

# Approaches for Testing Uniformity Hypothesis in Angular Data of Mega-Herbivores

Dr. Robert Mathenge Mutwiri<sup>1,3</sup>, Prof. Henry Mwambi<sup>1</sup>, Rob Slotow<sup>2</sup>

<sup>1</sup>School of Mathematics, Statistics and Computer Sciences, University of KwaZulu Natal, South Africa  
School of Pure and Applied Sciences, Kiriinyaga University College, Kenya

<sup>2</sup> Professor, School of Mathematics, Statistics and Computer Sciences, University of KwaZulu Natal, South Africa

<sup>3</sup>Professor, School of Life Sciences, University of KwaZulu Natal, South Africa

**Abstract:** *Circular statistics is an area not used very much by ecologists to describe animal movement patterns. Nevertheless, the connection between the evaluation of temporal recurring events and the analysis of directional data have converged in several papers, and show circular statistics to be an outstanding tool for understanding animal movement better. The aim of this chapter is to evaluate the applications of circular statistical tests to check uniformity hypothesis in animal movement and its potential interpretation within the general framework of movement ecology. Four methods of circular statistics: Rayleigh's, Watson's, Rao's spacing and Kuiper's test based on the mean resultant length are applied to examine the uniformity hypothesis of GPS derived telemetry data of elephant movement collected from Kruger National Park(KNP) South Africa. Overall, circular statistical uniformity tests methods represent a useful tool for evaluation of directionality elephant movement with applications including (i) assessment of animal foraging strategies; (ii) determination of orientation in response to landscape features and (iii) determination of the relative strengths of landscape features present bin a complex environment.*

**Keywords:** Circular statistics, animal movement, turn angles, uniformity hypothesis

## 1. Introduction

The elaboration of appropriate conservation management and protection of endangered species of vegetation cover should be based on accurate interpretation of data and knowledge of animal movement on its habitat. One metric in movement ecology critical to this understanding is animal turn angles. It is urged in the literature that animals turn more during the wet season and less during the dry season. This variation is attributed to heterogeneous food resources and landscape features (Duffy et al., 2011). The movement of animals in protected areas (PA) and distribution of artificial water points has great impact on the vegetation and ecological dynamics of the ecosystem in general.

Circular statistical test of uniformity provides an opportunity to ecologists for understanding the turning patterns of the animals. In most studies, unimodal orientation may be the expected outcome. In several situations where for instant species or population preferences have been studied or when compass cues were set in experimental conflict, bimodal or multi-modal orientation may be expected (Fisher and Lee, 1992). Simulation studies based on random walk and Lévy flight theory of animal movement assume that turn angles are uniformly distributed (Viswanathan et al., 1996). As in the case linear statistics, the main objective here is to draw objective, reliable and biologically meaningful statistical inference about the population parameters on the basis of samples. Observations are either geometric or temporal in nature, where time related distributions can be fitted into a circular or spherical pattern (Batschelet, 1981). Parametric and non-parametric statistical methods can be used to test hypotheses concerning angular data (Batschelet, 1981).

These methods are based on the assumption that observed angles are independent, a condition that may not be satisfied when multiple angles are recorded from a single individual. Among the distributions used to describe data on a circle are the circular uniform and Von Mises distributions. Goodness-of-fit tests exist for these distributions. Four widely known methods for carrying out such a task are the subject of the current paper. Biological applications related to the technique described in section 2 and 3 illustrate the earlier predominant role of one particular life science in predicting the potential utility of circular statistics in other domains of scientific and technical endeavour.

In recent years new data sources and GIS tools have been increasingly used in ecological studies. A peculiar characteristic of these data sources is that, often, only information about the locations of the animal trajectory/path have been recorded over some sampling resolution. Methodologies targeted especially for these data need to be developed. Motivation of this work arises from the a study about the movement patterns of Elephant in Kruger National park, South Africa and the apparent lack of tangible awareness of circular statistics in the movement ecology literature. This study was performed in order to support decisions for the management, in particular the conservation strategies for Elephants.

A limitation of the most common approaches is that they treat turn angles qualitatively, that is, converting angular data as either north, south, east or west which may lead to loss of valuable information (Dai et al., 2007). The specific aim here is to demonstrate the potential usefulness of circular statistics to animal movement analysis, in deciding whether GPS derived telemetry data justify the inference of uniformity in animal movement. Following a short presentation of

theoretical aspects, numerical examples illustrate calculations required for drawing proper conclusions about the animal movement patterns.

## 2. Theory of Hypothesis Testing and Notation

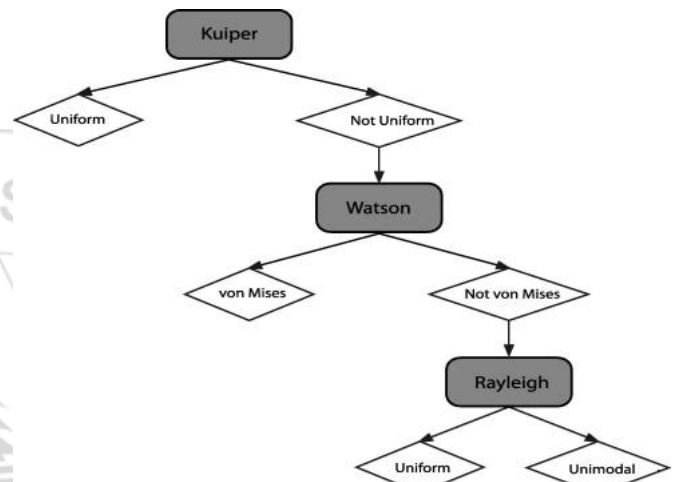
Circular hypothesis testing for uniformity is a valuable tool in movement ecology (Fisher and Lee, 1992). Statistical test should depend on previous assumptions supported by the descriptive or exploratory analysis (Fisher, Lewis and Embleton, 1993). In applied research, critical decisions based on data depends on objective and reliable assumptions. Due to technological advances and huge data collected on animal movement, testing such assumptions on animal movement data requires the knowledge of hypothesis testing.

A common question in circular statistics is whether a sample of data is uniformly distributed around the circle or has a common mean direction (Berens, 2009). A multiple of test statistics have been designed for testing this hypothesis. These methods includes: (i) Rao spacing test; (ii) Kuiper's test, (iii) Rayleigh test and (iv) Watson's test. These four tests can be used to assess the evidence for a uniform, unimodality and the goodness-of-fit for the von Mises distributions respectively (Mardia and Jupp, 2009). Previous investigations shows that the Rao's spacing method is more susceptible to rejecting  $H_0$  than the kuiper's test and the Rayleigh test in the face of a small data sample, unless data distribution is appreciably uniform at least in some of its sub-domains (Russell and Levitin, 1995). This implies that the Rao's spacing test carries a similar Type 1 error compared to other test; however, we point out that from the nature of the data, rejection of  $H_0$  cannot be absolutely certain (Mardia and Jupp, 2009).

In all the testing methods, the null hypothesis  $H_0$  states that the population samples are uniformly distributed around the circle and the alternative hypothesis  $H_A$  the population samples do not show a uniform (or random) circular distribution. Each method rejects  $H_0$  if its test statistic exceeds a critical value depending usually on sample size  $n$ , and level of significance  $\alpha$  (Jammalamadaka and SenGupta, 2001). In the theory of statistics,  $\alpha=0.05$  regarded as significant and  $\alpha=0.01$  is highly significant, meaning that a Type 1 error, is made by rejecting  $H_0$  is 5% or 1%, respectively. In the theory of statistics,  $\alpha=0.05$  is regarded as significant and  $\alpha=0.01$  is regarded as highly significant, meaning that a Type 1 error, is made by rejecting  $H_0$  is 5% level of significance, respectively. The current approach is more flexible by allowing the test statistic to determine the rejection of  $H_0$  on the basis of the p-value. According to Hogg and Craig (1978), the p-value is the magnitude of the error committed in rejecting  $H_0$  in face of the computed test statistic.

Rejecting the null hypothesis thus implies that deviations from uniformity are too large to assign them to chance factor; hence they are of deterministic origin Pewsey and Aurther (2013). Since rejection of a null hypothesis is statistically stronger than it's opposite, Rao's method is more inviting than any other test when at least medium-size deviations from

non-uniformity are expected from prior inspection descriptive statistics or exploratory analysis (Mardia and Jupp, 2009). Pewsey and Aurther (2013) emphasis the importance of testing uniformity hypothesis in circular. We note that if the data fits neither a von Mises distribution nor a uniform distribution and contains a single mode, then this data is said to follow a unimodal distribution (Fisher, Lewis and Embleton, 1993). In this case, although it is not possible to identify the actual distribution, the presence of a single mode not only indicates a preferred orientation in the sample, but also enables the use of non-parametric methods to estimate a mean direction with a confidence interval (Fisher, Lewis and Embleton, 1993). We then tested the uniformity hypothesis following the procedure outlined in figure 1.



**Figure 1:** Flowchart representing the sequence of hypothesis tests based on circular distributions

### 2.1 Rayleigh's Test

The Rayleigh's test is based on the intuitive idea of rejecting uniformity when the vector sample mean  $\bar{\theta}$ , is far from 0, when  $R$  is large (Fisher, Lewis and Embleton, 1993). The Rayleigh's test is the score test of uniformity within the von Mises model (Mardia and Lee, 2009). Put  $w = (\kappa \cos(\mu), \kappa \sin(\mu))'$ , the log likelihood Von Mises based on circular observations  $\theta_1, \dots, \theta_n$  is

$$l(\kappa; \theta_1, \dots, \theta_n) = n\bar{\kappa}\bar{\theta} - n \log I_0(\kappa),$$

where  $I_0(\kappa)$  is the modified Bessel function and

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^n (\cos \theta_i, \sin \theta_i)^T$$

is:

$$U = \frac{\partial l}{\partial \kappa^T} = n\bar{\theta} - nA(\kappa)(\cos \mu, \sin \mu)^T \quad (1)$$

From the moments properties of  $R$ , Mardia2009 it is possible to note that the score statistic is

$$U \text{var}(U)^{-1} U = 2nR^2. \quad (2)$$

From the general theory of score test, Mardia2009 the large sample asymptotic distribution of  $2nR^2$  under uniformity is a  $\chi^2$  with two degrees of freedom:

$$2nR^2 \sim \chi^2 \quad (3)$$

where n is the sample size. It has been demonstrated, also that the Rayleigh's test coincides with the likelihood test of uniformity within the von Mises family.

## 2.2 Kuiper's Test Statistic

Kuiper's test is used to determine if a given set of data can be a sample from a specific distribution. It is similar to the Kolmogorov-Smirnov (KS) test, as both compare cumulative distributions (Berens, 2009). For the one-sample test, the empirical cumulative distribution is compared to a theoretical cumulative distribution. As for circular case, this test measures the deviation between empirical distribution,  $S_n(x)$  and the Uniform cumulative distribution functions (cdf),  $F(x) = \frac{\theta}{2\pi}$ . In the case of circular data, the definition of cumulative distribution is not obvious and is quite different from the in line cdf (Pewsey, Neuhäuser and Ruxton, 2013). In the circular data case, in fact, we first have to choose the circle zero point and orientation, then we need to augment the ordered observations,  $\theta_1, \dots, \theta_n$  of  $x_0=0$  and  $\theta_{n+1}=2\pi$ . Then  $S_n$  is then defined by

$$S_n(\theta) = \frac{i}{n} \text{ if } \theta(i) \leq \theta \leq \theta(i+1) \quad i=0,1,\dots,n \quad (4)$$

Just as in Kolmogorov-Smirnov's test for in line distribution Durbin 1973, the following quantities are defined:

$$D_n^+ = \sup_{\theta} \{S_n(\theta) - F(\theta)\}, D_n^- = \sup_{\theta} \{F(\theta) - S_n(\theta)\}.$$

To overcome the dependence of  $D_n^+$  and  $D_n^-$  on the choice of the initial direction, (Kuiper, 1960) defined

$$V_n = D_n^+ + D_n^- \quad (5)$$

The statistic (5) has been demonstrated (Mardia and Jupp, 2009) to be invariant under the change of initial direction. The null hypothesis of uniformity is rejected for large values of  $V_n$ . Moreover, the Kuiper's test is consistent against all alternative to uniformity (Pewsey, Neuhäuser and Ruxton, 2013). For practical purposes, the following modification of  $V_n$  is used:

$$V_n^* = \frac{1}{2} V_n \left(1 + \frac{0.155}{\sqrt{n}} + \frac{0.24}{n}\right). \quad (6)$$

## 2.3 Watson's Test

Another common test of uniformity in circular statistics is the Watson  $U^2$  statistic (Watson, 1961) which is a modification of the Cramér-von Mises test (Durbin, 1973). This test is used as a goodness-of-fit statistics for the von Mises distribution Mardia 1972. As a test of goodness of fit for circular data, it is invariant to the choice of the origin. The Watson test statistic is defined as

$$U_n^2 = \int_0^{2\pi} \left[ (S_n(\theta) - F(\theta)) - \int_0^{2\pi} (S_n(\theta) - F(\theta)) dF \right] dF \quad (7)$$

It follows from this definition that the Watson's statistic is invariant under rotation and reflections. As for the Kuiper's test, it is useful to consider the following modified statistic:

$$U_n^{*2} = \left( U_n^2 - \frac{0.1}{n} + \frac{0.1}{n^2} \right) \left( 1 + \frac{0.8}{n} \right). \quad (8)$$

Stephens (1970) provides in a tabular form the quantiles of the Watson test statistic.

## 2.4 Rao's Spacing Test

Rao's Spacing test is a useful and powerful statistic for testing uniformity of circular data. As with other circular statistics, Rao's Spacing test is applicable for analysis of angular data, in studies of movement and spatial trends in geographical research (Pewsey, Neuhäuser and Ruxton, 2013). In many cases, particularly with an underlying distribution that is multimodal, it is more powerful than the popular Kuiper's Test and Rayleigh Test.

Rao's Spacing Test is based on the idea that if the underlying distribution is uniform, successive observations should be approximately evenly spaced, about  $360/N$  apart. Large deviations from this distribution, resulting from unusually large or unusually short spaces between observations, are evidence for directionality. It is related to the general class of linear statistical tests that are based on successive order statistics and spacing. The spacing tests sample arc lengths  $T_1, \dots, T_n$  defined as:

$$T_i = \theta(i) - \theta(i-1), \quad T_n = 2\pi - (\theta(n) - \theta(1)). \quad i=1, \dots, n-1, \quad (9)$$

Under uniformity  $E[T_i] = \frac{2\pi}{n}$ . Hence, it is reasonable to reject uniformity for large values of

$$L = \frac{1}{2} \sum_{i=1}^n \left| T_i - \frac{2\pi}{n} \right| \quad (10)$$

Large values of L indicate clustering of observations (Russell and Levitin, 1995). An extensive table of quantiles of L is given (Russell and Levitin, 1995), while (Sherman, 1950) shows that a suitable transformation of L is asymptotically standard normally distributed.

## 3. Application to Elephant Movement Data

### 3.1 Ethics Statement

Elephant capture and handling was conducted in strict accordance with ethical standards. Specific approval for this particular research project was obtained through the University of KwaZulu-Natal Animal Ethics sub-committee (Ref: 009/10/Animal). This research also forms part of a registered and approved SANParks project, in association with Kruger National Park and Scientific Services (Ref: BIRPJ743) (Birkett et al., 2012).

### 3.2 Study area and GPS Data on a elephants

The methods described were applied to GPS location data for three female elephant herds in Kruger national park of South Africa. Elephant movement have been previously studied by



Vanak2010. The elephant population in KNP was estimated to be 14,000 individuals during 2010 (SANParks, unpublished data). From 2006 to 2010, we collected geographical location data, downloaded from GPS/GSM Collars (Africa Wildlife Tracking cc., South Africa), fitted to three elephant cows from different herds. To ensure the independence of sampling, a single female in each herd was selected and collared. Because GPS coordinates were measured continuously at frequent time intervals, the trajectory we obtained was almost smoothly connected (Vanak et al., 2010). The GPS data are freely available upon request.

The turning patterns of elephants were monitored by computing the angle  $\theta$  between two consecutive relocations over 30 Minutes interval of time between May, 2006 and June 2009. The data set consists of 36395, 29221 and 29908 observations for herd AM108, AM307 and AM308 respectively recorded at an interval of 30 minutes. The turning angle ( $\theta$ ) was computed for the three herds as the change in the direction of movement made by an each individual elephant tagged from one location to the next. The turning angle is a right-hand turn that ranged from  $-\pi$  to  $\pi$ .

The GPS locations provided data every 30 minutes during an entire day with an accuracy of the locations within 50 meters. This information was sent via cellular phone (GSM) network to a website from where the information was downloaded.

#### 4. Results

Angular data of three elephant herds was used to construct circular histograms that depict the mean and frequency of movement orientations of animals Figure (2) and (3). The rose diagram indicates that elephants move in a non-random manner during the dry and the wet seasons. This variation is attributed to uneven distribution of resources during dry season than in the wet season. Similarly, the linear histograms in Figure (3) shows that elephants turn angles are concentrated around zero degrees which indicate that the movement is target oriented. The tendency is known as the unimodal movement pattern better described by unimodal distributions. In all the four herds, the assumption of Unimodality is fulfilled, and is statistically significant results are obtained based on the conventional circular statistical methods.

**Table 1:** Descriptive analysis of the turn angle data of three elephant herds collected from Kruger National Park, South Africa

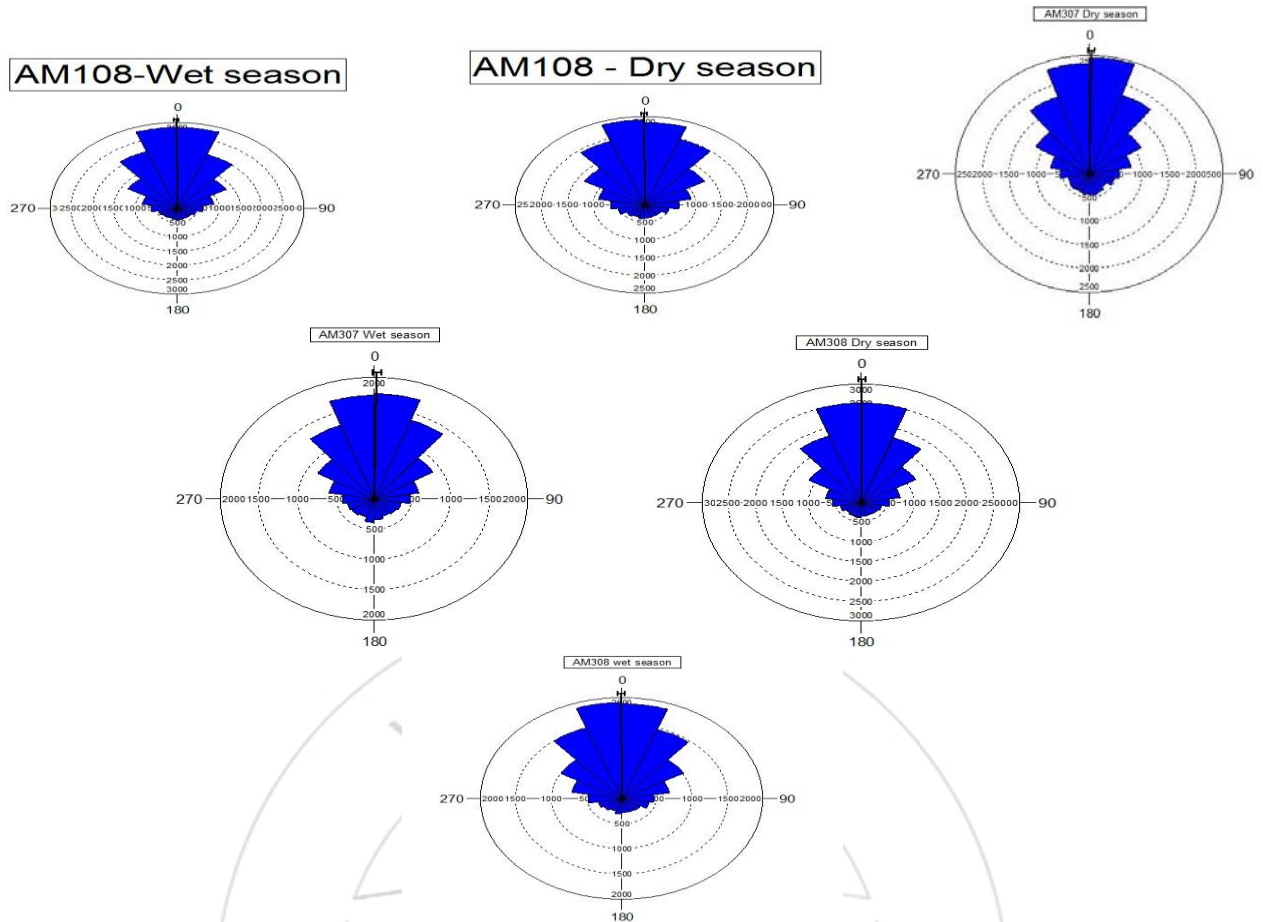
Statistic	AM108			AM307			AM308		
	Seasons			Seasons			Seasons		
Variable	Angle	Wet	Dry	angle	Wet	Dry	angle	Wet	Dry
Number of Observations	36395	17736	18659	29221	12652	16569	29908	13550	16358
Mean Vector ( $\mu$ )	359.425 <sup>o</sup>	359.434 <sup>o</sup>	359.418 <sup>o</sup>	0.706 <sup>o</sup>	0.84 <sup>o</sup>	0.609 <sup>o</sup>	359.957 <sup>o</sup>	359.7 <sup>o</sup>	0.16 <sup>o</sup>
Length of Mean Vector (r)	0.481	0.464	0.498	0.434	0.418	0.446	0.482	0.471	0.492
Concentration	1.096	1.046	1.144	0.962	0.92	0.995	1.099	1.065	1.127
Circular Variance	0.519	0.536	0.502	0.566	0.582	0.554	0.518	0.529	0.508
Circular Standard Deviation	69.292 <sup>o</sup>	71 <sup>o</sup>	67.686 <sup>o</sup>	74.044 <sup>o</sup>	75.65 <sup>o</sup>	72.833 <sup>o</sup>	69.19 <sup>o</sup>	70.336 <sup>o</sup>	68.248 <sup>o</sup>
Standard Error of Mean	0.414 <sup>o</sup>	0.618 <sup>o</sup>	0.556 <sup>o</sup>	0.519 <sup>o</sup>	0.821 <sup>o</sup>	0.668 <sup>o</sup>	0.455 <sup>o</sup>	0.695 <sup>o</sup>	0.602 <sup>o</sup>

**Table 2:** Tests for uniformity for three elephant herds turn angles derived from GPS tracking data.

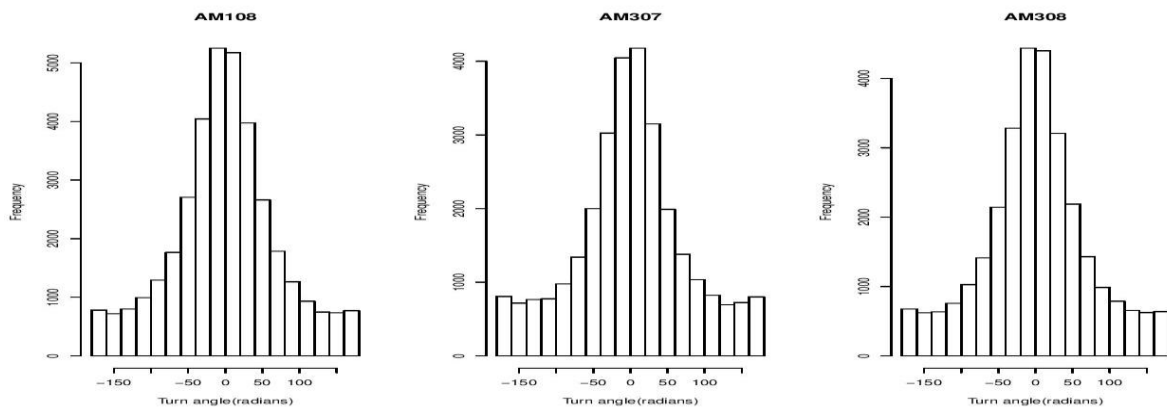
	Kuiper's		Rayleigh		Rao spacing		Watson's	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
AM108	61.6574	< 0.01	0.4813	0.000	166.2602	< 0.001	452.2261	< 0.01
AM307	50.7271	< 0.01	0.4339	0.000	162.4811	< 0.001	301.1729	< 0.01
AM308	56.3221	< 0.01	0.4823	0.000	166.2796	< 0.001	376.4143	< 0.01

The distribution of the turning angles between successive positions differed significantly from a uniform distribution for the three herds (Kuiper's test:  $p < 0.001$ ). Turning angles were backward oriented in the three herds at 359.43, 0.706 and 359.96 (mean direction of resultant vector, Rayleigh test:  $p < 0.001$ ) indicating a tendency of elephants to make a U-turns while foraging (Dai et al., 2007). There distribution was not different between the wet and dry season 1. Kuiper's and Rayleigh's test exhibit small p-values although the former is less realistic with respect to the powerful rejection of

uniformity by Rao's test. The results of the Rayleigh tests in Table 2 indicates that elephant movement turn angles are not evenly distributed in all directions and concentrated around 0<sup>o</sup> ( $\theta > 0$ ). The uniformity and von Mises provide more information as the Rayleigh and Rao's test confirm that azimuths are not uniformly distributed (see, Table 2). As azimuths do not follow a random distribution, we can look for a preferred direction; therefore, two tests for von Mises distribution were performed.



**Figure 2:** Circular histograms of elephant movement turn angles  $\theta$ . Circular bars indicate the number of observations within each class range and have been centred on  $0^\circ$ .



**Figure 3:** Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 - April 2009)

## 5. Discussion

This study focuses on the test of uniformity hypothesis for the circular animal movement data. The primary objective was to determine whether elephant orientation patterns were uniform. We also investigated the hypothesis that elephant orientation patterns would vary between seasons. Testing the null hypothesis can be accomplished by comparing any test statistic for uniformity to a reference distribution obtained by simulation. This is appealing in that precise distributions consistent with turn angle data need not be assumed for the procedure to have the proper type I error rate. Our results

suggest that elephant turn angle data is not uniformly distributed and showed no seasonal variation. The four statistical tests reject the uniformity hypothesis and conclude that animal movement turn angles data is not uniform. These applications, in summary entail calculation of animal orientation variables mean turn angle, concentration parameter, distribution, and determination of modality and testing hypothesis about uniformity of animal turn angles. It is important to note that the circular statistics applies to any level and scale of analysis, from individuals to several species of animals, and that the descriptive measures

calculated are easily compared statistically by means of a variety of two or more multi-sample tests.

## References

- [1] Batschelet, E. (1981). *Circular Statistics in Biology. Mathematics in Biology.* Academic Press.
- [2] Berens, P. (2009). *Circstat: A matlab toolbox for circular statistics.* *Journal of Statistical Software*, 31(10):1–21.
- [3] Birkett, P. J., Vanak, A. T., Muggeo, V. M. R., Ferreira, S. M., and Slotow, R. (2012). Animal perception of seasonal thresholds: Changes in elephant movement in relation to rainfall patterns. *PLoS ONE*, 7(6):e38363.
- [4] Dai, X., Shannon, G., Slotow, R., Page, B., and Duffy, K. J. (2007). Short-duration daytime movements of a cow herd of African elephants. *Journal of Mammalogy*, 88(1):151–157.
- [5] Duffy, K. J., Dai, X., Shannon, G., Slotow, R., and Page, B. (2011). Movement patterns of African elephants (*Loxodonta africana*) in different habitat types. *South African Journal of Wildlife Research*, 41(1):21–28.
- [6] Durbin, J. (1973). *Distribution theory for tests based on sample distribution function*, volume 9. Society for Industrial Mathematics.
- [7] Fisher, N. and Lee, A. (1992). Regression models for an angular response. *Biometrics*, pages 665–677.
- [8] Fisher, N., Lewis, T., and Embleton, B. (1993). *Statistical analysis of spherical data.* Cambridge University Press.
- [9] Hogg, R. V. and Craig, A. T. (1978). *Introduction to mathematical statistics.* New York: Macmillan.
- [10] Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics.* World Scientific, New York.
- [11] Kuiper, N. H. (1962). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A*, 63:38–47.
- [12] Mardia, K. (1972). *Statistics for directional data.* Academic press.
- [13] Mardia, K. and Jupp, P. (2009). *Directional statistics*, volume 494. Wiley.
- [14] Pewsey, A., Neuhauser, M., and Ruxton, G. D. (2013). *Circular statistics in R.* Oxford University Press.
- [15] Russell, G. S. and Levitin, D. J. (1995). An expanded table of probability values for rao's spacing test. *Communications in Statistics-Simulation and Computation*, 24(4):879–888.
- [16] Sherman, B. (1950). A random variable related to the spacing of sample values. *The Annals of Mathematical Statistics*, 21(3):339–361.
- [17] Stephens, M. A. (1970). Use of the kolmogorov-smirnov, cramér-von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 115–122.
- [18] Tracey, J. A., Zhu, J., and Crooks, K. (2005). A set of nonlinear regression models for animal movement in response to a single landscape feature. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(1):1–18.
- [19] Vanak, A., Thaker, M., and Slotow, R. (2010). Do fences create an edge-effect on the movement patterns of a highly mobile mega-herbivore? *Biological Conservation*, pages 2631–2637.
- [20] Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., and Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415.
- [21] Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, pages 109–114.