Bayesian Regression for Circular Data Using the Wrapping Approach

J. Cremers*, I. Jansen*, I. Klugkist

Utrecht University, Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences.

* These authors contributed equally to this work

Abstract

In this paper, a novel approach is introduced for circular regression based on the wrapping approach. Circular regression models are developed based on both the wrapped normal (WN) and wrapped Cauchy (WC) distribution. The scope of circular regression based on the wrapping approach is increased by making our methods applicable to models with an arbitrary number of continuous and dichotomous predictors. The resulting methods are validated by means of a simulation study, paying special attention to the differences between WN and WC. Simulation results show that under ideal conditions our methods work well. Additionally the methods react as expected to violations of assumptions in multiple regression and ANCOVA models. However, when the circular concentration of the data is lower the methods start to experience convergence problems due to multi-modality of the likelihood. Finally, to demonstrate the use of our techniques, a circular regression model is fit to a real dataset from the field of Cognitive Psychology.

Keywords: Circular data, Wrapping approach, Wrapped Normal distribution, Wrapped Cauchy distribution, Bayesian regression

Introduction

Circular data can be found both within and outside the social sciences. It is different from

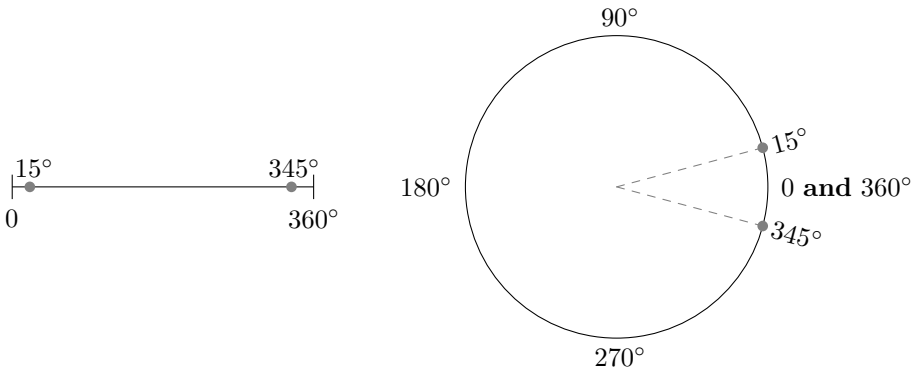


Figure 1. A linear (left) and circular (right) representation of the two data points 15° and 345°.

linear data in the sense that its natural representation is on the circle rather than on the real line. This is illustrated in Figure 1. If we would place the two directions 15° and 345° on the real line (left), they appear at the extremities of the interval $[0, 360°]$. However, if we place the same two points on the circle (right), it is clear that the two measurements actually point in almost the same direction.

There are different measurement instruments that provide circular data. The first instrument is a compass providing directions in angles. The resulting values can be represented in degrees with range $0° - 360°$ as in Figure 1, or in radians with range $0 - 2\pi$. Examples are research on the human sense of direction (see for example Brunyé, Burte, Houck, and Taylor (2015)) or on eye-hand coordination (f.e. Rentsch & Rand, 2014). Another instrument providing circular data is the clock. Time measured from $0 - 24$ hours can best be represented on the circle. But also other periodical measurements, like day of the year, provide circular data. Examples in the social sciences are Herrera, Resa, and Sordo (2010) who studied music listening patterns or Kimpton, Corcoran, and Wickes (2017) on crime timing. A third instrument regularly encountered in social sciences is the so-called circumplex. For an overview of circumplex instruments see Horowitz and Strack (2011). Examples of applications are, for instance, found in interpersonal psychology (e.g., in research by König, Onnen, Karl, Rosner, and Butollo (2016) on posttraumatic stress disorder) and in educational sciences (e.g., in a study by Pennings et al. (2018) on teacher behavior).

Conventional statistical methods cannot be used for circular data. As an example, consider the two data points provided in Figure 1. Using regular, linear, statistics the average of these two points would be $15 + 345 = 180$ and the distance (range) between the points $345 - 15 = 330$. However, the correct *mean direction* is 0° and the correct *distance on the circle* is 30°, as can

be seen in the righthand plot. In this paper, we investigate circular regression models, defined as models with a circular outcome variable and one or more predictors that may be linear, dichotomous (dummy variables), or circular. The analysis of such models requires the use of special statistical techniques.

Three approaches are available to deal with the periodic nature of circular data. The *intrinsic* approach makes use of distributions which are intrinsically defined on the circle, the most common being the Von Mises distribution (Von Mises, 1918). Key references for the intrinsic approach include Damien and Walker (1999); Nuñez-Antonio and Gutiérrez-Peña (2005b); Bhattacharya and SenGupta (2009) and Forbes and Mardia (2015). More recently, a Bayesian intrinsic approach for circular regression models was developed by Mulder and Klugkist (2017) and includes the R package **circglmbayes**. The *embedding* approach is based on *projecting* points in a two-dimensional space onto the circle. Any distribution can be assumed as the underlying distribution for the bivariate data but the most common choice is the bivariate normal. The resulting circular distribution is called the projected normal distribution (see, Nuñez-Antonio and Gutiérrez-Peña (2005a); Nuñez-Antonio, Gutiérrez-Peña, and Escarela (2011); Wang and Gelfand (2013)). Cremers, Mulder, and Klugkist (2018) extended these methods in the Bayesian framework and provided the R package **bpnreg** for circular regression models. The focus in this paper is on the third approach, the *wrapping* approach. The basic idea here is that the real line from minus to plus infinity is wrapped around the circle, such that observations that are a multiple of 2π apart end up on the same place on the circle. Any distribution on the line can be wrapped. If, for instance, the univariate normal distribution is assumed, the resulting distribution on the circle is called the *wrapped normal* (WN). Another probability density of interest is the *wrapped Cauchy* (WC) distribution.

Because of its intuitive relation to linear distributions, the wrapping approach could serve as an interesting alternative for circular regression. Yet so far its use in circular regression has hardly been investigated. Bayesian regression methods based on *data augmentation* were proposed (Coles, 1998), but only applied to the case of circular mean estimation (Ferrari, 2009) or to uni-variable circular regression (Ravindran, 2002; Ravindran & Ghosh, 2011). The only multi-variable model that was derived so far was based on maximum likelihood (ML) estimation for the WC (Abuzaid & Allahham, 2015).

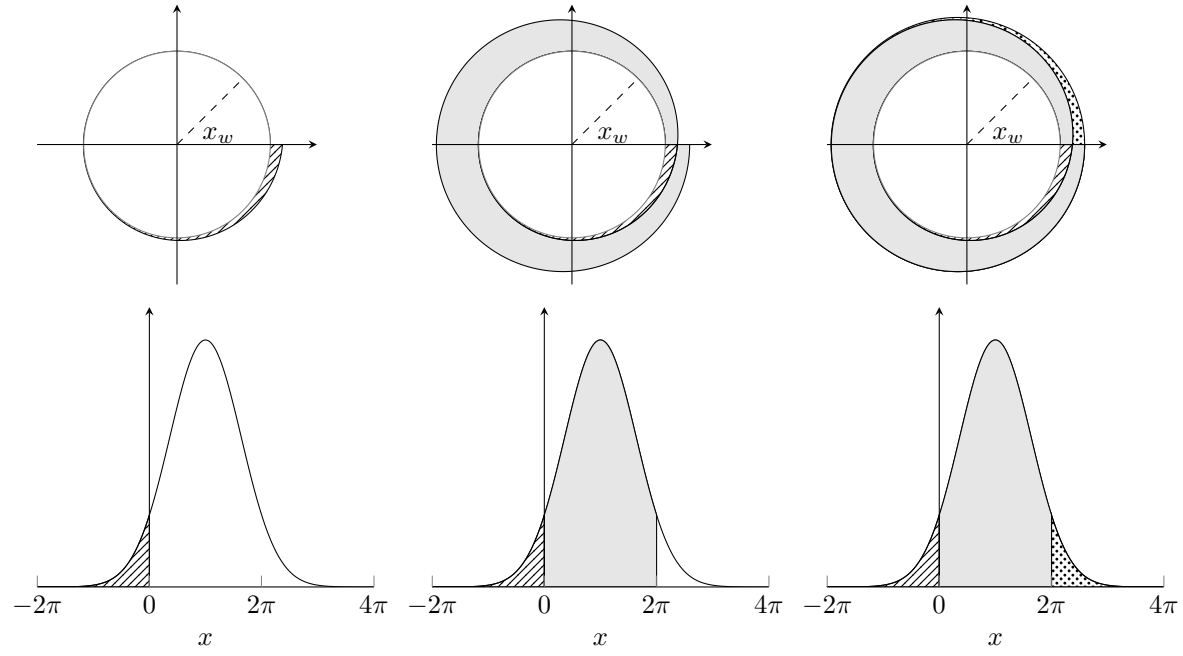
In this paper, we present the wrapping approach for circular multi-variable regression models using a Bayesian approach. In the next section, we will first elaborate on the wrapping approach using both the WN and WC distribution. Then, in the methods section we will derive the posterior distributions based on both WN and WC and describe how to sample from them. Our methods will be validated by means of a simulation study. The simulations include an investigation of performance under suboptimal circumstances that are often encountered in practice, like violated model assumptions and limited sample sizes. Subsequently, an application of the new regression techniques to an existing dataset from Cognitive Psychology is presented. The paper ends with a discussion of the performance of our methods.

The wrapping approach

As stated in the introduction, the idea behind the wrapping approach is simple (Fisher, 1995; Mardia & Jupp, 1999). The real line is wrapped around the circle in such a way that every two points that are a multiple of 2π apart end up on the same place. This wrapping technique can also be applied to probability distributions defined on the real line. In Figure 2, the process of wrapping a normal distribution around the circle is illustrated. The first step is to wrap the interval $[-2\pi, 0]$ (corresponding to the striped area in Figure 2a), followed by wrapping the interval $[0, 2\pi]$ (the solid grey area in Figure 2b) and $[2\pi, 4\pi]$ (the dotted area in Figure 2c).

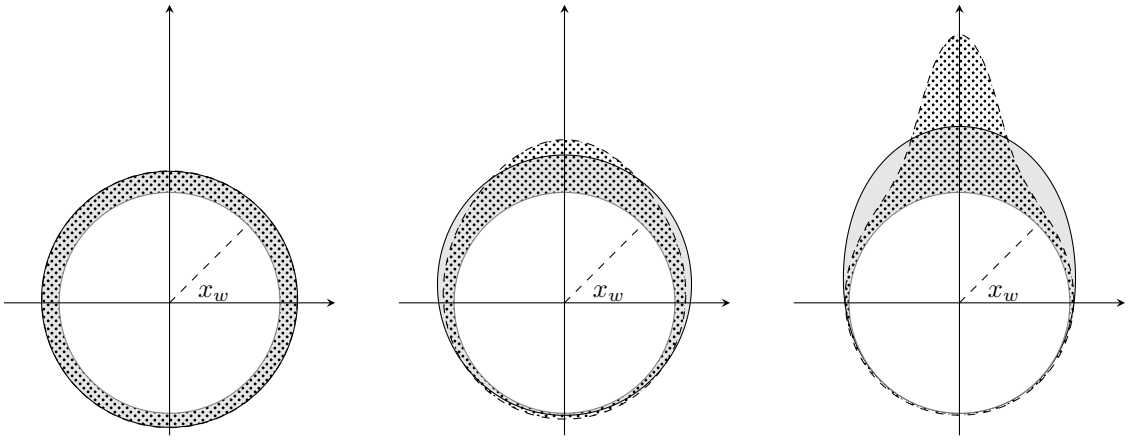
Another probability density of interest to circular regression is the *wrapped Cauchy* (WC) distribution. Neither the mean nor the variance of its linear counterpart are defined, which makes it unusable for linear regression. However, when wrapped around the circle a *close form* of this distribution can be derived with defined moments (Mardia & Jupp, 1999). In Figure 3, we compare the probability density function (pdf) of the WN (solid outlined, grey area) and the WC distribution (dashed outlined, dotted area) for increasing concentration. We observe that for very dispersed data (Figure 3a), both distributions resemble the wrapped Uniform distribution, whereas for more concentrated data (Figure 3b and Figure 3c) the WC tends to be more sharply peaked than the WN.

While in theory we should wrap from minus to plus infinity, in practice this is not feasible and it turns out to be unnecessary for both WN and WC. For WN, Kurz, Gilitschenski, and Hanebeck (2014) showed that for sufficiently small values of the standard deviation σ, without loss of accuracy, the density can be approximated by wrapping only a finite number of times.



(a) (b) (c)

Figure 2. Three consecutive steps of wrapping a $N(\pi, 2)$ distribution around the circle. On the bottom row the probability density is shown on the real line, whereas the top row shows its circular counterpart.



(a) (b) (c)

Figure 3. The pdf of the wrapped normal distribution (solid outlined, grey area) and the wrapped Cauchy distribution (dashed outlined, dotted area) for constant mean $\mu = \frac{\pi}{2}$ and increasing values of concentration.

Jona-Lasinio, Gelfand, and Jona-Lasinio (2012, p. 6) state that for values of the standard deviation $\sigma < 2\pi/3$, it is enough to wrap the density three times, which will result in a *truncated* version of WN. For WC it is not necessary to use an approximation, we can directly use its closed form .

Our method is based on the truncated version of the WN (Kurz et al., 2014) and on

the closed form of the WC distribution. Compared to prior work we extended the method to regression with a circular outcome measure and an arbitrary number of continuous and dichotomous predictors. Details of our Bayesian modeling approach are outlined in the next section.

Methods

In this section we will introduce a Bayesian circular regression technique based on WN and WC likelihoods for models with one circular outcome measure and an arbitrary number of dichotomous and continuous predictors. After defining the model, we will introduce prior distributions for the model parameters and derive the posterior distribution. Differences and similarities between the WN and WC method will be made explicit.

Model

We define a general circular regression model for a vector of circular outcomes $\mathbf{y}=(y_1, \cdots, y_n)$, with n being the sample size. We will predict the outcomes by the model matrix $\mathbf{X}=(\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_d, \mathbf{x}_{d+1}, \cdots, \mathbf{x}_p)$, with d the number of dichotomous independent variables and $p - d = c$ the number of continuous predictors. The number of rows in $\mathbf{x}$ equals the sample size. Note that categorical variables having more than two levels can be incorporated in this model as dummy variables. The circular residual error terms $\mathbf{R}$, are assumed to have a wrapped distribution (WD) with mean zero and circular concentration measure ρ:

$$\mathbf{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} \mod 2\pi \sim \text{WD}(\mu = 0,\ \rho \cdot \mathbf{I}_n), \tag{1}$$

where $\hat{\mathbf{y}}=(\hat{y}_1, \ldots, \hat{y}_n)$ represent the vector of predicted values of $\mathbf{y}$ and $\mathbf{I}_n$ denotes an n-dimensional identity matrix. The wrapped distribution WD can either stand for WN or WC.

The predicted values $\hat{\mathbf{y}}$ can be written as:

$$\hat{\mathbf{y}} = \mathbf{X} \cdot \boldsymbol{\beta} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \tag{2}$$

where $\mathbf{X}$ denotes the model matrix (of size n × p) and $\boldsymbol{\beta}$ together with ρ form the set of model parameters. If we split $\mathbf{X}$ in a dichotomous and a continuous part, we can rewrite this equation as:

$$\hat{\mathbf{y}} = \overbrace{\underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{d1} \\ 1 & x_{12} & \cdots & x_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{dn} \end{bmatrix}}_{\text{dichotomous predictors}} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix}}^{\text{circular intercept}} + \underbrace{\begin{bmatrix} x_{(d+1)1} & \cdots & x_{p1} \\ x_{(d+1)2} & \cdots & x_{p2} \\ \vdots & \ddots & \vdots \\ x_{(d+1)n} & \cdots & x_{pn} \end{bmatrix}}_{\text{continuous predictors}} \overbrace{\begin{bmatrix} \beta_{(d+1)} \\ \beta_{(d+2)} \\ \vdots \\ \beta_{(c+d=p)} \end{bmatrix}}^{\text{linear slopes}}, \tag{3}$$

where the first part can be interpreted as the circular intercept of the model, taking values between 0 and 2π.

In general, the likelihood of the data for any wrapped distribution WD can be denoted as:

$$\mathrm{L}(\mathbf{y} \mid \mathbf{X},\ \boldsymbol{\beta},\ \rho) = \prod_{i=1}^{n} \mathrm{WD}(y_i \mid \mu_i = \hat{y}_i,\ \rho), \tag{4}$$

If we use the truncated WN distribution as the wrapped pdf, we obtain the likelihood of a dataset with WN distributed error terms:

$$\mathrm{L}_{\mathrm{WN}}(\mathbf{y} \mid \mathbf{X},\ \boldsymbol{\beta},\ \rho) = \frac{1}{(\sqrt{-4\pi \log \rho})^n} \prod_{i=1}^{n} \sum_{k=-2}^{2} \exp\left(\frac{-(r_n + 2\pi k)^2}{4 \log \rho}\right), \tag{5}$$

where k is usually referred to as the *wrapping coefficient*. Note that the circular concentration parameter ρ of the WN is related to the standard deviation σ of the (linear) Normal distribution as $\sigma = \sqrt{-2 \ln(\rho)}$. If, in turn, we use the closed form of WC, we find the following likelihood:

$$\mathrm{L_{WC}}(\mathbf{y} \mid \mathbf{X},\ \boldsymbol{\beta},\ \rho) = \frac{1}{(2\pi)^n} \prod_{i=1}^{n} \frac{1-\rho^2}{1+\rho^2-2\rho\cos\left(r_n\right)}. \tag{6}$$

Prior distributions

For the model parameters we define the following set of uninformative prior distributions, similar to the priors used by Ravindran (2002). Note that we use the same priors for both WN and WC.

We use the circular uniform distribution for both the intercept β_0 and the model parameters $\beta_1, \cdots, \beta_d$:

$$\pi(\beta_0, \cdots, \beta_d) = \left(\frac{1}{2\pi}\right)^{d+1}. \tag{7}$$

For the slopes $\beta_{d+1}, \cdots, \beta_p \in \mathbb{R}$, we use:

$$\begin{aligned} \pi(\beta_{d+1}, \cdots, \beta_p) &= \mathrm{MVN}(\boldsymbol{\mu},\ \Sigma) \\ &= \exp\left(-\frac{\beta_{d+1}^2 + \cdots + \beta_p^2}{2\sigma_\beta^2}\right), \end{aligned} \tag{8}$$

which is a multivariate normal density with mean vector $\boldsymbol{\mu} = 0$ and covariance matrix Σ being equal to a c-dimensional identity matrix multiplied by the prior variance $\sigma_\beta^2 = 5^2$. A prior mean of 0 is a non-informative setting, the choice for a prior variance of 5 (also used by Ravindran (2002)) is explained later in the first simulation study. Note that we assume the prior variance to be equal for all continuous predictors in the model.

Finally, the circular concentration parameter ρ can take values between 0 and 1. We define the following prior density on this domain:

$$\pi(\rho) = \rho^{-0.5}(1-\rho)^{-0.5}. \tag{9}$$

Posterior distribution

Having expressions for both the prior densities and the density of the data, we find the joint distribution being proportional to the multiplication of Equations 4, 7, 8 and 9:

$$\mathrm{P}(\boldsymbol{\beta},\ \rho \mid \mathbf{y}, \mathbf{X}) = \mathrm{L}(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \rho) \cdot \pi(\beta_0, \cdots, \beta_d) \cdot \pi(\beta_{d+1}, \cdots \beta_p) \cdot \pi(\rho), \tag{10}$$

where the likelihood L can either be the WN (Equation 5) or the WC likelihood (Equation 6). The joint distribution cannot be written in terms of a standard distribution, therefore it is not possible to directly sample from it. This problem can be overcome by using slice sampling techniques (Neal, 2003). A conceptual introduction to slice sampling is available in Appendix A1. We will explicitly show how slice sampling can be used to sample the joint distribution for both WN and WC in Appendix A2.

Simulation studies

To validate our methods, we perform a number of simulation studies. The results of these studies are presented in this section. The main studies cover an analysis of covariance (ANCOVA) and a multiple regression (MR) setting. Some smaller studies are done to investigate how the sampler reacts to violations of assumptions. Note that all simulations, unless stated otherwise, are performed with a fixed number of samples of 500 and a number of observations (N) of 50. We use two chains with different initial values. The total number of iterations for each chain is 7000, of which 3000 are regarded as burn-in. For the slice sampler, a step-size $w = 1$ is used. Simulations with the WN sampler are performed with a wrapping coefficient k=1. Convergence and autocorrelations are checked by means of trace and autocorrelation plots. Unless stated otherwise, we did not find convergence or autocorrelation problems.

ANCOVA setting

The first simulation concerns a regression model with one circular outcome (y), one dichotomous (x_1) and one linear (x_2) predictor:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon. \tag{11}$$

Note that model parameters β_0 and β_1 are circular, while β_2 is linear. The dichotomous variable x_1 is sampled from a Binomial distribution with success probability $p = 0.5$. The linear variable x_2 is sampled from a standard Normal distribution. The error ϵ is sampled from either a WN or WC distribution with mean $\mu = 0$ and circular concentration ρ.

For the model parameters, we use all combinations of the following values:

- $\beta_0 = 0, 3, 6$

- $\beta_1 = 0, 3, 6$
- $\beta_2 = -5, -1, 0, 1, 5$
- $\rho = 0.5, 0.9$

The intercept parameters can take values around the whole circle. The range of values for the slopes are restricted. Due to the nature of wrapping, steep slopes will result in data following a divergent pattern. For slopes steeper than (minus) five, the resulting data is unsuitable for regression. Therefore, we only look at slope values between minus and plus five. The same argument holds for the values of ρ. Values of the circular concentration lower than one half result in (approximately) uniformly distributed residuals. Therefore, only the cases $\rho = 0.5$ and $\rho = 0.9$ will be considered.

Using the values above for the model parameters, for each observation the circular outcome y is randomly sampled according to Equation 11. The resulting datasets with WN distributed residuals are analysed with the WN sampler. Datasets with WC distributed residuals are analysed with the WC sampler. In Table 1, we summarize the results of the ten designs with $\beta_0 = \beta_1 = 3$ for both the WN and the WC sampler. Note that results from the other designs are rather similar.

We first discuss the right-most column of Table 1, which concerns a convergence score. The first 100 samples of each design were visually checked for convergence by means of trace plots. We counted the number of samples that did converge, resulting in a convergence score between 0 and 100. For designs where $\rho = 0.9$ all of the first 100 samples converge. For the designs where $\rho = 0.5$, however, we observe a somewhat different pattern. Most of the samples within each design still converge, but in some of the samples the sampler starts to make *jumps*. This is due to the fact that the posterior distribution, besides the absolute maximum has multiple local maxima. These local maxima tend to present a higher portion of the density when the value of ρ decreases. That is, for lower values of ρ the sampler will be more likely to jump to another (local) maximum.

The meaning of these local maxima can be understood by looking at Figure 4. On the left we see a dataset with a circular outcome y and a linear predictor x, together with a (perfectly) fitting regression line. Note that this plot in fact depicts the *net* of a cylinder cut along the dashed line. If we look at Figure 4b, however, we see the same data together with a regression

line that fits equally well. Put differently, due to circularity, more than one model might serve as a good fit to the data. This problem was observed, but not fully investigated in previous studies (Coles, 1998; Ferrari, 2009). A proposed workaround is the use of a relatively low value of the prior variance in the prior density of the slopes. This way, the sampler is *guided* towards the more parsimonious model.

From Table 1 we observe that despite the relatively low prior variance, for $\rho = 0.5$ the sampler cannot always be prevented from jumping to other local maxima, resulting in a low convergence score. The process of jumping is illustrated in Figure 5, where we plot the trace of the two chains (in red and blue) of β_2 for the design with true value $\beta_2 = 1$, analysed with the WC sampler for three different samples. The sample in Figure 5a converges well within the burn-in period (indicated in solid grey). Regarding the samples where the sampler does jump, we distinguish between two cases. Figure 5b shows a sample with *minor* jumps. That is, after a couple of iterations the sampler jumps back to the absolute maximum (matching the true parameter value). We do not expect those jumps to influence the summary statistics. In Figure 5c, on the other hand, we see a sample with *severe* jumps. In this case it is likely that the summary statistics are affected. In this specific design, a total of 41 of the first 100 samples show severe jumps (resulting in a convergence score of 59). From Table 1, it seems that the WN sampler is less sensitive for having severe jumps.

A valid question is whether the problems with convergence we encounter in this study will pose a problem for researchers who want to apply our methods to empirical data. We argue that this does not necessarily need to be the case. First of all, as shown in Figure 5, it is clear from the trace plot when severe jumps are present. In other words, if a trace plot is consulted in

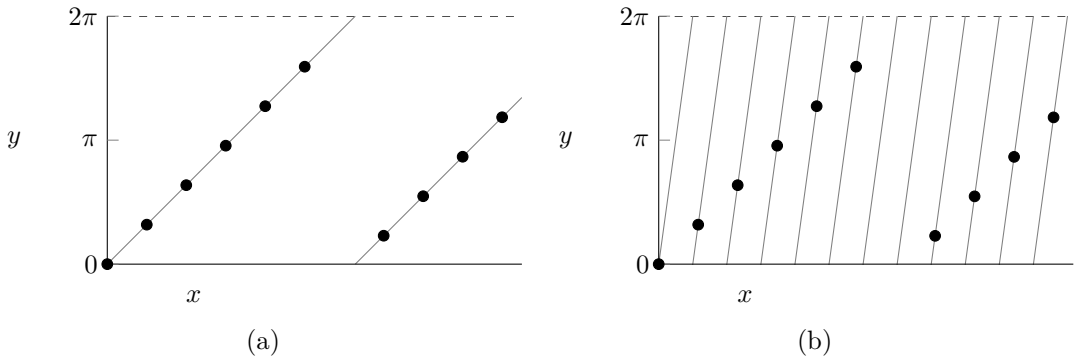


Figure 4. A dataset with a circular outcome y and a continuous predictor x together with two equally fitting regression lines.

addition to the summary statistics, the risk of making inferences based on a badly converged sample is minimal. Moreover, because we are dealing with a random process, it can simply help to rerun the analysis, optionally with different initial values.

In the remaining columns of Table 1, we report the summary statistics based on all samples. The reported absolute values of the biases in Table 1 of the mean (for β_0, β_1 and β_2) or the mode (for ρ, calculated as in Bickel and Frühwirth (2006)) are relative (in %), except for the case where β_2 equals zero. In that case, the absolute bias is reported. In terms of bias, we do not observe a clear difference in performance between the WN and WC samplers for $\rho = 0.9$. At lower circular concentration ($\rho = 0.5$), the WN sampler seems to perform better in some of the designs. Note, however, that those cases of better performance correspond to a higher convergence score. In other words, the results of the WC sampler at $\rho = 0.5$ are disturbed by severe jumps of the sampler.

The interval widths (IW) reported are the widths of the 95% credible intervals (CI) for the intercepts and slope parameters (e.g. β_0, β_1 and β_2), and the Highest Posterior Density (HPD) interval for the circular concentration measure (ρ). The CI of circular parameters are calculated with respect to the point on the circle opposite to the posterior mean. The interval widths of ρ can be regarded equal for both sampling methods under all designs. For the designs under $\rho = 0.9$ the interval widths of the other parameters seem to be more narrow for the WC sampling method. The opposite seems to be the case at $\rho = 0.5$, however, those values again are hard to compare because of the difference in convergence between the two samplers. The coverage probabilities (CP) of all designs range between .92 and .98, which is good.

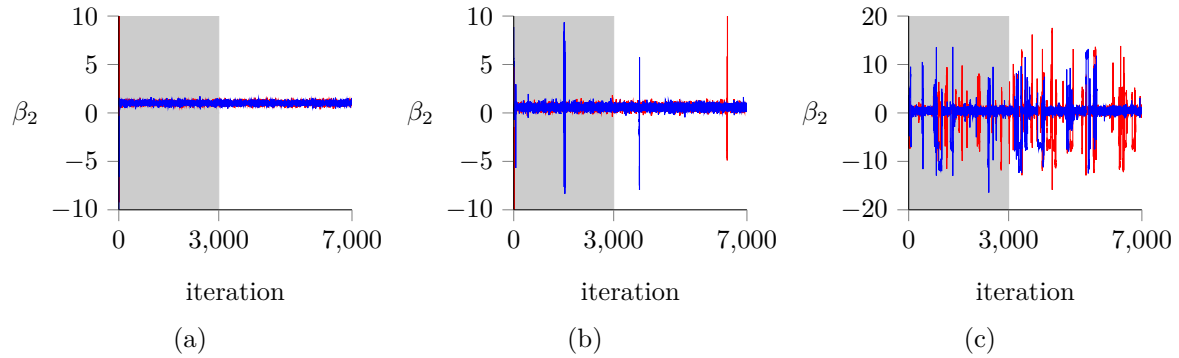


Figure 5. Trace plot of a sample with no jumps (left), some jumps (middle) and many jumps (right).

Multiple regression setting

The next part of our simulation studies concerns a linear regression setting in which a circular outcome (y) is predicted by two linear variables (x_1 and x_2), sampled from a standard bivariate normal distribution. Note that for now we assume the correlation between the two predictors to be zero. This results in a model equal to Equation 11, with the one difference that model parameters β_1 and β_2 now are both linear.

For the model parameters, we use all combinations of the following values:

- $\beta_0 = 3$
- β_1 = -5, -1, 0, 1, 5
- β_2 = -5, -1, 0, 1, 5
- ρ = 0.5, 0.9

Using those values, the circular outcomes y are randomly sampled assuming WN/WC distributed error terms.

In Table 2, we summarize the results of the ten designs with $\beta_0 = 3$, $\beta_1 = 1$ for both the WN and the WC sampler in the same way as with the ANCOVA setting. Again, in terms of bias, the only evident differences between the two sampling methods correspond to cases in which relatively many severe jumps occur (as can be observed from the convergence score). With regard to interval widths we observe a pattern similar to the ANCOVA setting: for $\rho = 0.9$ the WC sampler seems to perform better for all model parameters but ρ (for which the performance is equal with both sampling methods), while for $\rho = 0.5$ the results are hard to compare due to convergence problems. Results regarding coverage and convergence are in line with the ANCOVA setting.

Sensitivity of the sampler

The remainder of our simulations concerns a few smaller studies set up to investigate how sensitive our sampling methods are to input data deviating from the ideal situation. In particular, we will look at sensitivity to sample size, to wrong assumptions about the distribution of the error terms, to multicollinearity for models with two continuous predictors and to unequal group sizes for models with a dichotomous predictor. We will use relative

bias as a measure of *accuracy* of the parameter estimates. The *precision* is reflected in the posterior standard deviation (sd) of the parameter estimates (comparable to a standard error in a frequentist setting), where an increase in sd should be regarded as a *de*crease in precision.

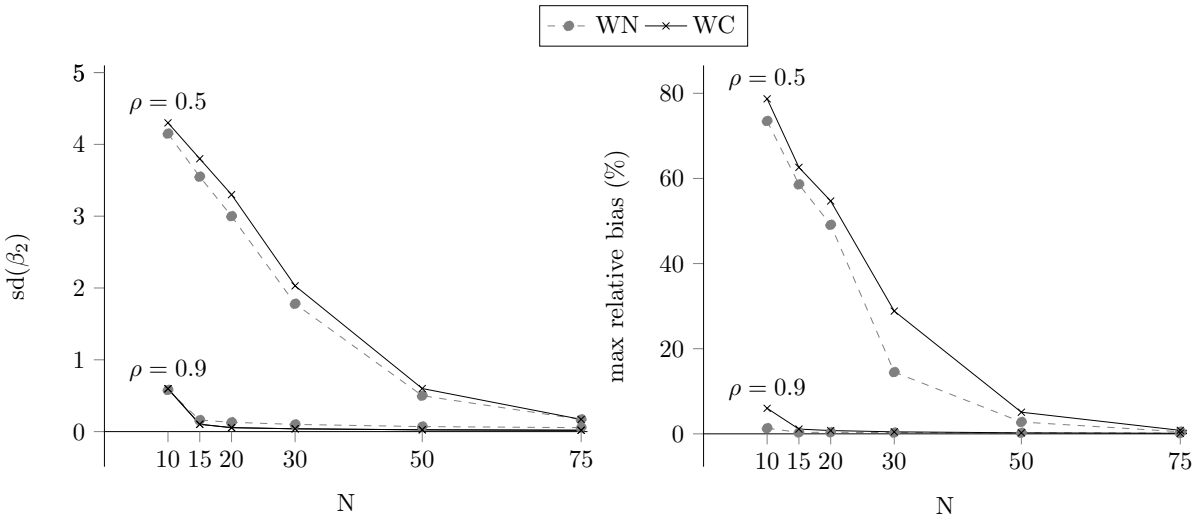
Sensitivity to sample size. We investigate sensitivity to sample size (N) for both the ANCOVA and MR setting. For the first type of model we use $\beta_0 = \beta_1 = 3$ and $\beta_2 = 1$, for the model with two continuous predictors we use $\beta_0 = 3$ and $\beta_1 = \beta_2 = 1$. Both the WN and WC models are analysed for $\rho = 0.5, 0.9$ with the corresponding sampling method. Finally, the sample size takes values of 10, 15, 20, 30, 50 and 75.

We observe an influence of sample size on both the precision and the accuracy of the parameter estimates. In Figure 6 the results are depicted for ANCOVA (top row) and MR models (bottom row) based on WN (dots) and WC (crosses) distributed errors. The two pairs of lines in each graph represent the two different values of ρ. In the left column the standard deviation of β_2[1] is plotted against the sample size. In the right column of Figure 6 we show the maximum observed relative bias (in %) of any of the model parameters (including ρ) against N. For $\rho = 0.5$, for all combinations of ANCOVA/MR and WN/WC, we observe a decrease in both the standard deviation and the relative bias with an increase in sample size. Standard deviations and biases are consequently higher than in the similar designs under $\rho = 0.9$. Note that for this higher circular concentration, both the standard deviation and the relative bias already tend to be constant for sample sizes of 15 and higher. Here, we have to remark that the results of $\rho = 0.5$ can be influenced by the severe jumps as described above.

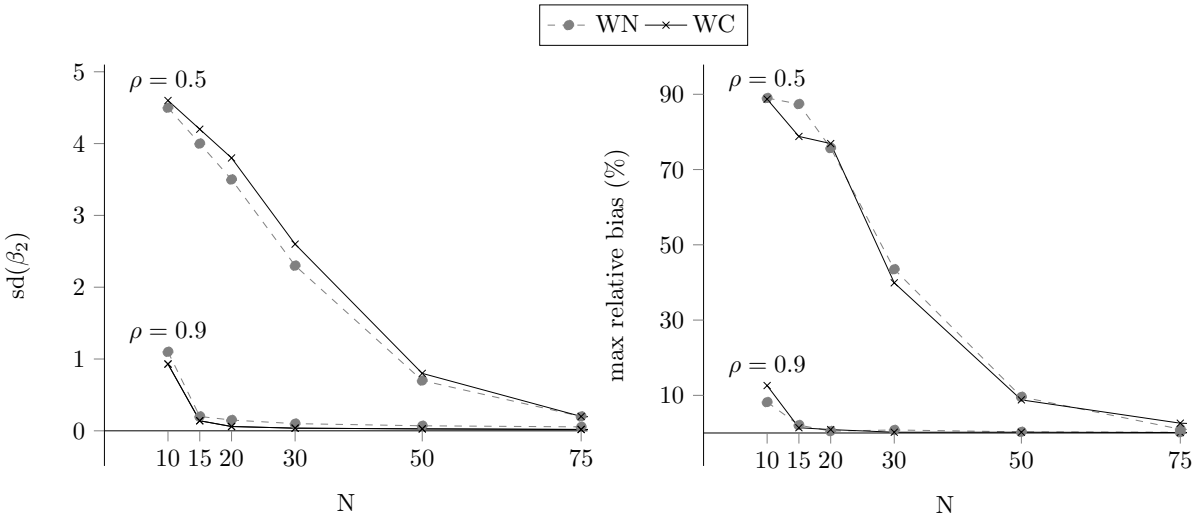
Based on this study, we conclude that the sensitivity to sample size is highly dependent on the value of ρ. For a large circular concentration, both sampling methods already worked well (in terms of precision and accuracy) for both kind of models for sample sizes of 15 and higher. For the lower value of ρ, it took a larger sample size to obtain reliable results. It is hard to draw a thorough conclusion as the results, to some extent, might be influenced by poor convergence, but we think it is safe to say that we would recommend a minimum sample size of 30 under this condition. This recommendation is based on a model with two predictors. We expect the minimal sample size to be higher for more extensive models, but will not investigate this here. In this study we also do not address the possible relation between convergence and sample size.

[1]Note that the pattern for the other model parameters was similar.

Sensitivity to wrong assumption about the distribution of the residuals. One of the most important assumptions of circular regression concerns the distribution of the error terms. Our methods assume either WN or WC distributed errors. To test the sensitivity of both samplers to a *wrong* assumption about the distribution of the residuals, we choose one of the ANCOVA ($\beta_0 = \beta_1 = 3$, $\beta_2 = 1$ and $\rho = 0.9$) and one of the MR settings ($\beta_0 = 3$, $\beta_1 = \beta_2 = 1$, $\rho = 0.9$) with either WN or WC distributed error terms, thus resulting in four datasets. We analyse these data with the wrong sampling method. Note that we only investigated the effect for $\rho = 0.9$. At this value of the circular concentration, we expect the effect to be larger than for



(a) Sensitivity to sample size in the ANCOVA setting.



(b) Sensitivity to sample size in the linear predictor setting.

Figure 6. Sensitivity to sample size as reflected in the standard deviation and the maximal relative bias.

lower values of ρ since WN and WC deviate the most (see Figure 3c).

The results of this simulation study can be found in Table 3. We observe no or little effect on the absolute value of the relative biases of the parameter estimates, except for ρ. The effect on accuracy of ρ is largest for the WN datasets analysed with the WC sampler, where we observe an underestimation in ρ of about 15 percent. Looking at Figure 3 this is as expected because the WN distribution for $\rho = 0.9$ looks like a WC distribution with lower concentration. Vice versa, that is WC data analysed with the WN method, the effect seems less severe (about 3%). Compared to the previous studies, the IW for the data with WN distributed residuals remain the same, while for the data with WC distributed residuals interval widths are wider. The coverage probabilities of the model parameters except ρ are lower than in the previous studies, but still acceptable. The CP of ρ are very low for all the datasets. All of the first 100 samples converged for all datasets.

In general, it can be concluded that a violation of the assumption about the distribution of the residuals mainly affects the parameter estimate of ρ. The summary statistics of the other parameters can be less precise in some cases, but this precision is still acceptable.

Sensitivity to multicollinearity. To investigate the sensitivity to correlated (continuous) predictors we fit models (WN and WC) with two predictors with β_0=0, β_1=β_2=1 and ρ=0.9. A similar study was performed under $\rho = 0.5$, but results were affected by poor convergence. The simulated correlation between the predictors takes values of -0.9, -0.5, -0.1, 0, 0.1, 0.5 and 0.9. Note that the cases with zero correlation correspond to the previously analysed designs.

We observe that parameter estimates remain stable under the variation of correlation. However, multicollinearity has an effect on the precision of the estimates. The precision, seems to remain constant up to correlations of 0.5, but rapidly decreases for more strongly correlated predictors. In Figure 7, we plotted the standard deviation of β_1 against the correlation between x_1 and x_2 for the WC (crosses) and the WN (dots) models. Note that the standard deviation of β_2 took similar values. For both WN and WC, the standard deviation for highly correlated predictors is about four times as large as the standard deviation for uncorrelated predictors.

Sensitivity to unequal group size. To investigate the sensitivity to unequal group size, we fit a model (WN and WC) with one dichotomous predictor for $\beta_0 = 0$, $\beta_1 = 1$ and

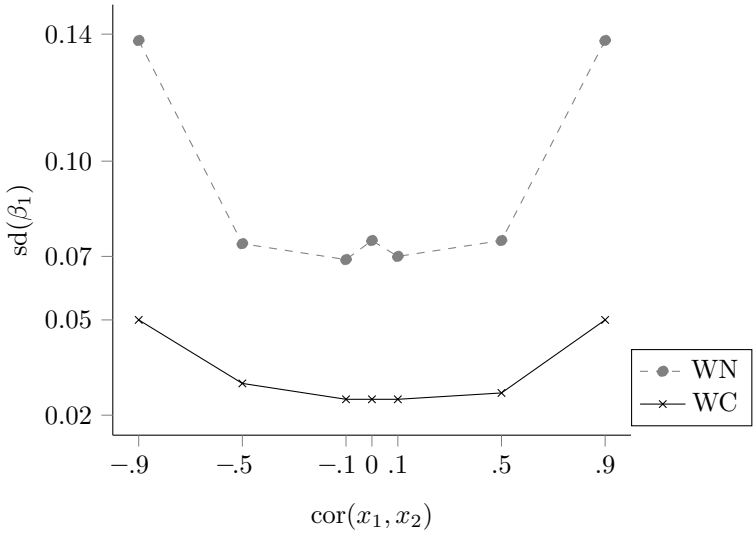


Figure 7. The effect of (multi)collinearity on the precision of the parameter estimates of the WN (blue) and WC (red) sampler for $\rho = 0.9$.

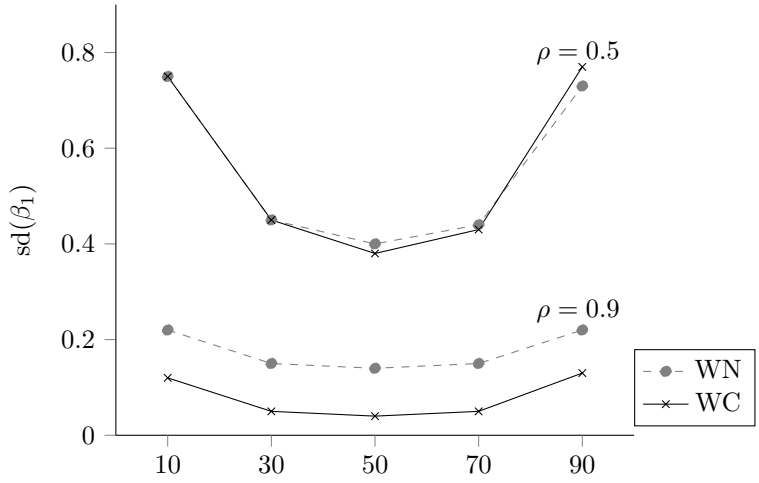


Figure 8. The effect of unbalancedness (on the x-axis) of a dichotomous variable on the standard deviation of its parameter estimate for the two sampling methods at three different values of ρ.

$\rho = 0.5, 0.9$. The group size of the experimental condition, indicated by the number of "ones" in the total of 50 observations, takes values of 5, 15, 25, 35 and 45.

Results are depicted in Figure 8, where dots indicate WN and crosses WC. The two pairs of lines are the results for the two different values of ρ. We observe an effect of group size on the precision of the parameter estimates. The most precise estimates seem to be obtained when group sizes are equal. Note that we did not find an effect of group size on the accuracy of the sampler.

Application to an empirical dataset

In this section we demonstrate how our methods can be used in practice. To this end we apply our regression techniques to an empirical dataset from Warren, Rothman, Schnapp, and Ericson (2017). The data concerns a subset of data from Warren et al. (2017) study on the geometry of humans' knowledge of navigation space. We will call this subset the cognitive maps data. The complete dataset can be found as the `Maps` data in the R-package `bpnreg` (Cremers, 2018) and has been analysed before in Cremers and Klugkist (2018) and Warren et al. (2017).

Description of the data

The cognitive maps data concerns an experiment in which a total of twenty participants used virtual reality headsets to navigate through a virtual maze. The twenty participants were split up in two groups that each had to navigate through a different type of maze (between-subjects factor). One version of the maze is a normal maze that the researchers refer to as the Euclidean maze. The other version of the maze, the non-Euclidean maze, has exactly the same layout as the Euclidean maze but it contains wormholes by which participants can be 'teleported' from one point in the maze to another. The experiment was made up of a training phase, in which the participants learned to navigate between different pairs of start and target objects in the maze, and a test phase. The number of trials each participant completed in the training phase was recorded. In the test phase of the experiment participants first walked to a start object. When they had reached the location of the start object, the maze disappeared and only a'textured groundplan' 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 it. The angular difference between the initial walking direction of a participant and the location of the target object, the angular error, was recorded as the outcome variable in the experiment.

In the test phase of the experiment all participants had to complete eight trials. In each of these trials participants had to walk to a specific target object. A pair of start and target objects could be one of two types: probe or standard (within-subjects factor). The probe objects were located near the entrance and exit of a wormhole 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 eight trials per participant. Table 4 shows descriptives for the cognitive maps data.

Note that the cognitive maps data is in fact repeated measures data and should not be analysed using a standard regression model. For the purpose of this paper avoid a violation of the independence of errors assumption and fit two regression models, one for each of the two levels of the within-person factor type of trial, probe and standard, to the cognitive maps data. In addition we aggregate over the four trials per participant for each trial type.

Analysis

For the cognitive maps data we are interested in 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 types of target object, probe and standard. By design the expected angular error was larger if a participant in the non-Euclidean maze used the wormhole to walk to the target object. We can thus use the angular error to differentiate between participants that used the wormhole and those that took another path to the target object. Additionally we would like to control for the amount of trials that a participant completed in the training phase.

Because the experiment contains a within-subjects factor we fit two separate regression models to the data, one for the probe trials and one for the standard trials. Each of the two regression models is fit using the WC and WN model resulting in a total of four regressions. The MCMC samplers for all four models converged within 7000 iterations (burn-in = 3000).

The results from these models are shown in Tables 5 and 6. In both the WN and WC regression for the probe trials there is an effect of the type of Maze (the CI does not overlap with that of the intercept). This means that the participants in the non-Euclidean maze had larger angular errors than those in the Euclidean maze and thus make use of the wormholes. For the standard trials there is no effect of the type of Maze for both the WN and WC regression. This is as expected since the objects in the standard trials are not located close to the wormholes and thus the participants in the non-Euclidean maze do not use them. The amount of training trials does not seem to affect the angular error. The 95% credible interval for the estimated regression coefficient contains zero in both regression models and for both the WN and WC sampler. This is a good sign since it means that the training phase of the experiment had the desired effect of

getting all participants at the right level.[2]

Discussion

In this paper we proposed a novel and promising alternative for circular regression using the wrapping approach. We developed models based on both the WN and WC distribution. Instead of using the data augmentation approach as adopted by Coles (1998) and Ravindran (2002), we based our techniques on the truncated version of the WN distribution as proposed by Kurz et al. (2014) and the closed form of the WC distribution (Mardia & Jupp, 1999). Where previous studies regarding circular regression based on the wrapping approach only focussed on mean estimation or models with one predictor (Ravindran, 2002; Ferrari, 2009), we increased the scope by making our methods applicable to models with an arbitrary number of continuous and dichotomous predictors.

Previous methods were only tested in rather limited simulation settings only covering concentrated data (see f.e.: Ravindran (2002)). Under these *ideal* conditions our methods turn out to work equally well. However, in our simulation studies we also addressed less ideal and probably more realistic cases, thus investigating the practical boundaries of our methods in particular and the wrapping approach in general.

We showed that when continuous predictors are present, due to the nature of wrapping, it can be that multiple models fit the data equally well (see Figure 4). This results in a posterior distribution with multiple local maxima. In our simulations we guided our sampler in the right direction by choosing a relatively small value of the prior variance for slopes as was previously proposed by Coles (1998), Ferrari (2009) and Ravindran (2002). However, for more dispersed data (e.g., smaller values of ρ), this turned out not to be enough to prevent all convergence problems. Especially the WC sampler seems to be sensitive for "jumping" to other local maxima. With respect to the convergence problems, we argue that it would be better to develop a sophisticated slice sampling scheme in which the absolute maximum of the posterior distribution is favoured over its local maxima. That is, the sampler should have a bias towards the most parsimonious model. We are confident that this should solve problems with convergence.

[2]We should not make a too strong conclusion about this effect since we only have 20 participants, meaning that the power of the test may be too small to detect a difference from 0. However, ρ is estimated to be quite large in all four regressions meaning that the small sample size is not expected to bias the results to a large extent and the standard deviation is expected to be small according to our simulation study (see Figure 6).

We investigated some of the problems a researcher can experience when using the method in practice. That is, multicollinearity for models with multiple continuous predictors or unequal group sizes in models containing one or more dichotomous predictors. For both issues we saw that our methods still gave reliable parameter estimates. However, when the deviation in the data from the ideal situation became severe, a decrease in precision was observed. Note that this is in line with expectation; in linear regression a similar pattern can be observed. We concluded that the minimal sample size needed to obtain viable results depends on the circular concentration of the residuals. In concentrated data a sample size of 15 already could be sufficient. However, in general, we would recommend a minimum of 30. Note that these findings are based on models with two predictors. More severe problems may arise in more extensive models. These should be included in future simulation studies. Although we made some efforts in investigating the influence of wrongly distributed residuals on the sampling results, this could be further investigated in future research.

Both methods were applied to the cognitive maps data. We found that both samplers performed well in terms of convergence and that the resulting parameter estimates were easy to interpret. Additionally, even though we ignored the repeated measures design of the data the results correspond to results in previous analyses of the data in Warren et al. (2017) and Cremers and Klugkist (2018). This bring us to another interesting opportunity for future research, which would be to make our methods applicable to more complex regression models, such as repeated measured data. In previous studies Bayesian analysis of complex regression models, such as multi-level regression models, was based on the embedding approach (see f.e.: Wang and Gelfand (2013); Nuñez-Antonio and Gutiérrez-Peña (2014)). To our knowledge, however, multi-level circular regression based on the wrapping approach is still non-existent.

In summary, this paper has introduced an adjusted and more extensive method for Bayesian regression using the wrapping approach. We provided an extensive simulation study showing that results are reliable as long as the trace plots show convergence. We showed that our methods are flexible and can be applied to a wide range of data resulting in parameter estimates that are easy to interpret.

Acknowledgements

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

References

Abuzaid, A., & Allahham, N. (2015). Simple circular regression model assuming wrapped cauchy error. *Pakistan Journal of Statistics*, *31*(4), 385–398.

Bhattacharya, S., & SenGupta, A. (2009). Bayesian inference for circular distributions with unknown normalising constants. *Journal of statistical planning and inference*, *139*(12), 4179–4192. doi: 10.1016/j.jspi.2009.06.008

Bickel, D. R., & Frühwirth, R. (2006). On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications. *Computational Statistics & Data Analysis*, *50*(12), 3500–3530. doi: 10.1016/j.csda.2005.07.011

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), e0135803. doi: 10.1371/journal.pone.0135803

Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, *8*(2), 105–113. doi: 10.1023/A:1008930032595

Cremers, J. (2018). bpnreg: Bayesian projected normal regression models for circular data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=bpnreg` (R package version 1.0.0)

Cremers, J., & Klugkist, I. (2018). One direction? a tutorial for circular data using r with examples in cognitive psychology. *Frontiers in Psychology*, *9*, 2040. doi: 10.3389/fpsyg.2018.02040

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

Damien, P., & Walker, S. (1999). A full Bayesian analysis of circular data using the von mises distribution. *Canadian Journal of Statistics*, *27*(2), 291–298.

Ferrari, C. (2009). *The Wrapping Approach for Circular Data Bayesian Modeling* (Unpublished doctoral dissertation). University of Bologna.

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Forbes, P. G., & Mardia, K. V. (2015). A fast algorithm for sampling from the posterior of a von

mises distribution. *Journal of Statistical Computation and Simulation*, *85*(13), 2693–2701. doi: 10.1080/00949655.2014.928711

Herrera, P., Resa, Z., & Sordo, M. (2010). Rocking around the clock eight days a week: an exploration of temporal patterns of music listening. In *1st workshop on music recommendation and discovery (womrad), acm recsys, Barcelona, Spain.*

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

Jona-Lasinio, G., Gelfand, A., & Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics*, *6*(4), 1478–1498. doi: 10.1214/12-AOAS576

Kimpton, A., Corcoran, J., & Wickes, R. (2017). Greenspace and crime: An analysis of greenspace types, neighboring composition, and the temporal dimensions of crime. *Journal of Research in Crime and Delinquency*, *54*(3), 303–337. doi: 10.1177/0022427816666309

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

Kurz, G., Gilitschenski, I., & Hanebeck, U. D. (2014). Efficient evaluation of the probability density function of a wrapped normal distribution. In *Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2014* (pp. 1–5). IEEE. doi: 10.1109/SDF.2014.6954713

Mardia, K. V., & Jupp, P. E. (1999). *Directional Statistics.* New York: Wiley.

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.

Nuñez-Antonio, G., & Gutiérrez-Peña, E. (2005a). 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., & Gutiérrez-Peña, E. (2005b). A Bayesian analysis of directional data using the von mises–fisher distribution. *Communications in Statisticsal Simulation and Computation*, *34*(4), 989–999. doi: 10.1080/03610910500308495

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

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

Ravindran, P. (2002). *Bayesian Analysis of Circular Data Using Wrapped Distributions* (Unpublished doctoral dissertation). North Carolina State University.

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

Rentsch, S., & Rand, M. K. (2014, 10). Eye-hand coordination during visuomotor adaptation with different rotation angles. *PLOS ONE*, *9*(10), 1–15. doi: 10.1371/journal.pone.0109819

Von Mises, R. (1918). Uber die ganzzahligkeit der atomgewichte und verwandte fragen. *Physikalische Zeitschrift*, *19*, 490–500.

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

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

Table 1

Results of the simulation with one dichotomous and one continuous predictor for $\beta_0 = \beta_1 = 3.0$

True value of:			Abs(Bias)				IW				CP				
β_2	ρ	Sampler	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	Convergence
-5	0.5	WN	0.15	0.05	5.13	0.38	1.41	1.79	2.46	0.31	.96	.96	.95	.94	86
		WC	0.02	0.58	4.77	0.70	1.28	1.62	2.48	0.33	.97	.96	.95	.94	82
	0.9	WN	0.16	0.15	0.11	0.04	0.41	0.54	0.28	0.08	.95	.96	.95	.95	100
		WC	0.08	0.08	0.03	0.04	0.15	0.19	0.10	0.08	.95	.95	.95	.95	100
-1	0.5	WN	0.05	0.17	2.32	0.79	1.32	1.70	1.62	0.30	.96	.96	.96	.96	90
		WC	0.03	0.13	0.46	0.56	1.28	1.62	2.07	0.32	.95	.94	.94	.98	83
	0.9	WN	0.00	0.08	0.77	0.07	0.41	0.54	0.27	0.08	.96	.95	.94	.94	100
		WC	0.00	0.10	0.06	0.42	0.15	0.20	0.11	0.08	.94	.95	.95	.95	100
0	0.5	WN	0.71	0.36	0.00	1.28	1.40	1.78	1.88	0.31	.93	.94	.94	.94	87
		WC	0.17	0.39	0.00	3.05	1.37	1.72	2.39	0.33	.93	.94	.94	.94	76
	0.9	WN	0.20	0.22	0.00	0.00	0.41	0.54	0.28	0.08	.93	.95	.94	.94	100
		WC	0.01	0.01	0.00	0.51	0.15	0.20	0.10	0.08	.95	.94	.95	.94	100
1	0.5	WN	0.22	0.32	2.60	0.41	1.33	1.73	1.64	0.30	.94	.94	.94	.95	91
		WC	0.14	0.25	6.26	3.2	1.38	1.74	2.48	0.33	.94	.94	.94	.98	59
	0.9	WN	.005	0.24	0.16	0.02	0.41	0.54	0.27	0.08	.95	.95	.93	.93	100
		WC	0.01	0.05	0.07	0.25	0.15	0.19	0.10	0.08	.95	.95	.95	.96	100
5	0.5	WN	0.02	0.33	3.27	1.52	1.34	1.74	2.09	0.30	.95	.95	.96	.94	88
		WC	0.04	0.08	6.11	2.09	1.36	1.70	2.84	0.33	.95	.96	.96	.92	79
	0.9	WN	0.13	0.27	0.05	0.01	0.41	0.54	0.27	0.08	.96	.95	.97	.96	100
		WC	0.12	0.11	0.11	0.42	0.15	0.19	0.10	0.08	.95	.96	.98	.94	100

Note. The bias shown is the absolute value of the relative bias (%), for $\beta_2 = 0$ (grey background) the absolute value of the absolute bias is given.
Interval width (IW) concerns the 95% CI for β_0, β_1 and β_2, and HPD interval for ρ.
Convergence is the number of the first 100 samples that converged within the burn-in period of 3000.
CP is the coverage probability.

Table 2

Results for the simulation with two continuous predictors for $\beta_0 = 3.0, \beta_1 = 1.0$

True value of:			Abs(Bias)				IW				CP				
β_2	ρ	Sampler	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	Convergence
-5	0.5	WN	0.71	8.73	7.44	0.94	1.04	2.72	2.89	0.31	.93	.95	.94	.95	90
		WC	0.02	13.83	14.71	4.68	1.19	4.33	4.73	0.35	.95	.95	.96	.87	69
	0.9	WN	0.09	0.02	0.01	0.34	0.26	0.28	0.28	0.08	.95	.95	.96	.94	100
		WC	0.02	0.10	0.02	0.11	0.09	0.10	0.10	0.08	.94	.96	.96	.95	100
-1	0.5	WN	0.15	6.07	8.12	0.60	1.04	2.51	2.52	0.31	.95	.97	.96	.93	85
		WC	0.02	9.39	13.23	1.40	0.98	3.03	3.13	0.34	.95	.95	.96	.91	78
	0.9	WN	0.11	0.37	0.77	0.01	0.27	0.28	0.27	0.08	.97	.95	.94	.95	100
		WC	0.02	0.16	0.14	0.30	0.09	0.10	0.10	0.08	.97	.95	.96	.96	100
0	0.5	WN	0.65	7.56	0.01	0.19	1.10	2.77	2.94	.31	.97	.96	.96	.93	82
		WC	0.15	14.67	0.00	0.91	1.00	3.09	3.12	0.34	.93	.94	.96	.91	83
	0.9	WN	0.16	0.49	0.00	0.10	0.27	0.27	0.27	0.08	.96	.96	.95	.95	100
		WC	0.05	0.52	0.01	0.51	0.10	0.10	0.10	0.08	.92	.95	.94	.95	100
1	0.5	WN	0.13	8.29	3.83	0.61	1.03	2.60	2.66	0.31	.97	.97	.96	.93	89
		WC	0.13	9.89	10.12	2.83	1.07	3.40	3.47	0.34	.96	.96	.96	.90	73
	0.9	WN	0.13	0.37	0.17	0.05	0.27	0.27	0.27	0.08	.95	.95	.95	.96	100
		WC	0.00	0.13	0.18	0.18	0.09	0.10	0.10	0.08	.94	.95	.95	.96	100
5	0.5	WN	0.16	1.33	1.21	1.91	1.24	3.70	3.99	0.32	.96	.97	.97	.92	86
		WC	0.08	4.76	12.46	3.01	1.10	3.79	4.12	0.34	.95	.95	.96	.89	74
	0.9	WN	0.09	0.05	0.02	0.14	0.26	0.27	0.28	0.08	.95	.95	.93	.94	100
		WC	0.09	0.04	0.01	0.37	0.10	0.11	0.11	0.08	.96	.96	.95	.96	100

Note. The bias shown is the absolute value of the relative bias (%), for $\beta_2 = 0$ (grey background) the absolute value of the absolute bias is given.
Interval width (IW) concerns the 95% CI for β_0, β_1 and β_2, and HPD interval for ρ.
Convergence is the number of the first 100 samples that converged within the burn-in period of 3000.
CP is the coverage probability.

Table 3

Regression results for the two types of models and the two types of residuals under the wrong assumption about the distribution of the error terms.

		Abs(Bias)			Bias	IW				CP				
Model	Residuals	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	β_0	β_1	β_2	ρ	Convergence
ANCOVA	WN	0.03	0.22	0.07	-15.41	0.41	0.54	0.27	0.16	.90	.90	.89	.00	100
	WC	0.08	0.17	0.11	- 3.57	0.48	0.64	0.32	0.10	.97	.98	.96	.47	100
MR	WN	0.02	0.05	0.20	-15.23	0.26	0.27	0.27	0.16	.89	.90	.92	.00	100
	WC	0.03	0.13	0.23	- 3.70	0.31	0.32	0.32	0.11	.97	.95	.96	.45	100

Note. The bias shown is the absolute value of the relative bias (%), except for ρ where the relative bias is reported.
Interval width (IW) concerns the 95% CI for β_0, β_1 and β_2, and HPD interval for ρ.
Convergence is the number of the first 100 samples that converged within the burn-in period of 3000.
CP is the coverage probability.

Table 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	Type of trial	$\bar{\theta}$	$\bar{R}$
Angular error	Euclidean	standard	-4.91°	0.89
		probe	4.46°	0.92
	non-Euclidean	standard	-17.59°	0.78
		probe	37.34°	0.93

Table 5

Summary statistics of the regression parameters in the model for the probe trials.

		Parameter estimate			Quantiles	
Method		Mean	Mode	SD	LB	UB
WN	**Intercept**	3.91°		0.17	344.68°	23.65°
	Maze	32.79°		0.10	21.58°	44.57°
	Training	0.04		0.01	-1.60	1.66
	ρ		.98		.97	.99
WC	**Intercept**	5.32°		0.20	341.93°	25.71°
	Maze	31.71°		0.10	20.83°	43.21°
	Training	0.06		0.01	-1.50	1.66
	ρ		.90		.82	.94

Note. For dichotomous variables (boldface) the circular mean and standard deviation are reported, together with the lower bound (LB) and upper bound (UB) of the circular 95% CI. Note that on the circle LB can be larger than UB due to the periodicity; e.g. 10° = 370°. For the linear parameters the mean, the standard deviation and the lower and upper bound of the 95% CI are shown. For ρ we report its mode and the 95% HPD interval.

Table 6

Summary statistics of the regression parameters in the model for the standard trials.

Method		Parameter estimate			Quantiles	
		Mean	Mode	SD	LB	UB
WN	**Intercept**	353.43°		0.35	313.12°	32.91°
	Maze	347.09°		0.22	321.98°	12.76°
	Training	0.14		0.03	-3.28	3.55
	ρ		.93		.85	.96
WC	**Intercept**	349.90°		0.28	319.03°	19.63°
	Maze	351.87°		0.17	333.42°	11.61°
	Training	0.41		0.02	-2.31	3.12
	ρ		.86		.73	.92

Note. For dichotomous variables (boldface) the circular mean and standard deviation are reported, together with the lower bound (LB) and upper bound (UB) of the circular 95% CI. Note that on the circle LB can be larger than UB due to the periodicity; e.g. 10° = 370°. For the linear parameters the mean, the standard deviation and the lower and upper bound of the 95% CI are shown. For ρ we report its mode and the 95% HPD interval.

Appendix

Slice sampling

A1 Conceptual introduction

In this appendix we provide a short and conceptual introduction to slice sampling Neal (2003). This technique can be used to obtain random samples from all (multivariate) probability distributions of which a pdf is known that is proportional to the density of the distribution itself. This way, it is possible to sample from non-standard probability distributions.

For a random variable x having a pdf proportional to $f(x)$, the first few steps of obtaining a random sample in are depicted in Figure A1. First, a starting value x_0 is chosen and the density $f(x_0)$ at this point is calculated (Figure A1a). The next step is to uniformly sample an auxiliary variable d from the interval $[0, f(x_0)]$. We define a slice S containing all values of x for which $d < f(x)$ (Figure A1b). From this slice, a next value x_1 is uniformly sampled, which then will be used to obtain a new value of d resulting in a new slice S (Figure A1c).

In Figure A2, three samples of increasing size, sampled from a standard normal distribution by means of slice sampling, are depicted. On the top row the pairs (x_n, d_n) are plotted together with the probability density function. On the bottom row, the obtained random samples are depicted as histograms. We observe that if the process as described above is repeated sufficient many times, the resulting sample indeed resembles the pdf.

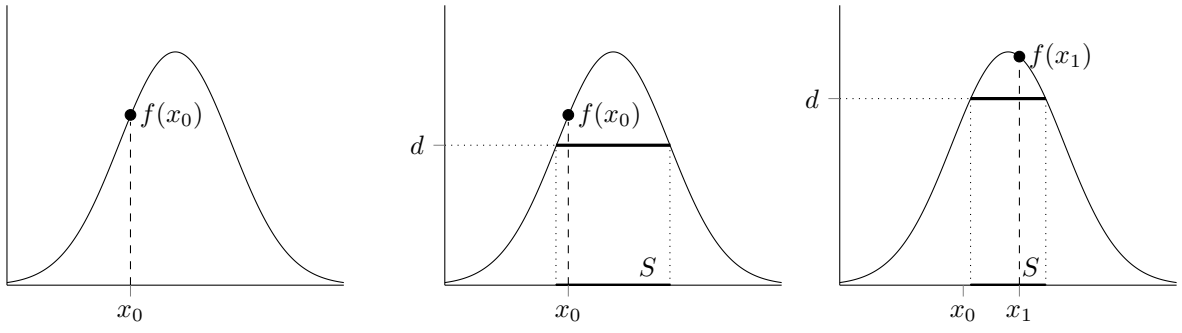
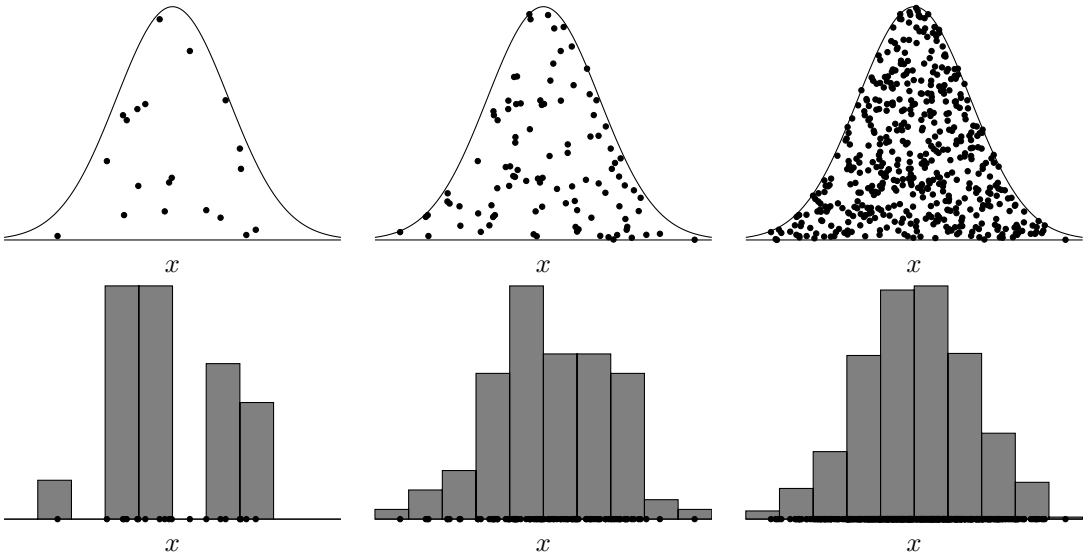


Figure A1. The first few steps in slice sampling a standard normal distribution.



(a) (b) (c)

Figure A2. Results of sampling respectively 20, 100 and 500 points using slice sampling.

A2 Multivariate slice sampling scheme to sample from the joint distribution

In this appendix we will derive a multivariate slice sampling scheme for a model with one circular outcome, an intercept and one continuous predictor. The rules as described below are also applicable to more advanced models, where model parameters of extra dichotomous predictors can be sampled like intercepts and model parameters belonging to extra continuous predictors can be sampled like slopes.

Definition of starting state. The first step in the slice sampling scheme is to define a *starting state* $\mathbf{s}_0$, or, stated differently, to assign starting values to the model parameters:

$$\mathbf{s}_0 = \begin{bmatrix} \beta_{0_0} & \in & [0, 2\pi] \\ \beta_{1_0} & \in & [0, 2\pi] \\ \beta_{2_0} & \in & \mathbb{R} \\ \rho_0 & \in & [0, 1] \end{bmatrix}. \tag{12}$$

Updating the auxiliary variable. We define the joint density at the current state of the sampler $\mathbf{s}_{cur}$ (in the initial step $\mathbf{s}_{cur} = \mathbf{s}_0$) for wrapped Cauchy distributed errors:

$$\mathrm{P}_{\mathrm{WC}}(\mathbf{s}_{cur}) \propto \frac{\exp\left(-\frac{\beta_{1_{cur}}^2}{2\sigma_\beta^2}\right) \cdot (\rho_{cur})^{-0.5}(1-\rho_{cur})^{n-0.5}(1+\rho_{cur})^n}{\prod_{i=1}^{n}\{1+\rho_{cur}^2 - 2\rho_{cur}\cos(r_{i_{cur}})\}}, \tag{13}$$

And for WN (note: $\sigma = \sqrt{-2\log(\rho)}$):

$$\mathrm{P}_{\mathrm{WN}}(\mathbf{s}_{cur}) \propto \prod_{i=1}^{n}\left\{\sum_{k=-1}^{1}\exp\left(\frac{(r_{i_{cur}}+2\pi k)^2}{4\log(\rho_{cur})}\right)\right\} \cdot \exp\left(-\frac{\beta_{1_{cur}}^2}{2\sigma_\beta^2}\right) \cdot (\rho_{cur})^{-0.5}(1-\rho_{cur})^{-0.5} \cdot (-\log\rho_{cur})^{-\frac{n}{2}}, \tag{14}$$

where:

$$r_{i_{cur}} = y_i - (\beta_{0_{cur}} + \beta_{1_{cur}} x_i) \mod 2\pi, \tag{15}$$

for both WC and WN.

Next, we sample a value z from a standard uniform distribution:

$$z \sim U(0,\ 1). \tag{16}$$

Using f we can update the auxiliary variable d as follows:

$$d = \begin{cases} z \cdot \mathrm{P_{WN}}(\mathbf{s}_{cur}), & for\ \ \mathrm{WN} \\ z \cdot \mathrm{P_{WC}}(\mathbf{s}_{cur}), & for\ \ \mathrm{WC} \end{cases}. \tag{17}$$

Updating intercepts. We draw a proposal $\beta_{0_{prop}}$ value for the intercept from its domain $[0, 2\pi]$:

$$\beta_{0_{prop}} \sim U(0,\ 2\pi). \tag{18}$$

We use the following acceptance rule to decide whether the proposal value is accepted. That is, we check if, given the current model parameters, the proposal value lies within the slice.

For WC:

$$\begin{gathered} d < \mathrm{P_{WC}}(\beta_{0_{prop}} \mid \beta_{1_{cur}},\ \rho_{cur}) \\ \frac{1}{z}\prod_{i=1}^{n}\{1+\rho_{cur}^2 - 2\rho_{cur}\cos\left(r_{i_{cur}}\right)\} > \prod_{i=1}^{n}\{1+\rho_{cur}^2 - 2\rho_{cur}\cos\left(y_i - \beta_{0_{prop}} - \beta_{1_{cur}}x_i \mod 2\pi\right)\}. \end{gathered} \tag{19}$$

For WN:

$$\begin{gathered} d < \mathrm{P_{WN}}(\beta_{0_{prop}} \mid \beta_{1_{cur}},\ \rho_{cur}) \\ z \cdot \prod_{i=1}^{n}\sum_{k=-1}^{1}\exp\left(\frac{\left(r_{i_{cur}} + 2\pi k\right)^2}{4\log(\rho_{cur})}\right) < \prod_{i=1}^{n}\sum_{k=-1}^{1}\exp\left(\frac{\left(\{y_i - \beta_{0_{prop}} - \beta_{1_{cur}}x_i \mod 2\pi\} + 2\pi k\right)^2}{4\log(\rho_{cur})}\right). \end{gathered} \tag{20}$$

This process is repeated until a proposal value for the intercept is found that is within the slice. In that case the current state of the sampler is updated using this proposal value.

Updating slopes. The domain of a slope parameter, in this case β_1, is not bounded, therefore we cannot directly sample a proposal value from this domain. We use a *stepping out procedure* (Neal, 2003, p. 713) to find a bounded domain $[\beta_{1\downarrow},\ \beta_{1\uparrow}]$ that contains S. That is, an interval of fixed size w is placed randomly around the current value of the slope $\beta_{1_{cur}}$. Each side of the interval is then extended in steps of size w until its border lie outside the slice (see Figure A3a).

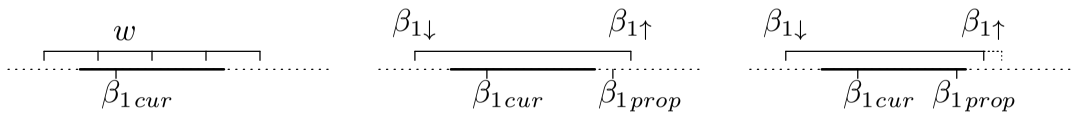
Next, we draw a proposal β_{1prop} value for the slope from $[\beta_{1\downarrow},\ \beta_{1\uparrow}]$ (see Figure A3b):

$$\beta_{1_{prop}} \sim U(\beta_{1\downarrow},\ \beta_{1_{\uparrow}}). \tag{21}$$

We use the following acceptance rule to decide whether the proposal value is accepted. That is, we check if, given the current model parameters, the proposal value lies within the slice.

For WC:

$$d < \mathrm{P_{WC}}(\beta_{1_{prop}} \mid \beta_{0_{prop}},\ \rho_{cur})$$
$$z \cdot \frac{\exp\left(-\frac{\beta_{1_{cur}}^2}{2\sigma_\beta^2}\right)}{\prod_{i=1}^{n}\{1+\rho_{cur}^2-2\rho_{cur}\cos(r_{i_{cur}})\}} < \frac{\exp\left(-\frac{\beta_{1_{prop}}^2}{2\sigma_\beta^2}\right)}{\prod_{i=1}^{n}\{1+\rho_{cur}^2-2\rho_{cur}\cos\left(y_i-\beta_{0_{cur}}-\beta_{1_{prop}}x_i \mod 2\pi\right)\}}. \tag{22}$$



(a) (b) (c)

Figure A3. Stepping out and shrinkage procedure.

For WN:

$$d < \mathrm{P_{WN}}(\beta_{1_{prop}} \mid \beta_{0_{prop}},\ \rho_{cur})$$
$$z \cdot \prod_{i=1}^{n} \left\{ \sum_{k=-1}^{1} \exp\left(\frac{(r_{i_{cur}} + 2\pi k)^2}{4\log(\rho_{cur})} \right) \right\} \cdot \exp\left(-\frac{\beta_{1_{cur}}^2}{2\sigma_\beta^2} \right) < \prod_{i=1}^{n} \left\{ \sum_{k=-1}^{1} \exp\left(\frac{(\{y_i - \beta_{0_{cur}} - \beta_{1_{prop}} x_i \mod 2\pi\} + 2\pi k)^2}{4\log(\rho_{cur})} \right) \right\} \cdot \exp\left(-\frac{\beta_{1_{prop}}^2}{2\sigma_\beta^2} \right). \tag{23}$$

If the proposal value lies outside the slice, its value is used to *shrink* the interval $[\beta_{1\downarrow},\ \beta_{1\uparrow}]$ (see Figure A3b). And the process is repeated. If, on the other hand, the proposal value for the slope is within the slice (see Figure A3c), the current state of the sampler is updated using this proposal value.

Updating the circular concentration. We draw a proposal ρ_{prop} value for the circular concentration from its domain $[0, 1]$:

$$\rho_{prop} \sim U(0,\ 1). \tag{24}$$

We use the following acceptance rule to decide whether the proposal value is accepted. That is, we check if, given the current model parameters, the proposal value lies within the slice.

For WC:

$$d < \mathrm{P_{WC}}(\rho_{prop} \mid \beta_{0_{prop}},\ \beta_{1_{cur}})$$
$$z \cdot \frac{(\rho_{cur})^{-0.5}(1-\rho_{cur})^{n-0.5}(1+\rho_{cur})^{n}}{\prod_{i=1}^{n} 1 + \rho_{cur}^2 - 2\rho_{cur}\cos(r_{i_{cur}})} < \frac{(\rho_{prop})^{-0.5}(1-\rho_{prop})^{n-0.5}(1+\rho_{prop})^{n}}{\prod_{i=1}^{n} 1 + \rho_{prop}^2 - 2\rho_{prop}\cos(r_{i_{cur}})}. \tag{25}$$

For WN:

$$d < \mathrm{P_{WN}}(\rho_{prop} \mid \beta_{0_{prop}},\ \beta_{1_{cur}})$$
$$z \cdot \prod_{i=1}^{n} \left(\sum_{k=-1}^{1} \exp\left(\frac{(r_{i_{cur}} + 2\pi k)^2}{4\log(\rho_{cur})} \right) \right) \cdot (\rho_{cur})^{-0.5}(1-\rho_{cur})^{-0.5} \cdot (-\log\rho_{cur})^{-\frac{n}{2}} < \prod_{i=1}^{n} \left(\sum_{k=-1}^{1} \exp\left(\frac{(r_{i_{cur}} + 2\pi k)^2}{4\log(\rho_{prop})} \right) \right) \cdot (\rho_{prop})^{-0.5}(1-\rho_{prop})^{-0.5} \cdot (-\log\rho_{prop})^{-\frac{n}{2}}. \tag{26}$$

This process is repeated until a proposal value for the circular concentration is found that is within the slice. In that case the current state of the sampler is updated using this proposal value.