One Direction? Modelling Circular Data in the Social Sciences using the Embedding Approach

Een richting? Het modelleren van circulaire data in de sociale wetenschappen met behulp van de inbedding methode (met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 24 mei 2019 des ochtends te 10.30 uur

door

Jolien Cremers

geboren op 8 februari 1992
te Venlo

Promotoren:

Prof.dr. I.G. Klugkist

Prof.dr. H.J.A. Hoijtink

Beoordelingscommissie:

Prof.dr. J.W.F. van Tartwijk

Prof.dr. J.G. Schouten

Prof.dr. J.K. Vermunt

Prof.dr. E.M. Wagenmakers

Dr. A. Maruotti

Cremers, Jolien
One Direction? Modelling Circular Data in the Social Sciences using the Embedding Approach.
Proefschrift Universiteit Utrecht, Utrecht. - Met lit. opg. - Met samenvatting in het Nederlands.

The studies in this thesis were funded by the Netherlands Organization for Scientific Research (project 452-12-010).

ISBN: 978-90-393-7103-9
Printing & Cover Design: Ridderprint BV

Contents

Introduction

Disappointing as it may be, the title of this dissertation does not in any way refer to a popular English-Irish boyband group[1]. Instead it refers to a hypothesis of interest in certain fields in the social sciences. Take for instance a fictional experiment in which we ask two groups of people at Times Square in New York City to point in the direction of a known landmark, e.g. the Brooklyn Bridge. Both groups have never visited New York City before and have only visited the Brooklyn Bridge once, right before either walking (group 1) or taking the subway (group 2) to Times Square. We are then interested in whether the two groups point in one and the same direction or whether they will point in two different directions and thus whether their sense of direction differs depending on the mode of transport, walking or taking the subway, they used. Figure 1 shows simulated data from this fictional experiment. Clearly, the group that walked to Times Square points in the correct direction on average (South) whereas the group that took the metro seem to point in all directions and there is no clear average direction.

Three types of circular data

The data that results from the type of experiment described above can be measured on a compass from 0 to 360 degrees (see Figure 1) and is considered one of three types of circular data. In the social sciences this first type of circular data arises in cognitive psychology in research on cognitive maps and the human sense of direction (Brunyé, Burte, Houck, & Taylor, 2015; Chrastil & Warren, 2017; Warren, Rothman, Schnapp, & Ericson, 2017), the visual perception of space (Matsushima, Vaz, Cazuza, & Ribeiro Filho, 2014), visual working memory (Heyes, Zokaei, & Husain, 2016), moving room experiments (Stoffregen, Bardy, Merhi, & Oullier, 2004) and research in eye-tracking (Rayner, 2009). In this dissertation examples from this type of circular data are shown in Chapters 1 and 3.

The second type of circular data that occur in the social sciences are those data that concern periodic measurements. For example, in situations where we are interested in the time of the day, the week, the year or any other prespecified time interval a certain event occurs. Also research in which we are interested in a phase difference belongs in this category. The second type of circular data is found among others in

[1] We refer those who do not understand this remark to: en.wikipedia.org/wiki/One_Direction

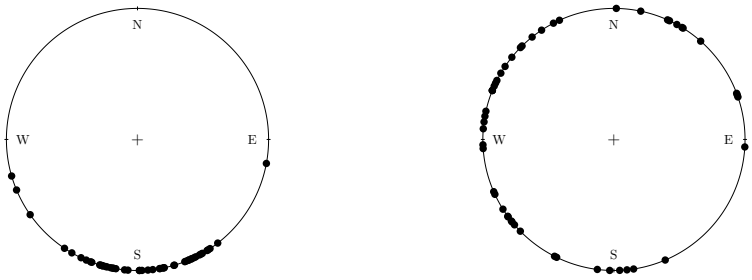

Figure 1: Data from participants in an experiment that were instructed to point towards the Brooklyn Bridge from Times Square. The plot on the left shows data from the group that walked to Times Square and the plot on the right shows data from the group that took the subway.

research on human motor resonance or movement synchronization (Ouwehand & Peper, 2015; Puglisi et al., 2017), sensorimotor synchronization in music making (Kirschner & Tomasello, 2009) and cyclical trends in the occurrence in gun crimes or political violence (Gill & Hangartner, 2010). Figure 2 shows density plots for the gun crime data from Gill & Hangartner (2010) on both a circular and a linear scale. The gun crime data contains instances of gun crimes and the time of day they took place measured on a 24 hour scale. As we can see in Figure 2, most gun crimes occur around midnight (the circular mean). There is however a big difference between the linear and circular density plot. The linear density plot is bimodal whereas the circular density plot is unimodal. This is caused by the fact that 1 am and 12 pm seem to be opposite ends on the linear scale whereas on the circular scale they lie close together. The circular scale thus provides a more natural representation of clock times. Using standard methods for statistical inference leads to problems in the gun crime data. For instance, the linear mean of 11 pm and 1 am equals 12 am ($23 + 1/2 = 12$) instead of 12 pm. We thus need special inference methods for circular data. In this dissertation an example of the second type of circular data can be found in Chapter 1 in which we describe an experiment investigating the phase difference between hand movements.

The last type of circular data occurs when a circular scale is constructed. Three scales of this type that occur in the social sciences are the interpersonal circumplex from interpersonal psychology (Horowitz & Strack, 2011), the right - left/conservative - progressive spectrum in which we can classify political parties' standpoints from political science (Gill & Hangartner, 2010) and the Basic Human Values scale (Schwartz, 1994) from sociology. Figure 3 shows the interpersonal circumplex as it is sometimes used in research on educational psychology (Brekelmans, Wubbels, & Van Tartwijk, 2005; Mainhard et al., 2011a; Pennings et al., 2018; Wubbels,

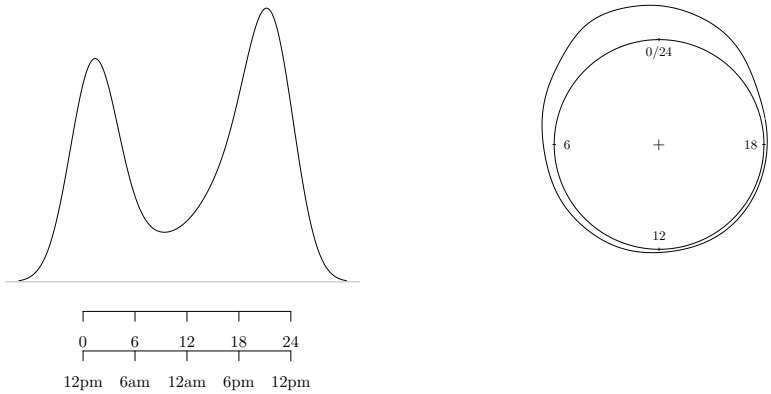

Figure 2: Density plots of a random subset (N = 400) from the gun crimes data on a linear (left) and circular (right) scale.

Brekelmans, Den Brok, & Van Tartwijk, 2006). In this figure we see two orthogonal dimensions: Agency and Communion that represent the interpersonal behavior of a teacher. Agency refers to the degree of power or control a person exerts in interaction with others. Communion refers to the degree of friendliness or affiliation a person conveys in interaction with others. We also see 8 different subscales (helpful, directing, imposing, etc.) that each characterize a different type of teacher behavior and represent a particular blend of Agency and Communion. In the literature, data from the interpersonal circumplex is mostly analysed as two separate scores on the Agency and Communion dimensions or as a score on a categorical variable comprised of the 8 subscales. We can however also regard the interpersonal circumplex as a circular scale (see Figure 3). Scores on the Agency and Communion dimension can be translated to a score on the interval from 0 to 360 degrees, a continuous measurement on the edge of the circle. We use data from the interpersonal circumplex in educational psychology throughout this dissertation. In Chapter 2 we use it as illustrative example and in Chapters 5 and 6 we show how methods for circular data can help us discover different patterns compared to existing methods for this type of data.

Modelling circular data

In the literature there are three approaches to modelling circular data. Firstly there is the 'intrinsic' approach where we model distributions directly on the circle. The most used distribution in this approach is the von Mises distribution. Secondly there is the 'wrapping' approach where a distribution on the real line, e.g. a normal distribution, is wrapped around the circle by computing modulo 2π for the distribution. In this dissertation however we choose to focus on the 'embedding' approach, more specifically a Bayesian embedding approach. In the embedding

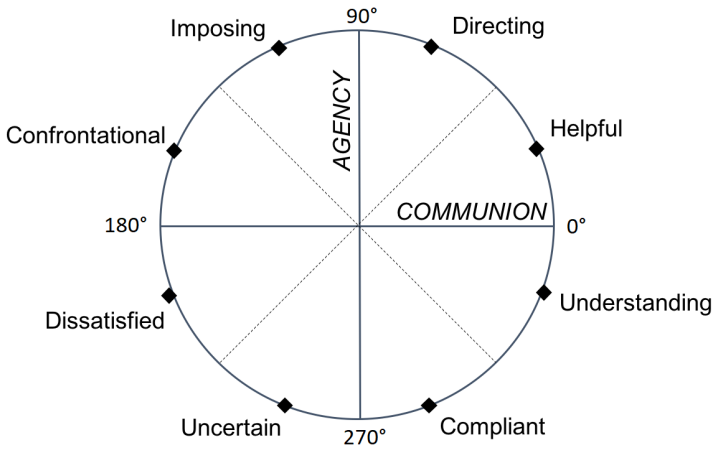

Figure 3: The interpersonal circle for teachers (IPC-T). The words presented in the circumference of the circle are anchor words to describe the type of behavior located in each part of the IPC.

approach circular datapoints are assumed to origin from a projection of a bivariate distribution, e.g. a bivariate normal distribution, onto the circle. Because we assume that the circular measurements arise from a bivariate distribution we can perform inference in bivariate space instead of directly working with the angles. This has an advantage because a lot of complex models for bivariate linear data already exist. We can use these existing models instead of developing completely new models for the circular data. A problem however concerns the fact that the results from a model fit to the underlying bivariate data are also defined in bivariate space. For example, if we estimate the mean of a set of circular values using the embedding approach the method will return two means to us, one for each component of the bivariate data. We thus need a way to translate the bivariate results to the circle. In this dissertation we have proposed a solution for this problem and introduced tools for translating the bivariate results back to the circle in a GLM (Chapter 3) and mixed-effects model (Chapter 5). We focus on a Bayesian embedding approach for a pragmatic reason. Within the Bayesian approach models have been developed that can handle more complex data. For example, the mixed-effects model that we use in Chapters 1 and 5 does not have a frequentist equivalent. Additionally, for the translation of bivariate results to the circle we make use of the advantages that the Bayesian approach gives us concerning the transformation of parameters and computation of uncertainty estimates.

This dissertation

Seeing that there are so many instances where circular data arise in the social sciences one would expect that the use of specialized methods and statistics for these kinds of data are also widespread. This is however not the case. Even though there

are textbooks dedicated to methods for circular data (Batschelet, 1981; Fisher, 1995; Jammalamadaka & Sengupta, 2001; Ley & Verdebout, 2017; Mardia & Jupp, 2000; Pewsey, Neuhäuser, & Ruxton, 2013), these works are not part of the 'standard' texts on statistical analysis in the social sciences nor are they very well known among social scientific researchers. The work that has been done is this dissertation is therefore directed towards three goals: making methodology for circular data available to social scientists, testing the performance of the available methods and extending the methods to serve the specific needs of social scientists. In Chapter 1 we present a tutorial on the analysis of circular data for cognitive psychologists and social scientists in general. As part of this tutorial, we have created an R package that allows social scientists to analyze both GLM and mixed-effects models with a circular outcome. In Chapter 2 we have tested the performance of the GLM model in a regression setting and in Chapter 3 and 4 we introduce new tools that allow for an easier interpretation of the parameters from the GLM. In Chapter 5 we illustrate the advantages of using the circular mixed-effects model to analyse longitudinal data from an interpersonal circumplex. In this Chapter we also develop new tools that allow for an easier interpretation of the parameters from the mixed-effects model. Finally, in Chapter 6, we illustrate how cylindrical models, in which we model both a circular and a linear outcome variable, might show even more advantages for modelling circumplex data than the methods presented in Chapter 5. More importantly, we extend four existing cylindrical models to include covariates to predict both the circular and linear outcome to make them more useful for social scientists.

Chapter 1

One direction? A tutorial for circular data using R with examples in cognitive psychology

by J. Cremers & I. Klugkist[1]

1.1 Introduction

Circular data arises in almost all fields of research, from ecology where data on the movement direction of animals is investigated (Rivest, Duchesne, Nicosia, & Fortin, 2015) to the medical sciences where protein structure (Mardia, Taylor, & Subramaniam, 2006) or neuronal activity (Rutishauser, Ross, Mamelak, & Schuman, 2010) is investigated using periodic and thus circular measurements. The most direct examples of circular data within the social sciences arise in cognitive and experimental psychology. For example, in experiments on cognitive maps the human sense of direction is investigated through asking participants in a study to point north (Brunyé et al., 2015) or to walk to a target object (Warren et al., 2017). The closer the participants' pointing or walking direction was to the actual north or target object, the better their sense of direction. Other examples include the visual perception of space (Matsushima et al., 2014), visual working memory (Heyes et al.,

[1]Published as Cremers, J. & Klugkist, I. (2018) One direction? A tutorial for circular data using R with examples in cognitive psychology. *Frontiers in Psychology: Cognitive Science, 9*, 2040. doi: 10.3389/fpsyg.2018.02040.

Author contributions: JC and IK designed the study, JC developed the associated R-package and analyzed, processed and interpreted the results of the empirical examples with feedback from IK, JC wrote the paper and IK gave feedback on the written work.

2016) and sensorimotor synchronization in music making (Kirschner & Tomasello, 2009).

However, despite the fact that circular data is being collected in different areas of cognitive and experimental psychology, the knowledge of this type of data is not well-spread. Circular data is fundamentally different from linear data due to its periodic nature. On the circle, measurements at 0° and 360° represent the same direction whereas on a linear scale they would be located at opposite ends of a scale. For this reason circular data require specific analysis methods. Some less technical textbooks on analysis methods for circular data have been written (Batschelet, 1981; Fisher, 1995; Pewsey et al., 2013). However these works are not part of the 'standard' texts on statistical analysis in psychology or the social sciences in general nor are they very well known among social scientific researchers.

Therefore, this paper aims at giving a tutorial in working with and analyzing circular data to researchers in cognitive psychology and the social sciences in general. The main goal of this tutorial is to explain how to inspect and analyse your data when the outcome variable is circular. We will discuss data inspection, model fit, estimation and hypothesis testing in general linear models (GLM) and mixed-effects models. We decide to mainly focus on one particular approach to the analysis of circular data, the embedding approach. We do so for the flexibility of this approach and the resulting variety in types of models that have already been outlined in the literature on circular data for this approach. Note that for an optimal understanding of the tutorial, the reader should ideally have some knowledge on `R` (R Core Team, 2017) and on GLM and mixed-effects models in the linear setting. The reader does not need to be familiar with circular data.

The structure of the tutorial is such that the reader is guided by two examples throughout the paper. One is an example for an ANOVA model and the other for a mixed-effects model. First however, we give a short introduction to circular data in general. Then we introduce the ANOVA example after which descriptive methods for circular data are explained through a section on data inspection for this example. After that we will continue with an analysis of the example datasets. First we analyse the ANOVA dataset using a method for circular GLM and give interpretation guidelines for this model. Subsequently we will introduce and analyse the mixed-effects example data. Again, we analyse this data and include guidelines for interpretation. The analyses of both datasets, the ANOVA and mixed-effects dataset, are performed using the `R` package `bpnreg` (Cremers, 2018). For both models, the GLM and the mixed-effects model for a circular outcome, we write a short technical section in which the mathematical details of the respective models are given. Lastly, we give a summary of the paper and additional references to literature on other models for circular data in the concluding remarks.

1.2 Circular data

In the introduction we have briefly mentioned that circular data is data of a periodic nature. The most intuitive form of circular data comes in the form of directions on a compass. For example, a participant in an experiment could be instructed to move

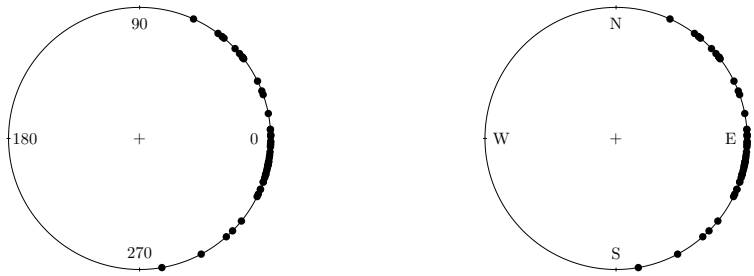

Figure 1.1: Data from participants in an experiment that were instructed to move East. The plot on the left shows the data on a 0° - 360° scale. The plot on the right shows the data on the compass.

or point to a certain target. We can then measure the direction, North, South, East or West on a scale from 0 to 360 degrees. A plot with simulated data containing such measurements for several participants is shown in Figure 1.1. In this plot we can easily see the periodicity of the data, 0° represents the same datapoint as 360°. Furthermore, we can see what happens if we would treat this data in the 'usual' linear way. Participants that moved North-East have a score of 45° and participants that moves South-East have a score of 315°. On the circle we can see that this is only 90° apart, while on a linear scale it is much further apart at 315° - 45° = 270°. More importantly however, there is a difference between the circular and linear means for this data. In Figure 1.1 we see that the circular mean direction is 0°. The linear mean however is 180° and is opposite to the actual mean direction of the data. Clearly, a linear treatment of the data in Figure 1.1 can lead to incorrect conclusions.

Clock times are another type of circular data. We might for instance be interested at what time of day a certain event takes place, e.g. the time of day at which positive affect is highest. Figure 1.2 shows simulated data for the time of day at which positive affect is highest for two groups of participants, e.g. two groups of psychiatric patients who are being treated for depression at different clinics. From the plots we clearly see that the peak of positive affect for the two groups is at roughly the same time of day, one slightly before 12 pm and one slightly after. However, if we were to analyze this data using standard statistics for linear data and we would compare the means of the two groups, 11 pm and 1 am we would reach a completely different conclusion. The two means are namely at the two opposite ends of a linear scale from 00.00 am to 12.00 pm, and we would conclude that the time of day at which positive affect is highest is different for the two groups.

The two examples of circular data that we have just given illustrate why it is

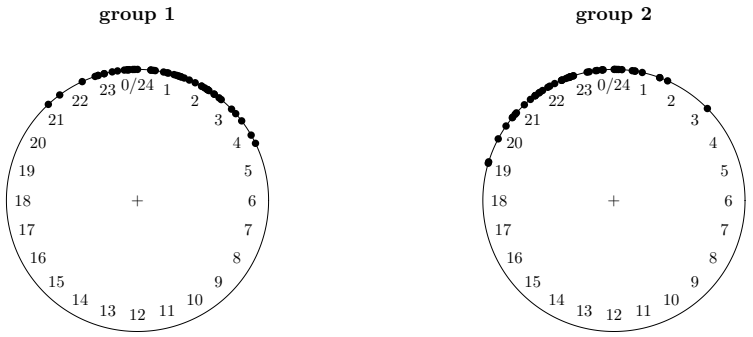

Figure 1.2: Data for the hour at which positive affect is highest for two groups of psychiatric patients who are being treated for depression at different clinics.

important to treat circular data differently from linear data. This goes both for describing your data, e.g. computing circular means, as well as analyzing them, e.g. testing whether the circular means of two groups differ. In the next section we will introduce an example dataset on which we will show several ways to inspect and compute descriptive measures for circular data.

1.3 Inspecting your data

In the previous section we have seen that the computation of a circular mean differs from that of a linear mean. Methods for data inspection, the computation of descriptive statistics and plotting methods, are different for circular data. Because data inspection should be done before performing inference of any kind we will outline a basic way to inspect circular data using the `R` packages `bpnreg` (Cremers, 2018) and `circular` (Agostinelli & Lund, 2017). We will discuss plots and several descriptive measures for circular data using an example dataset, the motor resonance data.

1.3.1 The motor resonance data

In this section we introduce data from an article by Puglisi et al. (2017) on human motor resonance. From now on we will call this data the motor resonance data. Motor resonance is a response in the brain in the primary motor cortex and spinal circuits that is caused by observation of others' actions. In their research Puglisi et al. (2017) conduct an experiment in which 'observers' are asked to either look at the movement of a hand of a 'mover' or at another object in order to evaluate the role of attention in motor resonant response. The experiment has three conditions: the

'explicit observation' condition (n = 14), where observers are explicitly instructed to observe the hand, the 'semi-implicit observation' condition (n = 14) where the observers have to perform a task that requires the implicit observation of the hand of the mover and the 'implicit observation' condition (n = 14), where observers have to perform a task that is independent of the observation of the hand of the mover. The idea of motor resonance is that the 'observer' starts moving his or her hand in the same manner as the 'mover' because he or she is implicitly or explicitly observing the hand of the 'mover'. This is the resonant response. This resonant response is hypothesized to be strongest and more synchronized with the hand movement of the mover in the explicit condition and smallest in the implicit condition. In each condition the hand movements of the observers were measured and the phase difference between movement of the observers' hand and the hand of the mover was calculated. A phase difference can be expressed in degrees or time and is formally defined as the difference at a specific point in time between two waves that have the same frequency. In the motor resonance data the phase difference is a measurement of the strength of the resonant response and a circular variable. It can thus be described and analyzed using circular statistics. In addition to the phase difference the average amplitude of the hand movement of the observer was computed. Note that in the original article there was also a baseline condition (n=14) without a mover. In this condition, observers had to look at an inanimate object that moves in an identical manner to the hand of the mover in other conditions. The baseline condition is however not included in the example data since no resonant response was observed in the observers' hand according to the original research.

The motor resonance data can be found in the package `bpnreg` as the dataframe `Motor`. `Motor` is a dataframe with 42 rows and 7 variables. The variable `Cond` indicates the condition (explicit, semi-implicit and implicit) a participant was placed in, the `AvAmp` variable contains the average amplitude, and the `PhaseDiff` and `Phaserad` variables contain the measured phase difference between 'observer' and 'mover' in degrees and radians respectively. Note that circular data can be represented both in degrees on a scale from 0° to 360° and in radians on a scale from 0 to 2π (1 degree = 1 * $\pi/180°$ radians).

1.3.2 Plots for circular data

The main question of interest for the motor resonance data is whether the phase difference between the three experimental conditions differs. To be more precise whether there is a smaller phase difference in the explicit condition than in the other two. Differences that are observed in the phase difference are interpreted as differences in the strength of the resonant response (Puglisi et al., 2017). A smaller phase difference indicates a stronger and more synchronized resonant response. A first step to investigating the question of interest is plotting the phase differences of the three conditions. We can do so using the package `circular`.

Figure 1.3 shows the plots of the phase differences of each condition. We see that the phase differences in the explicit condition are much less spread out on the circle than the phase differences in the other two conditions. Also the average phase

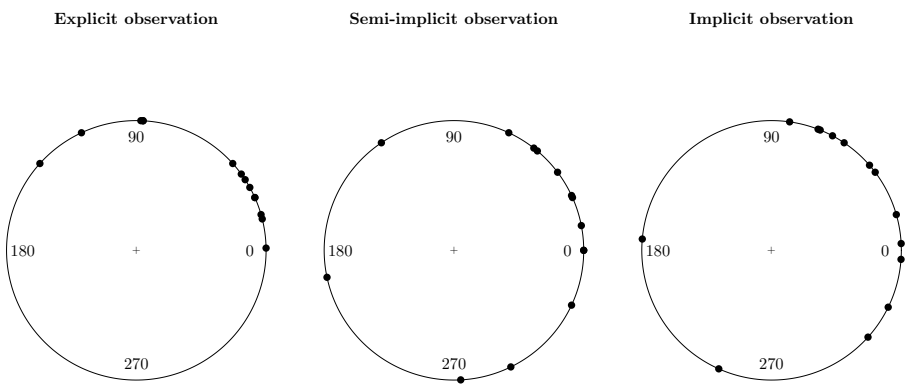

Figure 1.3: Plots of the phase differences for each condition of the motor resonance data.

Table 1.1: Descriptives for the motor resonance data with mean direction ($\bar{\theta}$), mean resultant length ($\bar{R}$), circular variance (V_m) and circular standard deviation (v) of the phase difference for each condition.

Phase difference	$\bar{\theta}$	$\bar{R}$	V_m	v
explicit	49.55°	0.77	0.23	41.39°
semi.implicit	18.45°	0.54	0.46	63.82°
implicit	31.94°	0.56	0.44	61.72°

differences seem to differ between the conditions. In the next section we show measures for the mean and variance of a sample of circular data.

1.3.3 Circular mean, resultant length and variance

Table 1.1 shows descriptives for the motor resonance data. For each group, the table contains sample statistics for the circular mean and mean resultant length of the phase difference. The circular population mean, μ, indicates the average direction of a certain variable in the population. The population mean resultant length, ρ, is a statistic between 0 and 1 that gives us information on the spread of a circular variable in the population. It is interpreted as a precision measure where 0 means that the spread is large and 1 means that all data are concentrated at a single value. Sample statistics for these values are $\bar{\theta}$ for the mean and $\bar{R}$ for the mean resultant length.

Graphically we can illustrate the computation of $\bar{\theta}$ and $\bar{R}$ as shown in Figure 1.4. On the left side of this figure we see two sets of circular data. We represent a circular datapoint as a vector composed of the cosine and sine of the datapoint

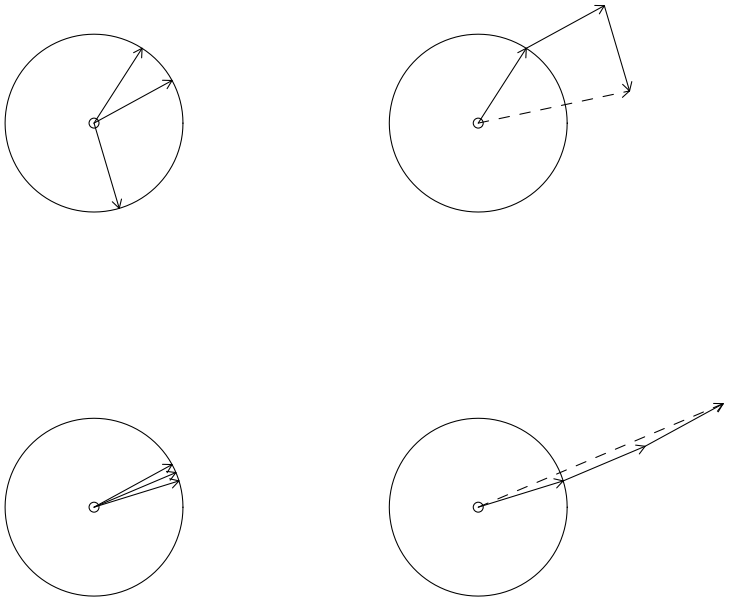

Figure 1.4: The computation of a circular mean and (mean) resultant length (right) for two sets of circular data (left). The solid lines are vectors representing the circular datapoints. The direction of the dotted vector is the mean direction and the length of the dotted vector is the resultant length.

instead of one value measured in degrees (or radians). E.g. for a score of 90° we have the following vector $(\cos(90°), \sin(90°))$. The solid vectors in Figure 1.4 each represent one circular datapoint. To compute the mean directions and resultant lengths for the datasets on the left we place the vectors head to toe, as in the right side of Figure 1.4. We then connect the toe of the first vector to the head of the last vector. This results in the dotted vectors on the right side of Figure 1.4. The direction of the dotted vector is the mean direction, $\bar{\theta}$ of the vectors from which it was created. The length of the dotted vector is the resultant length. The mean resultant length, $\bar{R}$ is the length of this vector divided by the number of vectors from which it was created. In Figure 1.4 we see that the data in the bottom left figure are much more concentrated on the circle than the data in the upper left figure. This translates to the resultant length in the bottom right being larger than the resultant length (length of the dotted vector) in the upper right. Formulae for the computation of $\bar{\theta}$ and $\bar{R}$ can be found in Fisher (1995).

In Table 1.1 we see that in the motor resonance data the circular mean of the phase difference for the explicit observation condition is highest with 49.55°. The mean phase differences for the semi-implicit and implicit observation conditions are lower at 18.45° and 31.94°. Moreover, the mean resultant lengths of the three groups differ. The phase differences of the individuals in the explicit observation condition are most concentrated with $\bar{R} = 0.77$. The phase differences differ more between the individuals in the semi-implicit and implicit observation conditions where the spread is larger at $\bar{R}$'s of 0.54 and 0.56 respectively.

Table 1.1 also shows a circular variance and standard deviation. The sample value, V_m for the circular variance is defined as $1 - \bar{R}$. Its interpretation is exactly opposite to the interpretation of the mean resultant length. A variance of one means that a variable has a very large spread and a variance of 0 means that all data are concentrated at one point. Note that unlike a linear variance the circular variance is bounded between 0 and 1. A sample circular standard deviation, v, can also be computed (Fisher, 1995). This deviation runs from 0 to infinity where higher values indicate a larger dispersion.

We have seen that both the average phase difference and variances of the phase difference seem to be different for the three conditions in the motor resonance data. To test whether these differences in circular means also exist in the population, we can use a a projected normal circular GLM. In the next section we will introduce this model and fit it to the motor resonance data.

1.4 A general linear model with a circular outcome

In this section we will introduce a projected normal circular regression model. Note that because it is a regression model we can also fit AN(C)OVA type models with it. We can thus refer to it as a projected normal (PN) circular GLM. The PN circular GLM falls within the embedding approach to circular data. The embedding approach is characterized by the fact that it takes an indirect approach to modelling

circular data. Instead of directly defining a model on the circular outcome θ we use a mathematical trick that allows us to define a model in bivariate linear space. The results of the model in bivariate linear space can then be translated back to the circle. Next, we will outline the theoretical background to the PN circular GLM and the embedding approach. Subsequently we will continue to fit an ANOVA to the motor resonance data. At the end of this section we will shortly consider different methods for circular ANOVA.

1.4.1 The embedding approach to circular data

In the previous section, at the computation of the circular mean, we have seen that a circular variable θ, e.g. the phase difference in the motor resonance data, can be expressed as a unit vector $\boldsymbol{u}$ composed of the sine and the cosine of an angle $\boldsymbol{u} = (\cos\theta, \sin\theta)$. If we translate this to bivariate real space the cosine is the x-component and the sine is the y-component of an angle. In the embedding approach we assume that $\boldsymbol{u}$ origins from a bivariate normal variable $\boldsymbol{y}$. This bivariate variable is not measured, and can thus be regarded as a latent variable. Figure 1.5 depicts the relation between $\boldsymbol{u}$ and $\boldsymbol{y}$ for a dataset with sample size $n = 10$. It shows that the circular datapoints could have originated from different sets of bivariate datapoints $\boldsymbol{y}$. Because the $\boldsymbol{y}$ are not observed we need special inference methods such as expectation maximization techniques in a frequentist approach or auxiliary variable techniques in a Bayesian approach to estimate a model. In this paper we will use a Bayesian approach. The reason to choose a Bayesian approach instead of a frequentist one is that it allows for the modelling of more complex data, e.g. there is no frequentist version of the circular mixed-effects model we will use in Section 1.5. More details on the Bayesian approach can be found in Nuñez-Antonio, Gutiérrez-Peña, & Escarela (2011) and Cremers et al. (2018b).

From the assumption that $\boldsymbol{y}$ has a bivariate normal distribution it follows that θ has a projected normal distribution (Presnell, Morrison, & Littell, 1998). When fitting a model using the PN distribution we model the mean vector $\boldsymbol{\mu}$ of the underlying bivariate data $\boldsymbol{y}$.

Because $\boldsymbol{y}$ is bivariate the mean vector $\boldsymbol{\mu}$ has two components, denoted with the superscripts I and II. These superscripts therefore refer to the x and y axis of the Cartesian plane or the cosine and sine component in $\boldsymbol{u}$ respectively. In a multiple regression model this $\boldsymbol{\mu}$ is specified as follows:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\beta}^I)^t \boldsymbol{x}_i^I \\ (\boldsymbol{\beta}^{II})^t \boldsymbol{x}_i^{II} \end{pmatrix}, \tag{1.1}$$

where $i = 1, \ldots, n$, $\boldsymbol{x}_i$ is a vector of predictor values for individual i and each $\boldsymbol{\beta}$ is a vector with intercept and regression coefficients. The superscript t for β_0^I and β_0^{II} denotes that the transpose of these vectors is taken. To be able to estimate an intercept, the first component of $\boldsymbol{x}_i$ equals 1. Note that the vectors $\boldsymbol{x}_i$ are allowed to differ for the two components I and II. In the next section we will fit this type of model to the motor resonance data.

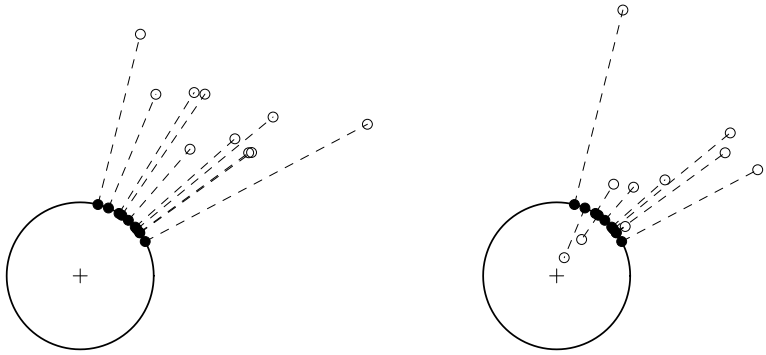

Figure 1.5: A set of datapoints in bivariate space $\boldsymbol{y}_i$, where $i = 1, \ldots, n$, projected onto the circle to produce a set of datapoints $\boldsymbol{u}_i$. The lines connecting the datapoints to the center of the circle represent r_i.

In terms of the interpretation of the circular effect of a variable the two component structure in (1.1) poses a problem.[2] The two components do not necessarily have a useful interpretation for each type of circular data, e.g. we cannot talk of a 12 o'clock (sine component) and 3 o'clock axis (cosine component) in Figure 1.2. To be able to interpret effects on the circle we transform the effects on the two components to an effect on the circle. This transformation was introduced by Cremers et al. (2018b) and will be applied in both example datasets.

1.4.2 Fitting an ANOVA model to the motor resonance data

In this section we will fit a circular ANOVA model to the motor resonance data using the PN circular GLM from the package `bpnreg`. Note that the model from this package is in fact a regression model that we can specify in such a way that it is mathematically equivalent to an ANOVA. First we will give the details on this model and subsequently we will discuss and interpret the results.

[2]Note that we could rotate and shift the components (axes) in bivariate space such that for a categorical predictor the x-component points to the mean of the reference category and the beta weights of the y-component refer to a deviation from this reference mean. This way we could test whether the means of the groups differ in bivariate space. However, this still does not lead to means or effects that are interpretable on the circle.

1.4.2.1 Fitting the model

To investigate the effect of condition on the phase difference we specify the prediction equation for the mean vector in the PN circular GLM as follows:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} \beta_0^I + \beta_1^I \text{semi.implicit}_i + \beta_2^I \text{implicit}_i \\ \beta_0^{II} + \beta_1^{II} \text{semi.implicit}_i + \beta_2^{II} \text{implicit}_i \end{pmatrix}, \tag{1.2}$$

where the variables semi.implicit and implicit are dummy variables indicating condition membership, β_0^I and β_0^{II} are the intercepts and β_1^I, β_2^I, β_1^{II} and β_2^{II} are the regression coefficients of the model. Note that in this model we take the explicit observation condition as the reference condition. When we translate this to the ANOVA context, the intercepts, β_0^I and β_0^{II}, represent the mean for the explicit condition, $\beta_0^I + \beta_1^I$ and $\beta_0^{II} + \beta_1^{II}$ are expressions for the mean of the semi-implicit condition and $\beta_0^I + \beta_2^I$ and $\beta_0^{II} + \beta_2^{II}$ are expressions for the mean of the implicit condition.

We use the package `bpnreg` to fit the model in (1.2). Because this is a Bayesian method we have to specify some parameters for the Markov Chain Monte Carlo (MCMC) sampler that estimates the parameters of the model. MCMC methods are iterative and we thus need to specify the number of iterations that we want to run. We choose a relatively high number of the output iterations (`its = 10000`) to make sure that the sampler converges and that we do not need to run it again in case it did not. We choose the burn-in period to consist of 100 iterations (`burn = 100`). This means that we throw away the first 100 iterations to make sure that the iterations we keep are those at which the sampler has reached its equilibrium. We also choose the lag, that is how many iterations we want to keep, in this case every third iteration (`n.lag = 3`). We set a lag to prevent possible auto-correlation between the parameter estimates. In the next section we will elaborate further on how to check convergence and choose these MCMC parameters wisely. But first, we fit the model:

```
fit.Motor <- bpnr(pred.I = Phaserad ~ 1 + Cond, data = Motor,
                  its = 10000, burn = 100, n.lag = 3, seed = 101)
```

1.4.2.2 Convergence

In a Bayesian model that uses MCMC sampling for estimation we always have to assess convergence of the MCMC chain for all parameters in the model. A traceplot is one way to assess the convergence of a parameter. As an illustration we only show the traceplot for the MCMC chains for one of the parameters of the model in Figure 1.6.

From the traceplot in Figure 1.6 we may conclude that the MCMC chain converged within 10000 iterations and a burn-in of 100. In general if the traceplot shows proper convergence there are no flat parts, where the chain stays in the same state for too long, or parts with too many steps in one direction. We want to see a pattern in which the chain moves from below a certain equilibrium to above and vice versa

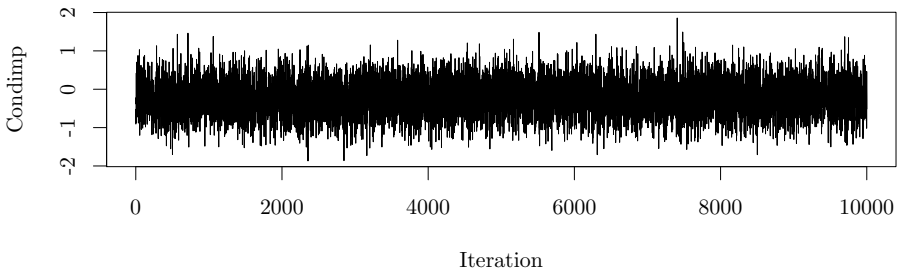

Figure 1.6: A traceplot showing convergence of the parameter β_2^I of the model fit to the motor resonance data.

in just a few iterations. The traceplot will then look like a so-called 'fat-caterpillar' meaning that the chain has reached an equilibrium around a particular value and has thus converged. In case an MCMC chain does not converge we could add more iterations, a larger `n.lag` or more burn-in iterations. We can also evaluate other convergence diagnostics. The focus of this paper however does not lie on Bayesian data analysis and therefore we refer to other works, e.g. Gelman et al. (2014), for more information on assessing convergence.

1.4.2.3 Results

To answer the question whether the phase differences in the three conditions of the motor resonance data differ we investigate their circular means. To do so we use methods from Cremers et al. (2018b) to transform the results from the two components of $\boldsymbol{\mu}$ to the circle. Note that investigating the regression coefficients on the two bivariate components separately might lead to wrong conclusions about the effect on the circle, the phase difference. This is due to the fact that even though there is an effect of a variable on each of the two components this does not mean that we can also see an effect on the circle. For a more detailed explanation we refer to Cremers et al. (2018b).

Because we use a Bayesian method we get the posterior distributions of the three circular means. Philosophically, in Bayesian statistics each parameter is said to have its own distribution. The posterior distribution is the result of the prior knowledge we have about a parameter before conducting a study, formalized as a 'prior' distribution (in this paper we choose non-informative priors for the parameters) and the information that lies in the data obtained from a study, formalized as the likelihood. The fact that we obtain the distribution of a parameter is convenient for inference purposes since this means that we do not just have a point estimate of a parameter (the mean or mode of the posterior distribution) but we also automatically get an uncertainty estimates (the standard deviation of the posterior distribution).

Table 1.2: Posterior estimates of the circular means of the phase difference for the three conditions of the motor resonance data.

Condition	Mode	Mean	sd	LB HPD	UB HPD
explicit	42.70°	45.56°	11.67°	22.26°	67.99°
semi-implicit	21.08°	19.40°	18.36°	-18.27°	55.22°
implicit	37.22°	33.47°	17.77°	-2.25°	68.22°

For more background on Bayesian statistics see e.g. Gelman et al. (2014).

Summary statistics for the posterior distributions of the circular means for each condition are shown in Table 1.2. This table shows the posterior mean, mode, standard deviation (sd) and the lower and upper bound of the 95% highest posterior density interval (HPD). The standard deviation of a posterior is an estimate for the standard error of the parameter. The HPD interval is the smallest interval in which 95% of the posterior distribution is located. In terms of interpretation, it is different from a frequentist confidence interval since HPD intervals allow for probability statements. For example, if the 95% HPD interval for a parameter μ runs from 2 to 4 we can say that the probability that μ lies between 2 and 4 is 0.95.

HPD intervals can also be used to test whether a parameter is different from a certain value or whether two parameter estimates are different. In Table 1.2 we see that the HPD intervals of the circular means for the three conditions in the motor resonance data overlap. The circular mean of the phase difference is estimated at 47.70° (22.26°; 67.99°) for the explicit condition, 37.22° (-2.25°; 68.22°) for the implicit condition and 21.08° (-18.27°; 55.22°) for the semi-implicit condition. Because the HPD intervals of these estimates overlap, we conclude that there is not enough evidence to reject the null hypothesis that the circular means for the three conditions do not differ and that there is no effect of condition on the average phase difference. Note that the fact that no difference was found may be due a lack of power caused by the relatively small sample size (N = 14).

In addition to testing whether the circular means of the three conditions are different, the circular ANOVA also allows us to test whether there is an effect of condition on the circular variances of the phase differences. Table 1.3 shows summary statistics for the posterior distributions of the circular variance for each condition. As expected the estimated circular variance for the explicit condition is lowest. However, the variances of the three groups do not significantly differ; their HPD intervals overlap. We thus conclude that there is no evidence for an effect of condition on the variance of the phase difference. Note that a function to compute these variances has not yet been implemented in version 1.0.0 of `bpnreg`. It is however possible to get the MCMC estimates from the fit object and subsequently use Equation 3 from Kendall (1974) on the estimated mean vector for each of the groups to compute the variances.

Table 1.3: Posterior estimates of the circular variances of the phase difference for the three conditions of the motor resonance data.

Condition	Mode	Mean	sd	LB HPD	UB HPD
explicit	0.21	0.26	0.09	0.09	0.45
semi-implicit	0.37	0.44	0.13	0.19	0.36
implicit	0.36	0.42	0.13	0.18	0.68

1.4.3 Other approaches to circular ANOVA

In the previous section we have tested whether the average phase differences of the three conditions of the motor resonance data differ in the population using a Bayesian PN circular GLM. We can also do this using a frequentist ANOVA for circular data that tests the hypothesis $H_0 : \mu_{explicit} = \mu_{semi-implicit} = \mu_{implicit}$. One of such tests is the Watson-Williams test. This test can be performed using the function `watson.williams.test` in the `circular` package and is similar to an ANOVA for linear data interpretation-wise. Note that the Watson-Williams test falls within a different approach to modelling circular data, the intrinsic approach. In this approach we directly model the circular data instead of making use of a mathematical trick that allows us to model the data in bivariate space and then translate the results back to the circle. For simple models, such as the ANOVA, this modelling approach works fine. However, for more complex data structures we have a much larger choice of models in the embedding approach. For example, a disadvantage of the Watson-Williams test is that it does not allow for the addition of covariates and thus cannot estimate AN(C)OVA models. The PN circular GLM does allow for the addition of covariates.

As in ANOVA models for linear data, we have to meet a set of assumptions for the Watson-Williams test to be valid. Firstly, the samples from the different conditions are assumed to be von-Mises distributed. Like the projected normal distribution this is a distribution for circular data. It is unimodal with mean μ and concentration κ. Secondly, the samples are assumed to have the same circular variance. This assumption of homogeneity of variance is tested within the `watson.williams.test` function itself. For the motor resonance data this assumption was met. The assumption of von-Misesness can be tested using e.g. the Watson's goodness of fit test for the von Mises distribution. If we perform this test on the phase differences of the three subgroups we conclude that only the phase differences from the `semi.implicit` and the `implicit` condition are von-Mises distributed (H_0 is not rejected). This means that it is not completely valid to perform the Watson-Williams test on the motor resonance data.

For educational purposes however we do decide to conduct this test. Similar to the projected normal circular GLM we conclude that the average phase differences of the three conditions are not significantly different: $F(2, 39) = 1.02, p > 0.05$.

1.5 Mixed-Effects models with a circular outcome

An advantage of employing the embedding approach to circular data over the intrinsic approach is that it is easier to model more complex data, e.g. repeated measures data, since we can 'borrow' methods from the bivariate linear context. In this section we will introduce such a method: the circular mixed-effects model. We will first introduce a new dataset, the cognitive maps data, and give descriptive statistics. Then, we will shortly outline the theoretical background to the mixed-effects model and fit it to the cognitive maps data.

1.5.1 The cognitive maps data

The cognitive maps data is a subset of data from a study by Warren et al. (2017) on the geometry of humans' knowledge of navigation space. In their study Warren et al. (2017) among others conduct an experiment in which a total of 20 participants used virtual reality headsets to navigate through one of two virtual mazes. The navigation task consisted of walking from a start object to a target object. In a training phase they had learned to navigate between different pairs of start and target objects in one of two versions of the maze. The number of trials each participant completed in this training phase was recorded. In the test phase of the experiment participants first walked to a start object. When they had reached this object the maze disappeared and only a 'textured groundplane' of the maze remained visible. The participants then turned toward the location of the target object that they had remembered during the training phase and started to walk toward the target. The angular difference between the initial walking direction of a participant from the start object and the location of the target object, that is, the angular error, was recorded as an outcome variable in the experiment.

The type of maze is a between-subjects factor, participants either had to navigate through a 'Euclidean' maze or a 'non-Euclidean' maze. The Euclidean maze is the standard maze and is a maze just as we know it in the real world. The other version of the maze, the non-Euclidean maze, has exactly the same layout as the standard maze but it has virtual features that do not exist in reality. It namely contains wormholes by which participants can be 'teleported' from one place in the maze to another.

In the test phase of the experiment all participants had to complete 8 trials. In each of these trials participants had to walk to a specific target object. A within-subjects factor is the type of target object. Pairs of start and target objects were of two types: probe and standard. The probe objects were located near the entrance and exit of a wormhole in the non-Euclidean maze whereas the standard objects were located at some distance from the wormholes. For each of these two types of objects participants had to find 4 different targets resulting in a total of 8 trials per participant.

For this experiment we could be interested in the question whether the participants in the non-Euclidean maze make use of the wormholes when navigating to the target objects and whether this is true for both the probe and standard target

Table 1.4: Descriptives for the cognitive maps data with mean direction ($\bar{\theta}$) and mean resultant length ($\bar{R}$) of the angular error for each condition.

Maze	Trial.type	$\bar{\theta}$	$\bar{R}$
Euclidean	standard	-4.91°	0.89
	probe	4.46°	0.92
non-Euclidean	standard	-17.59°	0.78
	probe	37.34°	0.93

objects. Due to the design of the mazes the expected angular error was larger if a participant used the wormhole to walk to the target object in the non-Euclidean maze. We can thus use the angular error, our outcome variable, to differentiate between participants that used the wormhole and those that took another path to the target object. Additionally we can control for the amount of trials that a participant completed in the training phase.

1.5.2 Descriptive Statistics

The cognitive maps data is incorporated in the package `bpnreg` as the dataframe `Maps`. This dataframe has 160 rows; there are 20 subjects that each completed 8 trials. The data contains an index variable for the subject `Subject` (N = 20) and trial number `Trial.no` (n = 8). It also includes variables indicating the type of maze `Maze`, a between-subjects factor, and type of trial `Trial.type`, a within-subjects factor. The variable `Learn` indicates the amount of learning trials completed. `L.c` is a centered version of this variable. The angular error is contained in the variables `Error` and `Error.rad` in degrees and radians respectively. Descriptives for this data are shown in Table 1.4. Note that we averaged over subjects and the trials of each type. The circular mean of the angular error for the standard trials in the Euclidean maze is thus an average over 10 participants and 4 trials. We see that the average angular errors, $\hat{\theta}$, for the non-Euclidean maze deviate more from 0° (direction of the target object) than for the Euclidean maze.

1.5.3 Fitting a mixed-effects model to the cognitive maps data

In this section we will first introduce a circular mixed effects model and fit this model to the cognitive maps data. Next we discuss the output produced by the `bpnreg` package. We will discuss the interpretation of fixed and random effects and model fit.

1.5.3.1 The embedding approach for mixed-effects models

The circular mixed-effects model from the package `bpnreg` is also based on the embedding approach to circular data. The basic idea behind this approach is

the same as outlined before. In a real dataset we have a set of outcome vectors $\boldsymbol{u}_{ij}$, one for each measurement j within a higher level observation i. We however estimate a model to the underlying bivariate data $\boldsymbol{y}_{ij}$. The Bayesian method used in the package `bpnreg` for estimating circular mixed-effects models is outlined in Nuñez-Antonio & Gutiérrez-Peña (2014).

For the cognitive maps data with $i = 1, \ldots, 20$ individuals and $j = 1, \ldots, 8$ measurements per individual we fit a mixed-effects model to investigate the influence of the type of Maze, type of trial and amount of learning trials on the angular error. The prediction for the mean vector in this model, $\boldsymbol{\mu}$, is specified as follows:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^{I} \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} \beta_0^I + \beta_1^I \text{Maze}_i + \beta_2^I \text{Trial.type}_{ij} + \beta_3^I \text{L.c}_i + b_{0i}^I \\ \beta_0^{II} + \beta_1^{II} \text{Maze}_i + \beta_2^{II} \text{Trial.type}_{ij} + \beta_3^{II} \text{L.c}_i + b_{0i}^{II} \end{pmatrix}, \tag{1.3}$$

where the variables `Maze` and `Trial.type` are dummy variables, β_0^I and β_0^{II} are the fixed intercepts, b_{0i}^I and b_{0i}^{II} are the random intercepts and β_1^I, β_2^I, β_3^I, β_1^{II}, β_2^{II} and β_3^{II} are the fixed regression coefficients of the model. Note that in this model we take the Euclidean maze and standard trials as reference conditions.

The interpretation problems caused by the two component structure in (1.3) is of a similar nature as the one in the GLM model. Cremers, Pennings, Mainhard, & Klugkist (2019) introduce new tools that solve the interpretation of circular effects in PN mixed-effects models. In this tutorial we will also use these tools.

1.5.3.2 Fitting the model

To fit the model in (1.3) we use the `bpnme()` function from the package `bpnreg`. We also need to specify some parameters for the MCMC sampler that estimates the model. We specify the output iterations (10000), the amount of burn-in (1000) and how many iterations we want to keep (`n.lag = 3`). Convergence was checked in the same manner as for the ANOVA model in the previous section and was reached using the settings for the MCMC algorithm we just specified.

```
fit.Maps <- bpnme(pred.I = Error.rad ~ Maze + Trial.type + L.c +
                  (1|Subject),
                  data = Maps,
                  its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Note that the syntax for the model specification in this function is similar to that of the package `lme4` for fitting (non-circular) mixed-effects models.

1.5.3.3 Fixed Effects

Next we investigate the coefficients of the fixed effects for this model. First we show results for the categorical variables type of maze (`Maze`) and type of trial (`Trial.type`).

Table 1.5: Posterior estimates of the circular mean of the angular error for each condition.

Maze	Trial.type	mode	mean	sd	LB HPD	UB HPD
Euclidean	standard	-12.97°	-13.48°	3.9°	-21.42°	-6.06°
	probe	11.38°	11.78°	3.29°	5.26°	18.30°
non-Euclidean	standard	-1.42°	-2.04°	6.68°	-15.75°	10.49°
	probe	31.04°	30.37°	4.31°	22.03°	38.92°

Table 1.5 shows summary statistics of the posterior of the average angular error for each of the categories. Note that because there is a continuous predictor in the model the posterior estimates represent a marginal effect, they are the effect for an individual with a 0 score on the continuous predictor `L.c`. Because we centered this predictor this means that this is the effect for an individual that has completed an average number of training trials.

By looking at the 95% HPD intervals of the angular errors in Table 1.5 we can test whether the type of maze and type of trial on average has an influence on the angular error and thus whether participants make use of the wormhole. For the standard trials we see that the HPD intervals of the angular error in the Euclidean and non-Euclidean overlaps and that thus the angular error is not different. This means that in the standard trials the participants on average did not make use of the wormholes in the non-Euclidean maze. For the probe trials however, the HPD intervals of the Euclidean and non-Euclidean do not overlap and thus the angular error is different. This means that in the probe trials, the participants on average did make use of the wormholes in the non-Euclidean maze.

For the continuous variable `L.c` we get a set of parameters, b_c, SAM and AS, describing its effect on the circle. How these parameters are computed is described in Cremers et al. (2018b) and Cremers et al. (2019). In this paper we will only focus on how to interpret them.

In Figure 1.7 a circular regression line for the effect of a predictor x on the circular outcome is shown. Because the outcome variable is measured on a circular scale, the slope of this line (the effect of x) is not constant but different for different x values. The regression line can be described using the three circular coefficients b_c, SAM and AS. The coefficient b_c represents the slope of the circular regression line at the inflection point (the square in Figure 1.7). However, this may not be a representative effect for each dataset as the inflection point can lie in the extremes of the data (as in Figure 1.7) or even completely outside the range of the predictor x. Therefore two additional circular coefficients were developed by Cremers et al. (2018b), the slope at the mean SAM and the average slope AS. The coefficient SAM represents the slope of the circular regression line at the average of the predictor ($\bar{x}$) and the coefficients AS represents the average slope over all values of x.

For the effect of L.c on the angular error in the cognitive maps data, the HPD intervals for all three circular coefficients, b_c, SAM and AS include 0 (see Table 1.6).

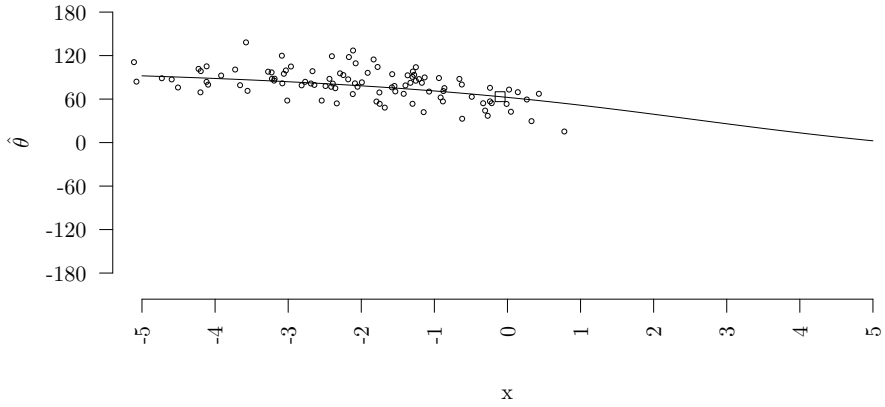

Figure 1.7: Predicted circular regression line for the relation between a linear predictor x and a predicted circular outcome θ together with the original datapoints. The square indicates the inflection point of the regression line.

Table 1.6: Posterior estimates of the coefficients of the effect of L.c on the angular error.

Coefficient	mode	mean	sd	LB HPD	UB HPD
b_c	-0.89°	-0.21°	1.73°	-2.84°	2.55°
SAM	0.58°	-0.84°	90.80°	-11.51°	12.73°
AS	-0.63°	-1.11°	92.44°	-13.17°	13.14°

Thus we do not find evidence that at the inflection point, at the average predictor value and on average the number of training trials (L.c) influences the angular error. Note that there not being evidence for influence is a good thing, since it indicates that the training phase of the experiment worked to get all participants at the same level. We do however need to be wary of making all to strong conclusions, we are essentially trying to find evidence for a null-hypothesis with a small sample (n = 20). If the sample had been larger we would have had more power to reject the hypothesis, possibly resulting in the opposite conclusion. For educational purposes we continue to give the interpretation of the coefficients. The SAM is interpreted as follows: at the average `L.c`, for a 1 unit increase in `L.c` the angular error increases with 0.58 degrees. The AS can be interpreted as: on average, for a 1 unit increase in `L.c` the angular error decreases with 0.63 degrees. The b_c can be interpreted as: at the inflection point, for a 1 unit increase in `L.c` the angular error decreases with 0.89 degrees.

1.5.3.4 Random Effects

In mixed-effects models we are also interested in evaluating the variance of the random effects. In the model for the cognitive maps data we included a random intercept. This means that we estimate a separate intercept for each participant. How to compute random effect variances on the circle is outlined in Cremers et al. (2019). For the cognitive maps data the posterior mode of the intercept variance on the circle is estimated at $3.5 * 10^{-5}$ and its HPD interval is $(4.2 * 10^{-6}; 1.4 * 10^{-3})$. This variance is very low which means that the participants do not differ a lot in their individual intercept estimates. Note that this is not necessarily problematic. In some cases we are not interested in the variances of the random effects but simply want to fit a mixed-effects model because we have within factors, such as `Trial.type`, that cannot be properly incorporated in a standard regression model.

1.5.3.5 Model Comparison

When fitting mixed-effects (or multilevel) models we often fit a set of nested models to our data and follow a model building strategy (Hox, 2002). We do this in case we have no specific model in mind that we want to test and want to explore the individual contributions of variables or groups of variables to the model. Such a model building strategy can be done top-down, starting with the most complex model, or bottom-up, starting with the simplest model. Here we use a bottom-up strategy and start with the so called intercept-only model, a model containing only a fixed and random intercept:

```
fit.MapsIO <- bpnme(pred.I = Error.rad ~ (1|Subject),
                    data = Maps,
                    its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

We then update this model with fixed effects for the predictors at the lowest level (within-subjects factors), in this case `Trial.type`. We do this to check whether

Table 1.7: Model fit criteria for several models fit to the cognitive maps data.

Criterion	Intercept-only	`Trial.type`	`Trial.type` `Maze`	`Trial.type` `Maze` `L.c`
DIC	304.61	267.91	253.97	257.94
DIC_{alt}	324.33	286.97	257.14	260.78
$WAIC_1$	308.41	271.61	255.00	258.41
$WAIC_2$	308.43	271.77	255.40	259.02

the set of predictors improved the fit of the model and can explain a part of the random intercept variance from the intercept-only model.

```
fit.Maps1p <- bpnme(pred.I = Error.rad ~ Trial.type + (1|Subject),
                    data = Maps,
                    its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

We then add fixed effects for the predictors at the higher level (between-subjects factors), in this case `Maze` and `L.c`. Again we do this to check whether they improve the fit of the model and whether they can explain a part of the random intercept variance.

```
fit.Maps <- bpnme(pred.I = Error.rad ~ Maze + Trial.type + L.c +
                  (1|Subject),
                  data = Maps,
                  its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Because we have already seen that the effect of `L.c` was not different from 0 we also fit the model with only the `Maze` and `Trial.type` predictors.

```
fit.Maps2 <- bpnme(pred.I = Error.rad ~ Maze + Trial.type +
                   (1|Subject),
                   data = Maps,
                   its = 10000, burn = 1000, n.lag = 3, seed = 101)
```

Additional steps, such as adding random slopes for first level predictors and cross-level interactions, can be taken. In this paper we will however restrict the analysis to the previous three models.

Model fit

To assess the fit of the models we look at 4 different model fit criteria: two version of the deviance information criterion (DIC and DIC_{alt}) and two versions of the Watanabe-Akaike information criterion ($WAIC_1$ and $WAIC_2$). We choose these four criteria because they are specifically useful in Bayesian models where MCMC methods have been used to estimate the parameters. All four criteria have a fit part consisting of a measure based on the loglikelihood and include a penalty in the form of an effective number of parameters. For all criteria lower values indicate better fit. Gelman et al. (2014) describes how to compute these criteria. Table 1.7 shows the results of these criteria for 4 different models.

In the results for the example we see that the fit improves in all 4 model diagnostics for each model except for the last one. This means that the predictor `Trial.type` improves the fit of the model over the intercept-only model and that the predictors `Maze` and `Trial.type` together improve the fit of the model over the model with only the `Trial.type` predictor. Because the variable `L.c` had no effect it is as expected that this predictor does not improve the fit of the most right model over the model with the `Maze` and `Trial.type` predictors. We conclude that the model with the predictors `Trial.type` and `Maze` fits best.

Explained variance

Apart from information about whether adding predictors improves the fit of the model we are also interested in whether these predictors explain a part of the random effect variances. For the cognitive maps data we are interested in whether the `Maze` and `Trial.type` predictors explain a part of the variance in individual intercepts. To assess this we compare the posterior estimates of the circular random intercept for the intercept-only model and the model with the `Maze` and `Trial.type` predictors.

The posterior mode of the intercept variance in the intercept-only model equals $6.61 * 10^{-5}(8.20 * 10^{-6}; 3.62 * 10^{-3})$. This means that there is almost no random intercept variance. The posterior mode of the circular variance is very close to 0. This also means that there is hardly any intercept variance that the `Maze` and `Trial.type` predictors can explain. For illustrative purposes however we continue to assess the intercept variance in the model with the `Maze` and `Trial.type` predictors. The posterior mode of the intercept variance in the model with `Maze` and `Trial.type` equals $3.25 * 10^{-5}(4.40 * 10^{-6}; 1.59 * 10^{-3})$. As expected, there is hardly any change in estimates for the variance in the model with `Maze` and `Trial.type` compared to the intercept-only model. Furthermore, their HPD intervals have a very large overlap. We thus conclude that the variables `Maze` and `Trial.type` did not explain any variance in the random intercepts.

1.6 Concluding remarks

In this paper we have given a tutorial for researchers in cognitive psychology on how to analyse circular data using the package `bpnreg`. We have covered data inspection in Section 1.3, the fitting of a Bayesian circular GLM in Section 1.4 and the fitting of a Bayesian mixed-effects model in Section 1.5. We have also given a short introduction into the theoretical background of these models in Section 1.4.1 and 1.5.3.1.

Apart from the embedding approach to circular data, as used in this tutorial, there are two other approaches to the analysis of circular data. In the wrapping approach the data on the circle is assumed to have originated from wrapping a univariate distribution on the real line onto the circle. In the intrinsic approach distributions, such as the von Mises distribution, are directly defined on the circle. For both approaches models have been described in the literature (Fisher & Lee, 1992; Gill & Hangartner, 2010; Lagona, 2016; Mulder & Klugkist, 2017; Ravindran & Ghosh,

2011). The regression model using the intrinsic approach from Fisher & Lee (1992) is a frequentist method and is implemented in the package `circular` and the circular general linear model from Mulder & Klugkist (2017) is a Bayesian method which is implemented in the package `circglmmbayes`. For neither approach however a detailed tutorial describing how to analyze circular data using the functions from their package has been written thus far. Furthermore, the PN approach to circular modelling has the additional advantage that it is relatively easy to fit more complex models, e.g. the mixed-effects model in this tutorial.

Chapter 2

Assessing a Bayesian embedding approach to circular regression models

by J. Cremers, M.T. Mainhard & I. Klugkist[1]

2.1 Introduction

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 et al. (2011a) 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 schoolyear. 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

[1]Published as Cremers, J., Mainhard, M.T. & Klugkist, I. (2018). Assessing a Bayesian Embedding Approach to Circular Regression Models. *Methodology, 14*(2), 69-81. doi: 10.1027/1614-2241/a000147.

Author contributions: JC and IK designed the study, JC designed and performed the simulation study and analyzed, processed and interpreted the results of the empirical example with feedback from IK, MTM provided the data for the empirical example, JC wrote the paper and IK and MTM provided feedback on the written work.

several octants, e.g. whether a teacher scored between 0° and 45°, 45° and 90°, etc. These octants or the two dimensions Agency and Communion were then analysed separately. However, according to Gurtman (2009) 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."* (p. 2). Wright, Pincus, Conroy, & 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 2.1 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 measuring instruments can be considered circular data.

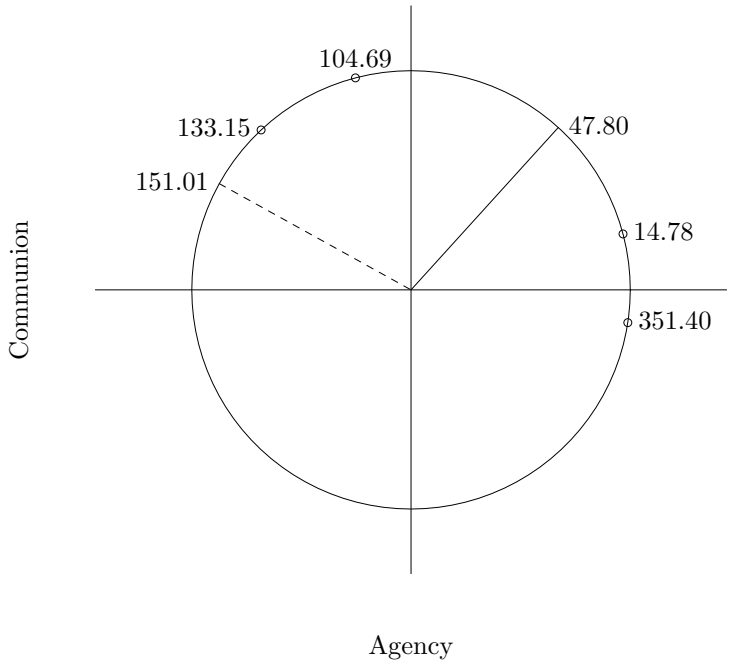

Figure 2.1: Interpersonal circumplex with axes Agency (vertical) and Communion (horizontal). Data points indicate the score of a teacher on the circumplex measured in degrees. The dashed line represents the direction of the linear mean and the dotted line represents the direction of the circular mean.

Circular data require 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 2.1, 151.01° is not the average direction of the scores. The correct 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 & Hangartner (2010), Lagona (2016) and Mulder & 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 & Ghosh (2011) provide an MCMC method for wrapped distributions. Nuñez-Antonio et al. (2011) and Wang & 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 are 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 & 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. 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 Section 2.2, the model for the Bayesian embedding approach to circular regression

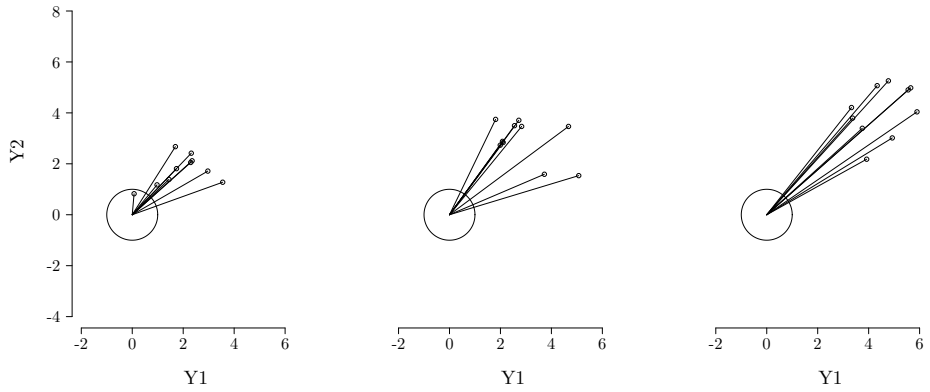

Figure 2.2: Three sets of bivariate normal data projected on the unit circle. From left to right mean vectors, $\boldsymbol{\mu}$ are: (1.5,1.5), (3,3), (4.5,4.5). The lengths of the lines drawn from each datapoint to the origin represent $\boldsymbol{r}$ and their intersection with the circle corresponds to individual values of the circular outcome vector, $\boldsymbol{\theta}$.

is introduced. Sections 2.3 and 2.4 contain the methodology and results of the simulation studies for one and multiple predictors. These are followed by Section 2.5 which presents the interpretation and results from the model fit on the Agency-Communion data. The paper ends with a discussion of results.

2.2 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.

2.2.1 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 works, imagine that we have one bivariate normal outcome variable, $Y \sim N_2(\boldsymbol{\mu}, \boldsymbol{I})$. Figure 2.2 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 $\boldsymbol{r} = (r_1 \dots r_N)$ where $i = 1, \dots, N$ and N is the sample size. Their intersections with the unit circle can be interpreted as datapoints of the circular outcome vector $\boldsymbol{\theta} = (\theta_1 \dots \theta_N)$. In Figure 2.2 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 $\boldsymbol{\theta}$ and a vector with distances to the origin $\mathbf{r}$. However, when we start with circular data the process is reversed. Imagine one circular outcome variable measured in angles, $\boldsymbol{\theta}$. 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 $\boldsymbol{r}$. The method that treats $\boldsymbol{r}$ as a latent variable and makes inference possible is introduced in Section 2.2.3.

2.2.2 Circular Regression

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

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

where $0 < \theta \leq 2\pi$ and $\boldsymbol{\mu}$ is the mean vector of the underlying bivariate normal distribution with identity variance-covariance matrix $\boldsymbol{I}$. Furthermore, $\mathbf{u}$ is the vector $(\cos\theta, \sin\theta)^t$, $\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}$ where $\boldsymbol{B} = [\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II}]$, $\boldsymbol{x}$ 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 $\boldsymbol{B}$, $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$, are vectors with regression coefficients and an intercept. Formally this notation is only correct when the predictors in $\boldsymbol{x}$ are equal for both components of $\boldsymbol{\mu}$. 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 $\boldsymbol{x}$, one for each component of $\boldsymbol{\mu}$. Lastly, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution.

2.2.2.1 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 schoolyear (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 2.1.

Table 2.1: 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. Note that TEX and EV were already centered in the original data.

	Mean/($\bar{\theta}$)	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(SCSSP)	0.34	0.62	-0.93	1.00	Continuous
cos(SCSSP)	-0.21	0.69	-0.99	0.99	Continuous

The resulting regression equations are:

$$\mu_1 = \beta_0^1 + \beta_1^1 \cos(\text{ACSSP}) + \beta_2^1 \sin(\text{ACSSP}) + \beta_3^1 \text{TEX}$$
$$\mu_2 = \beta_0^2 + \beta_1^2 \cos(\text{ACSSP}) + \beta_2^2 \sin(\text{ACSSP}) + \beta_3^2 \text{TEX} + \beta_4^2 \text{EV}$$

where μ_1 and μ_2 , 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, e.g. 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 Section 2.5.

2.2.3 Bayesian Estimation

In our paper the modelling and notation is like that of Section 4 in 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 $\boldsymbol{B}$:

$$N(\boldsymbol{\beta}^j | \boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j) \, \forall \, j = I, II,$$

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 $\boldsymbol{y}_i$, $\boldsymbol{y}_i \sim N_2(\cdot | \boldsymbol{\mu}_i = \boldsymbol{B}^t \boldsymbol{x}_i, I)$ we obtain the following posterior:

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

where $\boldsymbol{D} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ is a sample of independent bivariate normal observations, $\boldsymbol{\mu}_F^j = (\boldsymbol{\Lambda}_F^j)^{-1}(\boldsymbol{\Lambda}_0^j \boldsymbol{\beta}_0^j + (\boldsymbol{X}^j)^t \boldsymbol{y}^j)$, $\boldsymbol{\Lambda}_F^j = \boldsymbol{\Lambda}_0^j + (\boldsymbol{X}^j)^t \boldsymbol{X}^j$ 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 Θ, the observed

circular outcome, is defined by:

$$f(\theta, r|\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}) = (2\pi)^{-1}\exp\{-.5 \mid\mid \boldsymbol{\mu} \mid\mid^2\}\exp\{-.5[r^2 - 2r(\boldsymbol{u}^t\boldsymbol{\mu})]\} \mid J \mid,$$

where $|J| = r$ is the Jacobian of the transformation $\boldsymbol{y} \mapsto (\theta, r)$, where $\boldsymbol{y}$ is bivariate normal, and $\mathbf{u} = (\cos\,\theta, \sin\,\theta)^t$. 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

$$f(\boldsymbol{\beta}^j|\theta_1, \ldots, \theta_n, \boldsymbol{r}) = N(\cdot\,|\boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j)\;\forall\; j = 1, 2,$$

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

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

   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, & Van der Woerd (2017) and adapted for the regression situation (MCMC method 2). The two methods for sampling r_i are outlined in the appendix.
4. Steps 2 and 3 are repeated for a specified amount of iterations.

2.3 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 analysed. This amount 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 (2017). A data archive including all code for the simulations can be found in the supplementary material at https://econtent.hogrefe.com/doi/10.1027/1614-2241/a000147.

Convergence of parameter estimates in separate designs (populations) was checked for both sampling methods. First, a simulation was run for 3000 iterations of which 750 were considered burn-in. Then convergence was checked by looking at traceplots, as shown in 2.3, 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 (3000, 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 is

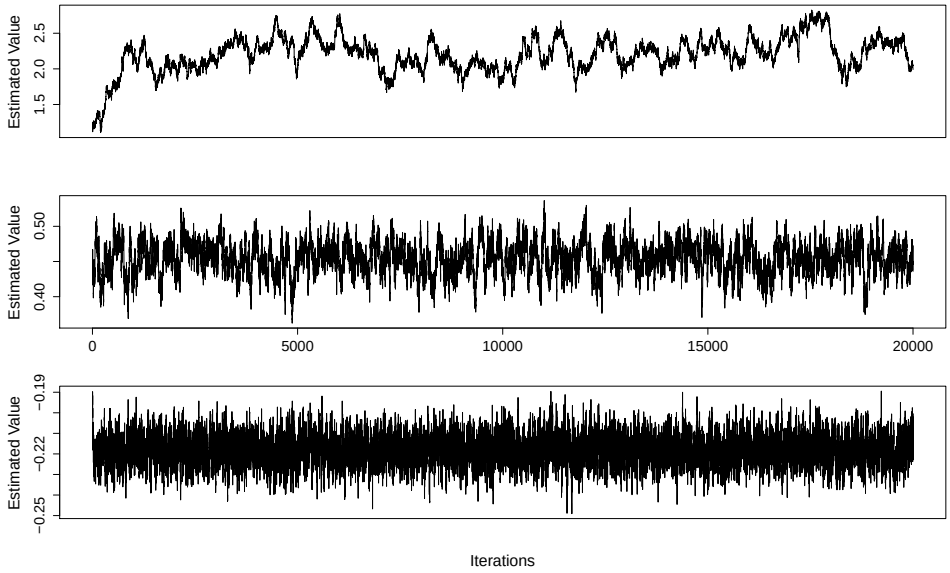

Figure 2.3: Example traceplots of the estimate for the regression coefficient β_2^I in the simulation studies showing no, acceptable and good convergence.

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 ($\mid Population - Estimated \mid / \mid Population \mid$). For parameters with population value zero we do not report the relative bias but the absolute bias ($\mid Population - Estimated \mid$). 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.

2.3.1 One Linear Predictor

2.3.1.1 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 (2.1). 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.

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 2.2 and 2.3 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 2.2). 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.

$$
\begin{aligned}
\theta = \text{atan2}\left(y^{II},\, y^{I}\right) &= \arctan\left(\frac{y^{II}}{y^{I}}\right) && \text{if} \quad y^{I} > 0 \\
&= \arctan\left(\frac{y^{II}}{y^{I}}\right) + \pi && \text{if} \quad y^{I} < 0 \qquad y^{II} \geq 0 \\
&= \arctan\left(\frac{y^{II}}{y^{I}}\right) - \pi && \text{if} \quad y^{I} < 0 \qquad y^{II} < 0 \\
&= \frac{\pi}{2} && \text{if} \quad y^{I} = 0 \qquad y^{II} > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad y^{I} = 0 \qquad y^{II} < 0 \\
&= \text{undefined} && \text{if} \quad y^{I} = 0 \qquad y^{II} = 0
\end{aligned} \tag{2.1}
$$

2.3.1.2 Results

Results are shown in Table 2.2 for the sampler with MH step and Table 2.3 for the slice sampler. The (relative) bias was rounded to two decimals. In Table 2.2 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 0. 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. We do see that, especially in designs with μ_x further from 0, the relation between N and coverage is less strong. In Table 2.3 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

Table 2.2: (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. A * or ** indicate that the amount of iterations was set to 20,000, 40,000 respectively.

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	2	-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									*

Table 2.3: (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. A *, ** or *** indicate that the amount of iterations was set to 20,000, 40,000 or 80,000 respectively.

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	2	-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									***

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.

2.3.2 One Circular Predictor

2.3.2.1 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 using (2.1). The two vectors $\boldsymbol{x}_{,2}$ and $\boldsymbol{x}_{,3}$ are the cosine and sine components of a vector sampled from VM($\mu_{circ,x}, \kappa_x$); a von Mises distribution with circular mean $\mu_{circ,x}$ 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 homogeneous the data (Fisher, 1995).

Parameters that were varied are: the sample size, N, the slopes of the cosine β_1^I and sine β_2^I component of the circular predictor in the regression equation predicting the first component of the outcome, the slopes β_1^{II}, β_2^{II} for the regression equation predicting the second component of the outcome, and κ_x. Tables 2.4 and 2.5 show the chosen values. Chosen N, β_0^I and β_0^{II} are the same as in Section 3.1.1. and $\mu_{circ,x} = 0$ 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.

2.3.2.2 Results

Results are shown in Table 2.4 for the sampler with MH step and Table 2.5 for the slice sampler. The (relative) bias was rounded to two decimals. In Table 2.4 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 2.5 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 worse performance, small N and large κ_x, the MH-sampler outperforms the slice sampler. Patterns for MCT are like those described in Section 2.3.1.2.

Table 2.4: (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. A * or **indicate that the amount of iterations was set to 20,000, 40,000 respectively.

Population values						(relative) Bias						Coverage						MCT	
β_1^I	β_1^{II}	β_2^I	β_2^{II}	κ_x	N	β_0^I	β_1^I	β_2^I	β_0^{II}	β_1^{II}	β_2^{II}	β_0^I	β_1^I	β_2^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	2	2	2	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	*

Table 2.5: (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. A *, ** or *** indicate that the amount of iterations was set to 20,000, 40,000 or 80,000 respectively.

Population values						(relative) Bias						Coverage						MCT	
β_1^I	β_1^{II}	β_2^I	β_2^{II}	κ_x	N	β_0^I	β_1^I	β_2^I	β_0^{II}	β_1^{II}	β_2^{II}	β_0^I	β_1^I	β_2^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	010	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	2	2	2	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	*

2.3.3 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 (p. 92, Gelman et al., 2014). We have seen that (relative) bias increases and coverage is worse in designs with high κ_x and μ_x further away from 0. 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.

2.4 Simulations for Multiple Predictor Models

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

2.5 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 5000 iterations the results in Table 2.6 were obtained.

The posterior means of the coefficients from the third column of Table 2.6 inform us about the linear relations between predictors and the components Agency and Communion as shown in Figures 2.4a, 2.4b, 2.4c and 2.4d. 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, e.g. 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 2.6 show the lower and upper bounds for the 95% credible intervals. Only for the coefficients Communion Self Perception for the Communion component and Teacher Experience for the Agency component the 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.

Table 2.6: 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 SCSSP	0.54	0.07	1.02
	Agency SCSSP	-0.13	-0.68	0.43
	TEX	-0.02	-0.06	0.02
Agency	Intercept	1.67	1.15	2.22
	Communion SCSSP	0.29	-0.31	0.90
	Agency SCSSP	0.50	-0.20	1.192
	TEX	0.10	0.04	0.15
	EV	0.34	-0.01	0.71

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 2.4e, 2.4f, 2.4g and 2.4h 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 (2.1) on the two linear outcomes. In the circular relation plots of Figure 2.4 the slope is not constant indicating that the relation between outcome and predictor is not linear. E.g. in Figure 2.4g 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 °.

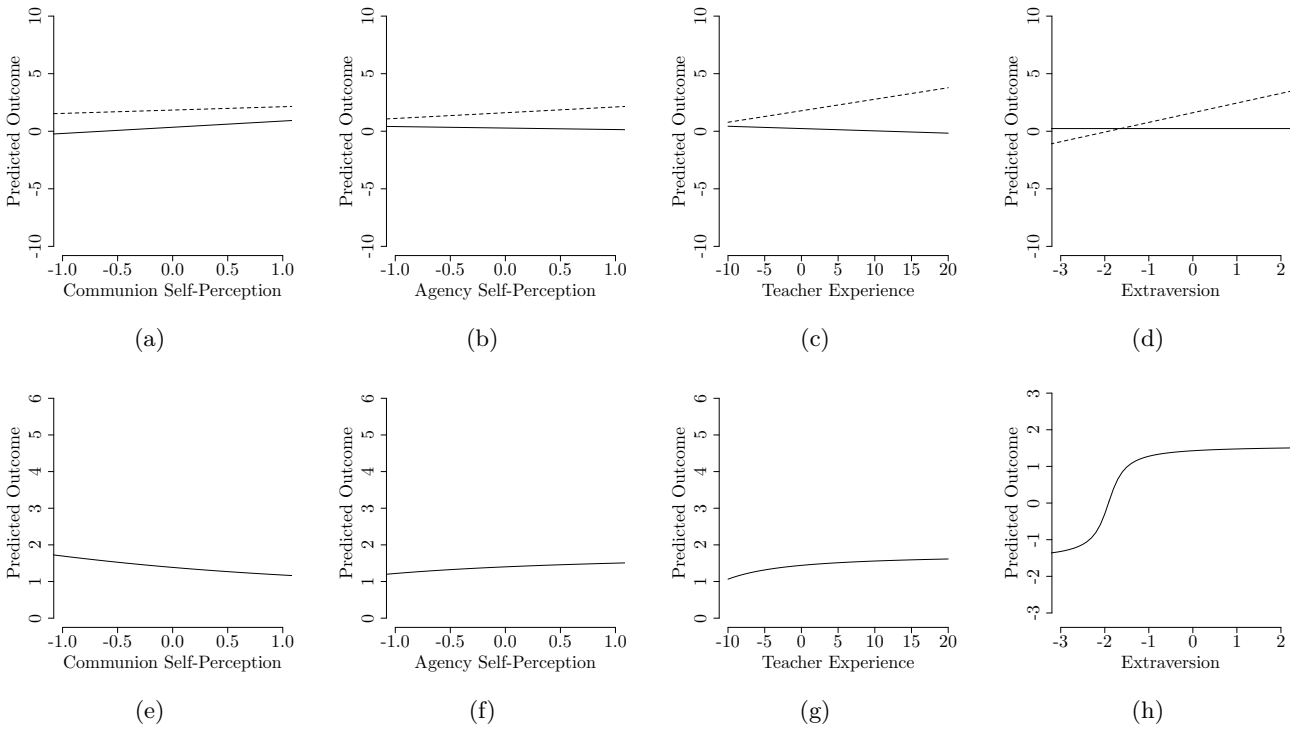

Figure 2.4: Predicted outcome for changes in 1 predictor. Upper subfigures are linear outcomes for the separate components (dotted=Agency, continuous=Communion). Lower subfigures are circular outcomes (measured in radians).

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 et al. (2005). Even though Brekelmans et al. (2005) subdivide the circumplex and therefore do not use circular statistics to analyse their data we can infer from their 8 profile types that teachers with less experience tend to score higher on the Communion component 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 come 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). Modelling circumplex data in a circular setting 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."* (p. 2, Gurtman, 2009).

2.6 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 example data does not (Mainhard et al., 2011a), as does other research on circumplex data. Instead, it considers statistical models for both dimensions separately (Mainhard et al., 2011b; Wubbels et al., 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, Brekelmans, 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, & 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 & Gutiérrez-Peña (2014) and Hernandez-Stumpfhauser, Breidt, & Van der Woerd (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 et al. (2018b). In their work (2.1) is reparameterized 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 3000 iterations. The designs that took longer to converge were those with a high κ_x, μ_x further from 0 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 & 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.

Chapter 3

Circular interpretation of regression coefficients

by J. Cremers, K. Mulder & I. Klugkist[1]

3.1 Introduction

Circular models are models for data with a circular outcome variable. A circular variable measures a direction in two dimensional space in degrees or radians and requires analysis methods that are different from standard methods for linear data. In the field of psychology circular data can be found in research on the visual perception of space (Matsushima et al., 2014), moving room experiments (Stoffregen et al., 2004), visual working memory experiments (Heyes et al., 2016), movement synchronization (Kirschner & Tomasello, 2009; Ouwehand & Peper, 2015) and cognitive maps (Brunyé et al., 2015). Measurements on the interpersonal circumplex can also be regarded as circular data (König, Onnen, Karl, Rosner, & Butollo, 2016; Santos, Vandenberghe, & Tavares, 2015; Wright et al., 2009; Zilcha-Mano et al., 2015). In general circular variables measure directions, are circumplex scales, or are a measure of periodic (weekly, daily, hourly, etc.) patterns. Wright et al. (2009) outline how circular data is used in research using circumplex measures. They show how to compute a circular mean and test for differences between groups. However, circular models could be used much more effectively in psychological science. Take for instance a study by Locke et al. (2016) on interpersonal characteristics of depressed outpatients. Circumplex profiles of patients were made and compared

[1]Published as Cremers, J. Mulder, K. & Klugkist, I. (2018) Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology, 71*(1), 77-95. doi:10.1111/bmsp.12108.

Author contributions: JC and IK designed the study, KM and JC developed the new measures, JC designed and performed the simulation study and analyzed, processed and interpreted the results of the empirical dataset with feedback from IK, JC wrote the paper and IK and KM gave feedback on the written work.

with those of normative samples. Although the authors do use circular means and variances and do compare groups, a more elaborate circular model, would allow for the simultaneous evaluation of multiple predictors. It would then be possible to assess whether there still is a difference in circumplex profile between depressed and normative samples when also accounting for other variables such as gender, age or a measure of depression severity.

Analyzing circular data is not straightforward and special methods are required. One of the approaches to circular data is the so called embedding approach. In this approach projected distributions are used. Projected distributions are bivariate distributions defined in $\mathbb{R}^2$ and projected onto the circle. Other approaches to modelling circular data are the wrapping and intrinsic approach (Mardia & Jupp, 2000). These approaches are based on respectively wrapping distributions defined in $\mathbb{R}$ onto the circle and using distributions defined on the circle itself, e.g. the von Mises distribution. A paper by Rivest et al. (2015) provides an overview of some of the circular regression models in the literature. Although various types of models for the three approaches have been described, we will only consider regression models of the embedding approach in this paper.

Nuñez-Antonio et al. (2011) and Wang & Gelfand (2013) have developed Bayesian methods for regression based on the projected normal and general projected normal distribution. These are both based on Presnell et al. (1998) who first used a projected normal distribution to analyze circular regression models. In the embedding approach it is relatively easy to fit more complex models because the distributions that are used are based on distributions in $\mathbb{R}^2$. Indeed, in the literature more complex regression type models have been introduced, including random effects models (Nuñez-Antonio & Gutiérrez-Peña, 2014) and spatial and spatio-temporal models (Mastrantonio, Lasinio, & Gelfand, 2016; Wang & Gelfand, 2014). We will however limit ourselves to the simpler multiple regression model.

Although the embedding approach is flexible with regard to model fitting, the interpretation of the effects of predictors in these models is not easy. According to Maruotti (2016) this is the major drawback of models based on projected distributions. Currently, when using the embedding approach we obtain two regression coefficients for each variable in the model. This is a result of using an underlying bivariate distribution. The two coefficients are however not interpretable as an effect on the circle. Additionally, they do not allow us to directly distinguish between an effect on the mean and an effect on the spread of the circular outcome. In the example of the study by Locke et al. (2016) we would not be able to distinguish between the differences between depressed and normative samples in average profiles and differences in the within group homogeneity of the profiles. There is no literature yet dealing with this problem. We will therefore introduce new interpretation tools that combine the bivariate coefficients into one circular coefficient. The new tools allow us to assess whether there is an effect of a predictor on the mean of the circular outcome and how large this effect is.

In Section 3.2 we describe the embedding approach. We distinguish between effects on the circular mean and spread in Section 3.3. Next, in Section 3.4, we combine two bivariate coefficients into one circular coefficient and introduce new tools for interpretation. These tools are then applied to the example dataset in Section 3.5.

Lastly, we perform a simulation study to assess the new tools in Section 3.6. The paper concludes with a discussion in Section 3.7.

3.2 Embedding approach

In this section we introduce the embedding approach and projected normal distribution. Both have been introduced and described previously (Nunez-Antonio & Gutierrez-Pena, 2005; Presnell et al., 1998). Therefore, a large part of this section focuses on the interpretation of the estimates from a projected normal regression model.

3.2.1 Projected Normal Distribution

There are several representations in which one can specify circular data: angles, polar coordinates or unit vectors in $\mathbb{R}^2$. In this paper we will refer to an observation of a circular outcome variable in angles as θ_i, where $i = 1, \ldots, n$ and to its unit vector representation as $\boldsymbol{u}_i$. We assume that the outcome variable can be represented by an unobserved column vector $\boldsymbol{y}_i$ in $\mathbb{R}^2$ as follows:

$$\boldsymbol{u}_i = \frac{\boldsymbol{y}_i}{r_i}.$$

where r_i is the length of the vector $\boldsymbol{y}_i$. The angular outcome variable originates from a projection onto the circle of a vector in $\mathbb{R}^2$. If we assume that the underlying $\boldsymbol{y}_i$ originate from a bivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{I}$, it follows from (3.1) that θ has a Projected Normal (PN) distribution with density function:

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

where θ is the circular outcome variable measured in radians $-\pi \leq \theta < \pi$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$ is the mean vector of the distribution, the variance-covariance matrix $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$. The terms $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution respectively. An identity variance-covariance matrix is chosen to identify the model. Due to this configuration the PN distribution is always unimodal and symmetric. Another configuration can be found in Wang & Gelfand (2013) who use different constraints on the matrix resulting in the general PN distribution that can also take a skewed and multimodal shape.

To fit a model on the mean of the projected normal distribution $\boldsymbol{\mu}$ we need r_i to obtain the unobserved bivariate normal vectors $\boldsymbol{y}_i$. The estimation of r_i is a missing data problem that is solved by treating the unobserved lengths r_i as latent or auxiliary variables in the model. We can then use existing techniques such as the EM algorithm (Presnell et al., 1998) or Bayesian methods (Nunez-Antonio & Gutierrez-Pena, 2005), to obtain inference on the $\boldsymbol{y}_i$.

3.2.2 Regression

In regression we have independent observations of a vector of linear predictors $\boldsymbol{x}_i$ for each individual $i = 1, \ldots, n$. The model for one of the bivariate normal vectors $\boldsymbol{y}_i$ has mean structure $\boldsymbol{\mu}_i = \boldsymbol{B}^t \boldsymbol{x}_i$, where $\boldsymbol{B} = [\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II}]$ and each $\boldsymbol{\beta}$ is a vector with intercept and regression coefficients. The first component of $\boldsymbol{x}_i$ equals 1 to be able to estimate an intercept. Formally this notation is only correct when the predictors in $\boldsymbol{x}_i$ are equal for both components of $\boldsymbol{\mu}_i$. The model then has the same structure as a multivariate regression model. The dimensions of $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$ are however allowed to differ.

Even though our main interest concerns effects on the circular mean, the mean structure of $\boldsymbol{y}_i$ can influence both the mean and spread of a circular outcome. Hereafter we call an effect on the mean of a circular outcome a location effect and an effect on the spread an accuracy effect. The size of the Euclidean norm of $\boldsymbol{\mu}$ influences the circular spread. The larger it is, the smaller the spread on the circle. The consequences of this property of $\boldsymbol{\mu}$ for the interpretation of results from a PN regression model will be considered later.

To estimate PN regression models a Bayesian Markov Chain Monte Carlo (MCMC) procedure is used in which r_i and $\boldsymbol{B}$ are sampled. The procedure is based on Nuñez-Antonio et al. (2011) and Hernandez-Stumpfhauser et al. (2017), a diffuse normal prior is used for $\boldsymbol{B}$ and the exact method of sampling is described in the Appendix.

3.3 Location and accuracy effects

There are several types of circular effects: a location effect, an accuracy effect, a mix of these or no effect. Because PN regression models are consensus models the location and the spread of the circular outcome are modeled simultaneously (Rivest et al., 2015). This means that we can create measures for checking whether a predictor has a location or an accuracy effect in a PN regression model. A measure to check for a location effect can be constructed by using a regression line in $\mathbb{R}^2$. Considering one predictor in a PN regression model, predicted values on the first and second bivariate component ($\hat{y}^I$ and $\hat{y}^{II}$) are determined as follows:

$$\hat{y}^I = \beta_0^I + \beta_1^I x$$

$$\hat{y}^{II} = \beta_0^{II} + \beta_1^{II} x$$

where β_0^I and β_0^{II} are intercepts, β_1^I and β_1^{II} are regression coefficients of a particular predictor on the two bivariate components and x is a predictor value. Whether this regression line runs through the origin determines the type of circular effect it represents. In Figure 3.1 we see two regression lines in $\mathbb{R}^2$ with a unit circle. The regression line on the left passes through the origin. The regression line on the right does not pass through the origin. We also see circular predicted values, the black dots on the unit circle. The circular predicted values lie very close together in the left plot, while in the right plot they move counterclockwise on the circle when the

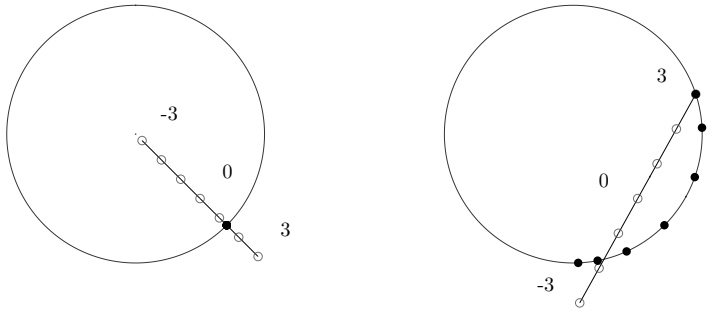

Figure 3.1: Two figures showing a bivariate regression line with linear predicted values (open dots) for x = $(-3, -2, -1, 0, 1, 2, 3)$ and a unit circle with circular predicted values (closed dots) on the circle for an accuracy (left) and location effect (right).

predictor value increases. The plot on the left represents an accuracy effect, the circular predicted values do not change with the predictor. The plot on the right represents a location effect.

To assess whether the regression line runs through the origin we first compute the shortest distance (SDO) from the regression line in bivariate space to the origin. The SDO is computed as:

$$SDO = \sqrt{\left(\beta_0^I + \beta_1^I a_x\right)^2 + \left(\beta_0^{II} + \beta_1^{II} a_x\right)^2}.$$

This is the Euclidean norm of the predictor value of the point where the line of predictions in $\mathbb{R}^2$ is closest to the origin (a_x). Because $SDO \geq 0$ we give it a sign such that its posterior is defined on the entire real line. We call this new parameter the signed shortest distance to the origin ($SSDO$). The following equation shows how to determine whether the sign should be positive or negative.

$$SSDO = \text{sign}[\sin\left(a_c - \text{atan2}(\beta_1^{II}, \beta_1^I)\right)]SDO. \tag{3.1}$$

The parameter a_c ($-\pi \leq a_c < \pi$) in this equation is the circular predicted value of a_x. How to compute a_x and a_c is outlined in Section 3.4.4. The function atan2() is defined in (3.2). An intuition for (3.1) is given in the Appendix. An example of how to use the $SSDO$ in practice will be given in Section 3.5. How well it performs at distinguishing accuracy and location effects is investigated in a simulation study in Section 3.6.

3.4 Quantifying location effects for continuous predictors

In this section we introduce how to compute circular predicted values and make predicted circular regression curves for a marginal effect. Subsequently we outline new tools for quantifying location effects.

3.4.1 Circular Predicted values

Using the two argument arctangent function, atan2, we compute predicted values on a circular scale, $\hat{\theta}$, as follows:

$$\begin{aligned}
\hat{\theta} = \text{atan2}\left(\hat{y}^{II},\, \hat{y}^{I}\right) &= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) && \text{if} \quad \hat{y}^{I} > 0 \\
&= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) + \pi && \text{if} \quad \hat{y}^{I} < 0 \qquad \hat{y}^{II} \geq 0 \\
&= \arctan\left(\frac{\hat{y}^{II}}{\hat{y}^{I}}\right) - \pi && \text{if} \quad \hat{y}^{I} < 0 \qquad \hat{y}^{II} < 0 \\
&= \frac{\pi}{2} && \text{if} \quad \hat{y}^{I} = 0 \qquad \hat{y}^{II} > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad \hat{y}^{I} = 0 \qquad \hat{y}^{II} < 0 \\
&= \text{undefined} && \text{if} \quad \hat{y}^{I} = 0 \qquad \hat{y}^{II} = 0 \qquad (3.2)
\end{aligned}$$

Here $\hat{y}^{I} = \boldsymbol{\beta}^{I}\boldsymbol{x}$ and $\hat{y}^{II} = \boldsymbol{\beta}^{II}\boldsymbol{x}$ are predicted values on the two components for a vector of predictor values $\boldsymbol{x}$.

3.4.2 Predicted regression curves

To visualize the circular effect, we compute a predicted circular regression curve for a marginal effect. For the marginal effect of one linear predictor with the values of the other predictors set to zero we specify $\hat{y}^{I} = \beta_0^{I} + \beta_1^{I}x$ and $\hat{y}^{II} = \beta_0^{II} + \beta_1^{II}x$. We fill out these functions for different values of x and the intercepts and coefficients are estimated. Figure 3.2 shows a regression curve for one predictor together with original data points. The y-axis of this plot contains the predicted outcome, $\hat{\theta}$, in degrees and the x-axis contains values for x with a range equal to the data range. This plot illustrates the effect of the predictor on the circular outcome.

By investigating a marginal effect all predictors except one are set to a specific value. In our case these are set to zero. For continuous variables we center the predictors and therefore zero refers to the mean value. For categorical variables zero refers to the baseline category. As in logistic regression the values to which the other predictors in the model are set influence the marginal effect we observe for the predictor of interest.

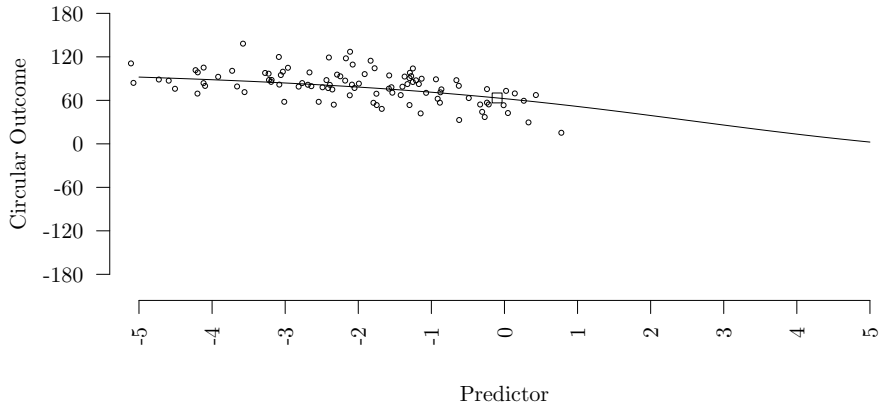

Figure 3.2: Predicted circular regression curve for the relation between a linear predictor and a circular outcome together with the original datapoints. The square indicates the inflection point of the regression curve.

3.4.3 A reparameterization for regression models

To quantify the slopes of circular regression curves we propose a reparameterization of (3.2). From this reparameterization we suggest three types of circular regression coefficients. The proposed reparameterization of (3.2) is as follows:

$$\begin{aligned}\hat{\theta} = \operatorname{atan2}\left(\hat{y}^{II},\, \hat{y}^{I}\right) &= \operatorname{atan2}\left(\beta_0^{II} + \beta_1^{II} x,\, \beta_0^{I} + \beta_1^{I} x\right) \\ &= a_c + \arctan\left[b_c\left(x - a_x\right)\right]. \end{aligned} \tag{3.3}$$

Here β_0^I and β_0^{II} are the linear intercepts and β_1^I and β_1^{II} are the linear coefficients of one continuous predictor variable x. The parameters, a_c, a_x and b_c describe a predicted circular regression curve, such as the one in Figure 3.2. The parameters a_c and a_x describe the location of the inflection point of the regression curve on the axis of the circular outcome and the axis of the predictor respectively. The inflection point occurs at the value of the predictor for which the regression line in $\mathbb{R}^2$ is closest to the origin. Hence, a_x is both the predictor value of the point on the regression line in bivariate space that lies closest to the origin as well as the location of the inflection point of the circular regression curve on the axis of the predictor. In Figure 3.2, the inflection point is indicated using a square. The parameter b_c describes the slope of the tangent line at the inflection point.

3.4.4 Parameter derivation

3.4.4.1 The x-coordinate of the inflection point (a_x)

To obtain a_x, the derivative of the Euclidean norm of the point where the line of predictions in $\mathbb{R}^2$ is closest to the origin is solved for 0. We find:

$$a_x = -\frac{\beta_0^I \beta_1^I + \beta_0^{II} \beta_1^{II}}{(\beta_1^I)^2 + (\beta_1^{II})^2}.$$

This is the location of the inflection point on the axis of the predictor x.

3.4.4.2 The y-coordinate of the inflection point (a_c)

To obtain a_c, we insert a_x into (3.3):

$$a_c = \text{atan2}\left(\beta_0^{II} + \beta_1^{II} a_x,\ \beta_0^I + \beta_1^I a_x\right).$$

This is the location of the inflection point on the axis of the predicted circular outcome.

3.4.4.3 The slope at the inflection point (b_c)

We may solve the reparameterization from (3.3) for b_c to obtain:

$$b_c = \frac{\tan\{\text{atan2}(\beta_0^{II} + \beta_1^{II} x,\ \beta_0^I + \beta_1^I x) - a_c\}}{x - a_x}, x \neq a_x. \tag{3.4}$$

Note that for $x = a_x$ this is undefined. We can simplify the formula by plugging in $x = 0$:

$$b_c = -\frac{\tan\{\text{atan2}(\beta_0^{II}, \beta_0^I) - a_c\}}{a_x}, a_x \neq 0. \tag{3.5}$$

This is the slope of the tangent line at the inflection point and a first circular regression coefficient. This coefficient is conceptually similar to the standard regression coefficient in von Mises regression models (Fisher & Lee, 1992). Like the SSDO this regression coefficient can be used as an indicator of a location effect. In a simulation study in Section 3.6 we investigate how these indicators, b_c or $SSDO$, perform at detecting location effects.

3.4.4.4 Additional quantification measures

The slope b_c is not necessarily a good measure for all datasets. For some data the inflection point of the regression curve does not lie near the data. In that case, b_c can take on a large range of different values while in the asymptotes the regression curve is still a good approximation to the data. This means that in some cases b_c represents a very unstable extrapolated effect. Then it is much more interesting to

investigate the slope of the regression curve near the data. We get the slope at a specific predictor value by taking the derivative of (3.3) for x and plugging in the value for which we want to know the slope:

$$\Delta\hat{\theta}(x) = \frac{d}{dx}\left\{a_c + \arctan[b_c(x - a_x)]\right\} = \frac{b_c}{1 + [b_c(x - a_x)]^2}.$$

An intuitive value for which we want to know the slope is the mean, $\bar{x}$. We obtain the slope at the mean (SAM) as:

$$SAM = \Delta\hat{\theta}(\bar{x}) = \frac{b_c}{1 + [b_c(\bar{x} - a_x)]^2},$$

where b_c is as computed in (3.5). We interpret this measure as: 'at $\bar{x}$, a one unit increase in x results in a SAM increase in $\hat{\theta}$.'

Another measure that we are interested in is the averaged slope (AS) over each datapoint. We compute slopes for all data values of x and average these slopes:

$$AS = \bar{\Delta}\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{b_c}{1 + [b_c(x_i - a_x)]^2}.$$

We interpret this measure as: 'On average, a one unit increase in the predictor, x, results in a AS increase in $\hat{\theta}$'.

3.5 Empirical Example

To illustrate the problems that occur when interpreting output from a PN model we fit a regression model to a dataset collected by Brunyé et al. (2015), the 'pointing north data'. In their study, 200 Tufts University students divided across 10 data collection sites were asked to point north. Pointing angles relative to the magnetic north (pointing errors) were recorded as outcome variable and several predictor variables were measured. The outcome variable was measured such that the real north was located at $\theta = 0°$. One of the predictor variables is self-reported spatial ability. This was measured using the Santa Barbara Sense of Direction questionnaire (SBSOD). Other predictors were age, experience with living on campus and gender (0 = male, 1 = female). Table 3.1 shows descriptives for these data.

Before analysis all continuous predictor variables were centered at 0. This affects the intercept but not the coefficients and will make the interpretation of individual effects more intuitive. When using the embedding approach we are predicting the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^t$ of the PN distribution for the pointing error. The prediction equations for the pointing north data are:

$$\hat{\mu}_1 = \beta_0^I + \beta_1^I Age + \beta_2^I Gender + \beta_3^I Experience + \beta_4^I SBSOD$$
$$\hat{\mu}_2 = \beta_0^{II} + \beta_1^{II} Age + \beta_2^{II} Gender + \beta_3^{II} Experience + \beta_4^{II} SBSOD.$$

Table 3.2 shows estimates from the regression model that was fit after careful

Table 3.1: Descriptives for the pointing north data with linear mean and standard deviation (SD) for continuous variables and mean direction ($\bar{\theta}$) and mean resultant length ($\bar{R}$) for circular variables.

		Mean/($\bar{\theta}$)	SD/$\bar{R}$	Minimum	Maximum	Type
Pointing error		19.57°	0.43	-	-	Circular
Age		19.68	1.29	18.00	23.00	Continuous
Experience		1.79	1.21	0.00	4.00	Continuous
SBSOD		4.10	1.06	1.47	6.80	Continuous
Gender		-	-	-	-	Categorical
	M	16.32°	0.56	-	-	
	F	24.33°	0.32	-	-	

Note that only absolute pointing errors are provided online with the original article. The pointing error used here was obtained by an iterative process multiplying the absolute errors by a random vector containing 1's and -1's and computing the mean resultant length ($\bar{R}$) until a dataset was found in which $\bar{R}$ was equal to 0.43, that of the original data. The mean direction was then set to 19.57, the mean direction of the original data.

Table 3.2: Posterior Modes and 95% HPD lower (LB) and upper bounds (UB) for the regression coefficients of the pointing north data. An asterisk (*) indicates that an HDP interval does not contain 0.

	Component I			Component II		
	Mode I	LB I	UB I	Mode II	LB II	UB II
Intercept	0.95	0.72	1.20 *	0.15	-0.07	0.41
Age	-0.07	-0.29	0.14	0.00	-0.23	0.21
Gender	-0.48	-0.73	-0.10 *	0.17	-0.18	0.47
Experience	0.22	-0.00	0.45	-0.02	-0.23	0.23
SBSOD	0.25	0.09	0.40 *	0.27	0.07	0.38 *

monitoring of convergence (iterations = 3000, burn-in = 1000). The Appendix shows histograms of the posterior distributions of the linear intercepts and coefficients. The interpretation of the regression coefficients is the same as in linear regression. For SBSOD that is: *'If the self-reported spatial ability increases by one unit, the predicted outcome on the first bivariate component increases by 0.25 and the predicted outcome on the second bivariate component increases by 0.27.'* For the categorical variable, gender, we may interpret the coefficients as follows: *'For females the predicted outcome on the first bivariate component is 0.48 lower than for males.'* Because our circular outcome concerns compass data the only interpretation we can give to the bivariate components is that of a North-South and an East-West axis. The coefficients do not give us the information on the change in the predicted pointing error on the circle, a circular effect, nor do they give us information on whether it is a location or accuracy effect. Next, we compute and interpret the circular regression coefficients introduced in Section 3.4 and the SSDO for the pointing north data.

Table 3.3: Posterior Modes (PM) and 95% HPD interval lower (LB) and upper bounds (UB) for the circular regression coefficients and $SSDO$ for the continuous variables of the pointing north data. An asterisk (*) indicates that an HPD interval does not contain 0.

	Age			Experience			SBSOD		
	PM	LB	UB	PM	LB	UB	PM	LB	UB
b_c	0.16	-2.57	2.80	-0.35	-7.04	6.50	0.52	0.09	3.48 *
SAM	-0.02	-0.22	0.24	-0.03	-0.31	0.18	0.17	0.00	0.40 *
AS	0.05	-0.21	0.26	-0.10	-0.36	0.27	0.23	0.01	0.42 *
$SSDO$	-0.76	-1.06	1.00	1.30	-1.87	2.07	-1.08	-2.11	-0.59 *

3.5.1 Circular effects

Before interpreting the circular regression coefficients we use the linear coefficients to determine whether there is any, location or accuracy, effect of a predictor on the pointing error. To do so we use the linear regression coefficients for a predictor, e.g. β_1^I and β_1^{II}. In case one or both regression coefficients of the two components are different from 0 there is an accuracy and/or location effect. To check whether this holds we can use the Highest Posterior Density (HPD) intervals of both linear coefficients. If they do not both contain zero we conclude that there is an effect of the predictor on the circle. From Table 3.2 we may thus conclude that only SBSOD and gender have an effect on the circle.

For the categorical variable gender we can compare a predicted angle for females, $\text{atan2}(\beta_0^{II} + \beta_2^{II}, \beta_0^I + \beta_2^I) = 34.25°$ and for males $\text{atan2}(\beta_0^{II}, \beta_0^I) = 8.97°$. Females thus have a higher pointing error than males on average. If we compute such predicted angles for each iteration of the MCMC sampler we can also compute HPD intervals for this effect.

For the continuous variables of the pointing north data, Table 3.3 shows the posterior modes (PM) and the upper (UB) and lower bounds (LB) of the HPD intervals for the circular coefficients and the $SSDO$. We use the $SSDO$ to determine whether the effect of SBSOD on the pointing error is an accuracy or a location effect. Again we use the HPD interval. If the HPD interval of the $SSDO$ of a predictor does not include 0 we conclude that the bivariate regression line does not run through the origin and that there is a location effect. For SBSOD the HPD interval does not include zero. This means that the effect of SBSOD on the pointing error is a location effect.

Next, we investigate the circular regression coefficients for SBSOD shown in Table 3.3. For SBSOD, all three circular coefficients have HPD intervals that do not contain 0. This means that SBSOD has a location effect and is what we expected after concluding that the $SSDO$ for SBSOD was different from zero. Figure 3.3 shows the relation between SBSOD and the pointing error. The posterior mode of the slope at the inflection point of the predicted circular regression curve, b_c, for SBSOD is equal to 0.52. This means that at the inflection point, as SBSOD increases by one unit the pointing error increases counterclockwise, with $0.52 \cdot 180/\pi \approx 29.8°$

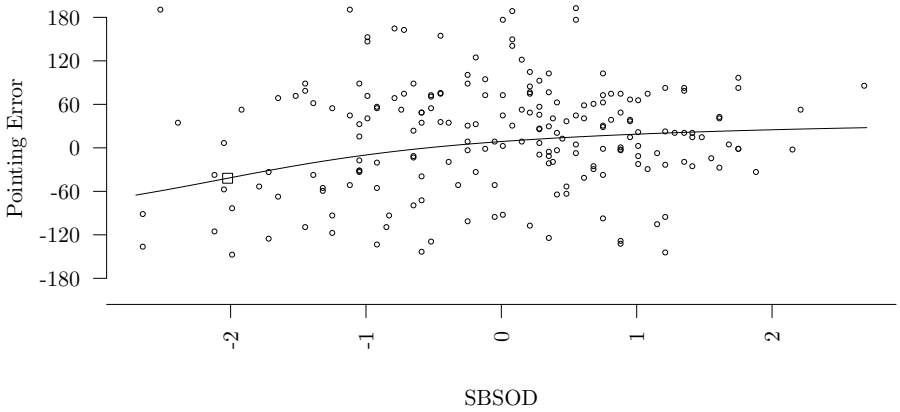

Figure 3.3: Predicted circular regression curve for the relation between SBSOD and the pointing error together with the original datapoints. The square indicates the inflection point of the regression curve.

keeping all other predictors at zero. However, the inflection point lies almost outside the data range so we would rather interpret the AS or SAM. On average an increase of 1 unit in SBSOD results in an counterclockwise increase in pointing error of $AS \cdot 180/\pi = 0.23 \cdot 180/\pi \approx 13.2°$ keeping all other predictors at zero. At the average SBSOD an increase of 1 unit results in an counterclockwise increase in pointing error of $SAM \cdot 180/\pi = 0.17 \cdot 180/\pi \approx 9.7°$ keeping all other predictors at zero.

Even though they do not have a circular effect, it is interesting to look at the parameter estimates of Age and Experience. The results show estimation issues for both predictors exemplified by the wide HPD intervals of b_c. When we look at the posterior histograms of the circular regression coefficients and $SSDO$ for Age and Experience in Figure A2 in the Appendix we also see estimation issues. Whereas the histogram of b_c shows that there are some posterior estimates with extreme positive or negative values the histograms for AS and SAM do not. These extreme values are probably the cause of the wide HPD intervals. Additionally, such estimation issues may occur when we try to estimate a location effect in a situation where there is no or just a very small location effect. Judging from the data plotted along the predicted regression curve in Figure 3.4 this may well be the case for the variable Experience. In Section 3.6 we will illustrate what happens when we try to estimate a location effect in data where there is no or almost no location effect. Note however that Experience does seem to influence the spread of the pointing error. Figure 3.5 shows the relation between Experience and the concentration, which is the reciprocal of the spread, of the predicted values on the circle. The concentration was computed using the formula for the mean resultant

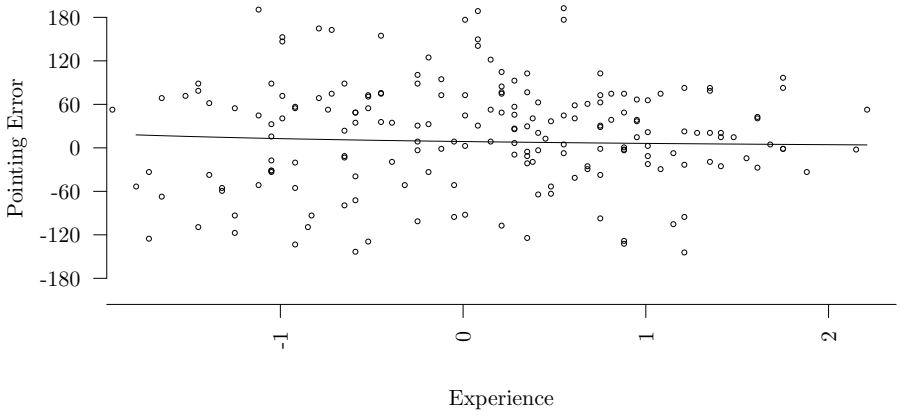

Figure 3.4: Predicted circular regression curve for the relation between Experience and the pointing error together with the original datapoints. The inflection point of the regression curve is not shown as this point lies outside the range of the data.

length from Kendall (1974). In the Appendix predicted circular regression plots as well as figures showing the effect on the concentration are shown for all continuous variables in the pointing north data.

3.6 Simulation study

To assess the performance of the circular coefficients b_c, SAM and AS and the ability to distinguish between location and accuracy effects we conducted a simulation study with 1225 designs with one predictor. Of these designs 1056 were classified as location designs, 144 as accuracy and 25 as having no effect. Because the last category is so small its results are excluded from the simulation study. A description of the designs is given in Section 3.6.1. In Section 3.6.2 a summary of the simulation results is given and in Section 3.6.3 we try to explain the causes of patterns observed in these results. A more detailed description of the simulation study and the results be found in the Appendix.

3.6.1 Design

In each design different population values were chosen for the linear intercepts β_0^I and β_0^{II} and the regression coefficients β_1^I and β_1^{II}. From these values, the population values of the parameters a_x, a_c, b_c, SAM, AS, SDO and $SSDO$ were computed. For each design 1000 datasets were simulated, 500 with N = 50 and

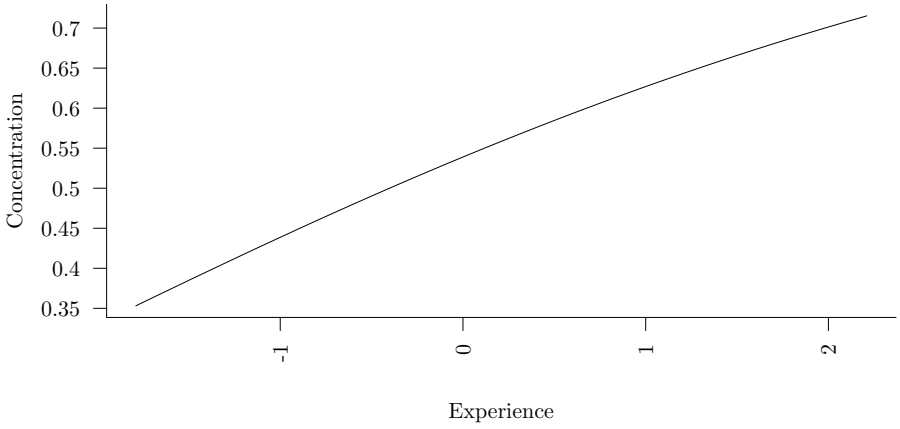

Figure 3.5: Figure showing the relation between Experience and the concentration of the predicted values on the circle.

500 with N = 200. Each dataset contains one circular outcome θ and one linear predictor $x \sim N(0, 1)$. The relation between predictor and outcome was determined by the chosen values for the linear intercepts and coefficients. Before analysis of a dataset the linear predictor was centered at 0. After analysis the relative bias, frequentist coverage of the HPD interval and average interval width (AIW) of the estimates for b_c, SAM, AS and $SSDO$ of each design were computed.

3.6.2 Results

In this section we shortly summarize the results from the simulation study with regards to how well we can detect location and accuracy effects and the performance of the MCMC sampler in estimating the circular coefficients. We will especially focus on the designs in which the measures did not perform optimal.

If there is any effect this can be found in more than 90% of the datasets of a design. This holds in all design categories, location and/or accuracy. The indicators for location effects that were tested, b_c and $SSDO$, work well for accuracy and location effects with $SDO > 1$. For location effects with $SDO \leq 1$ the indicators perform worse. All indicators perform better in designs with a larger sample size.

Concerning the performance of the MCMC sampler in estimating the circular regression coefficients we can say that designs with a larger sample size and larger $SSDO$ in most cases perform better in terms of relative bias, coverage and AIW. In accuracy designs the coefficients have smaller relative biases compared to location designs with an $SSDO$ close to 0. In general AS and SAM have lower relative bias than b_c. In terms of coverage the AS shows slight undercoverage. The SAM

has slight undercoverage in location designs and overcoverage in accuracy designs. The log AIW is largest for accuracy designs and the parameter b_c. This seems to correspond to the estimates for the example data.

3.6.3 Explaining patterns

To explain the patterns in relative bias, coverage and log AIW we will show what happens in the estimation of circular regression coefficients of an exemplary design. The exemplary design is a location design with an $SSDO$ of 0.24 and a sample size of 50. The results for this design are summarized in Table 3.4.

Figure 3.6 shows histograms for the posterior modes for a_c, b_c, AS and SAM for all 500 datasets. Notice that the histograms for a_c, b_c and AS are bimodal. The bimodality is caused by the estimated a_c that switches to the other side of the circle. How often a_c switches sides is determined by the SDO. When the SDO is zero, in an accuracy design, a_c is equally likely to switch to either side of the circle and the histograms of modes will all be bimodal and symmetrical. When the SDO is large, the a_c will almost never switch to the opposite side of the circle and the histograms of modes will all be unimodal and symmetrical. In both cases the symmetry of the histograms of modes around the true value results in little bias in the estimate of b_c and AS. In designs with a small SDO the histogram of modes is bimodal but not symmetrical. This causes bias in b_c and AS. The bimodalty problem does not occur in the SAM which explains the lower relative bias. Because of this property and its interpretation we prefer the SAM over b_c and AS.

In Section 3.5 we have observed that for some iterations of the MCMC sampler the estimated b_c is either extremely negative or extremely positive. The extremes are caused by the tangent in the formula for b_c in (3.4) and (3.5). The tangent function has asymptotes at $.5\pi$ radians and at $-.5\pi$ radians. If $\text{atan2}(\beta_0^{II}+\beta_1^I x, \beta_0^I+\beta_1^{II} x)-a_c$ or $\text{atan2}(\beta_0^{II},\ \beta_0^I) - a_c$ is close to either one of these asymptotes we get extreme estimates for b_c. These extremes cause the AIWs to be large as can be seen in Table 3.4. Large AIWs cause the coverage of the designs with small SDOs, to be as good or better than in designs with larger SDOs and lower the ability to detect location effects in the designs with smaller SDOs.

3.7 Discussion

The main contribution of this paper is to simplify the interpretation of effects in projected normal regression models. In previous literature only the bivariate coefficients for a predictor were given without much indication of how to properly interpret these. Therefore, we developed methods for assessing circular effects. These methods allow us to interpret and quantify the effect of a predictor on a circular outcome. A simulation study has shown that the performance of the methods is good in designs with an easily detectable location effect. If the location effect is harder to detect the performance of the methods worsens. We need to increase the sample size to get more power and a better performance. The slope at the mean,

Table 3.4: Simulation results of 500 simulated datasets (N = 50) from a population with $\beta_0^I = 3$, $\beta_1^I = 2$, $\beta_0^{II} = 1$ and $\beta_1^{II} = 0.5$. Mode refers to the estimated posterior mode and LB and UB refer to the averaged lower and upper bounds of the 95% HPD intervals.

Parameter	Population Value	Mode	Bias	LB	UB	Coverage
a_x	-1.53	-1.46	-0.07	-1.91	-1.17	0.94
a_c	1.82	1.80	0.02	-1.28	1.89	0.96
b_c	-8.50	-2.45	-6.05	-72.62	65.01	0.95
AS	-0.41	-0.29	-0.12	-0.75	0.45	0.95
SAM	-0.05	-0.05	0.00	-0.17	0.06	0.96
SSDO	0.24	0.27	-0.02	-0.17	0.81	0.95

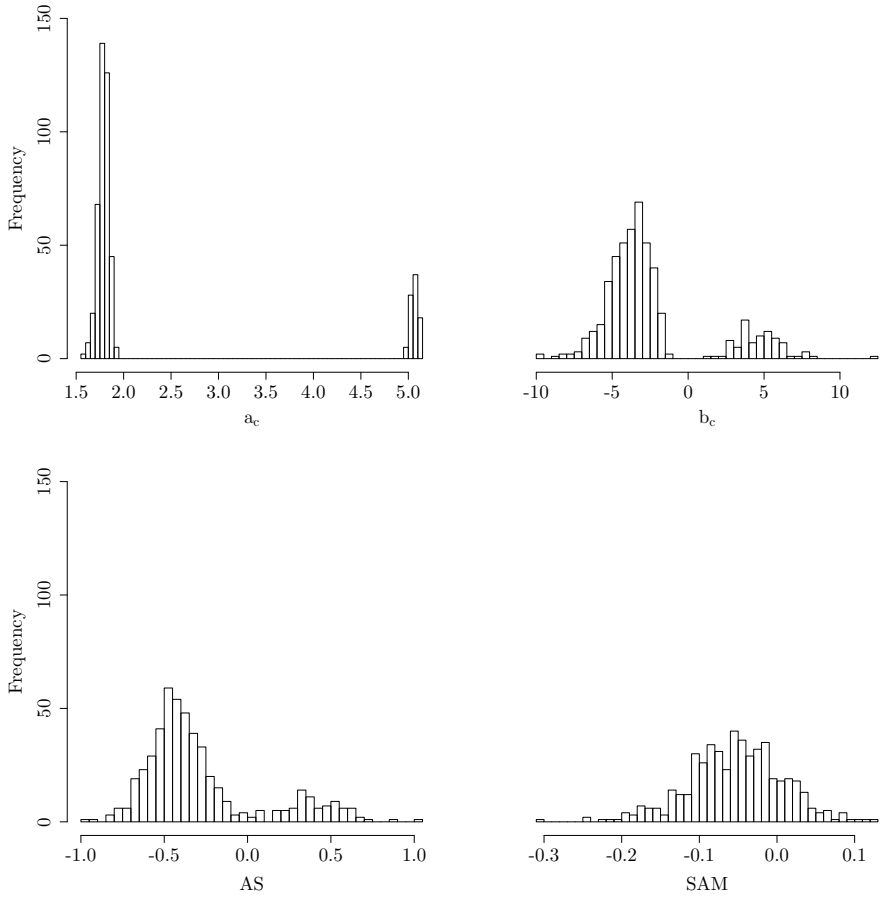

Figure 3.6: Histograms of the posterior modes of 500 simulated datasets (N = 50) from a population with $\beta_0^I = 3$, $\beta_1^I = 2$, $\beta_0^{II} = 1$ and $\beta_1^{II} = 0.5$ for the parameters a_c, b_c, AS and SAM.

(SAM) has an intuitive interpretation and performs best. The performance of the other circular coefficients seems to depend on the type of effect it is computed for. Therefore we recommend researchers to use the SAM and carefully investigate the circular regression plots and posterior histograms before using b_c or AS.

Additionally, we investigated the ability of our method to detect different types of circular effects. If there is any effect it can usually be picked up by the linear coefficients. The coefficient b_c and $SSDO$ perform equally well at detecting location effects. Assessing whether there is a location effect is harder in smaller samples, especially if it is a location effect with a small SDO. It is recommended that researchers make sure they have a large enough sample size to be able to detect the effect they are interested in. To be able to precisely say what sample size is needed more research needs to be done. From the present research we conclude that for a PN regression model with one continuous linear predictor a sample of 200 is large enough to be able to detect most location effects with small SDO.

The ability of the PN regression model to by default detect an effect, on the mean or spread, is an advantage over regression models of the intrinsic or wrapping approach to circular data because we do not need to fit two separate models. The model we use allows for investigation of the posterior distributions of the linear coefficients to check whether a location or accuracy effect is present in the data.

Although this paper focuses on regression models for the PN situation we may consider using the tools introduced here in more complex models or in models using the general projected normal (GPN) distribution. One possibility is to use them in models where the mean of the PN distribution is partly composed of basis functions of the covariates, e.g. polynomials. Because locally polynomials look like a straight line, our tools might be applied here as well. Another example of a more complex model is the mixed-effects model that Nuñez-Antonio & Gutiérrez-Peña (2014) propose. In this model all tools introduced in this paper can be used on fixed effects. For random effects we can also consider these tools, however, we have to be cautious on how to interpret circular random effects. Their interpretation depends on the point relative to which the circular random effects are computed; the individual random intercepts or the average intercept. We are in this case also interested in the spread of the random effects. To assess the spread new tools will have to be developed. In GPN models the tools introduced here will be more complex to interpret and to implement. For skewed data interpretation problems could be overcome, but for bimodal data we would e.g. have to choose at what mode of the data we want to compute the SAM. In a regression model this seems redundant as we would usually try to explain possible bimodality by including predictors in the model, e.g. having different means for men and women. This can already be done in a PN model. Tools for a GPN model would also be hard to implement. There is no analytic solution for obtaining the mean direction of GPN models. This complicates the computation of circular predicted values. It is possible to get these using Monte Carlo integration (Wang & Gelfand, 2014) and we may also be able to compute the slope of a circular regression curve and the tools proposed in this paper in a similar way. However, the behavior of these regression coefficients should then be investigated thoroughly. Especially their behavior and use in a model where the GPN distribution is bimodal and/or skewed.

In conclusion, this paper has contributed to our knowledge about interpreting effects in projected normal regression models. We have outlined how to assess whether a predictor has an effect on either the spread or the mean of the circular outcome. But most importantly we have found a way of quantifying an effect on the mean of a circular outcome. These methods allow us to directly assess the effect of a predictor on the circle. In our opinion this has removed the major drawback of the projected normal regression model.

Chapter 4

Indicators for effects on mean and variance in projected normal regression models for a circular outcome

by J. Cremers & D. Hernandez-Stumpfhauser[1]

4.1 Introduction

Circular regression models are those in which a circular variable $\theta \in [0, 2\pi)$ is regressed on a set of linear and/or circular predictors, $\boldsymbol{x}$. In the literature there are three approaches to circular data, the 'intrinsic', 'wrapping' and 'embedding' approach (Mardia & Jupp, 2000). In the intrinsic approach models are based on distributions directly defined on the circle whereas in the wrapping and embedding approach distributions are defined in respectively $\mathbb{R}$ and $\mathbb{R}^2$ and subsequently wrapped or projected onto the circle. Within each of these three approaches regression models for a circular outcome have been introduced (Fisher & Lee, 1992; Lagona, 2016; Mulder & Klugkist, 2017; Nuñez-Antonio et al., 2011; Presnell et al., 1998; Ravindran & Ghosh, 2011).

Another distinction that can be made in the literature on circular regression models is one between models with homogeneous errors and those with heterogeneous

[1]Manuscript submitted to Journal of Statistical Computation and Simulation
Author contributions: JC proposed the topic for the study, DH and JC developed the new measures, JC designed and performed the simulation study with feedback from DH, JC wrote the paper and DH gave feedback on the written work.

errors (Rivest et al., 2015). In a homogeneous error model predictor variables can only affect the mean of the circular outcome while in heterogeneous error models predictor variables can have an effect on both the mean (a location effect) and variance (an accuracy effect) of the circular outcome. The projected normal (PN) regression model, the model we focus on in this paper, is one of these heterogeneous error models for circular variables. This model was first introduced by Presnell et al. (1998) and adapted to the Bayesian context by Nuñez-Antonio et al. (2011).

Several measures for assessing the effect on the mean of the outcome, a location effect, in a PN regression model were introduced in Cremers et al. (2018b). Additionally they introduced an indicator for effects on both the variance and the mean. This indicator is called the $SSDO$. In this paper new indicators for effects on the variance and the mean in PN regression models will be introduced in Section 4.3. As for the $SSDO$ these new indicators will test H_0 : *There is only an accuracy effect.* The performance of the new and existing indicators will be assessed in a simulation study in Section 4.4. We will however first give a short introduction to the PN normal regression model in Section 4.2.

4.2 Projected Normal Regression Models for a Circular Outcome

In projected normal models we assume that the circular outcome θ results from a projection onto the circle of a bivariate normal variable $\boldsymbol{y}_i \sim N_2(\boldsymbol{\mu}_i, \boldsymbol{I})$ where $i, \ldots, n$. The relation between $\boldsymbol{y}$ and θ is defined as:

$$\boldsymbol{u} = \boldsymbol{y}/r \tag{4.1}$$

where $\boldsymbol{u} = (\cos\theta, \sin\theta)^t$ and $r > 0$. The idea behind this projection is that we do not have to conduct inference on θ directly but we can indirectly conduct inference on a bivariate normal variable $\boldsymbol{y}$. This makes for a flexible approach as a lot of different and complex models exist for bivariate normal data. However, both $\boldsymbol{y}$ and r cannot be directly obtained from θ. Instead the estimation of $\boldsymbol{y}$ and r is treated as a missing data problem. In this paper we use the approach used by Cremers et al. (2018b) to solve the missing data problem and fit the model.

The relation in (4.1) implies that θ has a projected normal distribution defined as:

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

where $-\pi \leq \theta < \pi$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}$ is the mean vector, the covariance matrix $\boldsymbol{I}$ is identity and $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$. The terms $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution. Note that we choose the covariance matrix to be identity for identification purposes. In Figure 4.1 we see that the shape of this density is rotationally symmetric about its mean direction $\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}$ and its concentration is dependent on $||\ \boldsymbol{\mu}\ ||^2$ (see Kendall (1974) for the exact form of this relation). A different way of parameterizing a projected normal distribution can be found in

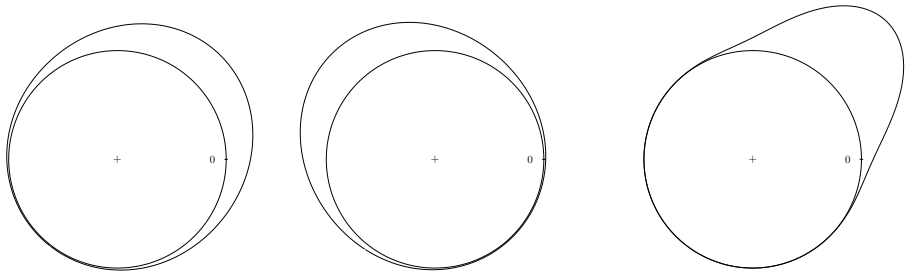

Figure 4.1: Projected normal densities with different $\boldsymbol{\mu}$. From left to right, the mean vector $\boldsymbol{\mu}$ is set to (2,2), (-2,2) and (4,4).

Wang & Gelfand (2013).

In a circular regression model $\boldsymbol{\mu}$ may have the following structure:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_i^I \\ \mu_i^{II} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\beta}^I)^t \boldsymbol{x}_i^I \\ (\boldsymbol{\beta}^{II})^t \boldsymbol{x}_i^{II} \end{pmatrix}, \tag{4.2}$$

where $i = 1, \ldots, n$, $\boldsymbol{x}_i$ is a vector of predictor values for individual i and each $\boldsymbol{\beta}$ is a vector with intercept and regression coefficients. To be able to estimate an intercept, the first component of $\boldsymbol{x}_i$ equals 1. In this paper we center the predictor x and estimate the PN regression model using MCMC methods also used in Cremers et al. (2018b). Because the structure in (4.2) is such that the predictor variables determine $\boldsymbol{\mu}$ the predictors can have effects on both the mean (a location effect) and the variance (an accuracy effect) of the circular outcome.

4.3 Two indicators for accuracy and location effects

In this section we will describe two different indicators that allow us to check whether the predictors in a projected normal regression model only have an effect on the variance (an accuracy effect) or also have an effect on the location (a location effect) of the circular outcome. The first indicator, the $SSDO$, has been introduced previously in Cremers et al. (2018b). The second indicator is new and actually comprises a set of indicators that we call angle based measures.

4.3.1 The signed shortest distance to the origin ($SSDO$)

In Cremers et al. (2018b) it is outlined that the effects of a variable x on both the variance and mean of the circular outcome in a PN regression model can be detected by looking at the shortest distance of the regression line in bivariate space to the origin (SDO). The regression line in $\mathbb{R}^2$ is defined as follows:

$$(\beta_0^I + \beta_1^I x,\ \beta_0^{II} + \beta_1^{II} x)^t.$$

Figure 4.2 shows two regression lines in $\mathbb{R}^2$ together with arrows representing predicted outcomes for $x = (x_{min}, 0, x_{max})$ and a line that represents the SDO. The intersections of the arrows with the circle represent the predicted outcomes on the circle. The left figure shows a situation with only an accuracy effect. In this situation the regression line runs through the origin ($SDO = 0$) and the arrows for the predicted outcomes on the circle are parallel and they intersect with the circle at one point, i.e.: the circular predicted values for different values of x are the same. The right figure shows a situation where there is also a location effect. In this situation the regression line does not run through the origin, $SDO > 0$ and the arrows representing the predicted outcomes on the circle are not parallel and do not intersect the circle at the same point, i.e.: the circular predicted values for different values of x are different.

Cremers et al. (2018b) introduced a measure, the signed shortest distance to the origin ($SSDO$), derived from the SDO, to detect accuracy and location effects. Just as for the SDO, H_0 : *There is only an accuracy effect* is true for the $SSDO$ when it equals 0. To test whether there is also a location effect with the $SSDO$ we can thus reformulate this null hypothesis as $H_0 : SSDO = 0$. The authors in Cremers et al. (2018b) make use of the highest posterior density (HPD) to test the null hypothesis. The null hypothesis is rejected at a significance level of α if the $100 * (1 - \alpha)$ HPD interval for SSDO does not include zero.

Cremers et al. (2018b) perform a simulation for the $SSDO$. This simulation is however limited to two sample sizes and does not adequately investigate the validity of the test. With validity we mean achieving a correct type-I error in the long run. We will therefore perform a new simulation with more sample sizes in Section 4.4. Additionally, we will set up this simulation such that it contains a larger variation of real SDO values than before. Finally, we also include the new angle based measures (introduced next) in this simulation. This new set-up will allow us to assess and compare the type-I errors of the different measures.

4.3.2 Angle based measures

In Figure 4.2 we saw that if the vectors representing predicted values for several x are not parallel there is a location effect. The new measures we introduce in this section are based on testing whether these vectors are parallel. A way to test whether two vectors are parallel is to compute the sine of the angle between the two vectors and check whether this is equal to 0. The sine of the angles λ and γ between the vectors $(\beta_0^I, \beta_0^{II})$ and $(\beta_0^I + \beta_1^I x_{min}, \beta_0^I + \beta_1^{II} x_{min})$ and $(\beta_0^I + \beta_1^I x_{max}, \beta_0^I + \beta_1^{II} x_{max})$

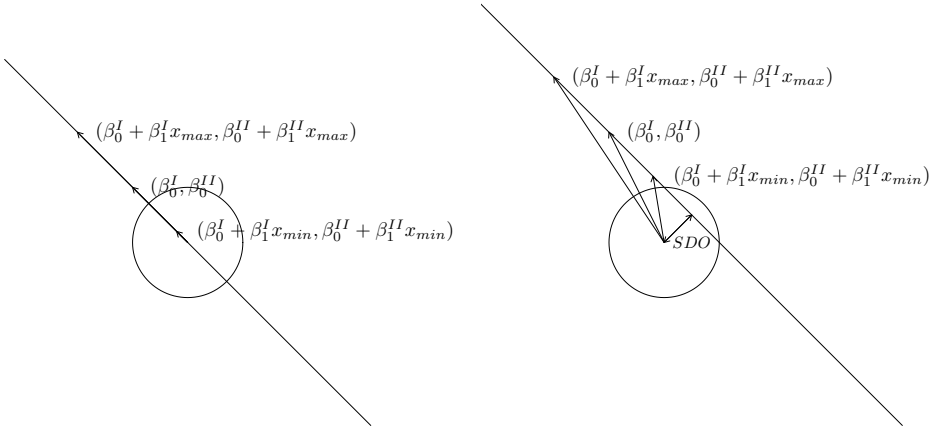

Figure 4.2: Two regression lines in $\mathbb{R}^2$ together with a unit circle and arrows representing predicted outcomes for $x = (x_{min}, 0, x_{max})$ and the SDO.

respectively from Figure 4.2 are computed as follows:

$$\sin(\lambda) = \frac{\beta_0^I(\beta_0^{II} + \beta_1^{II} x_{min}) - \beta_0^{II}(\beta_0^I + \beta_1^I x_{min})}{\sqrt{(\beta_0^I)^2 + (\beta_0^{II})^2}\sqrt{(\beta_0^I + \beta_1^I x_{min})^2 + (\beta_0^{II} + \beta_1^{II} x_{min})^2}}$$

$$\sin(\gamma) = \frac{\beta_0^I(\beta_0^{II} + \beta_1^{II} x_{max}) - \beta_0^{II}(\beta_0^I + \beta_1^I x_{max})}{\sqrt{(\beta_0^I)^2 + (\beta_0^{II})^2}\sqrt{(\beta_0^I + \beta_1^I x_{max})^2 + (\beta_0^{II} + \beta_1^{II} x_{max})^2}}$$

and is bounded between -1 and 1. From now on we will call these measures $\sin(\lambda)$ and $\sin(\gamma)$. Note that in this paper, because the predictor x has been centered, the vector $(\beta_0^I, \beta_0^{II})$ is the vector pointing at the data mean. We can thus reformulate H_0 : *There is only an accuracy effect* as $H_0 : \sin(\lambda) = 0$ or $H_0 : \sin(\gamma) = 0$. We can test this null hypothesis by using the HPD interval of the posterior distribution of $\sin(\lambda)$ or $\sin(\gamma)$. We could also measure the angle between the vectors for the predicted values at x_{min} and x_{max}. A reason to compute this other measure is that the angle is always larger than either λ or γ on its own, which means that it may also have larger power for rejecting H_0. We define this measure as follows:

$$\sin(\lambda + \gamma) =$$

$$\frac{(\beta_0^I + \beta_1^I x_{min})(\beta_0^{II} + \beta_1^{II} x_{max}) - (\beta_0^{II} + \beta_1^{II} x_{min})(\beta_0^I + \beta_1^I x_{max})}{\sqrt{(\beta_0^I + \beta_1^I x_{min})^2 + (\beta_0^{II} + \beta_1^{II} x_{min})^2}\sqrt{(\beta_0^I + \beta_1^I x_{max})^2 + (\beta_0^{II} + \beta_1^{II} x_{max})^2}}$$

The three measures we have just introduced, $\sin(\lambda)$, $\sin(\gamma)$ and $\sin(\lambda + \gamma)$, are all dependent on the shape of the data. For testing purposes we would rather consider a measure that is not dependent on the shape of the data. An intuitive way to construct such a measure is to consider the vector $(\beta_1^I, \beta_1^{II})$ and test whether this is parallel to $(\beta_0^I, \beta_0^{II})$:

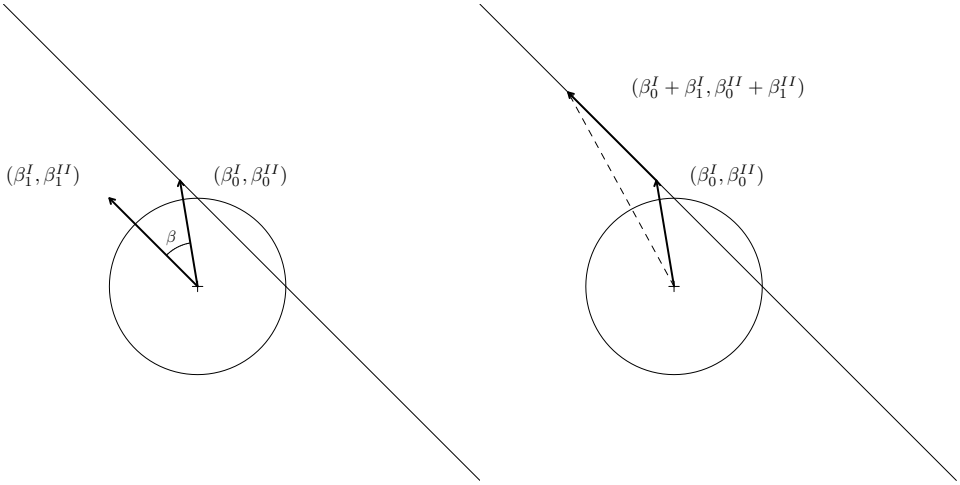

Figure 4.3: Visual representation of $\sin(\beta)$.

$$\sin\beta = \frac{\beta_0^I \beta_1^{II} - \beta_0^{II} \beta_1^I}{\sqrt{(\beta_0^I)^2 + (\beta_0^{II})^2}\sqrt{(\beta_1^I)^2 + (\beta_1^{II})^2}}$$

From now on we will call this measure $\sin(\beta)$. A visual representation of β is given in the left plot in Figure 4.3. The right plot shows how $(\beta_1^I, \beta_1^{II})$ is related to the regression line in bivariate space; the addition of $(\beta_1^I, \beta_1^{II})$ to $(\beta_0^I, \beta_0^{II})$ results in the vector $(\beta_0^I + \beta_1^I, \beta_0^I + \beta_1^{II})$.

Note that for all measures introduced above we could also use the cosine formula, $\cos(\theta) = \boldsymbol{a} \cdot \boldsymbol{b} / ||\boldsymbol{a}||||\boldsymbol{b}||)$, to compute the angle between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. However, the cosine of the angle between two vectors is -1 or 1 if they are parallel, which is not convenient for the purpose of hypothesis testing.

4.4 Simulation Study

To assess the performance of $SSDO$ and the angle based measures $\sin(\beta)$, $\sin(\gamma)$, $\sin(\lambda)$ and $\sin(\lambda + \gamma)$ we conducted a simulation study with 477 designs. In 72 of these designs, the accuracy designs, there was only an accuracy effect ($SDO = 0$). In the other designs, the location designs, there was also a location effect $SDO > 0$.

For each design 2500 datasets were simulated. These datasets had different sample sizes (N): 25, 50, 75, 100 and 150. These sample sizes were chosen because in earlier simulations for the $SSDO$ in Cremers et al. (2018b) only two sizes (50 and 200) were used and in this paper we want to see more detailed results regarding sample size. For each sample size 500 datasets were simulated. Each dataset contains one circular outcome θ and one linear predictor $x \sim N(0, 1)$ similar to the simulation in Cremers et al. (2018b). The relation between predictor and outcome was determined by the population values for the intercepts, β_0^I and β_0^{II}, and coefficients, β_1^I and

β_1^{II}. In each design different population values were chosen for the intercepts and regression coefficients. For the designs with a location effect pairs of population values $(\beta_0^I, \beta_0^{II})$ for the linear intercepts were:

$$\{(\cos(10°), \sin(10°)), (\cos(30°), \sin(30°)), (\cos(45°), \\ \sin(45°)), (\cos(60°), \sin(60°)), (\cos(80°), \sin(80°))\}$$

and pairs of population values for the regression coefficients $(\beta_1^I, \beta_1^{II})$ were:

$$\{(1,0), (0,1)(1,1), (0.5,0), (0,0.5), (0.5,0.5), (2,0), (0,2), (2,2)\}.$$

These population values were then multiplied by a multiplication factor between 1 and 5 at intervals of 0.5 to obtain regression lines with different SDO. Note that the coefficients were chosen to be largely similar to the coefficients that show good performance (in terms of bias and coverage) in previous simulation studies in Cremers et al. (2018b) and Cremers et al. (2018a). Datasets for each possible combination of the pairs of intercepts and regression coefficients and multiplication factors were simulated. These combinations led to a larger variation in real SDO values than in previous simulations. In the accuracy designs the population values for β_0^I and β_1^I and β_0^{II} and β_1^{II} were equal in each design. Their exact values and the simulation code is given in the supplementary material at https://github.com/joliencremers/IndicatorsPN.

From the population values of the intercepts and coefficients, the population values of the SDO, $SSDO$ and the angle based measures ($\sin(\beta)$, $\sin(\lambda)$ and $\sin(\gamma)$) were computed. For each dataset we determine whether x is predicted to have an accuracy effect by checking whether the 95% HPD intervals of the estimated $SSDO$ and the angle based measures included 0. We thus test H_0 : *There is only an accuracy effect.* If H_0 was not rejected the dataset was classified as having only an accuracy effect. For each design we then computed the proportion of datasets in which the $SSDO$ and the angle based measures indicated only an accuracy effect.

To display the results of the simulation in a concise manner, we grouped all simulation designs into 6 categories based on their population SDO: those where $SDO = 0$ (the 72 accuracy designs) and those where the $0 < SDO < 1$ (72 designs), $1 \leq SDO < 2$ (90 designs), $2 \leq SDO < 3$ (90 designs), $3 \leq SDO < 4$ (90 designs) and $4 \leq SDO$ (63 designs). For each of the 6 categories of our simulation we then averaged the proportion of datasets where only an accuracy effect was indicated over the designs of that category. For the location designs this proportion represents a type-II error; the proportion in which H_0 : *There is only an accuracy effect.* was not rejected even though it should have been. For the accuracy designs this proportion is equal to 1 minus the type-I error of the test or actually the proportion in which we correctly classify the effect as having only an accuracy effect. It also represents the coverage of the HPD interval; we are testing whether the interval includes 0 which is the real value in the accuracy designs. Thus, if the $SSDO$ and angle based measures perform well we expect the proportion to be high, around 0.95 because we set $\alpha = 0.05$, for the accuracy designs and low for the designs with a location effect.

Table 4.1: The proportion of datasets in which an accuracy effect is indicated by $SSDO$ grouped per real SDO value.

N	$SDO = 0$	$0 < SDO < 1$	$1 \leq SDO < 2$	$2 \leq SDO < 3$	$3 \leq SDO < 4$	$4 \leq SDO$
25	0.90	0.49	0.23	0.14	0.10	0.06
50	0.89	0.35	0.15	0.10	0.08	0.05
75	0.90	0.28	0.12	0.08	0.07	0.05
100	0.90	0.25	0.10	0.07	0.06	0.04
150	0.89	0.19	0.08	0.06	0.06	0.04

Table 4.2: The proportion of datasets in which an accuracy effect is indicated by $\sin(\beta)$ grouped per real SDO value.

N	$SDO = 0$	$0 < SDO < 1$	$1 \leq SDO < 2$	$2 \leq SDO < 3$	$3 \leq SDO < 4$	$4 \leq SDO$
25	0.91	0.52	0.27	0.17	0.13	0.08
50	0.92	0.39	0.18	0.12	0.09	0.06
75	0.94	0.33	0.14	0.09	0.07	0.05
100	0.95	0.29	0.12	0.08	0.07	0.05
150	0.94	0.23	0.10	0.07	0.06	0.04

4.4.1 Results

The results of the simulation are shown in Tables 4.1 and 4.2 for the performance of the $SSDO$ and $\sin(\beta)$ respectively. The tables show the proportion of datasets in which H_0 : *There is only an accuracy effect* could not be rejected grouped per real SDO value.

We see that for the location designs ($SDO > 0$) both measures perform better with increasing sample sizes and with increasing SDO. The proportion of datasets in which H_0 is not rejected ($SSDO = 0$ or $\sin(\beta) = 0$) becomes lower when sample size and SDO increase. For the accuracy designs there is no clear effect of sample size for the $SSDO$. The proportion of datasets in which H_0 is not rejected stays the same at around 0.9. It thus does not reach the correct type-I error level (of 0.05) within the current range of sample sizes. The fact that the actual alpha level consistently seems to lie at 0.1 is not promising of the tests ability to reach the correct alpha level asymptotically. There is an effect of sample size for $\sin(\beta)$ in the accuracy designs. This indicator reaches the correct type-I error level at a sample size of 75 meaning that in the long run this is a valid test for H_0.

Results for the other three angle based measures, $\sin(\gamma)$, $\sin(\lambda)$ and $\sin(\lambda + \gamma)$, are shown in the Appendix. In general these measures perform worse or equal to $\sin(\beta)$. Only the performance of $\sin(\gamma)$ stands out. This measure already reaches the correct type-I error level for a sample size of 25.

4.5 Discussion

In this paper we have introduced new angle based measures that allow testing of H_0 : *There is only an accuracy effect* in projected normal regression models for a circular outcome. We have investigated their performance and compared these measures to a previously introduced measure, the $SSDO$ in a simulation study.

The results of the simulation study lead us to conclude that in general the angle based measures perform better than the $SSDO$ since three of them reach the correct type-I error level in the long run. However, three of the four new measures that we introduced in this paper, $\sin(\gamma)$, $\sin(\lambda)$ and $\sin(\lambda + \gamma)$, are dependent on the distribution of the predictor variable. In further research we could focus on finding more angle based measures that are not dependent on the distribution of the predictor variable. To do so standardization of the predictor x instead of centering only could be an option. Additionally now that we have found tests that have valid type-I errors in the long run we can try to increase the power of these tests.

Lastly, we would like to note that in this paper we have only considered simulation designs in which there is an effect in the population, either only an accuracy effect or both an accuracy and a location effect. In reality however there are situations in which there is no effect at all. In further research it would be beneficial to investigate the performance of the $SSDO$ and the new measures introduced here in samples for which there is no effect in the population.

Chapter 5

Longitudinal circular modelling of circumplex measurements for interpersonal behavior

by J. Cremers, H.J.M. Pennings, M.T. Mainhard & I. Klugkist[1]

5.1 Introduction

Interpersonal behavior, both in personality psychology as well as in educational psychology, can be measured using circumplex measurement instruments (see Horowitz & Strack (2011) for an overview of many such instruments). In personality psychology these measurement instruments are for example used to characterize interpersonal subtypes of patients in research on posttraumatic stress disorder (König et al., 2016), borderline personality disorder (Wright et al., 2013), panic disorder (Zilcha-Mano et al., 2015) and other types of mental health problems (Salzer et al., 2008; Simon, Cain, Wallner Samstag, Meehan, & Muran, 2015). In this study we will however use example data from the field of educational science.

In educational science, students' interpersonal perceptions of their teachers' behavior is an indicator of classroom climate (Mainhard, Pennings, Wubbels, & Brekelmans, 2012; Wubbels et al., 2006). This is a major aspect of teaching and an extensive amount of research has shown that a positive classroom climate is central to student

[1]Manuscript submitted to Assessment

Author contributions: JC and IK designed the study, JC developed the new measures and analyzed, processed and interpreted the results with feedback from IK, HJMP and MTM provided the empirical datasets, JC wrote the paper and HJMP, MTM and IK gave feedback on the written work.

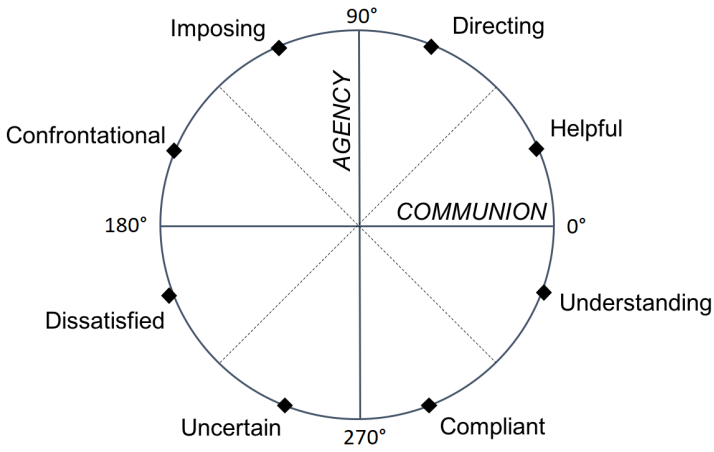

Figure 5.1: The interpersonal circle for teachers (IPC-T). The words presented in the circumference of the circle are anchor words to describe the type of behavior located in each part of the IPC.

academic and emotional outcomes (e.g. Church, Elliot, & Gable, 2001; Den Brok, Brekelmans, & Wubbels, 2004; Cornelius-White, 2007; Wubbels et al., 2006).

To study the student perceptions of their teachers' interpersonal behavior the Questionnaire on Teacher Interaction (QTI) can be used (Wubbels et al., 2006). The QTI is a measurement instrument that consists of items that all load on two orthogonal dimensions: Agency and Communion. Agency refers to the degree of power or control a person exerts in interaction with others. Communion refers to the degree of friendliness or affiliation a person conveys in interaction with someone else. The QTI results for these two dimensions can be placed within a two-dimensional space formed by Agency (vertical) and Communion (horizontal) "as a continuous order with no beginning or end" (Gurtman, 2009, p. 2). Such circular ordering is called a circumplex ordering and this two dimensional space is therefore often called the interpersonal circumplex (IPC).

In previous research employing longitudinal measurements of teachers on the QTI (Brekelmans et al., 2005; Mainhard et al., 2011a) the two axes Agency and Communion are analyzed separately. Also in more recent research employing repeated and time-series measures of Agency and Communion, these two dimensions are analyzed separately (Mainhard et al., 2012; Pennings et al., 2018; Pennings, Van Tartwijk, et al., 2014). However, interpersonal behavior represents a specific combination of Agency and Communion and therefore analyzing it as a blend of these two dimensions would be preferable. One way to do so is to subdivide the IPC into 8 subscales, see Figure 5.1. These subscales each characterize a different type of teacher behavior and represent a particular blend of Agency and Communion. Measurements on the two axes are obtained by combining scores of the items belonging to the 8 subscales. For example, Wright et al. (2013) used a latent class analysis to distinguish between these different interpersonal subtypes in patients with borderline personality disorder. This type of categorization can however prove

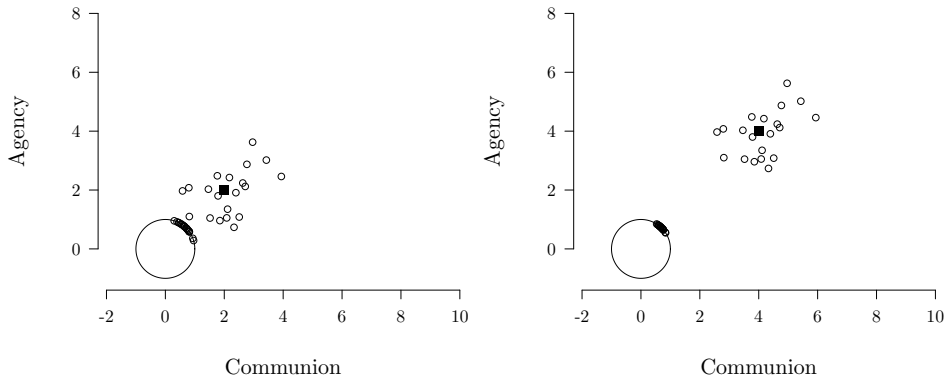

Figure 5.2: Communion (x-axis) and Agency scores (y-axis) and scores on the circumplex (circle) for two measurement occasions. The solid square indicates the average on Agency and Communion. Whereas on the Agency and Communion axes the mean changes, on the circumplex not the mean but the variance changes.

to be a disadvantage in longitudinal research where the interest lies in change on the circumplex. When data is categorized, small changes on the IPC are not automatically picked up as they do not necessarily imply a change from one category to the other.

Another way to study the IPC as a blend of the two dimensions is by using circular statistics. Instead of giving teachers a categorical score on the IPC (one of the eight subscales), we give them a continuous circular score, θ that lies between 0° and 360°[2]. A positive score on Communion and a 0 score on Agency represent a circular value of 0°, which falls between the subscales 'understanding' and 'helpful' of Figure 5.1. Wright et al. (2009) argue that the circumplex structure is amenable to circular statistical techniques and that such techniques can answer more precise questions than those that can be answered by analyzing the individual scales comprising circumplex instruments. Moskowitz & Zuroff (2004) also make this argument and have developed a method that makes use of circular descriptive statistics to investigate data from an interpersonal circumplex.

Figure 5.2 illustrates how interpretations of the data are different when analyzing Agency and Communion separately compared to analyzing Agency and Communion using circular statistics. It shows how the scores of individuals change over two measurement occasions (left and right plot) both on the two axes Agency and Communion and on the circumplex/circle. We see that from the left to the right measurement occasion the average score (solid square) on both the Agency and Communion axis increases but their variance stays the same. On the circumplex

[2]To convert a score of a teacher on Agency and Communion to a circular measurement, θ, we can make use of the double arctangent function in (5.1) where $\hat{y}_2$ is replaced by A and $\hat{y}_1$ by C. In this function A represents a score on Agency and C represents a score on Communion. Note that the order of the arguments in the function $atan2()$ determines what scores on Agency and Communion represent a circular value of 0°.

itself however, the average direction does not change whereas the variance decreases. Patterns in scores on the separate components over time may thus reflect different patterns in scores on the circle and circular statistics can thus give us a different type of information about data on the IPC. Additionally, a continuous circular score does not have the disadvantage that categorical scores have when the interest lies in change on the IPC. Therefore, it may be beneficial to consider a circular model for circumplex data.

In this paper we will fit a mixed-effects model for circular data (Nuñez-Antonio & Gutiérrez-Peña, 2014) to two datasets with repeated measures of teacher behavior on the IPC obtained from the QTI to illustrate the advantages of using circular statistics on real longitudinal datasets. This model falls within the embedding approach; one of three approaches to circular data. The embedding approach offers an advantage over the other two approaches in terms of flexibility. In the embedding approach we assume that the circular measurement origins from an unknown variable in two-dimensional (bivariate) space. This assumption makes it relatively easy to develop more complex models for circular data as we can base them on already existing models for bivariate linear data. A disadvantage of the embedding approach however concerns the interpretation of model parameters. The assumption that the circular variable origins from a variable in bivariate space effectively implies that in the modelling process the circular outcome is split into two components, based on the sine and cosine of θ. Consequently, the results from the mixed-effects model are not automatically on a circular scale. Therefore, we developed new interpretation tools that translate the results from the model back to the circle to solve interpretation problems. Note that in case the circular outcome in the mixed-effects model is measured on the IPC this does not mean that the sine and cosine components into which it is split in the embedding approach are the scores on Agency and Communion. There is a link between them but there is no one-to-one translation. This implies that using the circular mixed-effects model is not equivalent to fitting a bivariate linear mixed effects model. A more detailed explanation for this will be given when we introduce the method.

In Section 5.2 we will give more details on the embedding approach as well as present the circular mixed-effects model. Then we introduce the new measures that were developed for interpreting the effects from this model on the circle in Section 5.3. In Section 5.4 we will introduce the two datasets with teacher data, from now on called the teacher behavior datasets. We will analyze these datasets with the circular mixed-effects model from Section 5.2 and present the results of the analyses and their interpretation. The paper will be concluded with a discussion in Section 5.5.

5.2 Methodological background

The purpose of this section is to provide the theoretical background to both the embedding approach to circular data and the mixed-effects model for circular longitudinal data that we use in this paper. We will also give a more detailed description how to relate the sine and cosine components in the embedding approach

to the Agency and Communion dimension of the IPC. Whereas this section provides the methodological background, Section 5.3 will introduce the new tools we developed to translate the results from the circular mixed-effects model back to the circle. Note that for an understanding of Section 5.3 it is not necessary to comprehend the present section in full detail, especially for readers who are mainly interested in the application of the circular mixed-effects model.

5.2.1 The embedding approach

In the embedding approach we split the circular outcome in two parts and instead of using θ as outcome variable directly, we take its unit vector representation, $\boldsymbol{u} = (\cos\theta, \sin\theta)$ and thus split it up into a sine and cosine component. For data from the IPC the two components of $\boldsymbol{u}$, $\cos(\theta)$ and $\sin(\theta)$, can interpretation wise be referred to as a score on Communion and Agency. However, they do not represent exactly the same values since Agency and Communion scores range from $-\infty$ to ∞ while the sine and cosine range from -1 to 1. We assume that $\boldsymbol{u}$ can be represented by an unobserved vector $\boldsymbol{y}$ in bivariate real space ($\mathbb{R}^2$) as follows:

$$\boldsymbol{u} = \frac{\boldsymbol{y}}{r},$$

where r represents the length of the vector $\boldsymbol{y}$. The circular outcome, θ, thus originates from a projection onto the circle of a vector, $\boldsymbol{y}$, in $\mathbb{R}^2$. Note that when modelling circular data we have only observed $\boldsymbol{u}$, whereas $\boldsymbol{y}$ and r are considered latent variables.

In this paper we assume that the unobserved vector in $\mathbb{R}^2$, $\boldsymbol{y}$, is normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{I}$ ($\boldsymbol{y} \sim N_2(\boldsymbol{\mu}, \boldsymbol{I})$). It then follows that θ has a projected bivariate normal density:

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

where θ is the circular response variable $0 < \theta \leq 2\pi$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$, $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$. The objects $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution. Note that to be able to identify the model, the covariance matrix is fixed to be an identity matrix ($\boldsymbol{I}$). We have to do so because both the mean and the variance of a bivariate normal variable influence the spread of its projection onto the circle. This effect can be seen in Figure 5.3 where we have plotted three sets of data from different bivariate normal distributions (with components I and II) that are projected onto the circle. We see that the spread on the circle in the middle (higher mean) and right (smaller variance) plot is smaller than in the left plot. Thus both the mean and the variance of bivariate normal data influence projected scores on the circle. We need to fix either the mean or covariance matrix to identify the model and be able to estimate the model. This restriction also implies that we can not simply take scores on Agency and Communion as $\boldsymbol{y}$. Neither can we estimate a bivariate linear mixed-effects model to Agency and Communion scores and translate the results from this model to the circle easily due to the fact

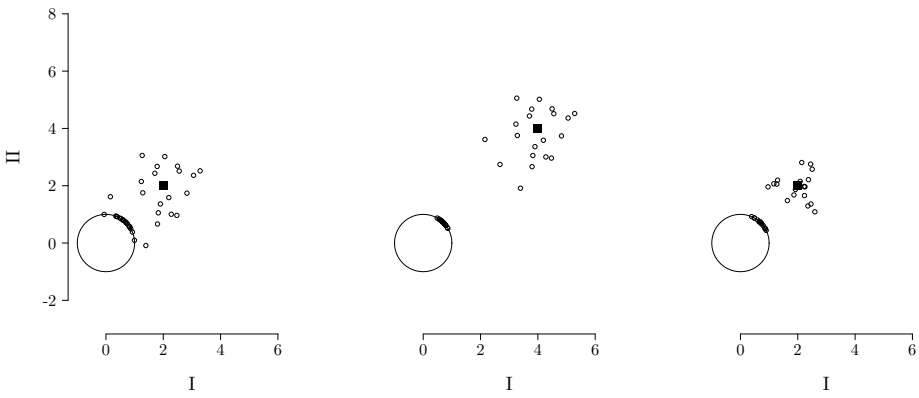

Figure 5.3: Three sets of bivariate normal datapoints with different mean vectors (from left to right $(2,2),(4,4),(2,2)$) and covariance matrices projected onto the circle (from left to right variances: 0.9, 0.9, 0.5, covariances are 0). We see that the spread on the circle in the middle (higher mean) and right (smaller variance) plot is smaller than in the left plot. Thus both mean and variance of the bivariate data influence the spread on the circle

the covariance matrix of such a model is unrestricted and thus both its mean and its variance influence the circular scores.

5.2.1.1 Estimation methods

In the embedding approach and thus also in the circular mixed-effects model, we have not observed the bivariate normal variable, $\boldsymbol{y}$, from which the circular outcome θ origins. This means that we need to use special techniques in the fitting of a model, e.g. an EM algorithm (Presnell et al., 1998) or auxiliary variables in a Bayesian setting (Nunez-Antonio & Gutierrez-Pena, 2005). In this paper we use a Bayesian Markov Chain Monte Carlo (MCMC) sampler to estimate the model. These methods are based on methods from Nuñez-Antonio & Gutiérrez-Peña (2014) and Hernandez-Stumpfhauser et al. (2017). The exact specifications of the priors and posteriors of the model and the MCMC sampler are given in the Appendix.

5.2.2 A mixed-effects model for circular data

A type of model that can be used to analyse longitudinal or repeated measures data is the so called mixed-effects model. In the social sciences a mixed-effects model may also be referred to as a multilevel model (Hox, 2002). For circular data, a mixed-effects model based on the projected normal distribution is developed by Nuñez-Antonio & Gutiérrez-Peña (2014). In this model, there are independent observations of a design matrix for the fixed effect and random effect predictors, $\boldsymbol{X}$ and $\boldsymbol{Z}$ in addition to a circular outcome vector $\boldsymbol{\theta}$ for each individual $i = 1, \ldots, n$.

The rows of the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ and the indexes of the vector $\boldsymbol{\theta}$ represent the measurement occasions $j = 1, \ldots, N$. The circular mixed-effects model can thus be regarded as a multilevel model with two levels. The model has the following mean structure:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^{I} \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\beta}^{I})^t \boldsymbol{x}_{ij}^{I} + (\boldsymbol{b}_i^{I})^t \boldsymbol{z}_{ij}^{I} \\ (\boldsymbol{\beta}^{II})^t \boldsymbol{x}_{ij}^{II} + (\boldsymbol{b}_i^{II})^t \boldsymbol{z}_{ij}^{II} \end{pmatrix}.$$

Here $\boldsymbol{\beta}^{I}$ and $\boldsymbol{\beta}^{II}$ are vectors with fixed effect coefficients and intercept and $\boldsymbol{b}_i^{I}$ and $\boldsymbol{b}_i^{II}$ are vectors with random effects for each individual. Note that each individual has two matrices of fixed effect predictors and two matrices of random effect predictors, one the sine and one for the cosine component (denoted I and II) of the mean vector of the model. The structure of the model is thus similar to that of a bivariate mixed-effects model.

For one individual i and component k the fixed and random effect matrices may look as follows:

$$\mathbf{X}_i^k = \begin{pmatrix} \text{Intercept} & x_1 & x_2 & \text{Measurement} \\ 1 & -28.31 & -10.40 & 0 \\ 1 & -13.94 & -10.40 & 2 \\ 1 & -6.79 & -10.40 & 3 \\ 1 & -1.29 & -10.40 & 5 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -55.85 & -10.40 & 15 \end{pmatrix},$$

$$\mathbf{Z}_i^k = \begin{pmatrix} \text{Intercept} \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, k = I, II.$$

In the fixed effect design matrix, $\boldsymbol{X}_i^k$, we have included an intercept, a predictor x_1 that varies with each measurement, a predictor x_2 that is constant for each measurement and a predictor that indicates the measurement occasion, e.g. week number. The design matrix for the random effects, $\boldsymbol{Z}_i^k$, only contains an intercept, meaning that only the intercept of the model but none of the slopes of the effects of the predictors are considered random. Note that the columns of the design matrices $\boldsymbol{X}_i^k$ and $\boldsymbol{Z}_i^k$ do not need to be of the same dimensions for both components, I and II, of the model. This means we can use different fixed effect predictors and specify different random effects for each component.

5.3 Assessing circular effects

In the embedding approach, model coefficients are estimated for the two bivariate components, I and II. For data from the QTI this means that although Agency and Communion are analyzed together in the circular mixed-effects model, the results of this model are interpreted separately on the two components. However, we also want to be able to interpret the effect of predictors on a circular scale and interpret effects on the 'blend' of Agency and Communion. In this section, we will propose methods to assess effects on a circular scale for both fixed-effect and random-effect predictors.

5.3.1 Fixed effects on the circle

In the circular mixed-effects model we can compute fixed effects on the circle in a manner that is similar to the computation of regression coefficients on the circle in Cremers et al. (2018b). We can distinguish between effects on the location (mean) of the circular outcome and on the spread (variance) of the circular outcome.

5.3.1.1 Effects on location

Predicted values on the circle, $\hat{\theta}$, can be computed as follows:

$$
\begin{aligned}
\theta &= \operatorname{atan2}(\hat{y}_2,\, \hat{y}_1) \\
&= \arctan\left(\frac{\hat{y}_2}{\hat{y}_1}\right) && \text{if} \quad \hat{y}_1 > 0 \\
&= \arctan\left(\frac{\hat{y}_2}{\hat{y}_1}\right) + \pi && \text{if} \quad \hat{y}_1 < 0 && \& \;\; \hat{y}_2 \geq 0 \\
&= \arctan\left(\frac{\hat{y}_2}{\hat{y}_1}\right) - \pi && \text{if} \quad \hat{y}_1 < 0 && \& \;\; \hat{y}_2 < 0 \\
&= +\frac{\pi}{2} && \text{if} \quad \hat{y}_1 = 0 && \& \;\; \hat{y}_2 > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad \hat{y}_1 = 0 && \& \;\; \hat{y}_2 < 0 \\
&= \text{undefined} && \text{if} \quad \hat{y}_1 = 0 && \& \;\; \hat{y}_2 = 0
\end{aligned} \tag{5.1}
$$

where $\hat{y}_1$ is the predicted value on the first bivariate component and $\hat{y}_2$ is the predicted value on the second bivariate component. To assess the fixed effect of a predictor x on the location in a mixed-effects regression model we set all predictors except x at 0 and define $\hat{y}_1 = \beta_0^I + \beta_1^I x$ and $\hat{y}_2 = \beta_0^{II} + \beta_1^{II} x$. If we center all continuous predictors and make sure that the scale of all categorical predictors includes 0, the interpretation of this marginal effect is straightforward, namely the effect of a predictor x for an individual from the reference group that has average scores on all other continuous predictors (see Sections 5.4.1.3 and 5.4.2.3). A reparameterization allows us to quantify the slopes of angular regression lines by

creating circular regression coefficients as follows:

$$\begin{aligned}\hat{\theta} &= \text{atan2}\left(\hat{y}_2,\ \hat{y}_1\right) \\ &= \text{atan2}\left(\beta_0^{II} + \beta_1^{II} x,\ \beta_0^{I} + \beta_1^{I} x\right) \\ &= a_c + \arctan\left(b_c\left(x - a_x\right)\right).\end{aligned}$$

The three parameters, a_c, a_x and b_c describe a predicted circular regression line such as the one in Figure 5.4. The parameter a_c describes the location of the inflection point, the point at which the increase or decrease of the line levels off, of the regression line (the square in Figure 5.4) on the axis of the predicted circular outcome $\hat{\theta}$ and a_x describes the location of the inflection point on the axis of the predictor x. The parameter b_c describes the slope of the tangent line at the inflection point. This is a regression coefficient that can be compared to the 'standard' regression coefficient from other types of circular models. Because the inflection point does not always lie within the data range, we can also compute the slope of the regression line for different values of x. The slope at the grand mean, SAM, is computed as:

$$SAM = \Delta\hat{\theta}(\bar{x}) = \frac{b_c}{1 + [b_c(\bar{x} - a_x)]^2}, \tag{5.2}$$

and the averaged slope over each datapoint, AS, as:

$$AS = \bar{\Delta}\hat{\theta}(x_{ij}) = \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\frac{b_c}{1 + [b_c(x_{ij} - a_x)]^2}. \tag{5.3}$$

The parameters b_c, SAM and AS are thus three types of circular regression coefficients. For a thorough investigation of these three coefficients see Cremers et al. (2018b)

5.3.1.2 Effects on spread

In a model from the embedding approach such as the circular mixed-effects model in this paper we automatically model an effect on the location and on the spread of the circular outcome. This is a remnant of the fact that the mean of the underlying bivariate distribution affects both the location and spread on the circle (see Figure 5.3 in Section 5.2.1). This means that each predictor in the model also has an effect on the spread of the circular outcome. In this paper we will use the circular variance, $1 - \rho$, as an indicator of the spread on the circle.

Kendall (1974) outlined how to compute a circular variance from parameters of a model based on the PN distribution. The equation to estimate the variance on the circle based on estimates from the circular mixed-effects model, is the following:

$$1 - \hat{\rho} = 1 - \sqrt{\pi\xi/2}\exp\left(-\xi\right)\left[I_0\left(\xi\right) + I_1\left(\xi\right)\right]$$

where $\xi = ||\boldsymbol{\mu}||^2/4$ and $I_\nu()$ is the modified Bessel function of the first kind and order ν. For the effect of a variable x on the variance $\boldsymbol{\mu}$ is defined as $(\beta_0^I + \beta_1^I *$

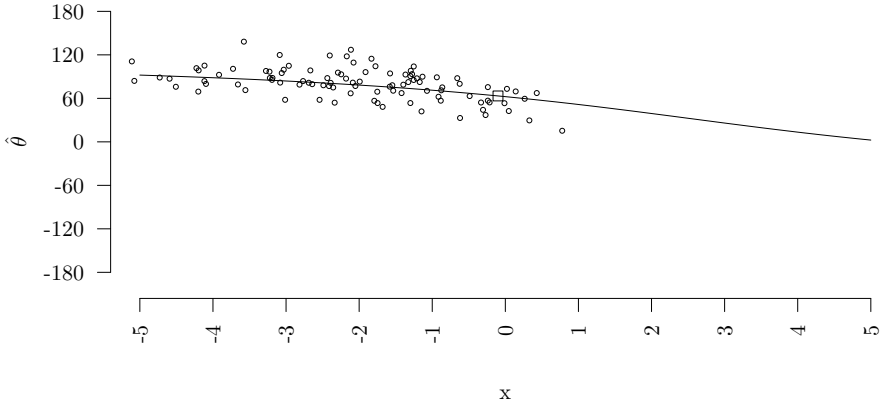

Figure 5.4: Predicted circular regression line for the relation between a linear predictor x and a predicted circular outcome θ together with the original datapoints. The square indicates the inflection point of the regression line.

$x, \beta_0^{II} + \beta_1^{II} * x$).

5.3.2 Random effects on the circle

To be able to assess random effects on the circle we first convert the estimated linear intercepts and slopes from the model to circular intercepts and slopes for each individual. The predicted linear intercepts for each individual are $\hat{y}_{1,i} = \beta_0^I + b_{i,1}^I$ and $\hat{y}_{2,i} = \beta_0^I + b_{i,1}^{II}$ for the two components. The parameters β_0^I and β_0^{II} are the fixed intercepts on both components and $b_{i,1}^I$ and $b_{i,1}^{II}$ are the random intercepts for individual i on each component. To convert the linear intercepts to circular ones we use (5.1) as follows:

$$\mu_{circ,i} = \text{atan2}\left(\beta_0^{II} + b_{i,1}^{II}, \beta_0^I + b_{i,1}^I\right).$$

A predicted circular slope can be computed in a similar fashion. We first compute predicted outcome values on the linear scale for the effect of a predictor x on the two components as follows: $\hat{y}_{1,i} = \beta_0^I + b_{i,1}^I + \beta_1^I x + b_{i,2}^I x$ and $\hat{y}_{2,i} = \beta_0^{II} + b_{i,1}^{II} + \beta_1^{II} x + b_{i,2}^{II} x$. The parameters β_1^I and β_1^{II} are the fixed slope coefficients and $b_{i,2}^I$ and $b_{i,2}^{II}$ are the random slope coefficients for individual i on each component. We then convert the linear predicted outcomes to circular ones using (5.1) and reparameterize this

equation to obtain circular slope coefficients for each individual:

$$\begin{aligned}\hat{\theta} &= \text{atan2}\left(\beta_0^{II} + b_{i,1}^{II} + \beta_1^{II}x + b_{i,2}^{II}x,\ \beta_0^{I} + b_{i,1}^{I} + \beta_1^{I}x + b_{i,2}^{I}x\right) \\ &= a_{c,i} + \arctan\left(b_{c,i}\left(x - a_{x,i}\right)\right),\end{aligned}$$

where $b_{c,i}$ is the circular random slope for individual i. Note that this approach is similar to the one for fixed effects in Section 5.3.1 except that we now compute separate slopes for each individual. We can also compute SAM and AS for each individual using (5.2) and (5.3). The vectors $\boldsymbol{b}_c$, $\boldsymbol{SAM}$ and $\boldsymbol{AS}$ then contain circular random slopes for each individual and $\boldsymbol{\mu}_{circ}$ contains the random intercepts.

5.3.2.1 Random effect variances

In mixed-effects models we are usually interested in the variance of the random intercepts and slopes and not in their values for each individual. Therefore, we compute the variance of the circular intercepts as follows:

$$1 - \hat{\rho} = 1 - \frac{\sqrt{\left(\sum_{i=1}^{N} \sin \mu_{circ,i}\right)^2 + \left(\sum_{i=1}^{N} \cos \mu_{circ,i}\right)^2}}{N}.$$

Note that this computation of the circular variance is different from the one in Section 5.3.1.2. We cannot use the present computation for fixed effects since we do not have multiple measurements per predictor value in that case. For the fixed effects we are looking at the 'average' effect and therefore do not have a value for each individual. Because the circular random slope coefficients are linear variates ($-\infty < b_{c,i} < \infty$) we can compute a linear variance of these to assess the circular slope variance of a model.

5.4 Teacher behavior data

The teacher behavior datasets are two datasets containing repeated measures on the IPC obtained with the QTI, for Dutch secondary school teachers. The first dataset was collected between 1982 and 2008 and contains over 7199 teachers at different stages of their teaching career. There are no further covariates in this dataset. A further description of the data, the models we will fit to them and the results will be given in Section 5.4.1. The second dataset was collected between 2010 and 2015 and contains 161 teachers. This dataset does contain additional covariates. For this dataset a further description of the data, the models we will fit to them and the results will be given in Section 5.4.2.

5.4.1 Dataset I

For this paper we selected teachers from dataset I whose data included at least 3 QTI measures during their career resulting in a dataset with 126 teachers. Descriptives of

Table 5.1: Descriptives for dataset I

Variable	mean/$\bar{\theta}$	sd/$\hat{\rho}$	Range	Type
IPC	25.21°	0.74	-	Circular
EX	5.61	4.68	0-29	Linear
Year	1992	6.31	1974-2006	Linear

this dataset are shown in Table 5.1. The variable `IPC` contains the circular outcome measurements. The variable `EX` indicates the experience (in years) that a teacher had at a specific measurement occasion. Note that this means that instead of taking the number of the measurement occasion (first, second, third, etc.) or the year of the measurement as the longitudinal variable of interest we take the experience of a teacher as the variable of interest. A teacher with 3 years of experience in 1990 thus has the same score on this time variable as a teacher with 3 years of experience in 2000. To control for possible biases due to the year in which the teacher started his or her career we include a control variable `Year` that indicates the year in which a teacher had 0 years of experience. The research question that can be answered using this dataset is how teachers' QTI scores change during their career.

5.4.1.1 The model

The teacher behavior data will be modeled using the circular mixed-effects model described in Section 5.2. We use this model to predict the score on the circumplex (`IPC`), our circular outcome, using the variables `EX` and `Year`. We also include a random intercept. We will use a model building procedure for this data. This means that we will start by fitting a so-called intercept-only model, where we include only the intercept and no other predictors. In the subsequent models, predictors are added one by one to the intercept-only model. The model that includes all predictors is:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^{I} \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (\beta_0^{I} + \beta_1^{I}\text{EX}_{ij} + \beta_2^{I}\text{Year}_i + b_{0i}^{I} \\ (\beta_0^{II} + \beta_1^{II}\text{EX}_{ij} + \beta_2^{II}\text{Year}_i + b_{0i}^{II} \end{pmatrix}.$$

To decide on the exclusion of predictors we will use four different model selection criteria: two version of the deviance information criterion (DIC and DIC_{alt}) and two versions of the Watanabe-Akaike information criterion (WAIC_1 and WAIC_2) (Gelman et al., 2014). We choose these four criteria because they are specifically useful in Bayesian models where MCMC methods have been used to estimate the parameters. All four criteria have a fit part consisting of a measure based on the loglikelihood and include a penalty in the form of an effective number of parameters. For all criteria lower values indicate better fit. Gelman et al. (2014) describes how to compute these criteria. We fit the models in `R` (R Core Team, 2017) using the package `bpnreg` (Cremers, 2018) and all continuous variables will be centered on their grand mean before inclusion in the analysis.

Before evaluation of the results the convergence of the MCMC samplers was checked

Table 5.2: Model fit statistics for dataset I

Statistic	Intercept-only	EX	EX + Year
DIC	1417	1375	1381
DIC_{alt}	2103	2046	2044
$WAIC_1$	1390	1351	1355
$WAIC_2$	1482	1438	1444

by means of traceplots. For all models the parameters showed proper convergence within 1000 iterations (burn-in = 1000, lag = 3).

5.4.1.2 Model fit

First we compare the fit of the intercept-only model, the model including the predictor `EX` and the model with the covariate `Year`. In Table 5.2 the model fit statistics are shown for these three models. From the statistics in this table we conclude that both models with predictors fit better than the intercept-only model. All four model fit statistics are smaller, e.g. the DIC decreases from 1417 (intercept-only) to 1375/1381 (`EX`/`EX + Year`). The addition of the covariate `Year` does not improve the fit of the model (slightly larger values on 3 of the 4 criteria). This indicates that the year of the first measurement does not have an effect on a teachers' average score on the IPC. We therefore decide to continue with the model that contains only the predictor years of experience (`EX`). First we show results for the fixed effects in Section 5.4.1.3. Subsequently we show results for the random effects in Section 5.4.1.5.

5.4.1.3 Fixed effects

In the R-package `bpnreg` the circular fixed effects are computed as described in Section 5.3.1. In Table 5.3, estimates of the posterior mean, mode, and standard deviation (sd) and the lower and upper bound of the 95% highest posterior density (HPD) interval for the three types of circular coefficient for the effect of `EX` are shown. The standard deviation of a posterior is an estimate for the standard error of the parameter. The HPD interval is the smallest interval in which 95% of the posterior mass is located. In terms of interpretation, it is different from a frequentist confidence interval since HPD intervals allow for probability statements. For example, if the 95% HPD interval for a parameter μ runs from 2 to 4 we can say that the probability that μ lies between 2 and 4 is 0.95.

All three coefficients are small but different from 0 as indicated by their HPD interval. That means that on average there is an effect of years of experience on a teachers' score on the circumplex. To be more precise, at the inflection point an increase of 1 unit in `EX` results in a counterclockwise move of $b_c = 0.10$ radians or $0.10 * (180/\pi) = 5.73^\circ$ on the circumplex. At the grand mean of `EX` an increase of 1 unit in `EX` results in a counterclockwise move of $SAM = 0.04$ radians or $0.04 * (180/\pi) = 2.29^\circ$ on the circumplex. On average an increase of 1 unit in `EX`

Table 5.3: Descriptives of the posterior distributions of several circular regression coefficients for the effect of EX in dataset I

Statistic	Mode	Mean	sd	LB HPD	UB HPD
b_c	0.10	0.09	0.04	0.05	0.15
SAM	0.04	0.04	0.01	0.02	0.05
AS	0.04	0.04	0.02	0.02	0.05

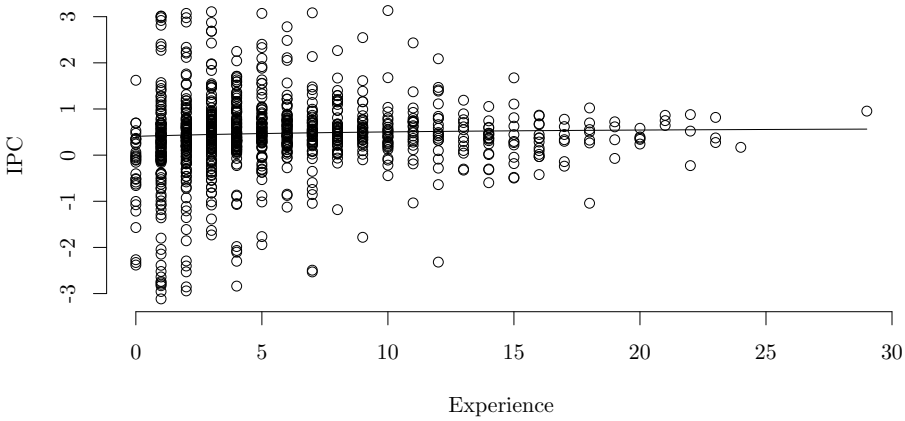

Figure 5.5: The effect of experience on the location of teachers' scores on the circumplex for dataset I.

results in a counterclockwise move of $AS = 0.04$ radians or $0.04 * (180/\pi) = 2.29°$ on the circumplex.

The relation between years of experience and the score on the IPC is plotted in Figure 5.5. The line in this figure represents the effect of `EX` on the location of the score of a teacher on the IPC and the dots represent the data. The predicted value at 0 years of experience is 23° and at 29 years of experience it is 32°. Both these values fall in the 'Helpful' category of the IPC. The line is not steep, reflecting the small values of the estimates for the circular coefficients. However, the data does show that although the location of a teachers' score does not change much over time, the variance of the scores of the teachers on the circumplex does seem to change. We see that the scores of teachers with a low amount of experience (`EX` = 0) are spread across the entire circle while the scores of teachers with a high amount of experience (e.g. `EX` = 20) are much more concentrated at a specific region on the circle. In Section 5.4.1.4 we will further investigate how the experience of a teacher influences the variance of scores on the circumplex.

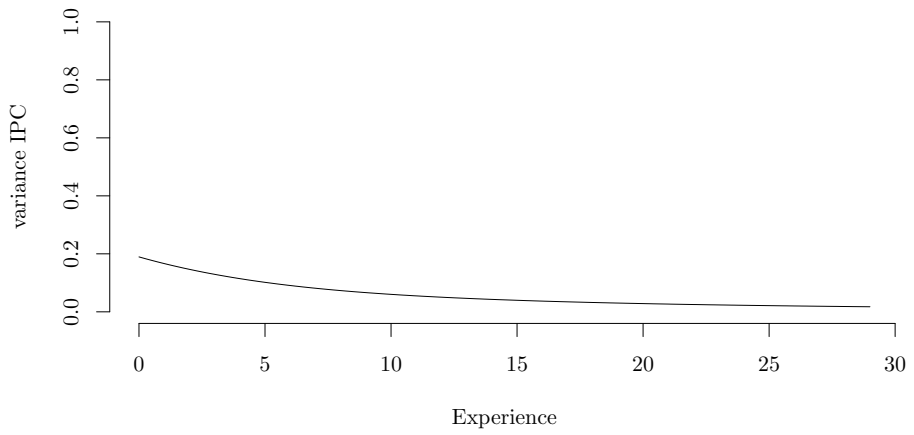

Figure 5.6: The effect of experience on the variance of teachers' scores on the circumplex for dataset I.

5.4.1.4 Effects on spread

In Section 5.4.1.3 we concluded from Figure 5.5 that there may be an effect of `EX` on the spread or variance of scores of teachers on the interpersonal circumplex. Scores from teachers with less experience are much less concentrated on the circumplex than scores from teachers with a lot of experience.

We compute values for this effect at different years of experience using the methods from Section 5.3.1.2. Figure 5.6 shows these estimates. This plots reflects, but now also quantifies, the pattern that we already observed in the data. For less experienced teachers, e.g. `EX` $= 0$, the variance of scores on the circle is larger at $\hat{\rho} \approx 0.2$ than for experienced teachers, $\hat{\rho} \to 0$.

5.4.1.5 Random effects

The R-package `bpnreg` also computes random effects variances following the methods described in Section 5.3.2 and 5.3.2.1. The variances of the linear random intercepts are estimated at 1.69 (HPD: 1.21,2.32) for the first or Communion component and 0.81 (HPD: 0.58, 1.09) for the second or Agency component. On the circle this translates to an estimated circular variance $(1 - \hat{\rho})$ of 0.25 (HPD: 0.19, 0.30). This means that there is variance in the teachers' score at 0 years of experience, and thus the teachers's scores differ on the circumplex at that point. This was to be expected from the plotted data in Figure 5.5. It is clear from this figure that at 0 years of experience (`EX` $= 0$) there is a lot of variance between teachers in their score on the circumplex. The intercept variance could be explained by additional teacher

Table 5.4: Descriptives for dataset II

Variable		mean/$\bar{\theta}$	sd/$\hat{\rho}$	Range	Type
IPC	male	32.56°	0.80	-	Circular
	female	39.08°	0.80	-	Circular
Time		1.38	1.19	0-3	Linear
EX		5.27	1.15	1.83-7	Linear
SE		4.77	0.85	1.75-6.5	Linear

characteristics. Unfortunately, dataset I does not contain additional covariates to further explore this.

5.4.2 Dataset II

This dataset contains several repeated measures on the IPC of 161 teachers gathered for the studies of Pennings et al. (2018), Claessens (2016) and Van der Want (2015). The measurements were obtained using the QTI and taken in different years and in different classes. For this paper we only consider the year and take the measurements for the largest class if data for multiple classes were available in one year. We ended up with a maximum of four measurements per teacher. In addition to a variable `Time` specifying the measurement occasion (year), the years of experience of the teacher at the first measurement (`EX`), a teachers' self-efficacy in dealing with student emotions (`SE`) , and the gender of the teacher (`Gender`) will be used as covariates in the analysis. Note that both `EX` and `Gender` are constant over measurements while `SE` is different for each measurement occasion. Table 5.4 shows descriptives for the circular outcome, `IPC`, and the predictors in the model.

5.4.2.1 The model

This second dataset will also be modeled using the circular mixed-effects model, predicting the score on the circumplex (IPC) using several predictor variables. We use a model building procedure in which we first fit an intercept-only model and then the model including the `Time` predictor. We also include a random intercept. The third model we will fit to dataset II includes the covariates `SE`, `EX` and `Gender`:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^{I} \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (\beta_0^I + \beta_1^I \text{Time}_{ij} + \beta_2^I \text{SE}_{ij} + \beta_3^I \text{EX}_{ij} + \beta_4^I \text{Gender}_{ij} + b_{0i}^I \\ (\beta_0^{II} + \beta_1^{II} \text{Time}_{ij} + \beta_2^{II} \text{SE}_{ij} + \beta_3^{II} \text{EX}_{ij} + \beta_4^{II} \text{Gender}_{ij} + b_{0i}^{II} \end{pmatrix}.$$

The fit of the models will be assessed using the same criteria as for dataset I and `SE` and `EX` will be centered at their grand mean before inclusion in the analysis.

Before evaluation of the results convergence was checked by means of traceplots. For all models the parameters showed proper convergence within 1000 iterations (burn-in = 1000, lag = 3).

Table 5.5: Model fit statistics for dataset II

Statistic	Intercept-only	Time	Time + SE + EX + Gender
DIC	462	469	460
DIC_{alt}	1476	1262	1354
$WAIC_1$	443	441	438
$WAIC_2$	513	511	510

5.4.2.2 Model fit

First we check the fit of the intercept-only model, the model including the predictor `Time` and the model with additional predictors `SE`, `EX` and `Gender`. In Table 5.5 the model fit statistics are shown for these three models. From the statistics in this table we conclude that the model including `Time` and the model including `Time` and three additional covariates fit better than the intercept-only model, the model fit statistics are smaller for respectively 3 and 4 of the criteria. The third model with three additional covariates shows improvement in 3 model fit criteria compared to the model with `Time`. We therefore decide to continue with that model.

5.4.2.3 Fixed effects

In Table 5.6 estimates of the posterior mean, mode and standard deviation (sd) and the lower and upper bound of the highest posterior density (HPD) interval for the three types of circular coefficient for the effects of `Time`, `SE` and `EX` are shown. Because the model contains a categorical covariate, `Gender`, we display a marginal effect for each variable at each of the levels of the covariate `Gender`. We do this because as in a logistic regression model the marginal effect of a predictor in a PN model, such as the circular mixed-effects model from this paper, is different for different levels or values of the other predictors in the model. This is caused by the fact that we are fitting a model in which the relation between outcome and predictors is non-linear.

Two coefficients, SAM and AS, for the effect of `SE` are small but different from 0 at both levels of `Gender` as indicated by their HPD intervals. That means that at `Time` = 0 and for a teacher with average experience there is an effect of `SE` on a teachers' score on the circumplex at the grand mean and on average. To be more precise, at the grand mean of `SE` an increase of 1 unit in `SE` results in a counterclockwise move on the IPC of $SAM = 0.12$ radians or $0.12 * (180/\pi) = 6.88°$ for `Gender` = male and $SAM = 0.17$ radians or $0.17 * (180/\pi) = 9.74°$ for `Gender` = female. On average, an increase of 1 unit in `SE` results in a counterclockwise move of $AS = 0.12$ radians or $0.12 * (180/\pi) = 6.88°$ for `Gender` = male and $AS = 0.18$ radians or $0.18 * (180/\pi) = 10.31°$ for `Gender` = female on the IPC. Figure 5.7 shows the effects of `SE` on the IPC for both male and female teachers. We see that the score on the circumplex slightly changes with increasing self-efficacy for both male and female teachers. Note that the difference between the circular regression coefficients

Table 5.6: Descriptives of the posterior distributions of several circular regression coefficients for the effect of the continuous predictors in dataset II

		Statistic	Mode	Mean	sd	LB HPD	UB HPD
Gender = male	Time	b_c	0.04	0.03	0.04	-0.05	0.09
		SAM	0.02	0.01	0.53	-0.15	0.19
		AS	0.02	0.03	0.70	-0.16	0.19
	SE	b_c	0.28	0.24	0.19	-0.32	0.50
		SAM	0.12	0.18	0.18	0.05	0.68
		AS	0.12	0.15	0.69	0.04	0.71
	EX	b_c	0.01	0.02	0.50	-0.01	0.02
		SAM	0.00	0.00	0.58	-0.15	0.15
		AS	0.00	-0.01	0.45	-0.10	0.17
Gender = female	Time	b_c	0.20	0.11	3.84	-0.28	0.46
		SAM	0.09	0.04	3.77	-0.12	0.34
		AS	0.07	0.06	1.42	-0.10	0.23
	SE	b_c	0.35	0.36	0.92	0.09	0.73
		SAM	0.17	0.14	0.87	0.05	0.35
		AS	0.18	0.17	0.14	0.07	0.34
	EX	b_c	0.18	0.24	1.79	-0.26	0.48
		SAM	0.10	0.18	1.76	-0.12	0.36
		AS	0.11	0.07	0.05	-0.07	0.12

for the two levels of `Gender` does not differ from 0; their HPD intervals overlap.

To investigate the effect of gender at `Time` = 0 for teachers with average scores on self-efficacy and experience we look at the predicted scores on the IPC for males and females. The predicted score on the IPC for males is estimated at $0.48\,(0.39; 0.61)$ and for females at $0.57\,(0.50; 0.70)$. Their HPD intervals overlap which means that on average, male and female teachers do not differ in their scores on the IPC. For `Time` and `EX` none of the coefficients are different from 0 according to their HPD interval in Table 5.6.

5.4.2.4 Random effects

Table 5.7 shows the posterior modes of the linear and circular random intercept variances for the three models that were fit to dataset II. We see that from the Intercept-only to the model with `Time` the posterior mode of the circular variance increases slightly. This is a phenomenon that occurs more often in mixed-effects models that are fit to longitudinal data (Hox, 2002). The increase in variance is caused by the fact that the model is based on the assumption that the measurements within an individual are random samples from a population of possible measurements and assumes a certain variance for these measurements. However, in a longitudinal model the repeated measures are usually fixed and therefore have a lot less variance than expected by the model. Therefore, it is possible that we see an increase in intercept variance from the model without the predictor `Time`, that indicates

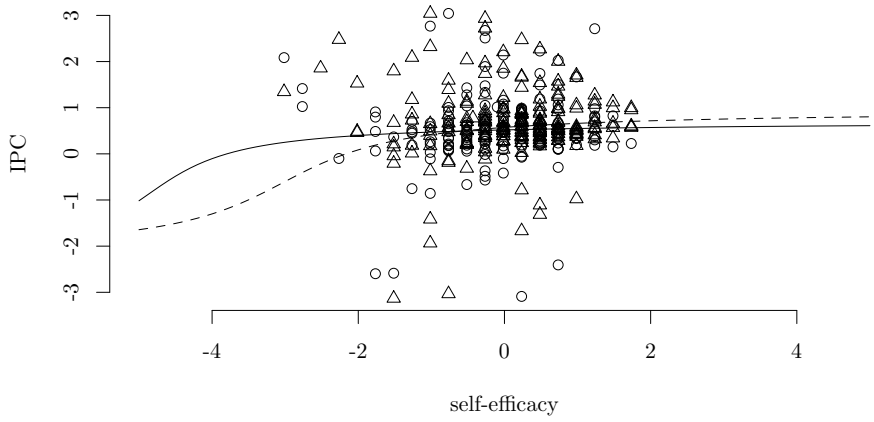

Figure 5.7: The effect of self-efficacy on the score on the IPC for dataset II. We distinguish between the effect for male (solid line and circular datapoints) and female teachers (dashed line and triangular datapoints).

measurement occasion, and the model with `Time`. However, from Table 5.7 we also see that the HPD intervals of the variances from these two models overlap. So, if there is any increase in variance that is different from 0 it is so small that it cannot be detected using the circular mixed-effects model on this dataset.

In contrast to the best fitting model for dataset I, the best fitting model for dataset II contains additional covariates (apart from `Time`). Therefore, we may try to explain some of the intercept-variance from the model with `Time` using those covariates. From the model with `Time` to the model with `Time`, `SE`, `EX` and `Gender` the posterior modes of all variances, linear and circular, decrease. However, their HPD intervals overlap meaning that the decrease in variance is not larger than 0. This means that the three additional covariates, `Gender`, `SE` and `EX` do not explain a part of the circular or linear intercept variance.

5.5 Discussion

In this paper we have shown how a circular mixed-effects model can be used to model repeated measures from an interpersonal circumplex for teacher behavior. We have also developed new interpretation tools that have solved the interpretation problems associated with the circular (PN) mixed-effects model. This model together with the new interpretation tools has allowed us to interpret the effect of covariates on a directional score on the IPC itself instead of on its two separate components Agency and Communion.

Table 5.7: Posterior mode estimates and their HPD interval for the linear and circular random intercept variances for dataset II.

Model	Component I (HPD)	Component II (HPD)	Circular (HPD)
Intercept-only	2.51 (1.85; 4.50)	0.79 (0.36; 1.22)	0.11 (0.08; 0.14)
Time	2.73 (1.84; 4.32)	0.79 (0.43; 1.25)	0.12 (0.08; 0.16)
Time + SE + EX + Gender	2.72 (1.84; 4.17)	0.59 (0.29; 0.96)	0.11 (0.06; 0.17)

For the first dataset we can compare the results from the analysis with the circular mixed-effects model to previous analyses. In Brekelmans et al. (2005) the dataset that we took a subset from to create dataset I is analysed using a similar model to the one in this paper. In their analysis, student perceptions of teachers split in a Proximity (Communion) and Influence (Agency) score are predicted by the teachers' experience in a multilevel growth model. Their findings are that students perceptions of a teachers' Communion remain stable over the teacher career while perceptions of a teachers' Agency grow in the first 6 years of a teacher career. In our analysis we have however combined the effects on Agency and Communion which means that we can characterize a change in teachers' behavior on a blend of these dimensions We have shown that there is a small effect of experience on a teachers' score on the IPC. They move from a score of 23° at 0 years of experience to a score of 34° at 29 years of experience. This means that over the course of the teaching career teachers move more towards the center (45°) of the 'preferred' styles of teacher behavior which is comprised of the 'Directing' and 'Helpful' types of Figure 5.1. Over time teachers thus develop their style of teacher behavior more towards the preferred style of teacher behavior. Such a conclusion could not have been reached when we had used an analysis on the 8 subtypes; 23° and 34° both fall within the 'Helpful' subtype. Using results for Agency and Communion from Brekelmans et al. (2005) could give us an indication of the effect we expect on the IPC itself but this effect could not have been quantified in the sense that we could test if it was different from zero. However, treating data on the IPC as circular as done in this paper does allow us to do so.

In addition to being able to investigate the effect of a covariate on the location of a score on the IPC itself instead of on Agency and Communion separately, the circular mixed-effects model allows us to investigate effects on the spread of scores on the IPC. This gives us additional insights compared to the analysis from Brekelmans et al. (2005). From the analysis in Section 5.4.1 we conclude that not only the location of the score on the IPC changes with experience but also the spread changes with experience. To be more precise, the scores of more experienced teachers are more concentrated in the preferred type of teaching behavior than the scores of less experienced teachers. Note that although this is an indication that all teachers on average move toward the preferred type of behavior this does not mean that each individual teacher necessarily does so. To reach a more formal conclusion about

this we have to look at the effect of experience on the score on the IPC for each individual separately and include a random slope for experience into the model.

Since in the first dataset no covariates were available we used a second dataset to show how covariates can be included in the models to explain variance in the circular data. This is the first time that the longitudinal circular data in this dataset have been analysed. Although the models including effects over time and those of time combined with gender, experience and self-efficacy improved model fit, unfortunately, the effects of the covariates were too small to explain any variance in the circular or linear intercepts. Still, including this dataset enabled us to show how covariates can be included in the analysis and attends to the need for an analysis method to analyze Agency and Communion blended together. It thus provides researchers with the means to study associations between covariates and the circular IPC data without analyzing the two dimensions separately. The associations found in this dataset may be quite interesting for practice. For example, we concluded from the analysis that gender does not matter for the quality of teacher-student relationships in dataset II, and that although the self-efficacy of teachers may vary, it does not necessarily affect the quality of teacher-student relationships to a great extent. This is important, because especially early career teachers may worry about their ability to teach. Fortunately, that does not necessarily seem to affect the quality of teacher-student relationships, which means that teachers can still develop towards establishing the preferred type of teacher behavior style (as found in dataset 1) even though they do not feel confident in their teaching yet.

In this paper we have used the embedding approach to circular data. Two other approaches to analyze circular data are the wrapping and intrinsic approach (Mardia & Jupp, 2000). In the intrinsic approach distributions that are directly defined on the circle, such as the von Mises distribution, are used to model the data. In the wrapping approach the data is modeled by wrapping a univariate distribution defined on the real line, for example the Normal distribution, onto the circle. Even though more complex models have been introduced for these approaches (Lagona, 2016; Wang & Gelfand, 2014), a mixed-effects model was not among them. An advantage of models from the intrinsic and wrapping approaches is that their results are easier to interpret compared to models from the embedding approach. Therefore, we have in this paper introduced new interpretation tools for the circular mixed-effects model based on the PN distribution that solved the interpretation problems associated with the embedding approach.

A possible critique on the way we have analyzed circumplex data in this paper is that we have only considered information on the angle resulting from the conversion of a score on Agency and Communion to the IPC. This includes directional information contained in the direction of the 2-dimensional vector of an Agency and Communion score but excludes information on the 'size' or 'intensity' of this vector. In contrast, modelling Agency and Communion separately allows us to model the 'intensity' but excludes directional information. Models that allow for simultaneous modelling of a circular and linear variable (direction and intensity) however do exist. In the literature Mardia & Sutton (1978), Abe & Ley (2017), Mastrantonio, Maruotti, & Jona-Lasinio (2015) and Mastrantonio (2018) have introduced models for these so-called cylindrical data. However, thus far these models only model

the relation between the linear and circular variable and their respective means but do not introduce a regression structure to predict their means using additional covariates. Additionally, the models do not allow for the modelling of multiple measurements over time. In future research it would be useful to extend existing models for cylindrical data with a regression structure and a structure that allows for longitudinal data. We could then apply these extended cylindrical models to the datasets analysed in this paper.

In this paper we have included methods that allow for the analysis of a circumplex outcome variable, however, in certain cases these variables also serve as a predictor variable. At first glance modelling circumplex predictors seems much easier than modelling circumplex outcomes. A circular predictor can namely be modeled in a standard model for 'linear' outcomes by including the cosine and the sine of the predictor as two predictor variables into the model. However, if we do this for a circumplex predictor this is equivalent to including the Agency and Communion component of the predictor into the model separately. If we are interested in the effect of a blend of Agency and Communion we should come up with a different modelling strategy for circumplex predictors. The reparameterization used in this paper for obtaining effects on the circle may provide a solution for this problem. Also different methods for including circular predictors have been developed in the literature (e.g. in Kim & SenGupta (2015)). In further research it may be worthwhile to explore the case of circumplex predictors further.

In conclusion, we have shown that modelling longitudinal data from the IPC using a circular model is possible. It offers us a different perspective to the data that is more in line with the original idea of the IPC as "as a continuous order with no beginning or end" (Gurtman, 2009, p. 2). In addition we have been able to solve interpretation issues of the specific model used in this paper.

Chapter 6

Regression models for cylindrical data in psychology

by J. Cremers, H.J.M. Pennings & C. Ley[1]

6.1 Introduction

Cylindrical data are data that consist of a linear variable and a directional variable. In this paper, the directional variable is circular, meaning that it consists of a single angle instead of a set of angles. A circular variable is different from a linear variable in the sense that it is measured on a different scale. Figure 6.1 shows the difference between a circular scale (right) and a linear scale (left). The most important difference is that on a circular scale the datapoints 0° and 360° are connected and in fact represent the same number while on a linear scale the two ends, $-\infty$ and ∞, are not connected and consequently the values 0° and 360° are located on different places on the scale. This difference requires us to use special statistical methods for circular variables (see e.g. Fisher (1995) for an introduction to circular data and Mardia & Jupp (2000), Jammalamadaka & Sengupta (2001) and Ley & Verdebout (2017) for a more elaborate overview). For instance, a notion as simple as the sample average needs to be re-defined. As is the case for circular data, the analysis of cylindrical data also requires special methods.

Cylindrical data occur in several fields of research, such as for instance in meteorology (Garcia-Portugués, Crujeiras, & González-Manteiga, 2013), ecology

[1]Manuscript submitted to Multivariate Behavioral Research

Author contributions: JC designed the study, JC developed the new models and analyzed, processed and interpreted the results with feedback from CL, HJMP provided the empirical dataset, JC wrote the paper (with CL) and HJMP gave feedback on the written work.

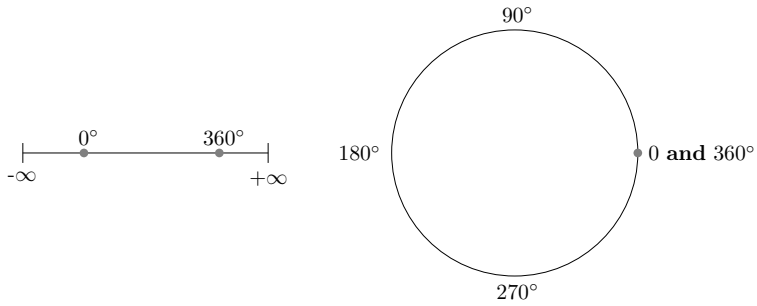

Figure 6.1: The difference between a linear scale (left) and a circular scale (right).

(Garcia-Portugués, Barros, Crujeiras, González-Manteiga, & Pereira, 2014) or marine research (Lagona, Picone, Maruotti, & Cosoli, 2015). However, also in psychology several types of cylindrical data can be found. For example, in research on human navigation in the field of cognitive psychology both the distance, a linear variable, and the direction, a circular variable, of movement are of interest (Chrastil & Warren, 2017). In eye-tracking research, saccade data are an example of cylindrical data, because both the direction (*i.e.*, the circular variable) and the duration (*i.e.*, the linear variable) of the saccades are of interest (for a review of eye-tracking research see Rayner (2009)). The type of data that is used in the present study are also psychological, namely data from circumplex measurement instruments. For instance, data from the interpersonal circumplex as used in personality psychology are by definition of a cylindrical nature (see Section 6.2 for a more detailed explanation).

In the literature, several methods have been put forward to model the relation between the linear and circular component of a cylindrical variable. Some of these are based on regressing the linear component onto the circular component using the following type of relation:

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

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

Our aim in this paper is to fill this gap in the literature by adapting four existing cylindrical models in such a way that they include a regression of both the linear and circular component of a cylindrical variable onto a set of covariates. From now on we will therefore refer to the components of the cylindrical variable as outcome

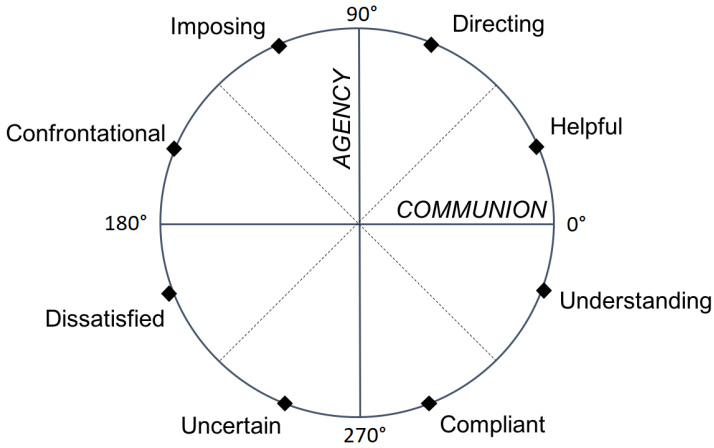

Figure 6.2: The interpersonal circle for teachers (IPC-T). The words presented in the circumference of the circle are anchor words to describe the type of behavior located in each part of the IPC.

components. Additionally, we will show how a correct statistical treatment of such cylindrical data can lead to new insights. We will do this for the teacher data, a dataset from the field of educational psychology. In the teacher data, apart from modelling the relation between the linear and circular component of a cylindrical variable we would also like to predict the two components from a set of covariates in a regression model.

The paper is organized as follows: Section 6.2 describes the teacher data, while Section 6.3 presents the four cylindrical models and our associated adaptations to new regression models. In that same section we also discuss the model fit criterion that we will use in Section 6.4 for the comparison of the four models. A detailed data analysis with interesting new insights is also provided in Section 6.4. We conclude the paper with a discussion in Section 6.5, and the Appendix collects the technical details of the MCMC procedures.

6.2 Teacher data

The motivating example for this article comes from the field of educational psychology and was collected for the studies on classroom climate of Van der Want (2015), Claessens (2016) and Pennings et al. (2018). An indicator of the quality of the classroom climate is the students' perception of their teachers' interpersonal behavior. These interpersonal perceptions, both in educational psychology as well as in other areas of psychology, can be measured using circumplex measurement instruments (see Horowitz & Strack (2011) for an overview of many such instruments).

The circumplex data used in this paper are measured using the Questionnaire on

Teacher Interaction (QTI) (Wubbels et al., 2006) which is one such circumplex measurement instrument. The QTI is designed to measure student perceptions of their teachers' interpersonal behavior and contains items that load on two interpersonal dimensions: Agency and Communion. Agency refers to the degree of power or control a teacher exerts in interaction with his/her students. Communion refers to the degree of friendliness or affiliation a teacher conveys in interaction with his/her students. The loadings on the two dimensions of the QTI can be placed in a two-dimensional space formed by Agency (vertical) and Communion (horizontal), see Figure 6.2. Different parts of this space are characterized by different teacher behavior, e.g. 'helpful' or 'uncertain'. This two-dimensional space is called the interpersonal circle/circumplex (IPC). The IPC is *"a continuous order with no beginning or end"* (Gurtman, 2009, p. 2). We call such ordering a circumplex ordering and the IPC is therefore often called the interpersonal circumplex. The ordering also implies that scores on the IPC could be viewed as a circular variable.

Cremers et al. (2018a) explain the circular nature of the IPC data and analyze them as such using a circular regression model. The two-dimension scores Agency and Communion can be converted to a circular score using the two-argument arctangent function in (6.1), where A represents a score on the Agency dimension and C represents a score on the Communion dimension

$$\theta = \operatorname{atan2}(A,\, C) = \begin{cases} \arctan\left(\frac{A}{C}\right) & \text{if} \quad C > 0 \\ \arctan\left(\frac{A}{C}\right) + \pi & \text{if} \quad C < 0 \;\&\; A \geq 0 \\ \arctan\left(\frac{A}{C}\right) - \pi & \text{if} \quad C < 0 \;\&\; A < 0 \\ +\frac{\pi}{2} & \text{if} \quad C = 0 \;\&\; A > 0 \\ -\frac{\pi}{2} & \text{if} \quad C = 0 \;\&\; A < 0 \\ \text{undefined} & \text{if} \quad C = 0 \;\&\; A = 0. \end{cases} \tag{6.1}$$

The resulting circular variable θ can then be modeled and takes values in the interval $[0, 2\pi)$. However, when two-dimensional data are converted to the circle we lose some information, namely the length of the two-dimensional vector $(A, C)^t$, *i.e.*, its Euclidean norm $||\ (A, C)^t\ ||$. This length represents the strength of the type of interpersonal behavior a teacher shows towards his/her students and can be considered as the linear variable in a cylindrical model, allowing us to model a circular variable θ together with the linear variable corresponding to $||\ (A, C)^t\ ||$. This leads to an improved analysis of interpersonal circumplex data as we take all information into account. In the next section we introduce several models that can be used for a more accurate and informative regression analysis on the teacher data. First however we will provide descriptives for our data set.

6.2.1 Data description

The teacher data was collected between 2010 and 2015 and contains several repeated measures on the IPC of 161 teachers. Measurements were obtained using the QTI and taken in different years and classes. For this paper we only consider one measurement, the first occasion (2010) and largest class if data for multiple classes were available. In addition to the score on the IPC, the circular outcome, and the strength of the score on the IPC, the linear outcome, a teachers' self-efficacy

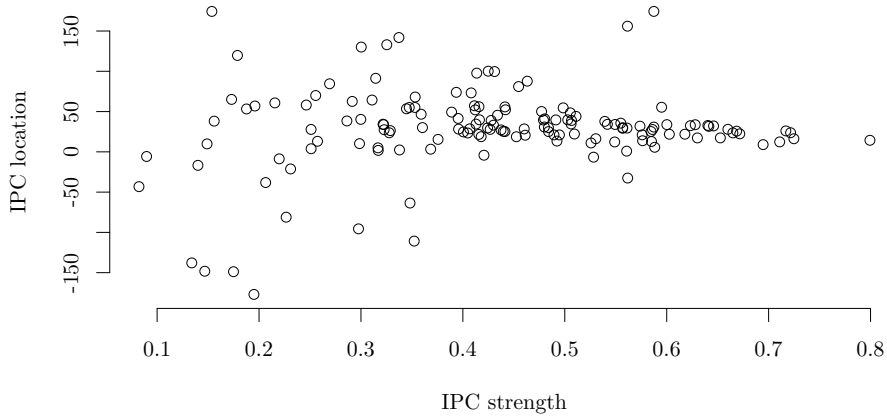

Figure 6.3: Plot showing the relation between the linear and circular outcome component (in degrees) of the teacher data.

(SE) concerning classroom management is used as covariate in the analysis. After listwise deletion of missings (3 in total, only for the self-efficacy) we have a sample of 148 teachers. Table 6.1 shows descriptives for the dataset. Here $\hat{\rho}$ is a sample estimate for the circular concentration where a value of 0 means that the data is not concentrated at all, *i.e.* spread over the entire circle, and a value of 1 means that all data is concentrated at a single point on the circle. Figure 6.3 is a scatterplot showing the relation between the linear and circular outcome of the teacher data.

Table 6.1: Descriptives for the teacher dataset.

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

6.3 Four cylindrical regression models

In this section we present four cylindrical models and adapt them such that they contain predictors for the linear and circular outcomes, Y and Θ. The first two models are based on a construction by Mastrantonio et al. (2015), while the other models are extensions of the models from Abe & Ley (2017) and Mastrantonio (2018).

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

Following Mastrantonio et al. (2015) we consider in this section two models where the relation between $\Theta \in [0, 2\pi)$ and $Y \in (-\infty, +\infty)$ and q covariates is specified as

$$Y = \gamma_0 + \gamma_{cos} * \cos(\Theta) * R + \gamma_{sin} * \sin(\Theta) * R + \gamma_1 * x_1 + \cdots + \gamma_q * x_q + \epsilon, \quad (6.2)$$

where the random variable $R \geq 0$ will be introduced below, the error term $\epsilon \sim N(0, \sigma^2)$ with variance $\sigma^2 > 0$, $\gamma_0, \gamma_{cos}, \gamma_{sin}, \gamma_1, \ldots, \gamma_q$ 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 outcome 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 teacher dataset, the regression equation for the linear outcome in the CL-PN and CL-GPN model is the following:

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

where SE_i is the self-efficacy score of one individual $i = 1, \ldots, n$ where n is the sample size.

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

$$\boldsymbol{S} = \begin{bmatrix} S^I \\ S^{II} \end{bmatrix} = R\boldsymbol{u} = \begin{bmatrix} R\cos(\Theta) \\ R\sin(\Theta) \end{bmatrix}, \quad (6.3)$$

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 $\boldsymbol{\mu} \in \mathbb{R}^2$, $\boldsymbol{\Sigma} = \begin{bmatrix} \tau^2 + \rho^2 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho \in (-\infty, +\infty)$ and $\tau^2 \geq 0$ (as in Hernandez-Stumpfhauser et al. (2017)). 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.

6.3.1.1 The modified CL-PN distribution

Following Nuñez-Antonio et al. (2011), the joint density of Θ and R for the PN distribution equals

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{I}) = \frac{r}{2\pi} \exp\left[-\frac{(r\boldsymbol{u} - \boldsymbol{\mu})^t (r\boldsymbol{u} - \boldsymbol{\mu})}{2}\right], \tag{6.4}$$

where $\boldsymbol{u} = \begin{bmatrix}\cos(\theta) \\ \sin(\theta)\end{bmatrix}$ and r is the same as in (6.2) and (6.3) and is defined in (6.3). In a regression setup the outcomes 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 (6.4). 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. Note however that the dimensions of $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$ need not necessarily be the same and we are thus allowed to have a different set of predictor variables and vectors $\boldsymbol{z}_i^I$ and $\boldsymbol{z}_i^{II}$ for the two components of $\boldsymbol{\mu}_i$. For the teacher dataset, the regression equation for the circular outcome in the CL-PN model is

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

6.3.1.2 The modified CL-GPN distribution

Following Wang & Gelfand (2013) and Hernandez-Stumpfhauser et al. (2017) the joint density of R and Θ for the GPN distribution equals

$$f(\theta, r \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{r}{2\pi\tau} \exp\left[-\frac{(r\boldsymbol{u} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (r\boldsymbol{u} - \boldsymbol{\mu})}{2}\right], \tag{6.5}$$

where we recall that $\boldsymbol{\Sigma} = \begin{bmatrix}\tau^2 + \rho^2 & \rho \\ \rho & 1\end{bmatrix}$. In a regression setup the outcomes θ_i and r_i for each individual are generated independently from (6.5). 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. Note that in contrast with the CL-PN model where $\boldsymbol{z}_i^I$ and $\boldsymbol{z}_i^{II}$ are allowed to differ we do need to have the same predictors for both components of $\boldsymbol{\mu}_i$ in the CL-GPN model. This is due to the fact that the variance-covariance matrix $\boldsymbol{\Sigma}$ is no longer identity. For the teacher dataset, the regression equation for the circular outcome in the CL-GPN model is the same as in the CL-PN model.

6.3.1.3 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. (2017) for the regression of the circular

outcome. A detailed description of the Bayesian estimation and MCMC samplers can be found in the Appendix.

6.3.2 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

$$f(\theta, y) = \frac{\alpha\beta^{\alpha}}{2\pi\cosh(\kappa)}(1 + \lambda\sin(\theta - \mu))y^{\alpha-1}\exp[-(\beta y)^{\alpha}(1 - \tanh(\kappa)\cos(\theta - \mu))], \tag{6.6}$$

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 $\beta > 0$ and circular location parameter $\mu \in [0, 2\pi)$, both of which we express in terms of covariates: $\beta_i = \exp(\boldsymbol{x}_i^t\boldsymbol{\nu}) > 0$ and $\mu_i = \eta_0 + 2\tan^{-1}(\boldsymbol{z}_i^t\boldsymbol{\eta})$. The parameter $\boldsymbol{\nu}$ is a vector of q regression coefficients $\nu_j \in (-\infty, +\infty)$ for the prediction of y where $j = 0, \ldots, q$ and ν_0 is the intercept. The parameter $\eta_0 \in [0, 2\pi)$ is the intercept and $\boldsymbol{\eta}$ is a vector of p regression coefficients $\eta_j \in (-\infty, +\infty)$ for the prediction of θ where $j = 1, \ldots, p$. The vector $\boldsymbol{x}_i$ is a vector of predictor values for the prediction of y and $\boldsymbol{z}_i$ is a vector of predictor values for the prediction of θ. In a regression setup the outcome vector $(\theta_i, y_i)^t$ for each individual is generated independently from the modified density (6.6).

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

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

For the teacher data, $\boldsymbol{z} = \boldsymbol{x}$ and the regression equations for the circular and linear outcomes in the Abe-Ley model are:

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

and

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

We can use numerical optimization (Nelder-Mead) to find solutions for the maximum

likelihood (ML) estimates for the parameters of the model.

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

This model is an extension of the cylindrical model introduced by Mastrantonio (2018) to the regression context. Both models contain m independent circular outcomes and w independent linear outcomes. The circular outcomes $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_m)$ are modeled together by a multivariate GPN distribution. The joint distribution of $\boldsymbol{\Theta}$ and $\boldsymbol{R}$ can thus be modeled as the product of (6.5) for each of the m circular outcomes. The linear outcomes $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_w)$ are modeled together by a multivariate skew normal distribution (Sahu, Dey, & Branco, 2003). Because the GPN distribution is modeled using a so-called augmented representation (as in (6.3) and (6.5)) it is convenient to use a similar tactic for modelling the multivariate skew normal distribution. Following Mastrantonio (2018) the linear outcomes are represented as

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

where $\boldsymbol{\mu}_y$ is a mean vector for the linear outcome $\boldsymbol{Y}$, $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ 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 $\boldsymbol{\mu}_y + \boldsymbol{\Lambda}\boldsymbol{D}$ and covariance matrix $\boldsymbol{\Sigma}_y$. The joint density for $(\boldsymbol{Y}^t, \boldsymbol{D}^t)^t$ is defined as:

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

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

$$f(\boldsymbol{\theta}, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{d}) = 2^w \phi_{2m+w}((\boldsymbol{s}^t, \boldsymbol{y}^t)^t \mid \boldsymbol{\mu} + (\boldsymbol{0}_{2m}^t, (\text{diag}(\boldsymbol{\lambda})\boldsymbol{d})^t)^t, \boldsymbol{\Sigma})\phi_w(\boldsymbol{d} \mid \boldsymbol{0}_w, \boldsymbol{I}_w) \prod_{j=1}^{m} r_j, \tag{6.7}$$

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

In our regression extension we have $i = 1, \ldots, n$ observations of m circular outcomes, w linear outcomes and g covariates. The mean in the density in (6.7) then becomes $\boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i$ 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.

For the teacher data, the regression equations for the circular and linear outcomes in the GPN-SSN model are

$$\hat{\boldsymbol{\mu}}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{SE}_i,$$

where $\hat{\boldsymbol{\mu}}_i = (\hat{\boldsymbol{\mu}}_{s_i}^t, \hat{\mu}_{y_i})^t$, $\boldsymbol{\beta}_0 = (\beta_{0_{sI}}, \beta_{0_{sII}}, \beta_{0_y})^t$ and $\boldsymbol{\beta}_1 = (\beta_{1_{sI}}, \beta_{1_{sII}}, \beta_{1_y})^t$. Note that because $m = 1$ and $w = 1$, $\boldsymbol{\mu}_{s_i}$ is a 2 dimensional vector and μ_{y_i} is a scalar. We estimate the model using MCMC methods. A detailed description of these methods is given in the Appendix.

6.3.4 Model fit criterion

For the four cylindrical models we focus on their out-of-sample predictive performance to determine the fit of the model. To do so we use k-fold cross-validation and split our data into 10 folds. Each of these folds (10 % of the sample) is used once as a holdout set and 9 times as part of a training set. The analysis will thus be performed 10 times, each time on a different training set.

A proper criterion to compare out-of-sample predictive performance is the Predictive Log Scoring Loss (PLSL) (Gneiting & Raftery, 2007). The lower the value of this criterion, the better the predictive performance of the model. Using ML estimates this criterion can be computed as follows:

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

where l is the model likelihood, M is the sample size of the holdout set, x_i is the i^{th} datapoint from the holdout set and $\hat{\boldsymbol{\vartheta}}$ are the ML estimates of the model parameters. Using posterior samples the criterion is similar to the log pointwise predictive density (lppd) (Gelman et al., 2014, p. 169) and can be computed as:

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

where B is the amount of posterior samples and $\boldsymbol{\vartheta}^{(j)}$ are the posterior estimates of the model parameters for the j^{th} iteration. Because the joint density and thus also the likelihood for the modified GPN-SSN model in (6.7) is not available in closed form (Mastrantonio, 2018) we compute the PLSL for the circular and linear outcome separately for all models. Note that although we fit the CL-PN, CL-GPN and GPN-SSN models using Bayesian statistics, we do not take prior information into account when assessing model fit with the PLSL. According to Gelman et al. (2014) this is not necessary since we are assessing the fit of a model to data, the holdout set, only. They argue that the prior in such case is only of interest for estimating the parameters of the model but not for determining the predictive accuracy.

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

of the PLSL in the teacher data:

- For the modified CL-PN model:

  $y_i \mid \mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$, where $\mu_i = \hat{y}_i$ and for θ_i we use (6.4).

- For the modified CL-GPN model:

  $y_i \mid \mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$, where $\mu_i = \hat{y}_i$ and for θ_i we use (6.5).

- For the modified Abe-Ley model:

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

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

- For the modified joint projected and skew normal model:

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

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

  where SSN is the skew normal distribution.

For each of the four cylindrical models and for each of the 10 cross-validation analyses we can then compute a PLSL for the circular and linear outcome by using the conditional log-likelihoods of the respective outcome. To evaluate the predictive performance we average across the PLSL criteria of the cross-validation analyses. We also assess the cross-validation variability by means of the standard deviations of the PLSL criteria.

6.4 Data Analysis

In this section we analyze the teacher data with the help of the four cylindrical models from Section 6.3. We will present the results, posterior estimates and their interpretation, per model and finish with a section comparing the fit of the different models.

6.4.1 Results & Analysis

In the Appendix we have described the starting values for the MCMC procedures, hence it remains to specify the starting values for the maximum likelihood based Abe-Ley model: $\eta_0 = 0.9, \eta_1 = 0.9, \nu_0 = 0.9, \nu_1 = 0.9, \kappa = 0.9, \alpha = 0.9, \lambda = 0$. The initial amount of iterations for the three MCMC samplers was set to 2000. After convergence checks via traceplots we concluded that some of the parameters of the GPN-SSN model did not converge. Therefore we set the amount of iterations of the MCMC models to 20,000 and subtracted a burn-in of 5000 to reach convergence. Note that we choose the same amount of iterations for all three Bayesian models to make their comparison via the PLSL as fair as possible. Lastly, the predictor `SE`

was centered before inclusion in the analysis as this allows the intercepts to bear the classical meaning of average behavior.

6.4.1.1 The modified CL-PN and CL-GPN models

First recall the regression equations predicting the linear outcome

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

and circular outcome

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

For both models, $\hat{y}$ is the predicted strength of interpersonal behavior, $\hat{\boldsymbol{\mu}}$ is the predicted mean vector of the type of interpersonal behavior and the γ's and β's are intercepts and regression coefficients.

Table 6.2 shows the results for the modified CL-PN and CL-GPN models fit to the teacher dataset. The estimates in this table are the averages and standard deviations of the 10 cross-validation estimates. We show both the estimated posterior mode and the 95% highest posterior density (HPD) interval for each parameter. The posterior estimates for the β's and γ's are quite similar for both models. We also see this similarity when we look at the predicted circular and linear means. The predicted linear mean is equal to γ_0 which is equal to 0.38 and 0.37 in the CL-PN and CL-GPN models, respectively. The predicted circular means can be computed from β_0^I and β_0^{II} using a double arctangent function $\text{atan2}(\beta_0^{II}, \beta_0^I)$, see (6.1). If we do this at each iteration of the MCMC sampler we get the posterior distribution of the circular mean in both models. Table 6.3 shows the posterior modes of the estimated circular means, which are very similar (32.29° and 33.54°).

Next we investigate the effect of self-efficacy. In the CL-PN model we can transform the regression parameters on the two components I and II to one circular regression coefficient b_c using the methods described by Cremers et al. (2018b). The dashed line in Figure 6.4 shows the predicted regression line from the CL-PN model. Here b_c is the slope of this line at the inflection point (black square) and its posterior mode is estimated at 1.67 (-24.66, 29.33)[2]. Even though the HPD of b_c includes 0, indicating there is no evidence to reject the null-hypothesis of no effect, we continue

[2]Note that this is a linear approximation to the circular regression line representing the slope at a specific point. Therefore it is possible for the HPD interval to be wider than 2π. In this case the interval is much wider and covers 0, indicating there is no evidence to reject the null-hypothesis of no effect.

Table 6.2: Results, cross-validation mean and standard deviation, for the modified CL-PN and CL-GPN models

Parameter	CL-PN			CL-GPN		
	Mode	HPD LB	HPD UB	Mode	HPD LB	HPD UB
β_0^I	1.76 (0.09)	1.50 (0.07)	0.09 (0.09)	2.42 (0.15)	1.90 (0.10)	3.05 (0.17)
β_1^I	0.65 (0.07)	0.42 (0.06)	0.08 (0.08)	0.85 (0.12)	0.45 (0.09)	1.30 (0.15)
β_0^{II}	1.15 (0.05)	0.92 (0.04)	0.04 (0.04)	1.47 (0.05)	1.17 (0.04)	1.79 (0.05)
β_1^{II}	0.58 (0.03)	0.38 (0.04)	0.04 (0.04)	0.70 (0.08)	0.46 (0.05)	0.96 (0.07)
γ_0	0.38 (0.01)	0.31 (0.01)	0.01 (0.01)	0.37 (0.01)	0.31 (0.01)	0.42 (0.01)
γ_{cos}	0.04 (0.00)	0.01 (0.00)	0.00 (0.00)	0.03 (0.00)	0.01 (0.00)	0.05 (0.00)
γ_{sin}	-0.01 (0.00)	-0.04 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.03 (0.00)	0.03 (0.00)
γ_1	0.03 (0.01)	-0.00 (0.00)	0.07 (0.01)	0.03 (0.01)	-0.00 (0.00)	0.06 (0.00)
σ	0.14 (0.00)	0.12 (0.00)	0.00 (0.00)	0.14 (0.00)	0.12 (0.00)	0.16 (0.00)
$\sum_{1,1}$	NA (NA)	NA (NA)	NA (NA)	3.02 (0.25)	1.83 (0.14)	5.02 (0.41)
$\sum_{1,2}$	NA (NA)	NA (NA)	NA (NA)	0.46 (0.12)	0.12 (0.12)	0.80 (0.10)
$\sum_{2,2}$	NA (NA)	NA (NA)	NA (NA)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 6.3: Posterior estimates (in degrees) for the circular mean in the CL-PN, CL-GPN and GPN-SSN models

	Mode	HPD LB	HPD UB
Cl-PN	32.29	24.81	39.71
CL-GPN	33.70	26.72	41.15
GPN-SSN	35.30	28.31	43.10

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

with its interpretation for educational purposes. The interpretation of b_c is that at the inflection point an increase of 1 unit in self-efficacy leads to an increase of $1.67 * (180/\pi) = 95.68°$ in the type of interpersonal behavior..However, as we can see in Figure 6.4 the inflection point lies almost outside the range of the actual data. Instead of looking at the inflection point we might compute the slope of the regression line at the average self-efficacy. This parameter we call the slope at the mean (SAM) (Cremers et al., 2018b) and it is estimated at 0.02 (0.01; 0.04) for our data. This means that at the average self-efficacy an increase of 1 unit only leads to an increase of $0.02 * (180/\pi) = 1.15°$ in the type of interpersonal behavior. The HPD of the SAM does not include 0 which means that the effect at the average self-efficacy can be distinguished from 0.

In the CL-GPN model we cannot compute circular regression coefficients such as the b_c and SAM computed above due to the fact that not only the mean vector $\boldsymbol{\mu}$ but also the covariance matrix $\boldsymbol{\Sigma}$ influences the predicted value on the circle. Instead, we will compute posterior predictive distributions for the predicted circular outcome of individuals scoring the minimum, maximum and median self-efficacy. The modes and 95% HPD intervals of these posterior predictive distributions are $\hat{\theta}_{SE_{min}} = 215.74°(147.36°, 44.49°)$, $\hat{\theta}_{SE_{median}} = 25.93°(337.02°, 138.59°)$, $\hat{\theta}_{SE_{max}} = 30.86°(8.63°, 72.19°)$. Note that we display the modes and HPD intervals for the posterior predictive distributions on the interval $[0°, 360°)$. Note that $44.49° = 404.49°$ due to the periodicity of a circular variable. The posterior mode estimate of 215.74° thus lies within its HPD interval (147.36°, 44.49°).The HPD intervals of the three posterior predictive distributions overlap. Had they not overlapped we could have concluded that as the self-efficacy increases, the score of the teacher on the IPC moves counterclockwise.

The effect of self-efficacy on the strength of interpersonal behavior is quantified by γ_1 in both the CL-PN and CL-GPN models. This parameter can be interpreted in the same way as in a usual regression model. The HPD interval of γ_1 includes 0 which means that there is not enough evidence for an effect of self-efficacy on the strength of interpersonal behavior. Note however that the lower bounds of the HPD interval of this parameter are very close to zero.

The relation between the circular and linear outcome, that is, between type of interpersonal behavior and strength of interpersonal behavior, is described by the

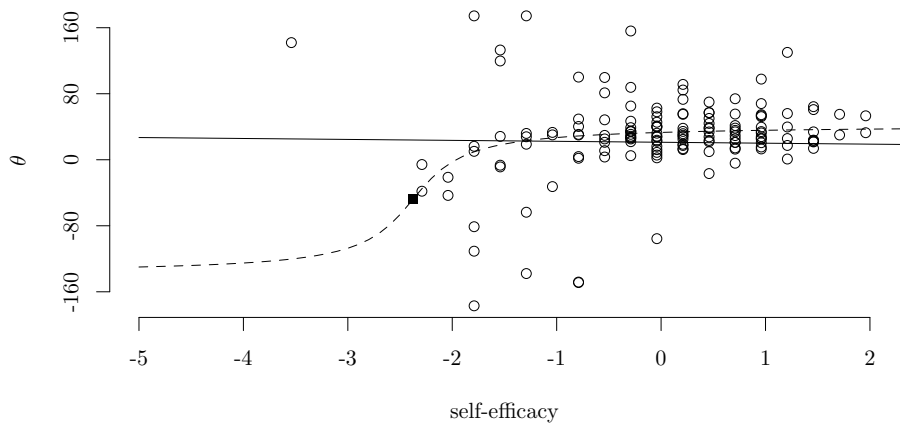

Figure 6.4: Plot showing circular regression lines for the effect of self-efficacy as predicted by the Abe-Ley model (solid line) and CL-PN model (dashed line). The black square indicates the inflection point of the circular regression line for the CL-PN model.

parameters $\gamma_{\cos}$ and $\gamma_{\sin}$. The HPD interval of $\gamma_{\cos}$ does not include 0 for both the CL-PN and CL-GPN models, meaning that the cosine component of the type of interpersonal behavior has an effect on the strength of interpersonal behavior. In the teacher data the sine and cosine components have a substantive meaning. In this case the Communion (cosine) component of the IPC positively effects the strength of a teachers' type of interpersonal behavior, in plain words: teachers exhibiting interpersonal behavior types with higher communion scores (e.g., 'helpful' and 'understanding' in Figure 2) are stronger in their behavior.

6.4.1.2 The modified Abe-Ley model

First recall the regression equations predicting the linear outcome

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

and circular outcome

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

Here $\hat{\mu}$ is the predicted mean vector of the type of interpersonal behavior and $\hat{\beta}$ is the predicted scale parameter of the strength of interpersonal behavior, while the η's and ν's are regression parameters.

Table 6.4 shows the results for the modified Abe-Ley model fit to the teacher

Table 6.4: Results, cross-validation mean and standard deviation, for the modified Abe-Ley model

Parameter	ML-estimate
η_0	0.36 (0.02)
η_1	-0.03 (0.01)
ν_0	1.17 (0.02)
ν_1	0.04 (0.02)
α	3.66 (0.12)
κ	1.51 (0.08)
λ	0.70 (0.05)

dataset. The estimates in this table are the averages and standard deviations of the 10 cross-validation estimates. The estimate of the circular mean at an average self-efficacy, η_0, equals 0.36 radians or 20.62°. For the Abe-Ley model we can also investigate the effect of self-efficacy on the type of interpersonal behavior. The solid line in Figure 6.4 shows this effect. Unlike for the regression line of the CL-PN model, the inflection point of the regression line of the Abe-Ley model lies outside the x-range of the figure.

For the Abe-Ley model the conditional distribution for the linear outcome is Weibull. This means that we can use methods from survival analysis to interpret the effect of self-efficacy. In survival analysis a 'survival' function is used in which time is plotted against the probability of survival of subjects suffering from a specific medical condition. In our data however we plot the strength on the IPC against the probability of a teacher having such a strength. This probability is computed using the 'survival-function' $\exp(-\alpha y_i^{\beta(1-\tanh(\kappa)\cos(\theta_i-\mu_i))^{1/\alpha}})$ with $\beta = \exp(\nu_0 + \nu_1 \mathrm{SE}_i)$. In Figure 6.5, we plot the survival function for the minimum, median and maximum value of self-efficacy. We conclude that stronger interpersonal behaviors are less probable. We also see that the relation between self-efficacy and the strength on the IPC is not linear; the probability of having a stronger interpersonal behavior is higher for both the minimum and maximum self-efficacy compared to a median self-efficacy score. Note however that the circular outcome also influences the survival function. The relation between the type of interpersonal behavior and the strength of interpersonal behavior may thus influence the shape of the survival function.

6.4.1.3 The modified GPN-SSN model

First recall the regression equation predicting the circular and linear outcome:

$$\hat{\boldsymbol{\mu}}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathrm{SE}_i,$$

where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{s_i}, \boldsymbol{\mu}_{y_i})^t$, $\boldsymbol{\beta}_0 = (\beta_{0_{sI}}, \beta_{0_{sII}}, \beta_{0_y})^t$ and $\boldsymbol{\beta}_1 = (\beta_{1_{sI}}, \beta_{1_{sII}}, \beta_{1_y})^t$. The

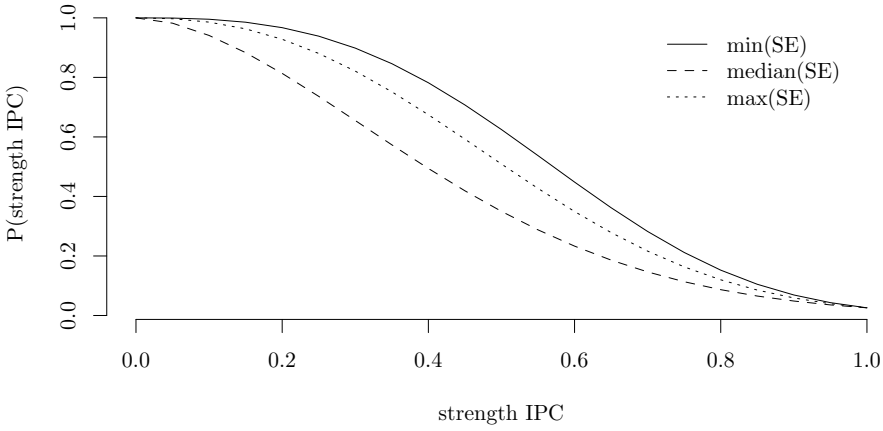

Figure 6.5: Plot showing the probability of having a particular strength of interpersonal behavior (survival plot) for the minimum, mean and maximum self-efficacy in the data.

parameters $\beta_{0_{sI}}$, $\beta_{0_{sII}}$, $\beta_{1_{sI}}$ and $\beta_{1_{sII}}$ are the intercepts and regression coefficients for the circular outcome and β_{0_y} and β_{1_y} are the intercept and regression coefficient for the linear outcome.

Table 6.5 shows the results for the modified GPN-SSN model fit to the teacher dataset. The estimates in this table are the averages and standard deviations of the 10 cross-validation estimates. We show both the estimated posterior mode and the 95% highest posterior density (HPD) interval for each parameter. The predicted circular means can be computed from $\beta_{0_{sI}}$ and $\beta_{0_{sII}}$ in a similar fashion as for the CL-PN and CL-GPN models. Table 6.3 shows the posterior mode of the estimated circular mean, which equals 35.30° and hence is very similar to those of the CL-PN and CL-GPN models.

For the same reason as in the CL-GPN model we cannot compute circular regression coefficients for the effect of self-efficacy on the type of interpersonal behavior such as the b_c and SAM. Instead, we will again compute posterior predictive distributions for the predicted circular outcome of individuals scoring the minimum, maximum and median self-efficacy. The modes and 95% HPD intervals of these posterior predictive distributions are $\hat{\theta}_{SE_{min}} = 206.87°(117.11°, 72.02°)$, $\hat{\theta}_{SE_{median}} = 24.68°(334.73°, 128.27°)$, $\hat{\theta}_{SE_{max}} = 29.81°(0.74°, 80.61°)$. On a circle the HPD intervals of the three posterior predictive distributions overlap. Had they not overlapped we could have concluded that as the self-efficacy increases, the score of the teacher on the IPC moves counterclockwise. The effect of self-efficacy on the strength of interpersonal behavior is quantified by β_{1_y}, and we learn that for a 1 unit increase in self-efficacy the strength of interpersonal behavior increases by 0.09. The average strength is quantified by β_{0_y} and equals 0.33.

Table 6.5: Results, cross-validation mean and standard deviation, for the GPN-SSN model

Parameter	Unconstrained			Constrained		
	Mode	HPD LB	HPD UB	Mode	HPD LB	HPD UB
$\beta_{0_s^I}$	0.30 (0.01)	0.26 (0.01)	0.34 (0.01)	2.11 (0.10)	1.75 (0.09)	2.50 (0.11)
$\beta_{0_s^{II}}$	0.19 (0.00)	0.17 (0.01)	0.21 (0.00)	1.33 (0.07)	1.10 (0.05)	1.57 (0.06)
β_{0_y}	0.33 (0.01)	0.30 (0.30)	0.36 (0.01)	0.33 (0.01)	0.30 (0.01)	0.36 (0.01)
$\beta_{1_s^I}$	0.09 (0.01)	0.05 (0.01)	0.13 (0.01)	0.60 (0.06)	0.33 (0.05)	0.90 (0.06)
$\beta_{1_s^{II}}$	0.07 (0.00)	0.04 (0.00)	0.09 (0.01)	0.48 (0.03)	0.30 (0.03)	0.66 (0.04)
β_{1_y}	0.09 (0.01)	0.06 (0.06)	0.12 (0.01)	0.09 (0.01)	0.06 (0.01)	0.12 (0.01)
$\sum_{s_{1,1}}$	0.05 (0.00)	0.04 (0.00)	0.06 (0.00)	2.43 (0.14)	1.72 (0.07)	3.46 (0.13)
$\sum_{s_{2,2}}$	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
$\sum_{y_{3,3}}$	0.03 (0.00)	0.02 (0.02)	0.04 (0.00)	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)
$\sum_{s_{1,2}}$	0.00 (0.00)	-0.00 (0.00)	0.01 (0.00)	0.07 (0.06)	-0.20 (0.06)	0.35 (0.06)
$\sum_{sy_{1,3}}$	0.03 (0.00)	0.02 (0.00)	0.04 (0.00)	0.23 (0.01)	0.17 (0.01)	0.32 (0.01)
$\sum_{sy_{2,3}}$	0.01 (0.00)	0.01 (0.01)	0.02 (0.00)	0.09 (0.01)	0.06 (0.01)	0.12 (0.01)
λ	0.16 (0.01)	0.14 (0.01)	0.18 (0.01)	0.16 (0.01)	0.14 (0.01)	0.18 (0.01)

Table 6.6: PLSL criteria, cross-validation mean and standard deviation, for the circular and linear outcome in the four cylindrical models

Model	Circular		Linear	
	mean	sd	mean	sd
CL-PN	82.96	(9.47)	-17.65	(3.70)
CL-GPN	85.22	(18.12)	-18.12	(3.57)
Abe-Ley	31.97	(22.07)	25.49	(17.46)
GPN-SSN	106.38	(8.84)	-2.14	(6.78)

To investigate the association between the linear and circular outcome we look at the covariances between the linear outcome and the sine and cosine of the circular outcome $\sum_{sy_{2,3}}$ and $\sum_{sy_{1,3}}$. Both covariances, $\sum_{sy_{2,3}} = 0.09$ and $\sum_{sy_{1,3}} = 0.23$, are different from zero, but the one of the cosine component is larger. This means the correlation with the Communion component is larger and that teachers scoring both high on Communion and Agency show stronger behavior. This is a slightly different conclusion from the one in the CL-PN and CL-GPN models.

6.4.1.4 Model fit

In this section we will assess the overall fit of the cylindrical models to the data via the PLSL criterion described in Section 6.3.4. Table 6.6 shows the values of this criterion for the linear and circular outcomes of the four models.

The CL-PN and CL-GPN models have the best out-of-sample predictive performance for the linear outcome. They show roughly the same performance because they model the linear outcome in the same way. Only the value of r in (6.2) differs. We should note that even though the predictive performance of the Abe-Ley model for the linear outcome is worst on average, the standard deviation of the cross-validation estimates is rather large. This means that in some samples, the Abe-Ley model shows a lower PLSL value than the average of 25.49.

The Abe-Ley model has the best out-of-sample predictive performance for the circular outcome. This would suggest that for the circular variable a slightly skewed distribution fits best. However, both the GPN-SSN and the CL-GPN models fit much worse even though the distribution for the circular outcome in these models can also take a skewed shape. It should be noted that the standard deviation of the cross-validation estimates is rather large for the Abe-Ley and the CL-GPN model. It is possible that these large standard deviations for the PLSL are caused by the fact that they are computed for a relatively small sample size, but this does not explain why the PLSL has a large standard deviation for only a few cylindrical models and not for all.

6.5 Discussion

In this paper we modified four models for cylindrical data in such a way that they include a regression of both the linear and circular outcome onto a set of covariates. Subsequently we have shown how these four methods can be used to analyze a dataset on the interpersonal behavior of teachers. In this final section we will comment on the differences between these models, the results from the analysis of the teacher data and how cylindrical models improve the analysis of such cylindrical data.

In terms of interpretability, the CL-PN and Abe-Ley models perform best. In the CL-GPN and GPN-SSN models the interpretation of the parameters of the circular outcome component is not straightforward, if at all possible. This is caused by the fact that in addition to the mean vector the covariance matrix of the GPN distribution affects the location of the circular data, making it difficult to compute regression coefficients on the circle. Wang & Gelfand (2013) state that Monte Carlo integration can be used to compute a circular mean and variance for the GPN distribution. In future research, this solution might be applied to the methods of Cremers et al. (2018b) in order to compute circular coefficients for GPN models.

In terms of flexibility the GPN-SSN model scores best. Multiple linear and circular outcomes can be included and we can thus apply the model to multivariate cylindrical data. In addition the GPN-SSN, the CL-GPN and CL-PN models are extendable to a mixed-effects structure and can thus also be fit to longitudinal data (see Nuñez-Antonio & Gutiérrez-Peña (2014) and Hernandez-Stumpfhauser et al. (2017) for hierarchical/mixed-effects models for the PN and GPN distributions respectively). For the Abe-Ley model this may also be possible but has not been done in previous research for the conditional distribution of its circular outcome (sine-skewed von Mises). Concerning asymmetry, both the GPN-SSN as well as the Abe-Ley model allow for non-symmetrical shapes of the distributions of both the linear and circular outcome, while the CL-GPN model permits an asymmetric circular outcome.

To investigate model fit for the teacher data, we assessed out-of-sample predictive performance for both the linear and circular outcome. The CL-PN and CL-GPN models have the best fit for the linear outcome while the Abe-Ley model fits best the circular outcome. Differences in fit, in addition to being a result of different distributional assumptions, may also be caused by the way in which the relation between the linear and circular outcome is modeled. Whereas in both the Abe-Ley and GPN-SSN models the distribution of the linear outcome is conditioned on the circular outcome and vice versa, the distribution of the circular outcome in the CL-PN and CL-GPN models is independent of the linear outcome. In these models the circular outcome is regressed onto the linear outcome.

The four cylindrical models that were modified to the regression context in this paper are not the only cylindrical distributions available from the literature. Other interesting cylindrical distributions have been introduced by Fernández-Durán (2007), Kato & Shimizu (2008) and Sugasawa, Shumizu, & Kato (2015) (for more references we refer to Chapter 2 of Ley & Verdebout (2017)). In the present study we have decided not to include these distributions for reasons of space, complexity

of the models and ease of implementing a regression structure. In future research however it would be interesting to investigate other types of cylindrical distributions as well in order to compare the interpretability, flexibility and model fit to the models developed in the present study.

We conclude the paper on a general note regarding cylindrical models. They offer new insights into data of a cylindrical nature in psychology. Concerning the example used in this paper, the advantage of cylindrical data analysis is that we were able to analyze all circular and linear information in the data simultaneously. In previous research, the two components of the interpersonal circumplex (*i.e.*, Agency and Communion) were analyzed separately. Such an approach also provides information about the strength of teachers' score on Agency and Communion, yet a large portion of information about the combination of Agency and Communion, which describes the kind of behavior that is observed, gets lost. A first solution to include both dimensions as a circular variable in data analysis was described by Cremers et al. (2018a). A downside of that analysis was that information about the strength of the specific type of interpersonal behavior could not be retained. In the present study, we have shown how using cylindrical models can simultaneously model the information about the type of and strength of interpersonal behavior and how these are influenced by teachers' self-efficacy in classroom management. The results of the present study therefore provide an answer to an often stated problem in data analysis of interpersonal circumplex data.

Appendices

Chapter 2

MCMC method 1

Note that MCMC method 1 is similar to the one Nuñez-Antonio et.al. (2011) use.

In a MH step (MCMC method 1) use the following proposal distribution:

$$p(r_i) = N(m_{0i}, v_{0i})$$

where $m_{0i} = \log\left(\frac{1}{2}[b_i + (b_i^2 + 8)^{\frac{1}{2}}]\right)$ and $v_{0i} = (2 + \exp(2m_{0i}))^{-1}$. Define a value, y_0 as the (natural) logarithm of the current value of r_i, $y_0 = \log(r_i)$. Next we complete the following steps 5 times:

- Sample a second value y_1 from the proposal distribution.
- Compute the kernel of $f(\ln(r_i)|\theta_i)$ for y_1:

$$f(y_1) = 2y_1 - 0.5\exp(y_1)(\exp(y_1) - 2b_i).$$

- Compute the natural logarithm of the density of the proposal distribtution for y_1.
- Compute the difference between $\log(f(y_1))$ and $\log(p(y_1))$:

$$w_1 = \log(f(y_1) - \log(p(y_1)).$$

- Repeat the previous three steps for y_0.
- Compute the acceptance ratio:

$$\alpha = w_1 - w_0.$$

  where $w_0 = \log(f(y_0) - \log(p(y_0))$.

- Sample a value k from a uniform distribution:

$$k \sim U(0, 1).$$

- If $\alpha \geq \log(k)$ we accept y_1, so $y_0 = y_1$. Else, we do not accept y_1 and $y_0 = y_0$.

After these five iterations $r_i = \exp(y_0)$.

MCMC method 2

In a slice sampler (Hernandez-Stumpfhauser et al., 2017; Neal, 2003) the joint density for an auxiliary variable v_i with r_i for regression is:

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

The full conditionals for v_i and r_i are:

$$p(v_i \mid r_i = r_i, \boldsymbol{\mu}_i, \theta_i) \sim U(0, \exp\{-.5(r_i - b_i)^2\})$$

$$p(r_i \mid v_i = v_i, \boldsymbol{\mu}_i, \theta_i) \propto r_i \mathbf{I}\left(b_i + \max\left\{-b_i, -\sqrt{-2 \ln v_i}\right\} < r_i < b_i + \sqrt{-2 \ln v_i}\right)$$

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

Chapter 3

This is supplementary material for the paper: 'Circular interpretation of projected normal regression coefficients'. Code for the simulations, empirical data analysis and plots can be found at the following GitHub archive: https://github.com/Circular-Data/Circular-Interpretation-BJMSP .

MCMC procedure for regression models

Consider the following model for a circular outcome variable:

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

where θ is the circular outcome variable measured in radians ($-\pi \leq \theta < \pi$), $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$ is the mean vector of this distribution, the variance-covariance matrix $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$ and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution. The model for the mean vector is: $\boldsymbol{\mu} = \boldsymbol{B}^t \boldsymbol{x}$, where $\boldsymbol{B} = [\boldsymbol{\beta}^I, \boldsymbol{\beta}^{II}]$ a matrix of regression coefficients and $\boldsymbol{x}$ is a vector of predictor variables.

A method to estimate this circular regression model is presented in Nuñez-Antonio et al. (2011). The MCMC procedure used in this paper is the same except for the sampling of the vector of latent lengths, $\boldsymbol{r} = r_1, \ldots, r_n$, where n is the sample size, see (3.1).

Simulation studies have show that the method of sampling used by Nuñez-Antonio et al. (2011) works well, the performance in terms of bias and coverage are reasonable to good in most cases. Using a slice sampler for the latent lengths instead of a Metropolis-Hastings sampler results in improved performance and efficiency. R code for the sampler and results for the simulation can be requested from the authors.

In this appendix we only shortly mention the priors and conditional posteriors for the regression coefficients. A normal prior is specified for each of the two components of $\boldsymbol{B}$:

$$\boldsymbol{\beta}^j \sim N(\boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j) \text{ for } j = I, II,$$

where $\boldsymbol{\beta}_0^j$ is a vector with prior values for the regression coefficients and intercept and $\boldsymbol{\Lambda}_0^j$ is the prior precision matrix of component j. The full conditional density of $\boldsymbol{\beta}^j$ equals:

$$\boldsymbol{\beta}^j \mid \boldsymbol{\theta}, \boldsymbol{r} \sim N(\boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j) \text{ for } j = I, II, \tag{6.8}$$

where $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ and $\boldsymbol{\mu}_F^j = (\boldsymbol{\Lambda}_F^j)^{-1}(\boldsymbol{\Lambda}_0^j \boldsymbol{\beta}_0^j + (\mathbf{X}^j)^t \mathbf{y}^j)$, $\boldsymbol{\Lambda}_F^j = \boldsymbol{\Lambda}_0^j + (\mathbf{X}^j)^t \mathbf{X}^j$, $\mathbf{X}^j$ is a design matrix. The latent lengths in $\boldsymbol{r}$, are given a prior that is uniform

between 0 and ∞. The full conditional density of one latent length r_i can be found in Nuñez-Antonio et al. (2011) and equals:

$$r_i \mid \theta_i, \boldsymbol{\mu}_i \propto r_i \exp(-0.5 r_i^2 + b_i r_i),$$

where $b_i = \boldsymbol{u}_i^t \boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_i = \boldsymbol{B}^t \boldsymbol{x}_i$. The sampler that can be used to obtain estimates for the vectors of regression coefficients $\boldsymbol{\beta^j}$ and values for the vector $\boldsymbol{r}$, contains the following steps:

1. The priors for $\boldsymbol{\beta}^j$ are specified by choosing values for $\boldsymbol{\beta}_0^j$ and $\boldsymbol{\Lambda}_0^j$. In this paper we use 0 for each of the elements in $\boldsymbol{\beta}_0^j$. We specify $\boldsymbol{\Lambda}_0^j$ as a diagonal matrix with diagonal values equal to 1×10^{-4}.

2. A starting value for $\boldsymbol{r}$ is chosen. We choose a vector of ones.

3. Using the starting value for an r_i and $\boldsymbol{u}_i$ computed from the data we may compute $\boldsymbol{y}_i = r_i \boldsymbol{u}_i$.

4. The $\boldsymbol{\beta}^j$ are sampled from their conditional posterior, Equation 6.8.

5. Using the estimates for the$\boldsymbol{\beta}^j$ new r_i are sampled from their conditional posterior, Equation 6.5, using slice sampling (Neal, 2003). The specifics of this slice sampler are presented by Hernandez-Stumpfhauser et al. (2017) and were adapted for the regression situation. The joint density for the auxiliary variable v_i with r_i for regression is:

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

   The full conditionals for v_i and r_i are:

$$p(v_i \mid r_i = r_i, \boldsymbol{\mu}_i, \theta_i) \sim U(0, \exp\{-.5(r_i - b_i)^2\})$$

$$p(r_i \mid v_i = v_i, \boldsymbol{\mu}_i, \theta_i) \propto r_i \mathbf{I}\left(b_i + \max\left\{-b_i, -\sqrt{-2\ln v_i}\right\} < r_i < b_i + \sqrt{-2\ln v_i}\right)$$

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

6. Using the new r_i new values for $\boldsymbol{y}_i$ are computed as $\boldsymbol{y}_i = r_i \boldsymbol{u}_i$.

7. Steps 4, 5 and 6 are repeated for a specified amount of iterations. After the iterations are completed convergence is checked. If convergence is not reached additional iterations are run.

Posterior Histograms pointing north data.

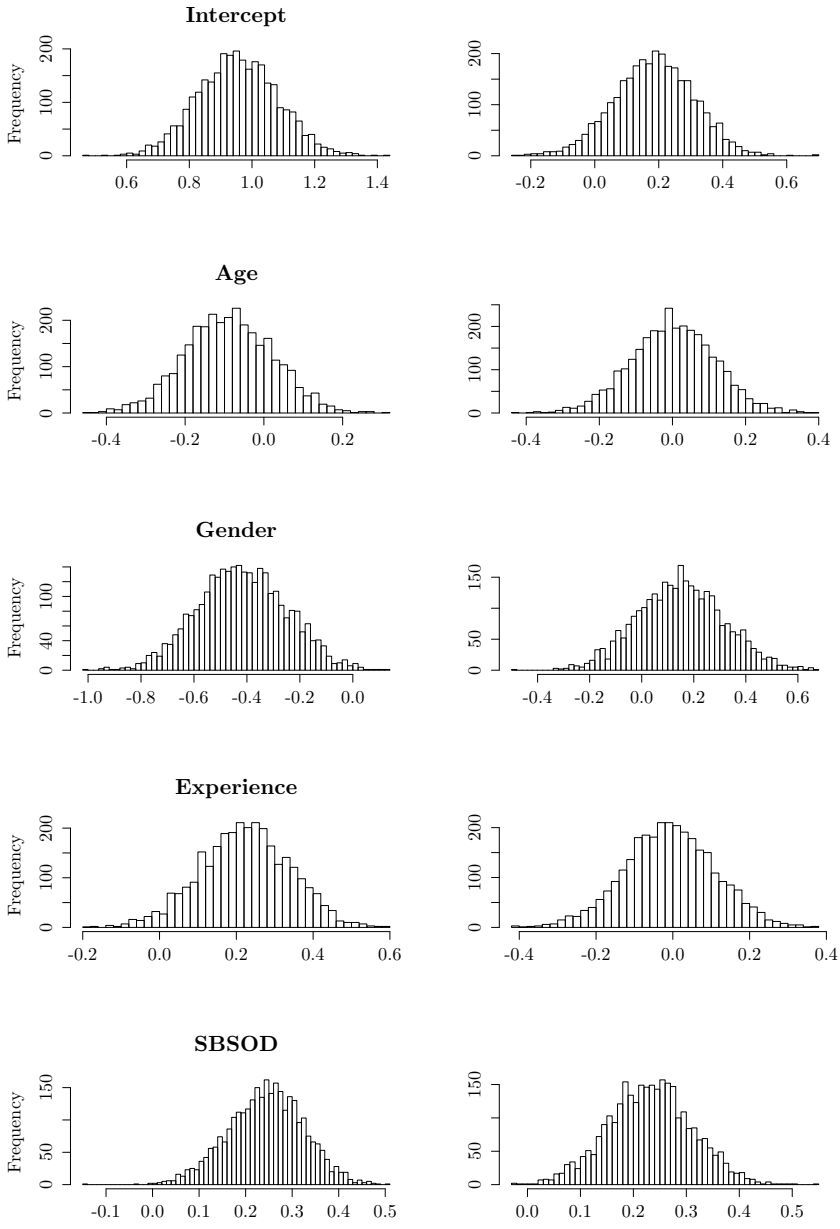

Figure A1: Posterior histograms for the linear intercepts, β_0^I and β_0^I and regression coefficients, β_1^I, β_2^I, β_3^I, β_4^I, β_1^{II}, β_2^{II}, β_3^{II}, and β_4^{II} for the pointing north data. Component I is shown on the left, component II is shown on the right.

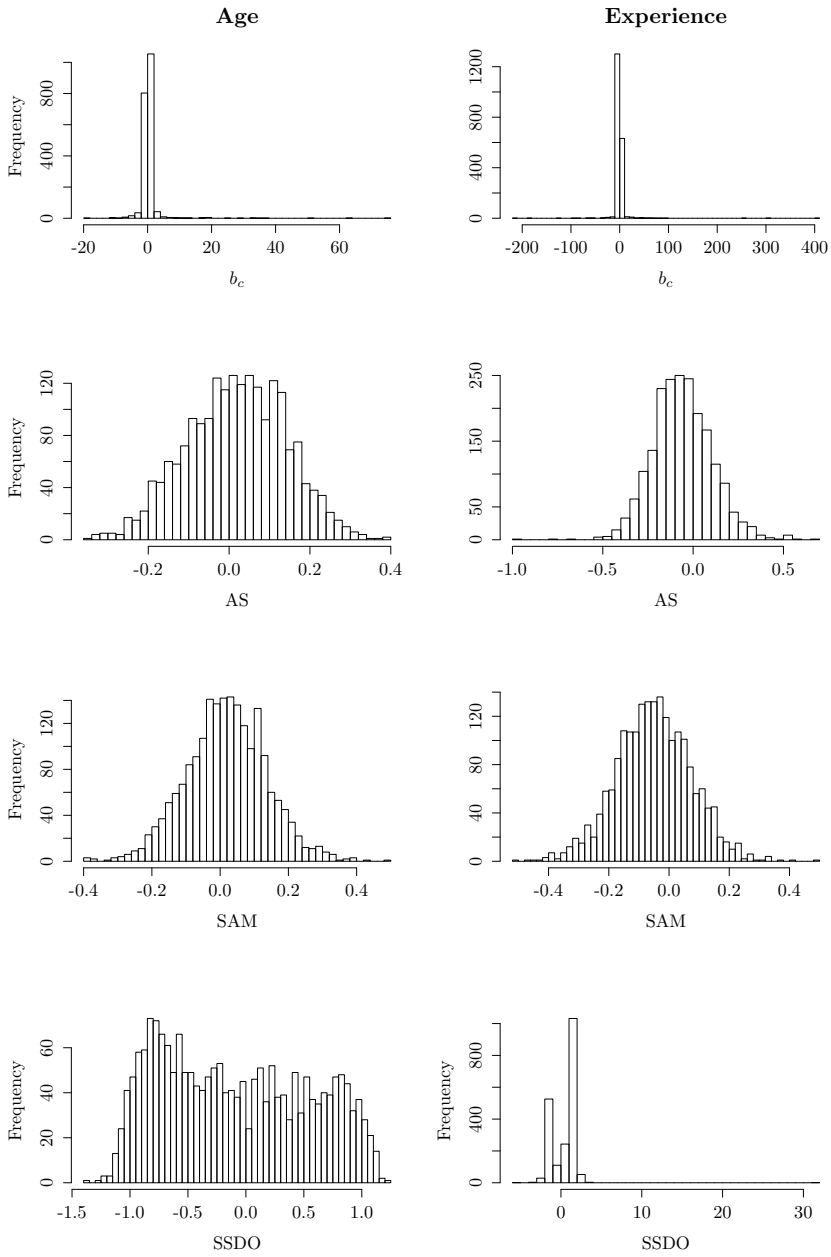

Figure A2: Posterior histograms for the circular regression coefficients and the *SSDO* for the variables Age (left) and Experience (right) of the pointing north data.

Signed Shortest Distance to the Origin

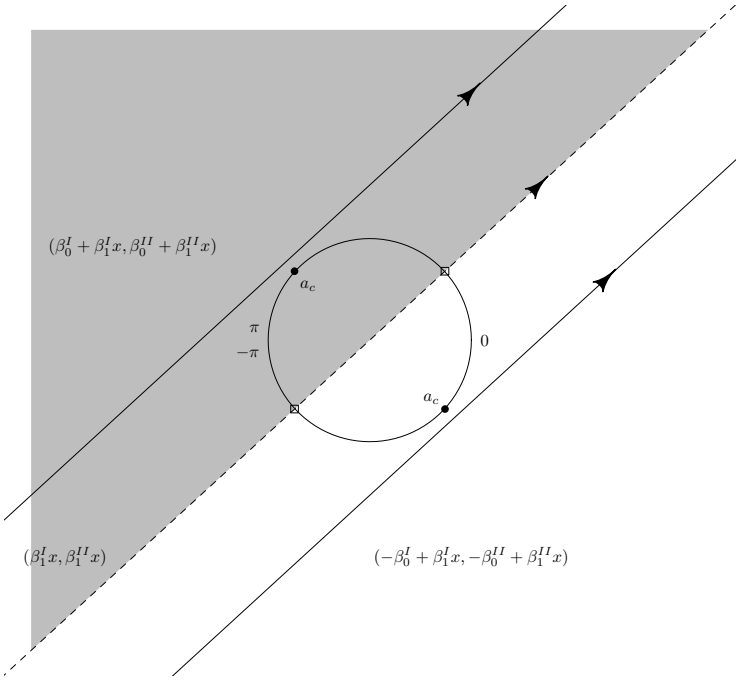

Figure A3: Plot illustrating the computation of the SSDO showing a unit circle and two regression lines in bivariate space. Squares are $\text{atan2}(\beta_1^{II}, \beta_1^I)$ and $\text{atan2}(\beta_1^{II}, \beta_1^I) - \pi$, dots are the a_c for the two regression lines. The dotted line is the regression line for which the linear intercepts, β_0^I and β_0^{II}, equal 0. The shaded area represents the area where the SSDO of a regression line has a positive sign.

To understand the intuition behind the sign of the shortest distance to the origin, SDO, we created Figure A3. In this figure we show two bivariate regression lines:

$$y^I = \beta_0^I + \beta_1^I x$$
$$y^{II} = \beta_0^{II} + \beta_1^{II} x.$$

and

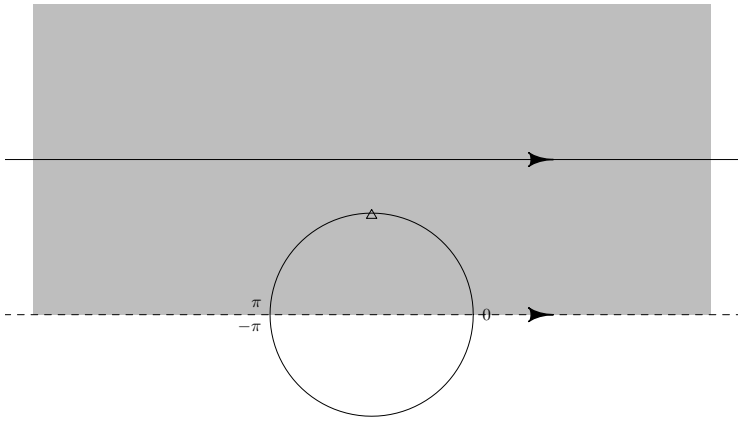

Figure A4: Plot illustrating the computation of the SSDO showing a unit circle and a regression line in bivariate space. The triangle represents $a_c - \text{atan2}(\beta_1^{II}, \beta_1^{I})$. The dotted line is the regression line for which the linear intercepts, β_0^I and β_0^{II}, equal 0. The shaded area represents the area where the SSDO of a regression line has a positive sign.

$$y^I = -\beta_0^I + \beta_1^I x$$
$$y^{II} = -\beta_0^{II} + \beta_1^{II} x.$$

The unit circle in Figure A3 is split into a positive (grey) and a negative side according to the line

$$y^I = \beta_1^I x$$
$$y^{II} = \beta_1^{II} x.$$

The intersection points of this line with the circle (squares) are equal to $\text{atan2}(\beta_1^{II}, \beta_1^{I})$ and $\text{atan2}(\beta_1^{II}, \beta_1^{I}) - \pi$. The direction of the line in (6.9) determines which side of the circle is the positive side. Imagine the direction of the line is North, then the West side is the positive side of the circle. To determine the sign of the SSDO of a regression line we compute $a_c - \text{atan2}(\beta_1^{II}, \beta_1^{I})$. In this case we can see this as turning the regression line anti-clockwise by $\text{atan2}(\beta_1^{II}, \beta_1^{I})$. See Figure A4 for the turning of regression line 6.9. By computing $\sin a_c - \text{atan2}(\beta_1^{II}, \beta_1^{I})$ we can then determine whether the regression line is on the positive or negative side of the unit circle. We can thus compute the SSDO as:

$$SSDO = \text{sign}[\sin\left(a_c - \text{atan2}(\beta_1^{II}, \beta_1^{I})\right)]SDO.$$

Plots for the Pointing North Data

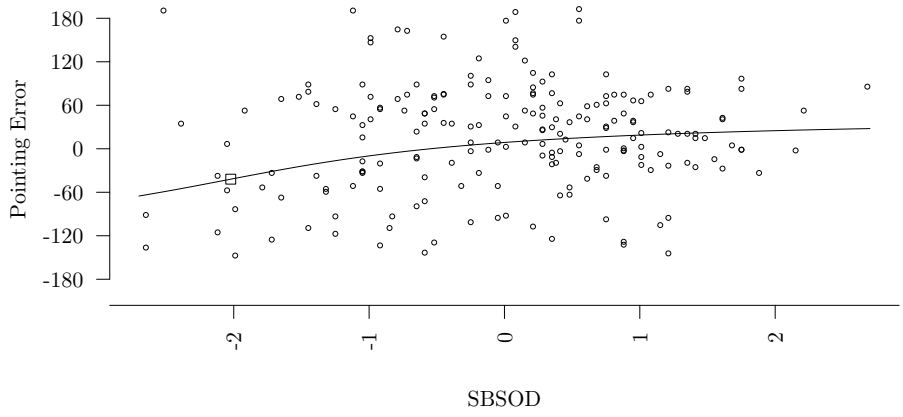

Figure A5: Predicted circular regression curve for the relation between SBSOD and the pointing error together with the original datapoints. The square indicates the inflection point of the regression curve.

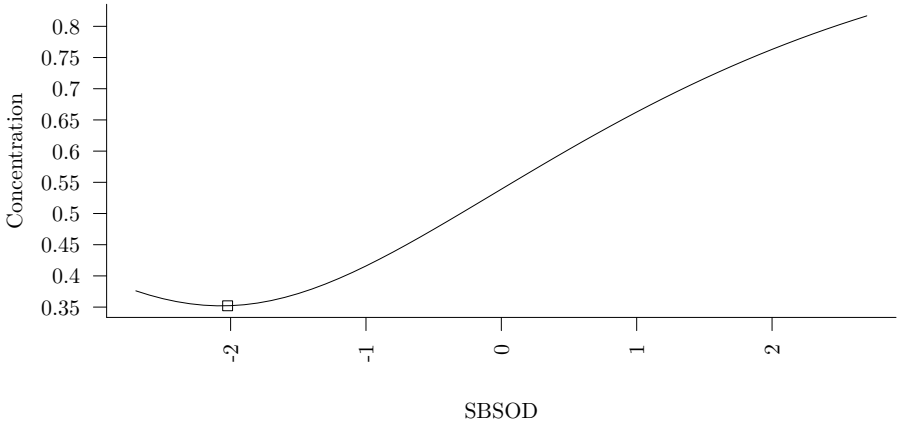

Figure A6: Figure showing the relation between SBSOD and the concentration of the predicted values on the circle.

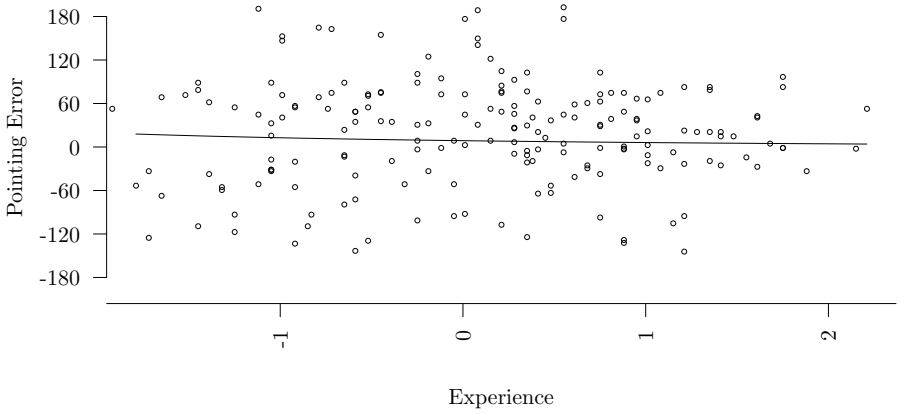

Figure A7: Predicted circular regression curve for the relation between Experience and the pointing error together with the original datapoints. The inflection point of the regression curve is not shown as this point lies outside the range of the data.

Simulation study

To assess the performance of the circular coefficients b_c, SAM and AS and the ability to distinguish between location and accuracy effects we conducted a simulation

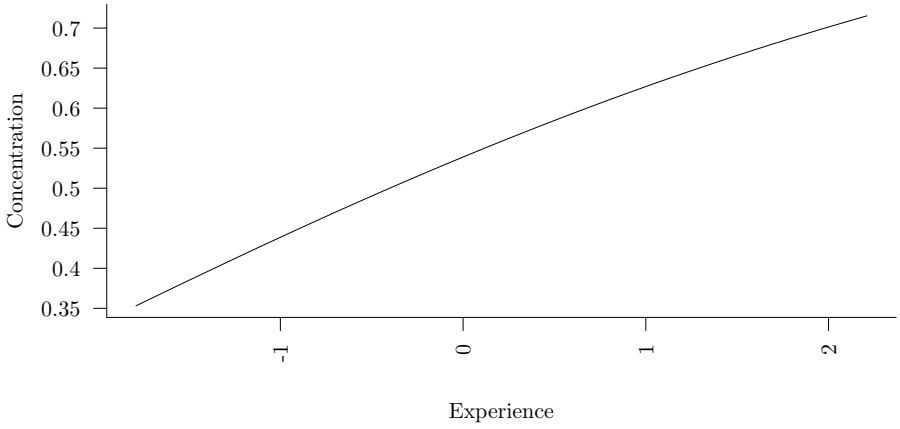

Figure A8: Figure showing the relation between Experience and the concentration of the predicted values on the circle.

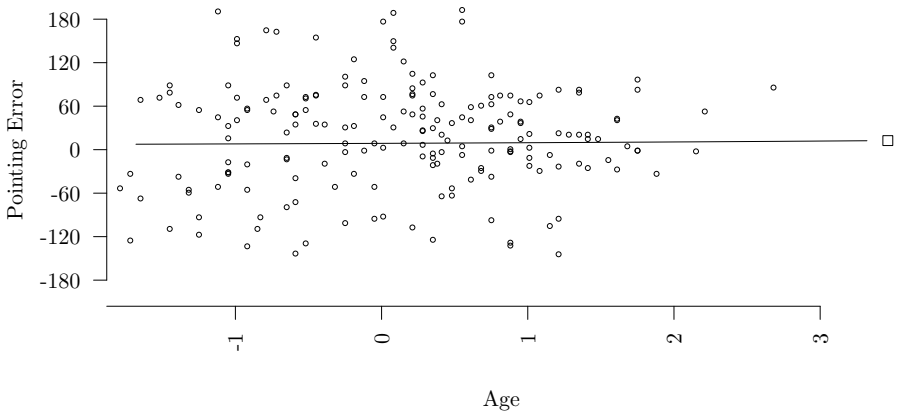

Figure A9: Predicted circular regression curve for the relation between Age and the pointing error together with the original datapoints. The square indicates the inflection point of the regression curve.

study with 1225 designs with one predictor. Of these designs 1056 were classified as location designs, 144 as accuracy and 25 as having no effect bassed on their population values.

Figure A10: Figure showing the relation between Age and the concentration of the predicted values on the circle.

Design

In each design different population values were chosen for the linear intercepts β_0^I and β_0^{II} and the regression coefficients β_1^I and β_1^{II}. Chosen values were: -3, -1, 0, 1, 3 for the linear intercepts and -2, -1, -0.5, 0, 0.5, 1, 2 for the linear regression coefficients. From these values, the population values of the parameters a_x, a_c, b_c, SAM, AS, SDO and $SSDO$ were computed.

We simulated 1000 datasets for each design. There are 500 datasets with N = 50 and 500 where N = 200 for each design. Each dataset contains one circular outcome θ and one linear predictor $x \sim N(0, 1)$. The relation between predictor and outcome was determined by the chosen values for the linear intercepts and coefficients. Before analysis of a dataset the linear predictor was centered at 0 such that the simulation study corresponds to the recommendations in Section 5 of the paper. For a random selection of datasets the convergence of the MCMC sampler used for estimation was checked. Typical posterior histograms are shown in Figure A11. For all of the selected datasets 5000 iterations, from which a burn-in of 1000 iterations was substracted, were deemed sufficient for convergence.

For each simulated dataset we computed several statistics from the posterior distributions of the parameters β_0^I, β_0^{II}, β_1^I, β_1^{II}, a_x, a_c, b_c, SAM, AS and $SSDO$. We estimated the posterior mode, the deviation defined as population value minus estimated mode and the 95% HPD interval. The posterior mode and HPD interval were estimated using the methods proposed by Venter (1967). We also recorded whether the population value of each parameter was located within the HPD interval. Furthermore, we checked whether the HPD intervals of the parameters b_c, $SSDO$, β_1^I and β_1^{II} contained 0.

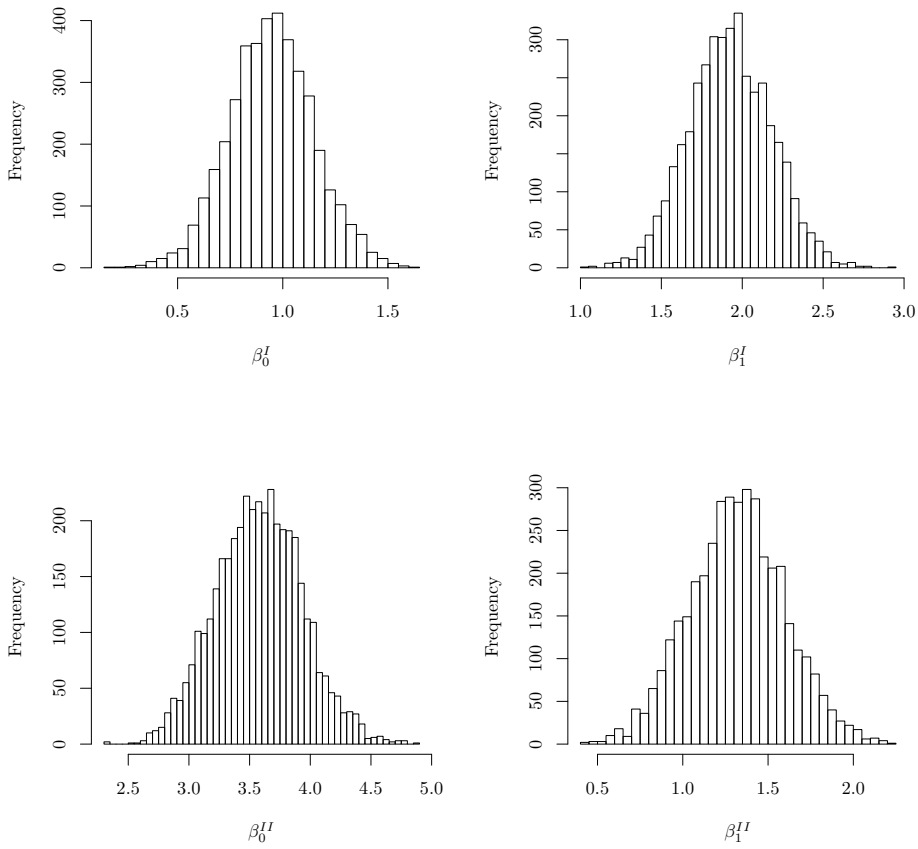

Figure A11: Typical posterior histograms for β_0^I, β_1^I, β_0^{II} and β_1^{II} of one dataset in the design with population values 0, 2, 3 and 1 respectively.

For datasets of the same sample size and from the same design we then computed summary statistics. We averaged the deviation of the estimated mode from the population value over all datasets to compute the bias. The relative bias for the location designs is computed as the ratio of the absolute population value of the coefficient divided by the absolute bias. The relative bias for the accuracy designs is the absolute difference between the population value and the bias. We computed the frequentist coverage of the HPD intervals by computing the percentage of datasets of which the HPD interval contained the population value. We computed the Average Interval Width (AIW) by averaging all lower and upper bounds of the HPD intervals and then computing the distance between these averaged bounds. Lastly, we computed the amount of datasets in which the HPD interval of the parameters b_c, $SSDO$, β_1^I and β_1^{II} contained 0.

Results

For each design we created cross-tabulations with the percentage of datasets in which a location effect was detected and we counted the percentage of datasets in which any effect, location or accuracy, was detected. The cross-tabulations contain two variables: a location effect detected by b_c (Yes/No) and a location effect detected by $SSDO$ (Yes/No). For the percentage of datasets in which any effect was detected we used β_1^I and β_1^{II} (Yes/No). Because we simulated datasets for a large amount of designs we decided to average the percentages over the designs of three different categories: the accuracy designs, the designs with a location effect and $SDO \leq 1$ and the designs with a location effect and $SDO > 1$. Because there were only 25 designs in which we did not simulate a circular effect we decided to not include results for this type. Tables A3 and A1 show the average percentages in combination with their standard deviations for designs with a sample size of 50. Tables A4 and A2 show the average percentages in combination with their standard deviations for designs with a sample size of 200.

Figures A12 and A13 show several plots with simulation results for location (grey) and accuracy (black) designs for the datasets with a sample size of 50. The three circular coefficients b_c, AS and SAM are shown from left to right. Note that each dot in these figures represents one design of our simulation. Again the no effect category was excluded. Figure A12 shows the relative bias plotted against the population value of $SSDO$. The relative bias for the location designs is computed as the ratio of the absolute population value of the coefficient divided by the absolute bias. The relative bias for the accuracy designs is the absolute difference between the population value and the bias. We see that for location designs the relative bias for all three parameters increases for $SSDO$s closer to zero. The coefficients AS and SAM have lower relative biases than b_c. In accuracy designs the coefficients have smaller relative biases compared to location designs with an $SSDO$ close to 0.

Figure A12 also shows the coverage plotted against the population value of the $SSDO$. For b_c the location designs show coverages around 0.95 or higher. For the accuracy designs coverages for this parameter also lie around 0.95 but sometimes they are slightly lower. For AS the location designs show some slight undercoverge, except those with an $SSDO$ close to 0. Most accuracy designs show undercoverage

for AS. For SAM location designs show a slight undercoverage. Accuracy designs show some overcoverage for SAM.

Figure A13 shows the log of the AIWs plotted against the population value of the $SSDO$. We see that the largest log AIWs occur in the accuracy designs of the parameters b_c and SAM. For the location designs b_c shows the largest and SAM shows the smallest log AIWs. The designs with smaller $SSDO$ perform worse. The log of the AIW is larger for these designs.

The results for datasets with a sample size of 200 are shown in Figures A14 and A15. These results show that the coverages are better and the biases are lower for larger datasets, as is to be expected.

Table A1: Cross-tabulations showing the average percentage of datasets with standard deviaton for which the $SSDO$ and b_c do (Yes) or do not (No) indicate a location effect per type of design with a sample size of 50.

			Location $SSDO$	
			Yes	No
Accuracy	Location b_c	Yes	5.51 (1.22)	1.40 (0.66)
		No	0.39 (0.79)	92.71 (1.38)
Location $SDO \leq 1$	Location b_c	Yes	68.34 (29.95)	2.47 (1.92)
		No	0.11 (0.33)	29.08 (28.53)
Location $SDO > 1$	Location b_c	Yes	92.27 (17.25)	0.46 (1.14)
		No	0.87 (1.95)	6.40 (15.15)

Table A2: Cross-tabulations showing the average percentage of datasets with standard deviaton for which the $SSDO$ and b_c do (Yes) or do not (No) indicate a location effect per type of design with a sample size of 200.

			Location $SSDO$	
			Yes	No
Accuracy	Location b_c	Yes	5.25 (1.06)	1.80 (0.62)
		No	0.04 (0.15)	92.91 (1.14)
Location $SDO \leq 1$	Location b_c	Yes	91.07 (17.91)	1.11 (1.94)
		No	0.00 (0.10)	0.90 (16.12)
Location $SDO > 1$	Location b_c	Yes	98.89 (6.81)	0.16 (0.87)
		No	0.05 (0.31)	7.82 (5.68)

Table A3: Average percentage of datasets with standard deviaton for which the any effect indicator indicates that there is (Yes) or is no effect (No) per type of design with a sample size of 50.

Accuracy		Location $SDO \leq 1$		Location $SDO > 1$	
Yes	No	Yes	No	Yes	No
91.14 (17.46)	8.86 (17.46)	99.53 (1.27)	0.47 (1.27)	95.27 (12.16)	4.73 (12.16)

Table A4: Average percentage of datasets with standard deviaton for which the any effect indicator indicates that there is (Yes) or is no effect (No) per type of design with a sample size of 200.

Accuracy		Location $SDO \leq 1$		Location $SDO > 1$	
Yes	No	Yes	No	Yes	No
99.17 (2.39)	0.83 (2.39)	100.00 (0.00)	0.00 (0.00)	99.67 (1.62)	0.33 (1.62)

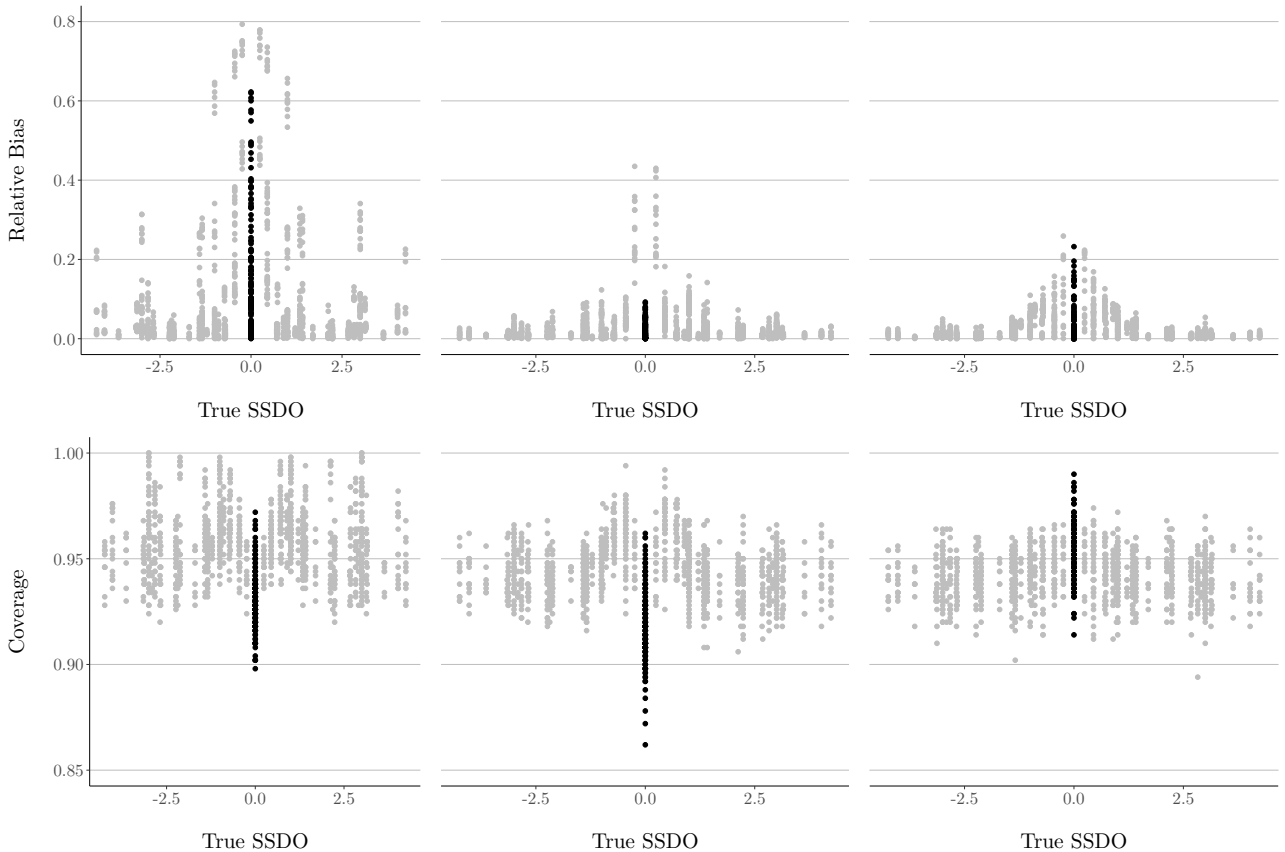

Figure A12: Relative bias and Coverage of the circular regression coefficients b_c, AS and SAM, from left to right, for location (grey) and accuracy (black) designs with a sample size of 50 plotted against the true $SSDO$.

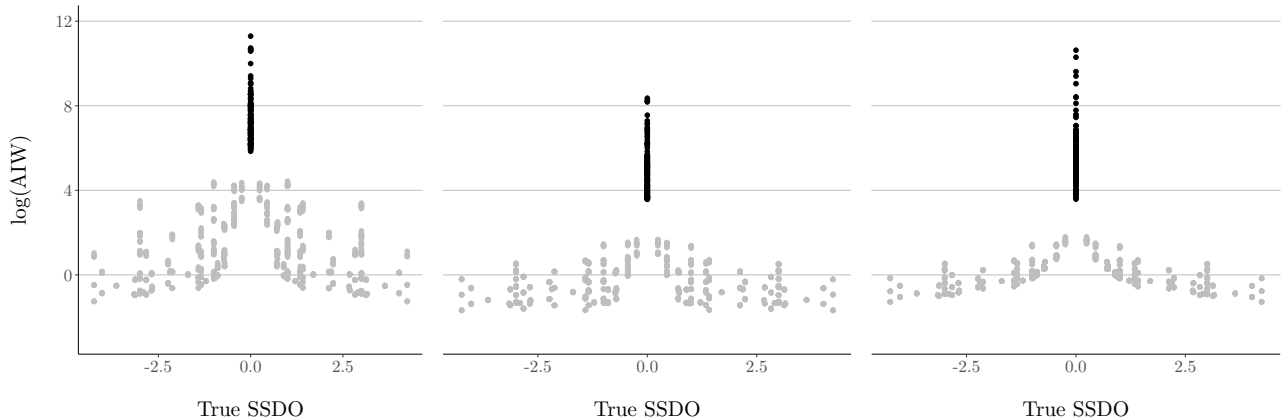

Figure A13: log(AIW) of the circular regression coefficients b_c, AS and SAM, from left to right, for location (grey) and accuracy (black) designs with a sample size of 50 plotted against the true $SSDO$.

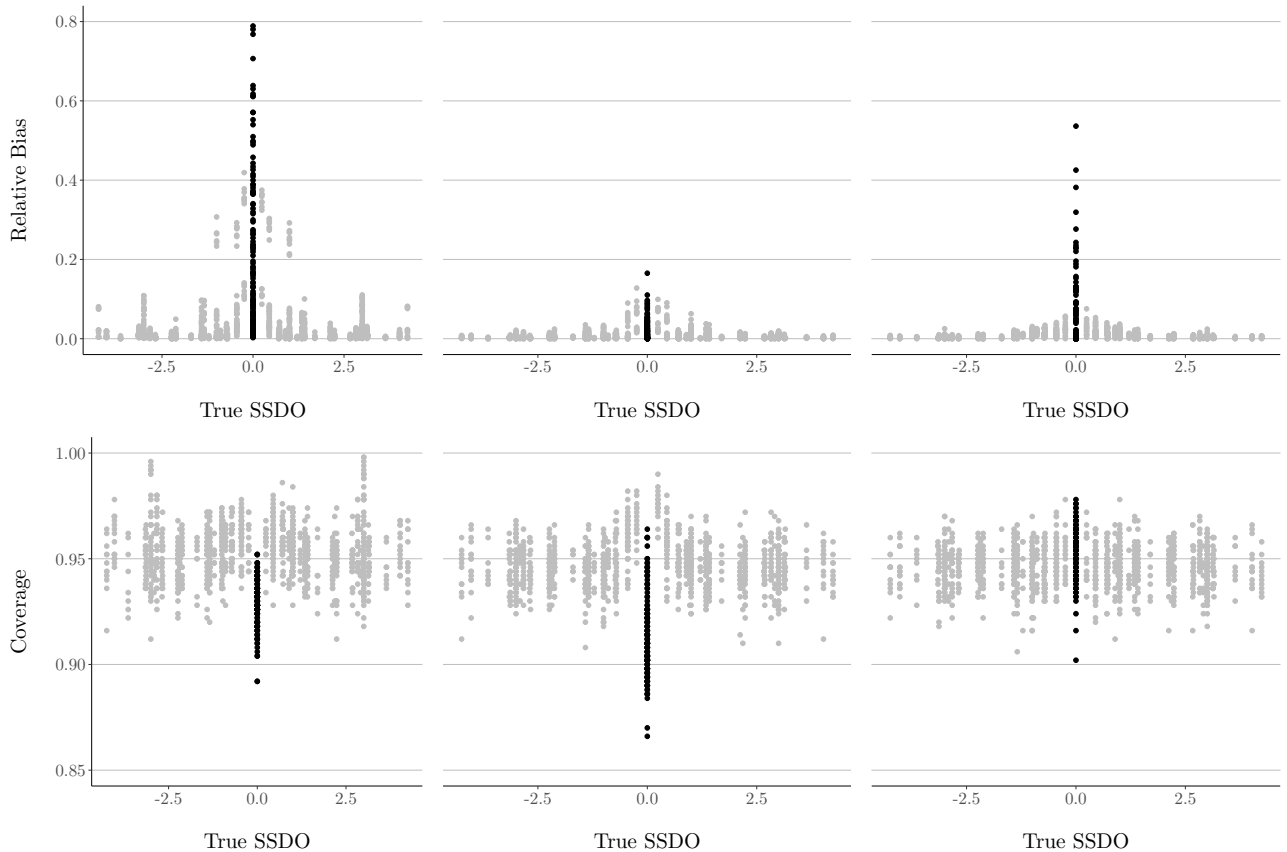

Figure A14: Relative bias and Coverage of the circular regression coefficients b_c, AS and SAM, from left to right, for location (grey) and accuracy (black) designs with a sample size of 200 plotted against the true $SSDO$.

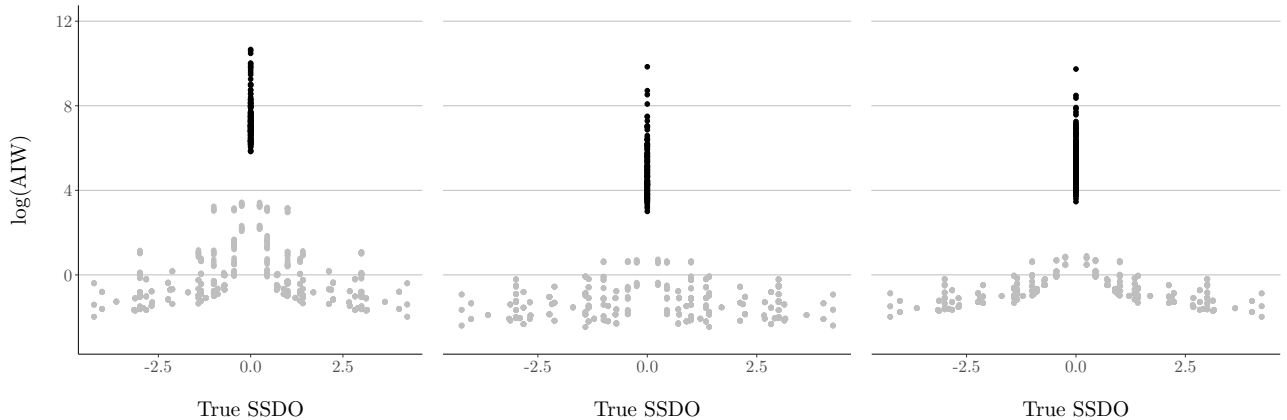

Figure A15: log(AIW) of the circular regression coefficients b_c, AS and SAM, from left to right, for location (grey) and accuracy (black) designs with a sample size of 200 plotted against the true $SSDO$.

Chapter 4

Table A5: The proportion of datasets in which an accuracy effect is indicated by $\sin(\lambda)$ grouped per real SDO value.

N	$SDO = 0$	$0 < SDO < 1$	$1 \leq SDO < 2$	$2 \leq SDO < 3$	$3 \leq SDO < 4$	$4 \leq SDO$
25	0.90	0.50	0.25	0.16	0.12	0.07
50	0.89	0.38	0.16	0.11	0.08	0.05
75	0.92	0.31	0.13	0.09	0.07	0.05
100	0.93	0.27	0.11	0.08	0.07	0.05
150	0.93	0.22	0.09	0.07	0.06	0.04

Table A6: The proportion of datasets in which an accuracy effect is indicated by $\sin(\gamma)$ grouped per real SDO value.

N	$SDO = 0$	$0 < SDO < 1$	$1 \leq SDO < 2$	$2 \leq SDO < 3$	$3 \leq SDO < 4$	$4 \leq SDO$
25	0.94	0.56	0.28	0.18	0.13	0.08
50	0.94	0.42	0.18	0.12	0.09	0.05
75	0.95	0.34	0.14	0.09	0.07	0.05
100	0.95	0.30	0.12	0.08	0.07	0.05
150	0.95	0.23	0.10	0.07	0.06	0.04

Table A7: The proportion of datasets in which an accuracy effect is indicated by $\sin(\gamma + \lambda)$ grouped per real SDO value.

N	$SDO = 0$	$0 < SDO < 1$	$1 \leq SDO < 2$	$2 \leq SDO < 3$	$3 \leq SDO < 4$	$4 \leq SDO$
25	0.91	0.53	0.25	0.16	0.12	0.07
50	0.92	0.39	0.17	0.11	0.08	0.05
75	0.94	0.32	0.13	0.09	0.07	0.05
100	0.95	0.28	0.11	0.08	0.07	0.05
150	0.95	0.23	0.09	0.07	0.06	0.04

Chapter 5

Bayesian Model and MCMC procedure

Consider the projected normal density for θ:

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

where θ is the circular response variable $0 < \theta \leq 2\pi$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^t \in \mathbb{R}^2$, $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{u}^t = (\cos\theta, \sin\theta)$. $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution. In the circular mixed-effects model, we have independent observations of a matrix of fixed effect $\boldsymbol{X}$ and random effect predictors $\boldsymbol{Z}$ in addition to a circular outcome vector $\boldsymbol{u}$ for each individual $i = 1, \ldots, n$. The rows of the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ and the indexes of the vector $\boldsymbol{u}$ represent the measurement occasions $j = 1, \ldots, N$. The longitudinal model has the following mean structure:

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^I \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\beta}^I)^t \boldsymbol{X}_{ij}^I + (\boldsymbol{b}_i^I)^t \boldsymbol{Z}_{ij}^I \\ (\boldsymbol{\beta}^{II})^t \boldsymbol{X}_{ij}^{II} + (\boldsymbol{b}_i^{II})^t \boldsymbol{Z}_{ij}^{II} \end{pmatrix},$$

where $\boldsymbol{\beta}^I$ and $\boldsymbol{\beta}^{II}$ are vectors with fixed effect coefficients and intercept and $\boldsymbol{b}_i^I$ and $\boldsymbol{b}_i^{II}$ are vectors with random effects for each individual. Note that each individual thus has two matrices of fixed effect predictors and two matrices of random effect predictors, one for each component of the mean vector of the model.

Bayesian Model and Priors

If we consider two components $k \in \{I, II\}$ the specification of the circular mixed-effects model is a standard bivariate mixed-effects model, a follows (Nuñez-Antonio & Gutiérrez-Peña, 2014):

- $\mathbf{Y}_i^k | \, \boldsymbol{\beta}^k, \{\mathbf{b}_i^k\} \sim N_{n_i}(\mathbf{X}_i^k \boldsymbol{\beta}^k + \mathbf{Z}_i^k \mathbf{b}_i^k, I_{n_i})$, $i = 1, \ldots, N$.
- $\boldsymbol{\beta}^k \sim N_{p_k}(\mathbf{0}, \mathbf{A}^k)$,
  $\mathbf{b}_i^k | \boldsymbol{\Omega}^k \sim N_{q_k}(\mathbf{0}, \boldsymbol{\Omega}^k), i = 1, \ldots, N$,
  where $\boldsymbol{\beta}^k$ and $\mathbf{b}_i^k$ are assumed independent and where p_k and q_k denote the dimensions of $\boldsymbol{\beta}^k$ and $\mathbf{b}_i^k$.
- $\boldsymbol{\Omega}^k \sim Wi(\nu^k, \mathbf{B}^k)$, $\nu^k \geq q^k$.

where $N_{p_k}(\mathbf{0}, \mathbf{A}^k)$ and $N_{q_k}(\mathbf{0}, \boldsymbol{\Omega}^k)$ are the priors for the fixed and random effect coefficient and $Wi(\nu^k, \mathbf{B}^k)$ is the prior for the variance of the random effects.

Conditional Posteriors

The conditional posteriors for the model parameters are the following for each of the components $k \in \{I, II\}$ (Nuñez-Antonio & Gutiérrez-Peña, 2014):

- $f(\boldsymbol{\beta}|\{\mathbf{b}_i\}, \boldsymbol{\Omega}, \mathbf{D}) = N_p\left(\boldsymbol{\beta}|\mathbf{C}^{-1}\sum_{i=1}^{N}\mathbf{X}_i^t\mathbf{e}_i, \mathbf{C}\right)$,
  where
  $$\mathbf{C} = \sum_{i=1}^{N}\mathbf{X}_i^t\mathbf{X}_i + \mathbf{A},$$
  $$\mathbf{e}_i = \mathbf{Y}_i - \mathbf{Z}_i\mathbf{b}_i$$
  and
  $$\mathbf{D}_i = \{(r_{11}, \theta_{11}), \ldots, (r_{Nn_N}, \theta_{Nn_n})\}.$$
  Here, $\mathbf{A}$ represents the prior precision matrix of $\boldsymbol{\beta}$ which indirectly also serves as a prior mean vector of $\boldsymbol{\beta}$. In this paper we use an uninformative prior $\mathbf{A} = \mathbf{0}$. This means that no prior information is taken into account for the precision matrix of $\boldsymbol{\beta}$ and the posterior mean vector does not include prior information either (zero prior information in posterior variance-covariance matrix so only likelihood information in posterior mean vector; includes inverse of posterior variance-covariance matrix).

- $f(\mathbf{b}_i|\boldsymbol{\beta}, \boldsymbol{\Omega}, \mathbf{D}) = N_q(\mathbf{b}_i|\mathbf{D}_i^{-1}\mathbf{Z}_i^t\tilde{\mathbf{e}}_i, \mathbf{D}_i)$,
  for $i = 1, \ldots, N$. Here, $\mathbf{C}$ is the same as in $f(\boldsymbol{\beta}|\{\mathbf{b}_i\}, \boldsymbol{\Omega}, \mathbf{D})$, $\mathbf{D}_i = \mathbf{Z}_i^t\mathbf{Z}_i + \boldsymbol{\Omega}$ and $\tilde{\mathbf{e}}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_i$.

- $f(\boldsymbol{\Omega}|\{\mathbf{b}_i\}, \mathbf{D}) = Wi\left(\nu + N, \mathbf{B} + \sum_{i=1}^{N}\mathbf{b}_i\mathbf{b}_i^t\right)$,
  where ν is the prior degrees of freedom and $\mathbf{B}$ is the prior scale matrix. In a wishart distribution the scale matrix is the sum of squares $\sum_{i=1}^{N}\mathbf{b}_i\mathbf{b}_i^t$ and thus $\mathbf{B}$ refers to the prior sum of squares. A small sum of squares represents a small effect.

- $f(r_{ij}|\{\theta_{ij}\}, \boldsymbol{\beta}, \mathbf{b}_i) \propto r_{ij}exp\left\{-0.5[r_{ij}^2 - 2b_{ij}r_{ij}]\right\}$,
  where $b_{ij} = \boldsymbol{u}_{ij}^t\boldsymbol{\mu}_{ij}$

the derivation of these conditionals can be found in Nuñez-Antonio (2010).

MCMC sampling

To sample from the posterior we use the following algorithm:

1. For each individual and each timepoint a starting value for r_{ij} is chosen. We choose 1.

2. Using the starting value for an r_i and $\boldsymbol{u}_i$ computed from the data we may compute $\boldsymbol{y}_{ij} = r_{ij}\boldsymbol{u}_{ij}$.

3. For each of the components $k \in \{I, II\}$ we:

- Sample $\boldsymbol{\beta}^k$ from

$$f(\boldsymbol{\beta}^k|\boldsymbol{\Omega}^k, \mathbf{D}) = N_{p_k}\left(\boldsymbol{\beta}^k|(\boldsymbol{\Lambda}_\beta^k)^{-1}\left\{\sum_{i=1}^{N}(\mathbf{X}_i^k)^t(\mathbf{V}_i^k)^{-1}\mathbf{Y}_i^k\right\}, \left(\mathbf{A}^k + \sum_{i=1}^{N}(\mathbf{X}_i^k)^t(\mathbf{V}_i^k)^{-1}\mathbf{X}_i^k\right)\right).$$

- Sample $\mathbf{b}_i^k$ from $f(\mathbf{b}_i|\boldsymbol{\beta}, \boldsymbol{\Omega}, \mathbf{D})$, for each $i = 1, \ldots, N$.
- Sample $\boldsymbol{\Omega}^k$ from $f(\boldsymbol{\Omega}^k|\{\mathbf{b}_i^k\}, \mathbf{D})$.

4. Sample r_{ij} from $f(r_{ij}|\{\theta_{ij}\}, \boldsymbol{\beta}, \mathbf{b}_i)$ for each $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. This is done using a slice sampler that was adapted from Hernandez-Stumpfhauser et al. (2017) for the mixed-effects regression situation. The joint density for the auxiliary variable v_{ij} with r_{ij} for mixed-effects regression is:

$$p(r_{ij}, v_{ij} \mid \theta_{ij}, \boldsymbol{\mu}_{ij}) \propto r_{ij}\mathbf{I}(0 < v_i < \exp\{-.5(r_{ij} - b_{ij})^2\})\mathbf{I}(r_{ij} > 0)$$

The full conditionals for v_{ij} and r_{ij} are:

$$p(v_{ij} \mid r_{ij}, \boldsymbol{\mu}_{ij}, \theta_{ij}) \sim U(0, \exp\{-.5(r_{ij} - b_{ij})^2\})$$

$$p(r_{ij} \mid v_{ij}, \boldsymbol{\mu}_{ij}, \theta_{ij}) \propto r_{ij}\mathbf{I}\left(b_{ij} + \max\left\{-b_{ij}, -\sqrt{-2\ln v_{ij}}\right\} < r_{ij} < b_{ij} + \sqrt{-2\ln v_{ij}}\right)$$

We thus sample v_{ij} from the uniform distribution specified above. Independently we sample a value m from $U(0, 1)$. We obtain a new value for r_{ij} by computing $r_{ij} = \sqrt{(r_{ij_2}^2 - r_{ij_1}^2)m + r_{ij_1}^2}$ where $r_{ij_1} = b_{ij} + \max\left\{-b_{ij}, -\sqrt{-2\ln v_{ij}}\right\}$ and $r_{ij_2} = b_{ij} + \sqrt{-2\ln v_{ij}}$.

5. Using the new r_{ij} new values for $\boldsymbol{y}_{ij}$ are computed as $\boldsymbol{y}_{ij} = r_{ij}\boldsymbol{u}_{ij}$.
6. Steps 3, 4, and 5 are repeated for a specified amount of iterations. After the iterations are completed convergence is checked. If convergence is not reached additional iterations are run.

Chapter 6

In this Appendix 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 $\boldsymbol{Z}^k_{n\times 2}$ (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.

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 $\boldsymbol{\mu}_{j_n} = \boldsymbol{\Sigma}_{j_n}\boldsymbol{\Sigma}^{-1}\left(-\sum_{i=1}^n z_{i,j-1}\sum_{l\neq j} z_{i,l-1}\boldsymbol{\beta}_l + \sum_{i=1}^n z_{i,j-1} r_i \begin{bmatrix}\cos(\theta_i)\\ \sin(\theta_i)\end{bmatrix}\right)$
     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 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. 2017):

- In a slice sampler the joint density for an auxiliary variable v_i with r_i $(r_i > 0)$ is

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

- 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.

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 $B_i = \begin{bmatrix}\cos(\theta_i) \\ \sin(\theta_i)\end{bmatrix}^t \boldsymbol{\Sigma}_{s_i|y_i}^{-1} \boldsymbol{\mu}_{s_i|y_i}$, $\boldsymbol{\mu}_{s_i|y_i} = \boldsymbol{B}_{s_i|y_i}^t \boldsymbol{x}_i$ 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 $(\boldsymbol{s}_i^t, y_i)^t \sim N_{2m+w}(\boldsymbol{\mu} + (\boldsymbol{0}_{2m}^t, \lambda d_i)^t, \boldsymbol{\Sigma})$. 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. 2017):

- In a slice sampler the joint density for an auxiliary variable v_i with r_i $(r_i > 0)$ is

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

- 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

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

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

Agostinelli, C., & Lund, U. (2017). *R package* `circular`*: Circular statistics (version 0.4-93)*. Retrieved from https://r-forge.r-project.org/projects/circular/

Batschelet, E. (1981). *Circular statistics in biology*. London: Academic Press.

Brekelmans, M., Wubbels, T., & van Tartwijk, J. (2005). Teacher–student relationships across the teaching career. *International Journal of Educational Research*, *43*(1), 55–71. doi:10.1016/j.ijer.2006.03.006

den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement*, *15*(3/4), 407–442. doi:10.1080/09243450512331383262

Brunyé, T. T., Burte, H., Houck, L. A., & Taylor, H. A. (2015). The map in our head is not oriented north: Evidence from a real-world environment. *PLoS ONE*, *10*(9), 1–12. doi:10.1371/journal.pone.0135803

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

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, *93*(1), 43–54. doi:10.1037/0022-0663.93.1.43

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

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, *77*(1), 113–143. doi:10.3102/003465430298563

Cremers, J. (2018). *Bpnreg: Bayesian projected normal regression models for circular data*. Retrieved from https://CRAN.R-project.org/package=bpnreg

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

doi:10.1027/1614-2241/a000147

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

Cremers, J., Pennings, H. J., Mainhard, M. T., & Klugkist, I. (2019). Longitudinal circular modelling of circumplex measurements for interpersonal behavior. *Manuscript Submitted for Publication*.

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

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

Fisher, N. I., & Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, *48*(3), 665–677.

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

Garcia-Portugués, E., Crujeiras, R. M., & González-Manteiga, W. (2013). Exploring wind direction and SO2 concentration by circular–linear density estimation. *Stochastic Environmental Research and Risk Assessment*, *27*(5), 1055–1067. doi:10.1007/s00477-012-06 42-5

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

Gill, J., & Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, *18*(3), 316–336. doi:10.1093/pan/mpq009

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

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

Hernandez-Stumpfhauser, D., Breidt, F. J., & van der Woerd, M. J. (2017). The general projected normal distribution of arbitrary dimension: Modeling and Bayesian inference. *Bayesian Analysis*, *12*(1), 113–133. doi:10.1214/15-BA989

Heyes, S. B., Zokaei, N., & Husain, M. (2016). Longitudinal development of visual working memory precision in childhood and early adolescence. *Cognitive Development*, *39*, 36–44. doi:10.1016/j.cogdev.2016.03.004

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

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Hove: Routledge.

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

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

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

Kendall, D. G. (1974). Pole-seeking brownian motion and bird navigation. *Journal of the Royal Statistical Society. Series B*, *36*(3), 365–417.

Kim, S., & SenGupta, A. (2015). Inverse circular–linear/linear–circular regression. *Communications in Statistics-Theory and Methods*, *44*(22), 4772–4782. doi:10.1080/0361092 6.2013.804561

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *102*(3), 299–314. doi:10.1016/j.jecp.2008.07.005

König, J., Onnen, M., Karl, R., Rosner, R., & Butollo, W. (2016). Interpersonal subtypes and therapy response in patients treated for posttraumatic stress disorder. *Clinical Psychology & Psychotherapy*, *23*(2), 97–106. doi:10.1002/cpp.1946

Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von mises distribution. *Environmental and Ecological Statistics*, *23*(1), 89–113. doi:10.1007/s10651-015-0330-y

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

Ley, C., & Verdebout, T. (2017). *Modern directional statistics.* Boca Raton, FL: Chapman & Hall/CRC.

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*, *25*(8), 988–1000. doi:10.1177/1073191116672330

Mainhard, M. T., Brekelmans, M., den Brok, P., & Wubbels, T. (2011a). The development of the classroom social climate during the first months of the school year. *Contemporary Educational Psychology*, *36*(3), 190–200. doi:10.1016/j.cedpsych.2010.06.002

Mainhard, M. T., Brekelmans, M., & Wubbels, T. (2011b). Coercive and supportive teacher behaviour: Within-and across-lesson associations with

the classroom social climate. *Learning and Instruction*, *21*(3), 345–354. doi:10.1016/j.learninstruc.2010.03.003

Mainhard, M. T., Pennings, H. J., Wubbels, T., & Brekelmans, M. (2012). Mapping control and affiliation in teacher–student interaction with state space grids. *Teaching and Teacher Education*, *28*(7), 1027–1037. doi:10.1016/j.tate.2012.04.008

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

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

Mardia, K., Taylor, C. C., & Subramaniam, G. K. (2006). Protein bioinformatics and mixtures of bivariate von mises distributions for angular data. *Biometrics*, *63*(2), 505–512. doi:10.1111/j.1541-0420.2006.00682.x

Maruotti, A. (2016). Analyzing longitudinal circular data by projected normal models: A semi-parametric approach based on finite mixture models. *Environmental and Ecological Statistics*, *23*(2), 257–277. doi:10.1007/s10651-015-0338-3

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., Lasinio, G. J., & Gelfand, A. E. (2016). Spatio-temporal circular models with non-separable covariance structure. *Test*, *25*(2), 331–350. doi:10.1007/s11749-015-0458-y

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

Matsushima, E. H., Vaz, A. M., Cazuza, R. A., & Ribeiro Filho, N. P. (2014). Independence of egocentric and exocentric direction processing in visual space. *Psychology & Neuroscience*, *7*(3), 277–284. doi:10.3922/j.psns.2014.050

Moskowitz, D., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology*, *86*(6), 880–893. doi:10.1037/0022-3514.86.6.880

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

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, *31*(3), 705–741. doi:10.1214/aos/1056562461

Nunez-Antonio, G., & Gutierrez-Pena, E. (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics*, *32*(10), 995–1001. doi:10.1080/02664760500164886

Nuñez-Antonio, G. (2010). *Análisis Bayesiano de modelos lineales para datos direccionales considerando la distribución normal bajo proyección.* (PhD thesis). Universidad Autonoma Metropolitana.

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

Nuñez-Antonio, G., Gutiérrez-Peña, E., & Escarela, G. (2011). A Bayesian regression model for circular data based on the projected normal distribution. *Statistical Modelling*, *11*(3), 185–201. doi:10.1177/1471082X1001100301

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

Ouwehand, P. E. W., & Peper, C. L. E. (2015). Does interpersonal movement synchronization differ from synchronization with a moving object? *Neuroscience Letters*, *606*, 177–181. doi:10.1016/j.neulet.2015.08.052

Pennings, H. J., Brekelmans, M., Sadler, P., Claessens, L. C., van der Want, A. C., & van Tartwijk, J. (2018). Interpersonal adaptation in teacher-student interaction. *Learning and Instruction*, *55*, 41–57. doi:10.1016/j.learninstruc.2017.09.005

Pennings, H. J., Brekelmans, M., Wubbels, T., van der Want, A. C., Claessens, L. C., & van Tartwijk, J. (2014). A nonlinear dynamical systems approach to real-time teacher behavior: Differences between teachers. *Nonlinear Dynamics, Psychology, and Life Sciences*, *18*(1), 23–45.

Pennings, H. J., van Tartwijk, J., Wubbels, T., Claessens, L. C., van der Want, A. C., & Brekelmans, M. (2014). Real-time teacher–student interactions: A dynamic systems approach. *Teaching and Teacher Education*, *37*, 183–193. doi:10.1016/j.tate.2013.07.016

Pewsey, A., Neuhäuser, M., & Ruxton, G. D. (2013). *Circular statistics in R*. Oxford: Oxford University Press.

Presnell, B., Morrison, S. P., & Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, *93*(443), 1068–1077. doi:10.1080/01621459.1998.10473768

Puglisi, G., Leonetti, A., Landau, A., Fornia, L., Cerri, G., & Borroni, P. (2017). The role of attention in human motor resonance. *PloS One*, *12*(5), e0177457. doi:10.1371/journal.pone.0177457

Ravindran, P., & Ghosh, S. K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice*, *5*(4), 547–561. doi:10.1080/15598608.2011.10483731

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

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rivest, L.-P., Duchesne, T., Nicosia, A., & Fortin, D. (2015). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *63*(3), 445–463. doi:10.1111/rssc.12124

Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, *464*(7290), 903–907. doi:10.1038/nature08860

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

Salzer, S., Pincus, A. L., Hoyer, J., Kreische, R., Leichsenring, F., & Leibing, E. (2008). Interpersonal subtypes within generalized anxiety disorder. *Journal of Personality Assessment*, *90*(3), 292–299. doi:10.1080/00223890701885076

Santos, G. C., Vandenberghe, L., & Tavares, W. M. (2015). Interpersonal interactions in the marital pair and mental health: A comparative and correlational study. *Paidéia (Ribeirão Preto)*, *25*(62), 373–382. doi:10.1590/1982-43272562201511

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45. doi:10.1111/j.1540-4560.1994.tb01196.x

Simon, S., Cain, N. M., Wallner Samstag, L., Meehan, K. B., & Muran, J. C. (2015). Assessing interpersonal subtypes in depression. *Journal of Personality Assessment*, *97*(4), 364–373. doi:10.1080/00223891.2015.1011330

Stoffregen, T. A., Bardy, B. G., Merhi, O. A., & Oullier, O. (2004). Postural responses to two technologies for generating optical flow. *Presence: Teleoperators and Virtual Environments*, *13*(5), 601–615. doi:10.1162/1054746042545274

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

Venter, J. (1967). On estimation of the mode. *The Annals of Mathematical Statistics*, *38*(5), 1446–1455.

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

Wang, F., & Gelfand, A. E. (2014). Modeling space and space-time directional data using projected gaussian processes. *Journal of the American Statistical Association*, *109*(508), 1565–1580. doi:10.1080/01621459.2014.934454

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

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition*, *166*, 152–163. doi:10.1016/j.cognition.2017.05.020

Wright, A. G., Hallquist, M. N., Morse, J. Q., Scott, L. N., Stepp, S. D., Nolf, K. A., & Pilkonis, P. A. (2013). Clarifying interpersonal heterogeneity in borderline

personality disorder using latent mixture modeling. *Journal of Personality Disorders*, *27*(2), 125–143. doi:10.1521/pedi.2013.27.2.125

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*(4), 311–322. doi: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). Malwah, NJ: Lawrence Erlbaum Associates.

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*(4), 509–536. doi:10.1080/15298868.2010.497376

Zilcha-Mano, S., McCarthy, K. S., Dinger, U., Chambless, D. L., Milrod, B. L., Kunik, L., & Barber, J. P. (2015). Are there subtypes of panic disorder? An interpersonal perspective. *Journal of Consulting and Clinical Psychology*, *83*(5), 938–950. doi:10.1037/a0039373

Nederlandse Samenvatting

Binnen de sociale wetenschappen worden zogenoemde circulaire data verzameld in verschillende vakgebieden. Circulaire data is data die gemeten kan worden binnen het domein $(0°, 360°]$. Het meest directe voorbeeld van circulaire data is de zogenaamde kompas data. Dit is data die wordt verkregen door middel van metingen met een kompas. Binnen de sociale wetenschappen komt kompas data vooral voor binnen de cognitieve psychologie, o.a. in onderzoek naar het menselijk gevoel voor richting (Brunyé et.al., 2015; Chrastil & Warren, 2017; Warren et.al., 2017), de visuele perceptie van ruimte (Matsushima et.al., 2014) en in onderzoek naar oogbewegingen (Rayner, 2009). Deze dissertatie bevat voorbeelden van kompas data in Hoofdstuk 1 en 3. Het tweede voorbeeld van circulaire data binnen de sociale wetenschappen betreft metingen met periodieke eigenschappen. Dit kan betrekking hebben op onderzoek waarin we geïnteresseerd zijn in op welk moment van de dag, de week, het jaar of een andere specifieke tijdsperiode een bepaalde gebeurtenis plaatsvind. Ook onderzoek waarin we geïnteresseerd zijn in een faseverschil valt binnen de tweede categorie van circulaire data. Het tweede type circulaire data komt voor binnen onderzoek naar menselijke bewegingscognitie of de synchronisatie van bewegingen of geluid (Ouwehand & Peper, 2015; Puglisi et al., 2017), bijvoorbeeld in het onderzoek van Kirschner & Tomasello (2009) naar geluidssynchronisatie tijdens het maken van muziek door kinderen. Maar ook binnen de criminologie en politicologie komt het tweede type van circulaire data voor, bijvoorbeeld in onderzoek waarin we geïnteresseerd zijn in het tijdstip waarop het meeste politieke geweld of de meeste misdaden met wapens plaatsvinden (Gill & Hangartner, 2009). In deze dissertatie gebruiken we een voorbeeld van het tweede soort circulaire data in Hoofdstuk 1 waar we een experiment beschrijven waarin de afhankelijke variabele het verschil in fase tussen handbewegingen is. Het derde type circulaire data komt voor in onderzoek waar gebruikt wordt gemaakt van een circulaire schaal. Binnen de sociale wetenschappen komen drie van dit soort schalen voor: de interpersoonlijke circumplex (Horowitz & Strack, 2011) uit de interpersoonlijke psychologie en onderwijswetenschappen, het politieke spectrum met de assen links-rechts en progressief-conservatief (Gill & Hangartner, 2009) en de 'Basic Human Values' schaal (Schwartz, 1994) binnen de sociologie. In deze dissertatie gebruiken we data van de interpersoonlijke circumplex in de onderwijswetenschappen in Hoofdstuk 2, 5 en 6.

In deze dissertatie maken we voornamelijk gebruik van één specifieke methode voor het analyseren van circulaire data, de inbedding methode. In deze methode nemen

we aan dat een circulaire variabele is ontstaan door de projectie van een latente bivariate variabele op de cirkel. Het voordeel van deze methode in vergelijking met de andere twee methoden voor circulaire data, de intrinsieke en omwind methode, is dat ze flexibeler is. Doordat we uitgaan van een projectie van de bivariate ruimte naar de cirkel kunnen we gebruik maken van reeds bestaande methoden voor lineaire data (op een schaal van -∞ tot ∞) en dus een model schatten voor bivariaat lineaire data in plaats van circulaire data. Het modelleren van circulaire data met deze methode is daardoor flexibeler dan met de andere methodes. De inbedding methode heeft echter ook een nadeel. Doordat we een model schatten in de bivariate ruimte zijn de resultaten ook bivariaat. Dit nadeel wordt verholpen in deze dissertatie door het introduceren van nieuwe maten die de resultaten in bivariate ruimte transformeren naar de cirkel. We doen dit in Hoofstuk 3 en 4 voor een regressiemodel en in Hoofdstuk 5 voor een 'mixed-effects' model. Verder hebben we in deze dissertatie de pragmatische keuze gemaakt om gebruik te maken van Bayesiaanse methoden. We doen dit vanwege het feit dat er voor complexere modellen (bijvoorbeeld het mixed-effects model) alleen Bayesiaanse methoden beschikbaar waren in de literatuur. Verder maakt een Bayesiaanse methode het makkelijk om onzekerheidsschattingen voor en transformaties van parameters te verkrijgen. Hiervan hebben we onder andere gebruik gemaakt voor het ontwikkelen van nieuwe maten in Hoofstuk 3.

Gezien het feit dat circulaire data relatief vaak voorkomt binnen de sociale wetenschappen zou men verwachten dat het gebruik van speciale methoden voor dit soort data wijdverspreid is binnen de sociale wetenschappen. Dit is echter niet het geval. Deze dissertatie heeft daarom drie doelstellingen: het beschikbaar maken van methoden voor circulaire data voor de toegepaste onderzoeker, het testen van beschikbare methoden voor circulaire data en het aanpassen (en uitbreiden) van deze methoden om ze geschikt te maken voor gebruik door sociale wetenschappers. Hoofstuk 1 is een 'tutorial' over de analyse van circulaire data voor cognitief psychologen en sociale wetenschappers in het algemeen. Als onderdeel van deze tutorial hebben we een R package ontworpen waarmee sociale wetenschappers regressiemodellen (GLMs en mixed-effects modellen) met een circulaire afhankelijke variabele kunnen schatten. In Hoofdstuk 2 testen we het functioneren van een circulair regressiemodel met behulp van een simulatiestudie. In Hoofdstuk 3, 4 en 5 introduceren we nieuwe maten die de interpretatie van circulaire modellen vereenvoudigt. In Hoofdstuk 5 illustreren we ook de voordelen van een circulair model voor het analyseren van longitudinale data van een interpersoonlijke circumplex. Als laatste ontwikkelen we in Hoofdstuk 6 nieuwe regressie modellen voor zogenaamde cilindrische data. Dit zijn multivariate data met een lineaire én circulaire afhankelijke variabele. Modellen voor cilindrische data hebben wellicht nog meer voordelen voor het modelleren van data van de interpersoonlijke circumplex dan modellen voor circulaire data.

Acknowledgements

A dissertation is not finished without thanking those people who have in some way contributed to the completion of it.

The first and most important person to thank is Irene, without whom this project would not have been possible. A good supervisor knows how to help her students and advise them what to do, but a great supervisor also makes sure to give her students the freedom to explore and execute their own ideas. I would also like to thank Herbert, who as a co-supervisor has been around whenever he was needed.

Many thanks also go to my co-authors. Heleen and Tim, without your ideas and discussions on circumplex data this dissertation would not be complete. Also thanks to Mieke for providing data for Chapter 5. Christophe, thanks for the invitation to work together on my idea for cylindrical regression but also for the social atmosphere in Gent. Daniel, thanks for our efficient skype meetings and all your ideas for new measures in Chapter 4.

I am grateful for having had such great colleagues and work environment at the M&S department and elsewhere. First of all, thanks to my fellow M&S research master students. You were there at the start of my journey into statistics in 2012 and are still part of that today. Second, thanks for all the (PhD) lunches at the M&S department that either resulted in food or statistics themed discussions. Oisín, you'll be glad to hear that 'knifey' has found a new and loving foster home. Fayette and Kees, thanks for being roommates and fellow PhDs that are willing to throw eggs out of windows, have discussions on (circular) data analysis and chit-chat about other not so very work related topics. Til sidst vil jeg takke Andreas, David og alle andre fra SODAS i København, without whom I would not have learned so many new things during the last part of my PhD.

Finally, I would like to thank friends and family for being part of and enriching my life outside of work. Lianne, Ilona en Irene, bedankt voor alle gezelligheid in het algemeen maar vooral ook voor: de 'culinaire hoogstandjes' van het campingfornuis tijdens de Nieuw-Zeelandse winter, de halve boekwerken aan tinder/date evaluaties en het (bijna) om de hoek komen wonen waardoor we weer gewoon plat kunnen praten. Mam, pap, bedankt voor het beantwoorden van mijn eindeloos aantal vragen en het stimuleren van mijn nieuwsgierigheid. Zonder een honger naar kennis word je geen wetenschapper. Vinnie, 'Nee, jíj bent de allerliefste M&Ser ooit.'

About the author

Jolien Cremers was born on February 8^{th}, 1992 in Venlo, the Netherlands. In 2012 she obtained her BA in liberal arts and sciences with a focus on political science, human geography and methods and statistics. She continued to pursue her interest in applied statistics by completing a research master (MSc) in methodology and statistics for social and behavioral sciences in 2014. In September 2014 she started her PhD at Utrecht University.

During her PhD she focused on developing and evaluating Bayesian methods for circular data in the social sciences. In addition to her research activities Jolien worked as a junior policy officer at the Netherlands Organization for Scientific Research (NWO), where she assisted in the review process of scientific research grant proposals. Currently Jolien works at the University of Copenhagen and Danmarks Statistik where she is doing research on modeling life trajectories from registry data.

Publications:

Cremers, J. & Klugkist, I. (2018). One direction? A tutorial for circular data using R with examples in cognitive psychology. *Frontiers in Psychology: Cognitive Science, 9*, 2040. doi: 10.3389/fpsyg.2018.02040.

Evers, C., Marchiori, D., Junghans, A., **Cremers, J.** & de Ridder, D. (2018). Citizen approval of nudging interventions promoting healthy eating: The role of intrusiveness and trustworthiness. *BMC Public Health, 18*(1), 1182. doi: 10.1186/s12889-018-6097-y.

Klugkist, I., **Cremers, J.** & Mulder, K. (2018). Bayesian analysis of circular data in social and behavioral sciences. In Ley, C & Verdebout, T. *Applied Directional Statistics: Modern Methods and Case Studies* (pp. 156 - 199). Chapman & Hall/CRC Press.

Cremers, J., Mainhard, M.T. & Klugkist, I. (2018). Assessing a Bayesian Embedding Approach to Circular Regression Models. *Methodology, 14*(2) 69-81. doi: 10.1027/1614-2241/a000147.

Cremers, J. Mulder, K. & Klugkist, I. (2018) Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology, 71*(1), 77-95. doi: doi:10.1111/bmsp.12108.