

## **Bringing the analysis of animal orientation data full circle: model-based approaches with maximum likelihood**

Robert R. Fitak\*<sup>1</sup> and Sönke Johnsen<sup>1</sup>

<sup>1</sup>Department of Biology, Duke University, Durham, NC 27708 USA

\*Corresponding author

Robert R. Fitak

Duke University

Box 90338

Durham, NC 27708

Tel: 919-684-7188

Fax: 919-660-7293

rfitak9@gmail.com

*Key words:* circular statistics, unimodal, bimodal, von Mises distribution, Rayleigh test

*Summary Statement:* A new statistical package was developed in R for the maximum likelihood analysis of animal orientation data. Both simulated and empirical examples are provided for demonstration.

## ABSTRACT

In studies of animal orientation, data are often represented as directions that can be analyzed using circular statistical methods. Although several circular statistical tests exist to detect the presence of a mean direction, likelihood-based approaches may offer advantages in hypothesis testing – especially when data are multimodal. Unfortunately, likelihood-based inference in animal orientation remains rare. Here, we discuss some of the assumptions and limitations of common circular tests and report a new R package called CircMLE to implement the maximum likelihood analysis of circular data. We illustrate the use of this package on both simulated datasets and an empirical example dataset in Chinook salmon (*Oncorhynchus tshawytscha*). Our software provides a convenient interface that facilitates the use of model-based approaches in animal orientation studies.

## INTRODUCTION

There are numerous instances where biological information can be projected onto the unit circle rather than in linear space. Circular biological data often take the form of directions or orientations (e.g., compass bearings), but can include any data functionally periodic in nature. Generally, circular data are described statistically using parameters intuitively and semantically identical to those of linear data (e.g., mean, variance, correlation), but their mathematical derivation can be quite different (reviewed in Lee, 2010). Numerous statistical tests have been developed to determine the presence of a preferred direction within a group or differences in preferred directions between groups (see Batschelet, 1981). However, as suggested by Ruxton (2017), biologists often are not familiar with or lack specific training in circular data, thus rendering the statistical treatment of these data many times inadequately explored.

A majority of studies of circular animal orientation utilize frequentist-based statistical tests. These tests are based upon the calculation of a test statistic and the probability of observing that statistic ( $p$  value) under the null hypothesis. It is well known that that use of  $p$  values and frequentist-based testing procedures often suffer from flaws in experimental design, logic, and/or interpretation of results (Steel et al., 2013). The most widely criticized mistakes include pseudoreplication (Hurlbert, 1984), failure to correct for multiple testing (Noble, 2009; Pike, 2011; Verhoeven et al., 2005) and low statistical power (Jennions and Møller, 2003; Nakagawa, 2004). Similarly, the improper statistical treatment of circular datasets may contribute to the lack of robust and reproducible results often encountered in animal orientation studies (e.g., Busse and Trocińska, 1999; Kirschvink et al., 2010). As an

alternative to the frequentist-based tests, both maximum likelihood (e.g., Schnute and Groot, 1992) and Bayesian (e.g., Ozarowska et al., 2013) methods have been developed as model-based approaches to describe orientation. These methods rely on the probability of data given a particular model, or likelihood, and are preferred for quantifying strength of evidence in favor of a hypothesis (Burnham and Anderson, 2002; Burnham et al., 2011). These likelihood-based approaches, however, are rarely applied in animal orientation studies.

In this article, we focus on the circular statistical treatment of the most fundamental question in animal orientation studies – the presence or absence of a preferred direction(s). Although we limit our discourse to animal orientation, the principles discussed are applicable to other examinations of circular data (e.g., studies of circadian rhythm). Furthermore, we only evaluate the statistical treatment of a single sample (within a group), and relegate two-sample (between groups) comparisons to future investigations. We begin by briefly describing the most frequent statistical tests used and discuss their assumptions and limitations. Next, we highlight the importance of model-selection procedures as an alternative or complement to the frequentist approach and enable such approaches by providing a convenient set of functions for use in R (package *CircMLE*; R Core Development Team, 2017). We conclude with a set of recommendations in the context of improving the statistical analysis of animal orientation data.

## TESTING FOR A PREFERRED DIRECTION

### Rayleigh's test

Despite the availability of numerous statistical tests for departure from uniformity in circular data, Rayleigh's test is by far the most common (Batschelet, 1981). With the Rayleigh test, the null hypothesis is uniformity (i.e., no preferred direction), whereas the alternative is unimodality (i.e., a single preferred direction) (Mardia, 1972). At the core of the Rayleigh test is the von Mises distribution (Mardia, 1972) – the circular equivalent of the normal distribution. This distribution is described by a mean direction  $\varphi$  and concentration parameter  $\kappa$ , which are analogous to the mean  $\mu$  and reciprocal of the variance  $\frac{1}{\sigma^2}$  of the normal distribution, respectively. The Rayleigh test compares the likelihood of the data assuming a uniform von Mises distribution ( $\kappa = 0$ ) with the likelihood assuming a non-uniform von Mises distribution ( $\kappa > 0$ ).

## V test

Although the Rayleigh test can be performed for an unknown preferred direction, a variant of the test, the V test, is more powerful at detecting departures from uniformity when a preferred direction is specified *a priori* (Durand and Greenwood, 1958). The interpretation of the V test is equivalent to that of the Rayleigh test and rejecting the null hypothesis does not indicate the population is oriented in the predicted direction. In other words, when using the V test the user must be warned that the null hypothesis again is uniformity and that the mean direction may differ from the predicted direction (Aneshansley and Larkin, 1981; Ruxton, 2017).

## Multimodality and the Rao spacing test

Although in many biological scenarios a single preferred direction is expected, multiple preferred directions are also plausible but violate the assumptions inherent in the Rayleigh and V tests. In these cases, researchers must use alternative statistical approaches to analyze multimodal distributions that are less well described than their unimodal counterparts. In animal orientation studies, the only widely accepted case of multimodality is when two principle directions are opposite ( $180^\circ$  apart), or often termed “axial” (Jammalamadaka and SenGupta, 2001). When data are axial the angles can be doubled to produce a unimodal distribution that can be analyzed using a Rayleigh test (Batschelet, 1981). Although common in the animal orientation literature, doubling the angles has been questioned (Arnold and SenGupta, 2006). Because procedures for accurately identifying axial data remain poorly characterized (although see Arnold and SenGupta, 2006), studies often make this determination based upon either visual inspection (e.g., Takebe et al., 2012) or whether doubling increases the mean vector length (e.g., Pinzon-Rodriguez and Muheim, 2017). This qualitative assessment, if incorrect, may result in misleading statistical results when doubling of the angles is performed.

When data are multimodal the non-parametric Rao spacing test can be used (Rao, 1972). Aside from including multimodal distributions in the alternative hypothesis, the Rao spacing test is oftentimes more powerful than both the Rayleigh and V tests – especially for small sample sizes (i.e.,  $n \leq 10$ ; Bergin, 1991; Rao, 1972). Because of these advantages, Bergin (1991) described the Rao spacing test as “indispensable” for orientation datasets influenced by a multitude factors like nest orientation in birds. However, when the null hypothesis of uniformity is rejected, the Rao spacing test provides no information regarding

whether or not the data are unimodal or multimodal. Additional tests for multimodal circular data have been described (e.g., Pycke, 2010), but are either rarely used in the literature or are less well-characterized.

## MODEL-SELECTION APPROACHES

The preceding section described statistical tests based on the central paradigm of accepting or rejecting a null hypothesis based upon an arbitrary significance threshold (often a  $p$  value  $< 0.05$ ). Although ubiquitous in the life sciences, a  $p$  value does not represent a formal quantitative measure for strength of evidence (Royall, 1997). Alternatively, information-theoretic (likelihood-based) approaches, such as those derived from the Kullback-Leibler divergence, inherently quantify the strength of evidence for a hypothesis. Rather than testing a single hypothesis, these methods facilitate the simultaneous testing of null and multiple alternative hypotheses (reviewed by Burnham et al., 2011). The Akaike Information Criterion (AIC; Akaike, 1973) is the perhaps the best known and most common of these methods. The AIC is the sum of the deviance, or the model fit, and twice the number of parameters  $k$ . The AIC, in effect, penalizes models with more parameters to avoid overfitting. Theoretically, the user calculates the AIC for each model and selects the one with the smallest AIC, or least amount of information loss, as the “best” model. In general, models with a difference in AIC relative to the best model ( $\Delta\text{AIC}$ )  $< 2$  are well supported, models with  $\Delta\text{AIC}$  between 2 – 7 should rarely be dismissed, and models with  $\Delta\text{AIC} > 9$  have little support (Burnham and Anderson, 2004; Burnham et al., 2011). Early recommendations often excluded models with  $\Delta\text{AIC} > 2$ , but any arbitrary cutoff such as this should be interpreted cautiously and in the context of other parameters such as the likelihood, model probabilities, and/or evidence ratios (Burnham and Anderson, 2002; Burnham et al., 2011). Other common methods include a corrected version of AIC (AICc; Hurvich and Tsai, 1989) and the Bayesian Information Criterion (BIC; Schwarz, 1978). The AICc is appropriate for small sample sizes (i.e., when  $k$  is a large fraction of  $n$ ; Hurvich and Tsai, 1989), and BIC is used only when the true model is included in the set otherwise BIC is a poor choice for empirical data (Burnham and Anderson, 2004).

When performing a model-based approach, the limiting factor is generating a different mathematical model to describe each possible hypothesis. In circular data analysis, both maximum-likelihood and Bayesian models have been proposed (Fernández-Durán and Gregorio-Domínguez, 2016; Fu et al., 2008; Ożarowska et al., 2013; Ravindran and Ghosh, 2011; Schnute and Groot, 1992). Of these, Schnute and Groot (1992) provide perhaps the

most thorough mathematical treatment of circular animal orientation models. Schnute and Groot described 10 models of orientation based upon the bimodal, or mixed, von Mises distribution (Table 1; see **MATERIALS AND METHODS** for a complete description). In general, these models fall into three categories: i) a uniform model (M1) is that of random orientation, ii) unimodal models (M2A, M2B, M2C) with a single preferred direction, and iii) bimodal models (M3A, M3B, M4A, M4B, M5A, M5B) with two preferred directions. The bimodal models can further be split into axial (M3A, M3B, M4A, M4B) and non-axial (M5A, M5B) types. Each model is described by up to five parameters: a mean direction ( $\varphi_1$ ) and concentration parameter ( $\kappa_1$ ) for the first mode, a mean direction ( $\varphi_2$ ) and concentration parameter ( $\kappa_2$ ) for the second mode, and the proportional size of the first distribution ( $\lambda$ , the second distribution is thus fixed at size  $1 - \lambda$ ). Rather than subjectively selecting models to test, Schnute and Groot (1992) demonstrated that a likelihood-based approach using these models is a practical method to simultaneously examine multiple orientation hypotheses. Despite the mathematical descriptions of these models, their implementation in studies of animal orientation remains rare.

## MATERIALS AND METHODS

### Orientation models and the likelihood function

Schnute and Groot (1992) originally proposed 10 models of animal orientation (Table 1). The uniform (M1) and unimodal (M2A, M2B, M2C) models can be described by the von Mises distribution (Mardia, 1972) with a principle direction  $\varphi$  and concentration parameter  $\kappa$

$$M(\theta|\varphi, \kappa) = \frac{e^{\kappa \cos(\theta-\varphi)}}{2\pi I_0(\kappa)} \quad (1)$$

where  $I_0(\kappa)$  is a modified Bessel function of order 0. The probability distribution of angles  $\theta$  thus depends on the parameters  $\varphi$  and  $\kappa$ . When  $\kappa = 0$ , the distribution is uniform, whereas increasing values of  $\kappa$  represent increased concentration of angles around  $\varphi$ . For convenience, all angles hereafter are reported in degrees modulo  $360^\circ$ .

In many situations, the distribution of angles in a dataset may contain multiple modal directions and is referred to as a mixture. When a mixture of two distributions exists, it is represented by the bimodal, or mixed, von Mises (equation 2.5 in Schnute and Groot, 1992)

$$B(\theta|\varphi_1, \kappa_1, \lambda, \varphi_2, \kappa_2) = \lambda M(\theta|\varphi_1, \kappa_1) + (\lambda - 1)M(\theta|\varphi_2, \kappa_2) \quad (2)$$

where each distribution  $i$  in  $i = 1, 2$  has a mean direction  $\varphi_i$  and concentration parameter  $\kappa_i$ , and  $\lambda$  denotes the proportional size of the first distribution. When the data are axial then  $\varphi_2 = \varphi_1 \pm 180^\circ$ . The models described by Schnute and Groot (1992) can be depicted by the bimodal von Mises (Eqn 2) with fixed and free parameters provided in Table 1. The negative log likelihood function of the bimodal von Mises is then calculated as

$$L(Q) = \sum_{i=1}^n -\ln[B(\theta_i|Q)] \quad (3)$$

for  $n$  observed angles in  $\theta$  given the vector of model parameters  $Q = (\varphi_1, \kappa_1, \lambda, \varphi_2, \kappa_2)$ .

### **CircMLE: a new R package**

It is possible that model-based approaches to investigate circular orientation are rarely applied because they lack convenient execution in conventional statistical software. To encourage their use, we implemented the models of Schnute and Groot (1992) into a new package CircMLE v0.2 for use in the statistical software R and available on CRAN (R Core Development Team, 2017). CircMLE calculates the maximum likelihood of the 10 models for a given set of data and compares them using either AIC, AICc, or BIC with a single command. CircMLE requires the ‘circular’ package (Agostinelli and Lund, 2013), also available on CRAN in order to run. The input data is a vector of angular measurements, preferably in radians and modulo  $2\pi$ , and of class ‘circular’ (see the ‘circular’ package description). Although the function ‘check\_data’ will attempt to coerce a vector into the appropriate format, it is recommended to set these attributes beforehand.

In CircMLE there are 10 separate functions corresponding to the models described by Schnute and Groot (1992). These functions separately report the likelihood, estimated parameters, and optimization diagnostics for the respective model. The functions are named to match those of Schnute and Groot (i.e., M1, M2A, M2B, M2C, M3A, M3B, M4A, M4B, M5A, M5B). Specific information and a description for each function are available in the package manual (available on CRAN). Alternatively, all models can be run simultaneously and compared using the function ‘circ\_mle’. The ‘criterion’ parameter (default is AIC) can be set to compare models with either AIC (Akaike, 1973), AICc (Hurvich and Tsai, 1989), or BIC (Schwarz, 1978). The models are ranked according to the criterion and the differences relative to the best model, or  $\Delta$  values (e.g.,  $\Delta$ AIC), are reported in the results table. The

relative likelihoods, model weights, and evidence ratios are also provided based on the criterion chosen (Wagenmakers and Farrell, 2004). The user can select from various likelihood-optimization procedures (default is the “BFGS” method; Nocedal and Wright, 1999) and run multiple chains (default is five) from randomized starting points to reduce the chance of identifying local maxima. In some cases, the optimization will not converge, and it may be useful to increase the number of steps, or iterations, with ‘niter’ (default = 5000). If the user is only interested in comparing two models, then their likelihoods can be compared using a likelihood ratio test with the function ‘lr\_test’. This function requires the observed data, a null model, and an alternative model. The alternative model must have more parameters than the null model and be nested within the null model to fit the testing assumptions.

We avoided instances where  $\kappa$  may tend towards infinity by restricting  $\kappa$  to  $0 < \kappa \leq 227$ . This was calculated by first assuming a maximum discernible angular resolution of  $15^\circ$  ( $\sim 0.26$  rad; Ożarowska et al., 2013) then applying the normal distribution approximation of the von Mises so that (equation 4.1 in Schnute and Groot, 1992)

$$\frac{1}{\sqrt{\kappa}} \approx \sigma \quad (4)$$

Under these assumptions, we expect 95% of angles to be concentrated within the range (equation 4.2 in Schnute and Groot, 1992)

$$2 \times 1.96 \times \frac{1}{\sqrt{\kappa}} = 0.26 \text{ rad} \quad (5)$$

Additional restrictions include mean directions ( $0 < \varphi \leq 360^\circ$ ), proportional group size ( $0.25 < \lambda \leq 0.75$ ), and bimodal difference (difference between  $\varphi_1$  and  $\varphi_2 \geq 45^\circ$ ). We restricted  $\lambda$  within bounds  $0.25 < \lambda \leq 0.75$  to minimize the convergence on bimodal distributions with very few individuals oriented in one direction and the rest in another. This effect is greater when sample sizes are large, but these bounds restrict the group sizes to biologically meaningful interpretations. The minimum  $\lambda$  can be customized using the ‘lambda.min’ parameter. The minimum bimodal difference can also be customized with the ‘q.diff’ parameter.

Overall, the default parameters are currently set to provide a robust maximum-likelihood search applicable to most animal orientation datasets. For comparison purposes,

the Rayleigh test statistic and p value are also reported, and a native plotting function, ‘plot\_circMLE’, is available to visualize the observed and fitted data. The use of CircMLE is illustrated in both the **Simulated examples** and **Empirical example** sections below.

### Simulated examples

We determined the statistical power of the model selection procedure in CircMLE to correctly identify both uniform (M1) and unimodal (M2A) models from simulated data. We simulated a total of 1000 datasets for each combination of sample size  $n = 20, 50, 100, 200,$  or 500 and concentration parameter  $\kappa = 0, 0.1, 0.25, 0.5, 1, 2, 5,$  or 10 using the ‘rvonmises’ function in the R package ‘circular’ v0.4-7 (Agostinelli and Lund, 2013). We ran the ‘circ\_mle’ function in CircMLE v0.2.0 using default parameters and counted the proportion of datasets that returned the correct model ( $\Delta\text{AIC} < 2$ ).

Next, we simulated 100 angles from the mixed bimodal von Mises (Eqn 2) with a  $\varphi_1 = 0^\circ, \varphi_2 = 90^\circ,$  concentration parameters  $\kappa_1 = \kappa_2 = 5,$  and  $\lambda = 0.5$  using the ‘rmixedvonmises’ command in the package ‘circular’. This simulated dataset corresponded with model M5A, a special case of the bimodal model M5B with equal concentration parameters. We compared models with the maximum likelihood procedure implemented in CircMLE v0.2.0 using the default parameters. We included all 10 models in the example calculations for purpose of demonstration, but encourage users to restrict the set of models *a priori* to those most plausible or feasible (see recommendations in Dochtermann and Jenkins, 2011) because the models of Schnute and Groot include models that are special cases of other models.

### Empirical example

We provide an example application of our model-selection procedure using the orientation data from Putman et al. (2014). Briefly, the authors examined juvenile Chinook salmon (*Oncorhynchus tshawytscha*) orientation in three different magnetic fields: 1) a field from the northern periphery of their range, 2) a field from the southern periphery, and 3) the ambient magnetic field. The authors reported, based on Rayleigh test results, that when exposed to a magnetic field from either the northern or southern periphery, the salmon oriented in a direction similar to the direction of their marine feeding grounds. We reexamined their data using CircMLE also with the default parameters. Again, we included all models for purpose of demonstration. For brevity we focused attention on models with  $\Delta\text{AIC} < 2,$  although in no way do we advocate the use of arbitrary  $\Delta\text{AIC}$  cutoff thresholds.

## RESULTS AND DISCUSSION

### Simulated examples

We simulated a total of 40 datasets, each replicated 1000 times, to determine the power of our method to correctly identify a uniform or unimodal model of orientation (Fig. S1). We used an 80% threshold as a minimum cutoff for sufficient statistical power as recommended (Cohen, 1988). For uniform models ( $\kappa = 0$ ), power was consistently above 80%, and increased with sample size. For unimodal models with a small sample size ( $n = 20$ ), a  $\kappa$  greater than  $\sim 1$  was necessary to confidently return the correct model. For larger sample sizes (e.g.,  $\geq 100$ ), a  $\kappa > 0.4$  provides sufficient statistical power. Using Eqn 5, this corresponds with a 95% angular confidence interval of  $355^\circ$ . This indicates that weakly oriented datasets can be detected using the approach in CircMLE. Overall, the sample sizes chosen here are typical of those encountered in animal orientation studies and serve as a guide for researchers designing experiments.

In the simulated bimodal example, the default model selection procedure (AIC) selected model M5B as the best ( $\Delta\text{AIC} = 0$ ; Table S1 and Fig. S2). However, model M5A also provided an excellent fit ( $\Delta\text{AIC} = 0.23$ ). The remaining eight models poorly described the simulated data ( $\Delta\text{AIC} \geq 11.6$ ). The model weights, or probabilities, for M5B and M5A were 53% and 47%, respectively. Model M5B predicted parameters  $\varphi_1 = 90^\circ$ ,  $\varphi_2 = 360^\circ$  ( $0^\circ$ ),  $\kappa_1 = 4.6$ ,  $\kappa_2 = 4.9$ , and  $\lambda = 0.48$ . These parameters were nearly identical to those used to simulate the dataset, demonstrating relatively high accuracy of the likelihood search.

### Empirical example

The results from both a traditional Rayleigh test and from CircMLE are shown in Table 2. For the Northern group, an axial model (M3B;  $\Delta\text{AIC} = 0$ ) was the best fit. This model suggested two, equally sized distributions with mean directions of  $18^\circ$  and  $198^\circ$  and concentration parameters of 0.59 and 1.30, respectively (Fig. 1A). Other models, including both unimodal (M2A, M2B) and bimodal (M4A, M4B, M5A) models, were also supported by the data. In the Southern group, the unimodal model (M2A) was the best fit with a mean direction of  $17^\circ$  and concentration of 0.33 (Fig. 1B). A second unimodal model (M2B) also appeared as a plausible choice. In this model (M2B), the individuals are a mixture of half oriented in the same direction ( $17^\circ$ ) as in the “best” model, whereas the other half are uniformly oriented. This is inherently similar to the best-fit unimodal model. In the Ambient

group, the best model was a uniform distribution (M1), yet an axial model (M3A) was also supported (Fig. 1C). In both the Southern and Ambient groups, the best model was in agreement with the conclusions presented by the authors. In the Northern group, however, the best model (M3B) was a type of axial model with half the individuals oriented similar to the mean direction and the second half 180° different. A similar axial model (M3A) was reasonably supported in the Ambient group as well. Thus, the model-based results not only supported the authors' conclusions, but also suggested that in each group bimodal orientation in Chinook salmon cannot be excluded. In all cases, the highest model probability is only 39%, which advocates for a reasonable amount of model uncertainty (see Technical Issue #6 in Burnham et al., 2011).

## **Conclusions**

Proper statistical treatment of biological data is critical for making the appropriate conclusions regarding whether or not an effect exists. With CircMLE, we make a model-based approach to the analysis of circular biological data, especially in animal orientation, conveniently available for use in R. In the future, research characterizing effect sizes relative to biological significance (Martínez-Abraín, 2008) in circular statistics in addition to extending orientation models to cross-sample comparisons are needed. Nevertheless, the use of model-based approaches will improve our understanding of how animals produce oriented responses.

## **ACKNOWLEDGEMENTS**

We thank the Duke Shared Cluster Resource for providing computational resources, N. Putman for providing the example dataset, and L. Schweikert and E. Caves for comments on previous drafts.

## **COMPETING INTERESTS**

No competing interests declared.

## **AUTHOR CONTRIBUTIONS**

R.R.F. developed the software and wrote the manuscript, S.J. reviewed the manuscript and provided support.

## **FUNDING**

This work was supported by the Air Force Office of Scientific Research [FA9550-14-1-0208 to S.J.].

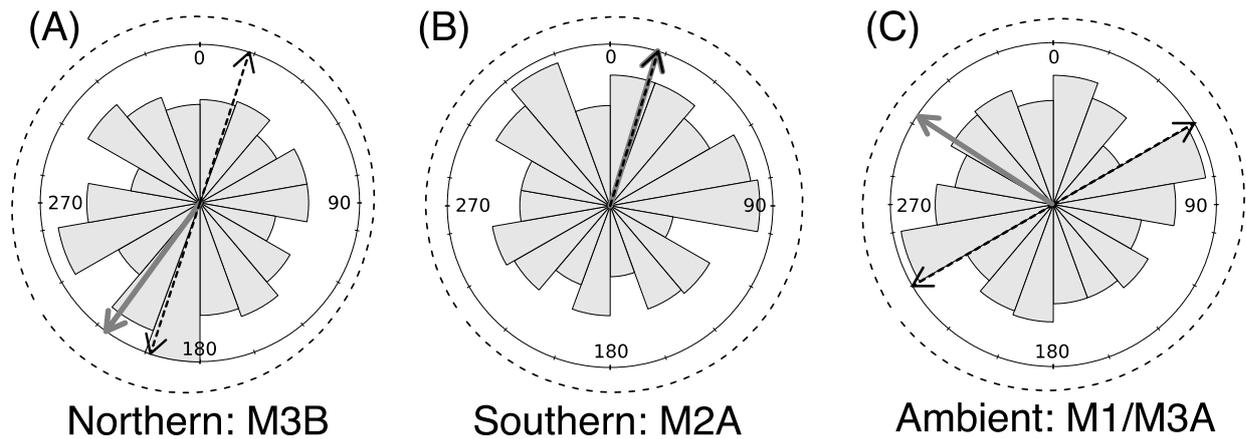
## REFERENCES

- Agostinelli, C. and Lund, U.** (2013). *R package 'circular': Circular Statistics (version 0.4-7)*. Retrieved from: <https://r-forge.r-project.org/projects/circular/>.
- Akaike, H.** (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (ed. B. N. Petrov and F. Csaki), pp. 267-281. Budapest: Akadémiai Kiadó.
- Aneshansley, D. J. and Larkin, T. S.** (1981). V-test is not a statistical test of homeward direction. *Nature*, **293**, 239-239.
- Arnold, B. C. and SenGupta, A.** (2006). Probability distributions and statistical inference for axial data. *Environ. Ecol. Stat.*, **13**, 271-285.
- Batschelet, E.** (1981). *Circular statistics in biology*. New York: Academic Press.
- Bergin, T. M.** (1991). A comparison of goodness-of-fit tests for analysis of nest orientation in western kingbirds (*Tyrannus verticalis*). *The Condor*, **93**, 164-171.
- Burnham, K. P. and Anderson, D. R.** (2002). *Model selection and multimodel inference: a practical and information-theoretic approach*. New York: Springer.
- Burnham, K. P. and Anderson, D. R.** (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.*, **33**, 261-304.
- Burnham, K. P., Anderson, D. R. and Huyvaert, K. P.** (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.*, **65**, 23-35.
- Busse, P. and Trocińska, A.** (1999). Evaluation of orientation experiment data using circular statistics - doubts and pitfalls in assumptions. *Ring*, **21**, 107-130.
- Cohen, J.** (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Erlbaum Associates.
- Dochtermann, N. A. and Jenkins, S. H.** (2011). Developing multiple hypotheses in behavioral ecology. *Behav. Ecol. Sociobiol.*, **65**, 37-45.
- Durand, D. and Greenwood, J. A.** (1958). Modifications of the Rayleigh test for uniformity in analysis of two-dimensional orientation data. *J. Geol.*, **66**, 229-238.
- Fernández-Durán, J. J. and Gregorio-Domínguez, M. M.** (2016). Bayesian analysis of circular distributions based on non-negative trigonometric sums. *J. Stat. Comput. Simul.*, **86**, 3175-3187.
- Fu, Y. J., Chen, J. H. and Li, P. F.** (2008). Modified likelihood ratio test for homogeneity in a mixture of von Mises distributions. *J. Stat. Plan. Inference*, **138**, 667-681.

- Hurlbert, S. H.** (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, **54**, 187-211.
- Hurvich, C. M. and Tsai, C. L.** (1989). Regression and time-series model selection in small samples. *Biometrika*, **76**, 297-307.
- Jammalamadaka, S. R. and SenGupta, A.** (2001). *Topics in circular statistics*. New York: World Scientific.
- Jennions, M. D. and Møller, A. P.** (2003). A survey of the statistical power of research in behavioral ecology and animal behavior. *Behav. Ecol.*, **14**, 438-445.
- Kirschvink, J. L., Winklhofer, M. and Walker, M. M.** (2010). Biophysics of magnetic orientation: strengthening the interface between theory and experimental design. *J. R. Soc. Interface*, **7**, S179-S191.
- Lee, A.** (2010). Circular data. *Wiley Interdiscip. Rev. Comput. Stat.*, **2**, 477-486.
- Mardia, K. V.** (1972). *Statistics of directional data*. London: Academic Press.
- Martínez-Abraín, A.** (2008). Statistical significance and biological relevance: a call for a more cautious interpretation of results in ecology. *Acta Oecol.*, **34**, 9-11.
- Nakagawa, S.** (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav. Ecol.*, **15**, 1044-1045.
- Noble, W. S.** (2009). How does multiple testing correction work? *Nat. Biotechnol.*, **27**, 1135-1137.
- Nocedal, J. and Wright, S. J.** (1999). *Numerical optimization*. New York: Springer.
- Ożarowska, A., Ilieva, M., Zehtindjiev, P., Åkesson, S. and Muś, K.** (2013). A new approach to evaluate multimodal orientation behaviour of migratory passerine birds recorded in circular orientation cages. *J. Exp. Biol.*, **216**, 4038-4046.
- Pike, N.** (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods Ecol. Evol.*, **2**, 278-282.
- Pinzon-Rodriguez, A. and Muheim, R.** (2017). Zebra finches have a light-dependent magnetic compass similar to migratory birds. *J. Exp. Biol.*, **220**, 1202-1209.
- Putman, N. F., Scanlan, M. M., Billman, E. J., O'Neil, J. P., Couture, R. B., Quinn, T. P., Lohmann, K. J. and Noakes, D. L. G.** (2014). An inherited magnetic map guides ocean navigation in juvenile Pacific salmon. *Curr. Biol.*, **24**, 446-450.
- Pycke, J. R.** (2010). Some tests for uniformity of circular distributions powerful against multimodal alternatives. *Can. J. Stat.*, **38**, 80-96.

- R Core Development Team.** (2017). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: <https://cran.r-project.org/>.
- Rao, J. S.** (1972). Some variants of chi-square for testing uniformity on circle. *Z. Wahrscheinlichkeitstheor. verw. Geb.*, **22**, 33-44.
- Ravindran, P. and Ghosh, S. K.** (2011). Bayesian analysis of circular data using wrapped distributions. *J. Stat. Theory Pract.*, **5**, 547-561.
- Royall, R.** (1997). *Statistical evidence: a likelihood paradigm*. London: Chapman & Hall.
- Ruxton, G. D.** (2017). Testing for departure from uniformity and estimating mean direction for circular data. *Biol. Lett.*, **13**, 20160756.
- Schnute, J. T. and Groot, K.** (1992). Statistical analysis of animal orientation data. *Anim. Behav.*, **43**, 15-33.
- Schwarz, G.** (1978). Estimating dimension of a model. *Ann. Stat.*, **6**, 461-464.
- Steel, E. A., Kennedy, M. C., Cunningham, P. G. and Stanovick, J. S.** (2013). Applied statistics in ecology: common pitfalls and simple solutions. *Ecosphere*, **4**, 115.
- Takebe, A., Furutani, T., Wada, T., Koinuma, M., Kubo, Y., Okano, K. and Okano, T.** (2012). Zebrafish respond to the geomagnetic field by bimodal and group-dependent orientation. *Sci. Rep.*, **2**, 5.
- Verhoeven, K. J. F., Simonsen, K. L. and McIntyre, L. M.** (2005). Implementing false discovery rate control: increasing your power. *Oikos*, **108**, 643-647.
- Wagenmakers, E. J. and Farrell, S.** (2004). AIC model selection using Akaike weights. *Psychon. Bull. Rev.*, **11**, 192-196.

**Figure 1. Observed and modeled orientation in Chinook salmon.** Mean directions (solid gray arrows) and circular histograms (gray bars) represent observed orientation of salmon in 20° bins exposed to (A) Northern, (B) Southern, or (C) ambient magnetic fields (Putman et al., 2014). The density (dashed line) and mean direction(s) (dashed arrows) of the best model are provided. In (B), the arrows overlap. In (C), the best model is a uniform distribution (M1), which has no mean direction, therefore the second-best model is shown.



## Tables

**Table 1. Ten models of animal orientation as described by Schnute and Groot (1992).**

Model Code	Modes	Name	P	Q ( $\varphi_1, \kappa_1, \lambda, \varphi_2, \kappa_2$ )
M1	0	Uniform	0	(na, 0, 1, na, na)
M2A	1	Unimodal	2	( $\varphi_1, \kappa_1, 1, na, na$ )
M2B	1	Symmetric modified unimodal	2	( $\varphi_1, \kappa_1, 0.5 na, 0$ )
M2C	1	Modified unimodal	3	( $\varphi_1, \kappa_1, \lambda, na, 0$ )
M3A	2	Homogenous symmetric bimodal	2	( $\varphi_1, \kappa_1, 0.5, \varphi_1 + \pi, \kappa_1$ )
M3B	2	Symmetric bimodal	3	( $\varphi_1, \kappa_1, 0.5, \varphi_1 + \pi, \kappa_2$ )
M4A	2	Homogenous axial bimodal	3	( $\varphi_1, \kappa_1, \lambda, \varphi_1 + \pi, \kappa_1$ )
M4B	2	Axial bimodal	4	( $\varphi_1, \kappa_1, \lambda, \varphi_1 + \pi, \kappa_2$ )
M5A	2	Homogenous bimodal	4	( $\varphi_1, \kappa_1, \lambda, \varphi_2, \kappa_1$ )
M5B	2	Bimodal	5	( $\varphi_1, \kappa_1, \lambda, \varphi_2, \kappa_2$ )

P = number of free parameters.

na = parameter not included in the model

**Table 2. Salmon orientation data from Putman et al. (2014).** The sample size ( $n$ ), mean vector magnitude ( $r$ ), and  $p$ -value ( $p$ ) from the Rayleigh test are shown in addition to the  $\Delta$ AIC for the 10 orientation models. Models with  $\Delta$ AIC  $< 2$  are shown in bold, and the best model is indicated by an asterisk.

Analysis		Northern	Southern	Ambient
Rayleigh test	$n$	233	234	240
	$r$	0.135	0.163	0.048
	$p$	0.014	0.002	0.581
Model $\Delta$ AIC (probabilities)	<b>M1</b>	5.81 (0.01)	8.56 (0.00)	<b>0.0*</b> (0.39)
	<b>M2A</b>	<b>1.24</b> (0.12)	<b>0.0*</b> (0.35)	2.91 (0.09)
	<b>M2B</b>	<b>0.69</b> (0.16)	<b>0.07</b> (0.34)	2.96 (0.09)
	<b>M2C</b>	5.40 (0.02)	20.1 (0.00)	20.4 (0.00)
	<b>M3A</b>	4.84 (0.02)	12.4 (0.00)	<b>1.30</b> (0.20)
	<b>M3B</b>	<b>0.0*</b> (0.23)	2.07 (0.13)	3.08 (0.08)
	<b>M4A</b>	<b>0.20</b> (0.21)	2.38 (0.11)	3.09 (0.08)
	<b>M4B</b>	<b>1.96</b> (0.09)	16.0 (0.00)	4.94 (0.03)
	<b>M5A</b>	<b>0.94</b> (0.14)	4.00 (0.05)	7.15 (0.01)
	<b>M5B</b>	20.6 (0.00)	6.15 (0.02)	6.00 (0.02)

**Table S1. Output from CircMLE v0.2.0.** The output from the ‘circ\_mle’ function for all 10 models of orientation using default parameters. The input data consisted of 100 angles simulated from a bimodal distribution with a  $\varphi_1 = 0^\circ$  (0 rad),  $\varphi_2 = 90^\circ$  (1.57 rad), concentration parameters  $\kappa_1 = \kappa_2 = 5$ , and  $\lambda = 0.5$ .

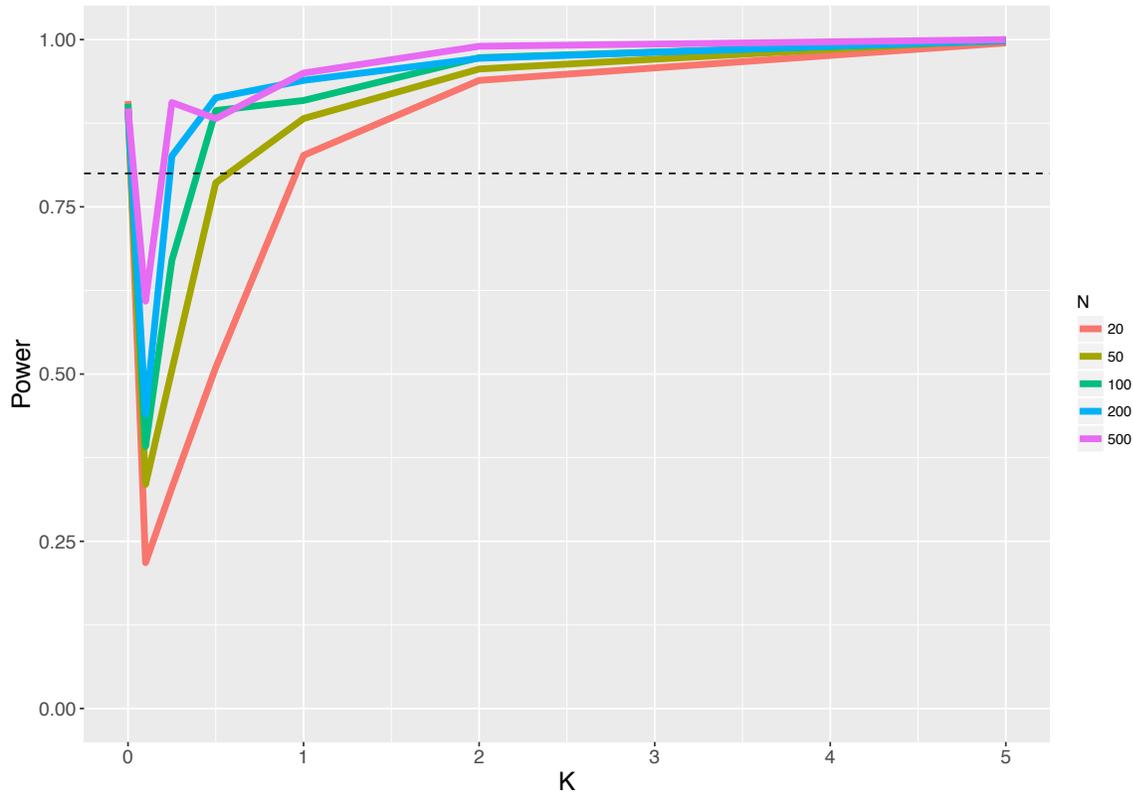
Model	M5B	M5A	M2A	M4B	M2B	M2C	M3B	M4A	M1	M3A
Params <sup>1</sup>	5	4	2	4	2	3	3	3	0	2
$\varphi_1$	1.57	1.53	0.83	4.01	0.83	0.83	3.97	1.31	NA	0.82
$\kappa_1$	4.63	6.01	1.82	0.00	2.33	3.82	0.00	1.05	0.00	0.87
$\lambda$	0.48	0.49	1.00	0.25	0.50	0.75	0.50	0.75	1.00	0.50
$\varphi_2$	6.28	6.28	NA	7.16	NA	NA	7.11	4.46	NA	3.97
$\kappa_2$	4.94	6.01	0	2.05	0	0	2.18	1.05	0	0.87
Likelihood <sup>2</sup>	123.9	125.0	132.7	140.6	150.4	150.2	150.5	163.3	183.8	183.1
Convergence <sup>3</sup>	0	0	0	0	0	0	0	0	0	0
AIC	257.8	258.1	269.5	289.1	304.9	306.4	307.0	332.6	367.6	370.3
AICc	258.5	258.5	269.6	289.5	305.0	306.6	307.3	332.8	367.6	370.4
BIC	270.9	268.5	274.7	299.5	310.1	314.2	314.8	340.4	367.6	375.5
$\Delta$ AIC	0	0.23	11.63	31.27	47.02	48.55	49.19	74.73	109.74	112.45
$\Delta$ AICc	0	0.01	11.12	31.05	46.51	48.17	48.80	74.35	109.10	111.94
$\Delta$ BIC	2.38	0	6.20	31.04	41.59	45.72	46.36	71.90	99.09	107.02
Relative likelihoods	1	0.89	0.003	0	0	0	0	0	0	0
AIC weights	0.53	0.47	0	0	0	0	0	0	0	0
Evidence Ratio	NA	1.12	3.36e <sup>-2</sup>	6.16e <sup>6</sup>	1.62e <sup>10</sup>	3.50e <sup>10</sup>	4.80e <sup>10</sup>	1.69e <sup>16</sup>	6.74e <sup>-23</sup>	2.62e <sup>24</sup>

<sup>1</sup>Number of free parameters in the model

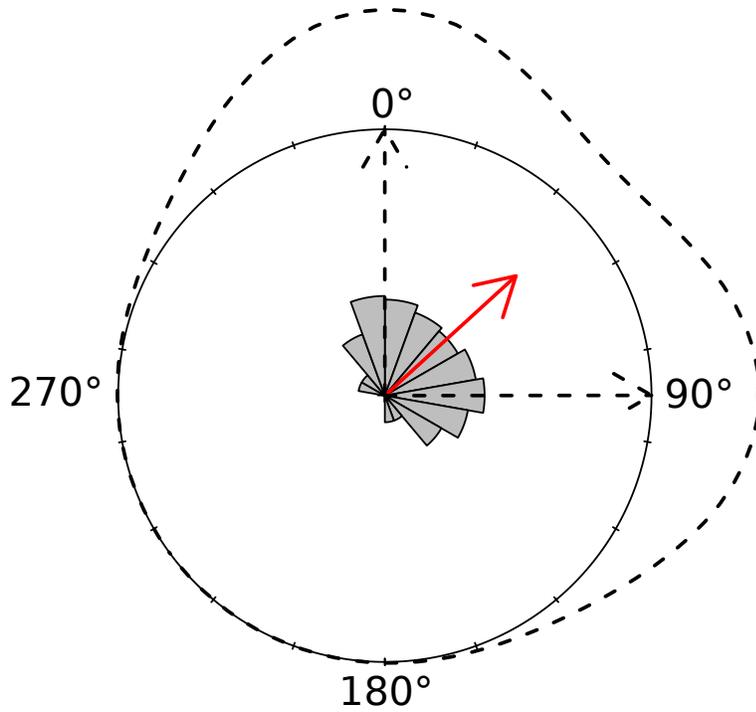
<sup>2</sup>Negative log likelihood

<sup>3</sup>Whether or not the maximum likelihood search converged, 0 = convergence.

**Figure S1. Power analysis of CircMLE.** The power of the maximum likelihood procedure in CircMLE v0.2.0 to correctly identify uniform (M1) and unimodal (M2A) models across various sample sizes (colored lines). Power is calculated as the proportion of 1000 simulated datasets that correctly returned the M1 (concentration parameter  $K = 0$ ) or M2A ( $K > 0$ ) models as best fit ( $\Delta AIC < 2$ ). The recommended minimum power of 80% is shown (dotted black line).



**Figure S2. Best fit model for the simulated bimodal dataset.** Circular histogram (gray bars) and mean vector (red arrow) of the simulated bimodal dataset. The density (dotted line) and mean angles (dotted arrows) for the best fit model, model M5B, are shown. The plot was produced using the 'plot\_circMLE' function in the CircMLE v0.2.0 package in R.



**Table S1. Output from CircMLE v0.2.0.** The output from the ‘circ\_mle’ function for all 10 models of orientation using default parameters. The input data consisted of 100 angles simulated from a bimodal distribution with a  $\varphi_1 = 0^\circ$  (0 rad),  $\varphi_2 = 90^\circ$  (1.57 rad), concentration parameters  $\kappa_1 = \kappa_2 = 5$ , and  $\lambda = 0.5$ .

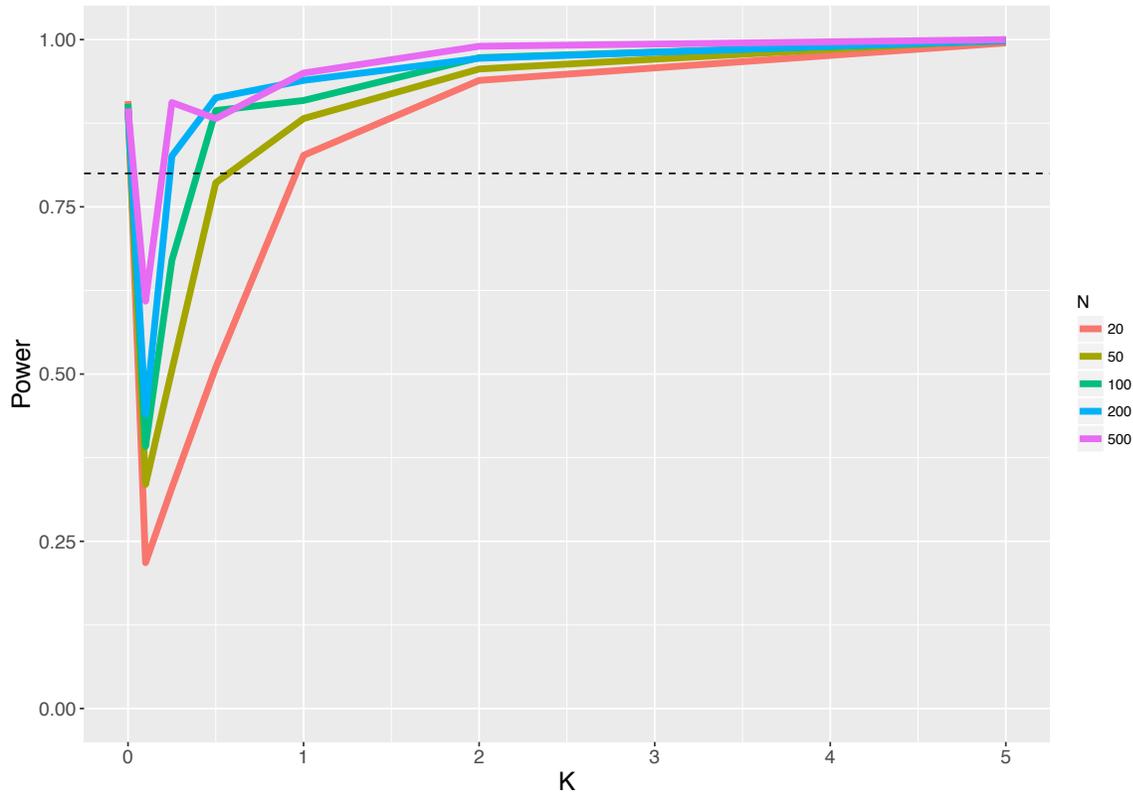
Model	M5B	M5A	M2A	M4B	M2B	M2C	M3B	M4A	M1	M3A
Params <sup>1</sup>	5	4	2	4	2	3	3	3	0	2
$\varphi_1$	1.57	1.53	0.83	4.01	0.83	0.83	3.97	1.31	NA	0.82
$\kappa_1$	4.63	6.01	1.82	0.00	2.33	3.82	0.00	1.05	0.00	0.87
$\lambda$	0.48	0.49	1.00	0.25	0.50	0.75	0.50	0.75	1.00	0.50
$\varphi_2$	6.28	6.28	NA	7.16	NA	NA	7.11	4.46	NA	3.97
$\kappa_2$	4.94	6.01	0	2.05	0	0	2.18	1.05	0	0.87
Likelihood <sup>2</sup>	123.9	125.0	132.7	140.6	150.4	150.2	150.5	163.3	183.8	183.1
Convergence <sup>3</sup>	0	0	0	0	0	0	0	0	0	0
AIC	257.8	258.1	269.5	289.1	304.9	306.4	307.0	332.6	367.6	370.3
AICc	258.5	258.5	269.6	289.5	305.0	306.6	307.3	332.8	367.6	370.4
BIC	270.9	268.5	274.7	299.5	310.1	314.2	314.8	340.4	367.6	375.5
$\Delta$ AIC	0	0.23	11.63	31.27	47.02	48.55	49.19	74.73	109.74	112.45
$\Delta$ AICc	0	0.01	11.12	31.05	46.51	48.17	48.80	74.35	109.10	111.94
$\Delta$ BIC	2.38	0	6.20	31.04	41.59	45.72	46.36	71.90	99.09	107.02
Relative likelihoods	1	0.89	0.003	0	0	0	0	0	0	0
AIC weights	0.53	0.47	0	0	0	0	0	0	0	0
Evidence Ratio	NA	1.12	3.36e <sup>-2</sup>	6.16e <sup>6</sup>	1.62e <sup>10</sup>	3.50e <sup>10</sup>	4.80e <sup>10</sup>	1.69e <sup>16</sup>	6.74e <sup>-23</sup>	2.62e <sup>24</sup>

<sup>1</sup>Number of free parameters in the model

<sup>2</sup>Negative log likelihood

<sup>3</sup>Whether or not the maximum likelihood search converged, 0 = convergence.

**Figure S1. Power analysis of CircMLE.** The power of the maximum likelihood procedure in CircMLE v0.2.0 to correctly identify uniform (M1) and unimodal (M2A) models across various sample sizes (colored lines). Power is calculated as the proportion of 1000 simulated datasets that correctly returned the M1 (concentration parameter  $K = 0$ ) or M2A ( $K > 0$ ) models as best fit ( $\Delta AIC < 2$ ). The recommended minimum power of 80% is shown (dotted black line).



**Figure S2. Best fit model for the simulated bimodal dataset.** Circular histogram (gray bars) and mean vector (red arrow) of the simulated bimodal dataset. The density (dotted line) and mean angles (dotted arrows) for the best fit model, model M5B, are shown. The plot was produced using the 'plot\_circMLE' function in the CircMLE v0.2.0 package in R.

