regression plots. When researchers are theoretically interested in a circular interpretation of the effects this qualitative approach offers them the possibility to do so. A quantitative assessment of the slope in circular regression plots has been given by Cremers, Mulder, and Klugkist (2018). In their work (7) is reparametrized in such a way that it contains a parameter to describe the slope. This is a first step to make hypothesis and model testing of circular effects possible in the embedding approach.

From the results of the simulation studies, we may reach several conclusions regarding performance. Because performance depends on what kind of data is investigated, researchers should inspect their data carefully before using the method described in this paper. It will produce biased estimates that have a low coverage if data is too highly concentrated on the circle ($\bar{R} > 0.95$) or has a too small sample size. In cases of extremely high concentration, disregarding the circular nature of the data and using linear estimation methods may give better results. To answer the question whether the investigated method will work on real data we need a better overview of what circular data looks like in practice. Regarding efficiency, in most designs the Bayesian sampler that was used converged well within 3,000 iterations. The designs that took longer to converge were those with a high κ_x, μ_x further from zero or small N. MCT is reasonably low overall and in all cases much lower for the slice sampler when taking the number of iterations into account.

Other methods to estimate parametric circular regression models exist. For methods based on the von Mises distribution (Gill & Hangartner, 2010; Lagona, 2016) no extensive simulation studies are performed in the literature. A known problem with this intrinsic approach is that the likelihood is not globally concave (Mulder & Klugkist, 2017) leading to estimation problems. The model presented by Ravindran and Ghosh (2011) is based on a wrapping approach and extensive simulation studies for the model were done and showed good performance. However, a measure of spread in the circular outcome was not systematically varied. For the present research, the influence of the spread in the circular outcome was investigated and found to affect the (relative) bias and coverage of the estimates. One aspect in which both the intrinsic approach and the wrapping approach differ from the embedding approach is that both only estimate one set of regression coefficients. This might prove to be an advantage in both hypothesis testing, and interpretation since results are obtained on a circular scale directly. In further research, it may thus be interesting to compare different methods for circular regression.

Acknowledgments

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

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1614-2241/a000147

ESM 1. Text (.txt)
Readme – Useful information.
ESM 2. Text (.pdf)
Detailed description of MCMC methods.
ESM 3. R code (.R)
Regression sampler MH.
ESM 4. R code (.R)
Regression sampler Slice.
ESM 5. Data file (.sav)
Empirical data.
ESM 6. Syntax file (.sps)
Syntax empirical data.
ESM 7. R code (.R)
Analysis empirical data.

References

Brekelmans, M., Wubbels, T., & Van Tartwijk, J. (2005). Teacher – student relationships across the teaching career. *International Journal of Educational Research, 43*, 55–71. https://doi.org/10.1016/j.ijer.2006.03.006

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

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

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

Gill, J., & Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis, 18*, 316–336. https://doi.org/10.1093/pan/mpq009

Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass, 3*, 601–619. https://doi.org/10.1111/j.17519004.2009.00172.x

Hernandez-Stumpfhauser, D., Breidt, F. J., & van der Woerd, M. J. (2017). The general projected normal distribution of arbitrary dimension: Modeling and bayesian inference. *Bayesian Analysis, 12*, 113–133. https://doi.org/10.1214/15-BA989

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

Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *Journal of the Royal Statistical Society. Series B, 37*, 97–133.

Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von Mises distribution. *Environmental and Ecological Statistics, 23*, 89–113. https://doi.org/10.1007/s10651-015-0330-y

Locke, K. D., Sayegh, L., Weber, C., & Turecki, G. (2016). Interpersonal self-efficacy, goals, and problems of persistently depressed outpatients: Prototypical circumplex profiles and distinctive subgroups. *Assessment*, 1–13. https://doi.org/10.1177/1073191116672330

Mainhard, M. T., Brekelmans, M., Brok, P. D., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary Educational Psychology, 36*, 190–200. https://doi.org/10.1016/j.cedpsych.2010.06.002

Mainhard, M. T., Brekelmans, M., & Wubbels, T. (2011). Coercive and supportive teacher behaviour: Within-and across-lesson associations with the classroom social climate. *Learning and Instruction, 21*, 345–354. https://doi.org/10.1016/j.learninstruc.2010.03.003

Mulder, K., & Klugkist, I. (2017). Bayesian estimation and hypothesis tests for a circular Generalized Linear Model. *Journal of Mathematical Psychology, 80*, 4–14. https://doi.org/10.1016/j.jmp.2017.07.001

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

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

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

Pennings, H. J., Brekelmans, M., Wubbels, T., van der Want, A. C., Claessens, L., & van Tartwijk, J. (2014). A nonlinear dynamical systems approach to real-time teacher behavior: Differences between teachers. *Nonlinear Dynamics, Psychology, and Life Sciences, 18*, 23–45.

R Core Team. (2014). *R: A language and environment for statistical computing* (Technical Report). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Ravindran, P., & Ghosh, S. K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice, 5*, 547–561. https://doi.org/10.1080/15598608.2011.10483731

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

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

Wubbels, T., Brekelmans, M., den Brok, P., & van Tartwijk, J. (2006). An interpersonal perspective on classroom management in secondary classrooms in the Netherlands. In J. Fagerberg, D. Mowery, & R. Nelson (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1161–1191). Mahwah, NJ: Erlbaum.

Zeigler-Hill, V., Clark, C. B., & Beckman, T. E. (2011). Fragile self-esteem and the interpersonal circumplex: Are feelings of self-worth associated with interpersonal style? *Self and Identity, 10*, 509–536. https://doi.org/10.1080/15298868.2010.497376

Received July 20, 2015
Revision received February 5, 2018
Accepted February 16, 2018
Published online June 21, 2018

Jolien Cremers
Department of Methodology and Statistics
Utrecht University
Utrecht
The Netherlands
j.cremers@uu.nl

Jolien Cremers is a PhD Candidate at the Department of Methodology and Statistics, Utrecht University. Her research interests are circular data modelling, multilevel models and modelling social scientific data in general.

Tim Mainhard is an Associate Professor at the Department of Education, Utrecht University. His research interests are (modelling) social dynamics in educational settings and the effect of these dynamics on student and teacher outcomes (socio-emotional and academic).

Irene Klugkist is a full Professor at the Department of Methodology and Statistics, Utrecht University. In addition, she serves as extra-ordinary Professor at the Behavioural Sciences Department at the University of Twente. Her research interests are Bayesian statistics, evaluation of inequality constrained hypotheses, accumulation of knowledge through the use of informative priors, and circular data analysis.