❐   155

A comparison on two discordancy tests to detect outlier in von mises (VM) sample

Fatin Najihah Badarisam, Adzhar Rambli, Mohammad Illyas Sidik
Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, Malaysia

Article Info	ABSTRACT
	This paper focuses on comparing two discordancy tests between robust and non-robust statistic to detect a single outlier in univariate circular data. So far, to the best author knowledge that there is no literature make a comparison between both tests of RCDu Statistic and $G1$ Statistic. The test statistics are based on the circular median and spacing theory. In addition, those statistics can detect multiple and patches outliers. The performance tests of RCDu Statistic and $G1$ Statistic are tested in outlier proportion of correct detection, masking and swamping effect. At the beginning stage, we obtained the cut-off points for the RCDu Statistic and $G1$ Statistic by applying Monte Carlo simulation studies. Then, generated sample from von Mises (VM) with the combination of sample size and concentration parameter. The estimating process of cut-off points for both statistics is repeated 3000 times at 10%, 5% and 1% upper percentiles. As a result, the RCDu Statistic perform well in detecting a correct single outlier. Moreover, the RCDu Statistic has a lower masking rate compared to $G1$ Statistic. However, the $G1$ Statistic is better than RCDu Statistic for swamping effect due to a lower swamping rate. Thus, RCDu Statistic performs better than $G1$ Statistic in detecting a single outlier for von Mises (VM) sample. As an illustration, both statistics were applied to the real data set from a conducted experiments series to investigate the northen cricket frogs homing ability. *Copyright © 2020 Institute of Advanced Engineering and Science.* *All rights reserved.*

Corresponding Author:

Fatin Najihah binti Badarisam,
Faculty of Computer and Mathematical Science,
Universiti Teknologi MARA (UiTM),
40450 Shah Alam, Selangor Darul Ehsan, Malaysia.
Email: fatin010475@gmail.com

1. INTRODUCTION

Univariate data is described as a dataset with single variable. For instance, weight of cat, length of rope and others. Circular data is defined as points that lie on the circumference of a circle. It can be measured in principal circular measuring instruments of compass and clock. For example, the typical observations of flight passengers' arrival time to the airport is measured by the clock (on a 24-hour clock), the wind and migrating birds' directions data are measured by a compass. Circular data is usually measured in degrees and radians [1]. Besides, linear data is the data that can be represented on a real line such as stopping distance of cars (km), depth of water (m) and length of rope (cm). Linear data differ from circular data in sense it cannot be measured in a periodical sample space [2].

The existence of outlier in circular data is one of the problem that need to be considered. "Outlier" is defined as extremely high or low data value as compared to the rest of the data values [3]. While [4] defined outlier as observations that are not consistent with the rest of the data set and appear out of line are more descriptively an "outright liar".

The existence of the outlier in the data is a common problem in the statistical analysis and can affect the estimation of parameter and sometimes it may be interesting for certain studies. Some studies would retain

the outlier and others would remove it from the data. The outlier could happen because of various factors such as the typing, human or measurement error. Some outliers might happened due to the natural phenomena occurrence. So, the appropriate statistical method is needed in order to detect the outlier according to the nature of the data. The outlier can exist in linear, circular or other type of data.

For instance, in natural science studies, an experiment on migrating bird direction have been conducted. The results find out that from 10 birds, only one bird was migrating too far than others. Then, it would be very interesting to find out the reasons. So, it is important to detect and retain the outlier because it could lead to new scientific discoveries. However, if the aim of the studies is about to explain the overall pattern of the population, then removing the outlier and redoing analysis without them is a good idea. It would help in producing good results and interpretations.

There are many statistical methods that can be used to detect the outlier in linear data. For instance, [5] proposed box and whisker plot while [6] introduced Grubbs test and [7] introduced Mahalanobis distance. However, there are only few literature that concern about detecting the outlier in circular data [8]. Previous study, [9] proposed M Statistic to detect the outlier in circular data. Next, [10] introduced statistics of C and D while A Statistic is introduced by [11]. The latest method is based on circular median and known as RCDu Statistic has been applied by [12]. The test of RCDu Statistic basically used robust technique RCDu Statistic and it can detect the multiple outliers. The test of RCDu Statistic performance test have been compared with the other statistics of M, A and Chord. Moreover, the Ga Statistic coined by [13] have advantages in detecting the patches of outliers and Ga Statistic have compared with the D, C and A Statistic. From all of these methods, the superior statistical test to detect the outlier in circular data are Ga Statistic and RCDu Statistic. So far, to the best author knowledge that there is no literature make a comparison between both tests of RCDu Statistic and Ga Statistic. In addition, RCDu Statistic and Ga Statistic can detect multiple and patches outliers

Next, there are many distributions that can be used to detect the outlier in linear and circular data. The distributions for linear data are Gamma, Cauchy, Exponential, Chi-square and Normal distributions. While the distribution for circular data are Wrapped Normal, Wrapped Cauchy (WC), Cardiod and von Mises (VM) distributions. The use of VM distribution is to describe the circle prediction, confidence interval and sampling distribution on estimation of parameter. The VM or other known as circular normal distributions is the continuous probability distributions on a circle.

The procedure of outlier identification is its robustness against two misclassification error which are masking and swamping effect. Masking effect is an outlier is classified as "non-outlier" and swamping effect is "non-outlier" is classified as an outlier. Unfortunately, for every process of identifying masking and swamping effect, it can be interfered by the outlier itself. Moreover, in identifying the outlier, it is depend on the measure of the robustness. The robustness measures are i) proportion of correct detection to detect the outlier, ii) masking and iii) swamping rate. In [12] stated that in the robustness literature, this measure of robustness is very popular in order to evaluate the particular method to detect the outlier. The discordancy tests should have a good robustness property of high proportion of correct detection, and low masking and swamping rate.

Median known as robust central location parameter. Robust method is an approach to overcome the outlier problem by reducing the outlier weightage. Median will not or less affected, if there is a suspicious observation in the data. Therefore, in this study, the RCDu Statistic is used to detect the outlier in univariate circular data. The RCDu Statistic is based on robust approach which is used a circular median and it will be more efficient if the outlier exists in the circular data [12]. Next, the second method that will be used in this study is Ga Statistic. This method is based on spacing theory [13] and can be defined as non-robust method. This method detect the outlier by measuring the gap between one observations to another observations.

In this paper, we are focused on detecting a single outlier in univariate circular data that following VM distribution by using Monte Carlo simulation studies. The performance of RCDu Statistic and Ga Statistic in proportion of correct detection to detect the outlier, masking and swamping rate are investigated. Both test statistics are compared in order to determine the best method between them. Then, the RCDu Statistic and Ga Statistic will be applied in the real data set to detect the outlier.

2. RESEARCH METHOD
2.1. The von mises distribution (VM)

The VM distribution is an analogue to the normal distributions on the line and introduced by [14]. This distribution denoted by VM (μ,κ). The density function of VM is given by:

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_o(\kappa)} exp\ [\kappa\ cos\ (\theta - \mu)],\ 0 < \theta, \mu \leq 2\pi, \kappa \geq 0 \tag{1}$$

where μ is a mean direction, κ is a concentration parameter and I_o is the modified Bessel function of the first kind and order 0, can be estimated by:

$$I_o(\kappa) \approx \frac{1}{2\pi} \int_0^{2\pi} exp(\kappa cos\theta) d\theta \tag{2}$$

Below are summarization properties of density in VM [14]:
(i) symmetrical about the mean direction μ,
(ii) has mode at μ, and
(iii) has antimode at $\mu \pm \pi$.

The range of the concentration parameter, κ is between 0 and ∞. Thus, the higher the value of concentration parameter, κ the less the dispersion and the circular sample will be concentrated towards mean, μ.

2.2. Robust circular distance *(RCDu)* statistic

The method of *RCDu Statistic* has been proposed by [15]. The single or multiple outlier in circular data can be identified by using this method. There are two main points in this method: first, the extreme values in circular data would not be the outlier and second, the symmetric in mean direction is the essential property of the von Mises distribution (VM). However, [16] cited that when the outlier exist in the circular data, the circular median is better than the mean direction. According to Mahmood [12] and He, X [17] suggested that when the data do not follow von Mises (VM) distribution, the circular median is more robust than the mean direction. Therefore, in order to detect single and multi-outliers, the circular distance between any observation and circular median can be used as a statistic. Let circular observations that placed on around of a unit circle as $\vartheta 1, \vartheta 2,\ldots, \vartheta n$. There are possible ways to apply proposed procedure in calculating the circular distance (i) between θi and the circular median, med. The properties are as follows:

i) If $0 \leq med \leq \pi$

$$dist_{(i)} = \begin{matrix} |\vartheta_i - med| & if \ |\vartheta_i - med| \leq \pi \\ 2\pi - \vartheta_i + med & if \ |\vartheta_i - med| > \pi \end{matrix} \tag{3}$$

ii) If $0 \leq med \leq 2\pi$

$$dist_{(i)} = \begin{matrix} |\vartheta_i - med| & if \ |\vartheta_i - med| \leq \pi \\ 2\pi - \vartheta_i + med & if \ |\vartheta_i - med| > \pi \end{matrix} \tag{4}$$

The $_{(i)}$ is expected to be relatively large if ϑ_i is an outlier. Therefore, the *RCDu* value is given by $RCDu$ = max $(dist)$. Consequently, if the $_{(i)}$ exceeds the cut-off points, the ϑ_i is identified as an outlier. The proposed method will be assessed depending on these two measures of robustness: i) masking rate and ii) swamping rate. The consideration of a good robustness properties for any outlier detection statistic are high proportion of outlier detected, and low masking and swamping rates.

2.2.1. Masking rate

According to Barnett [18] defined masking as to mask the extreme observations that do not considered as outliers. According to Wang [19] stated that masking is the undetected outlier because the existence of the adjacent ones [20]. In [21], they stated the effect of masking happened when the estimated mean and covariance have been skewed by the outlying observations cluster and the result of outlying point distance from the mean is small.

2.2.2. Swamping rate

According to Maimon [21] stated that swamping is the second observation swamped by the first outlier that have been declared as outlier. In other words, the "non-outlier" is classified as an outlier. They also stated when the swamping occurs, the distance result from the other non-outlying instances to the mean is large and they are potential look like outliers. This is due to the mean and the covariance estimates have been skewed toward and away from other non-outlying instances by a group of outlying instances. However, the robustness properties of masking and swamping rates are required for detecting more than one outlier. But, for this study, the robustness properties is used for detecting a single outlier in von Mises (VM) sample.

2.3. Spacing theory

According to Pyke [22] have a good review in a spacing theory. From previous literature, [13] have discussed and developed new discordancy test of G_a *Statistic* based on spacing theory by applying VM distribution. The (iid) circular observations placed on around of a unit circle have been given as $\theta_1, \theta_2,\ldots, \theta_n$ while $\theta_{(1)}, \theta_{(2)},\ldots, \theta_{(n)}$ are the sequenced circular observations. Then, the notation from [13, 23] have defined the ith ordered observations of one-step spacing as:

$$G_{ai} = {}_{(i+a)} - \theta_i, \ i = 1,2..\,, n, \quad \text{and} \tag{5}$$

$$G_{1n} = 2\pi - (n) + \theta(1) \tag{6}$$

note that $\{G_{ai}, i = 1, 2,\ldots, n\}$ provide distances order between observations on the circumference. Moreover, the choice of zero direction won't affected the order. So, for the ith sequenced observations of a-step spacing, $a = 1, 2, 3\ldots$ and $i = 1, 2,\ldots,$ n where:

$$G_{ai} = {}_{(i+a)} - \theta_i, \ i = 1, 2,\ldots, n - \text{a and} \tag{7}$$

$$G_{ai} = 2\pi - {}_{(ai)} + \theta_{(i+a)-n} \text{ for } i = (n + 1) - a, \ (n+2) - a,\ldots, n \tag{8}$$

this paper used the statistic (7) in implementation of the G_a *Statistic* to detect a single outlier (G_1) in univariate circular data.

2.4. The *Ga Statistic*

Assume (iid) circular observations as $\theta_1, \theta_2,\ldots, \theta_n$ from VM distribution. According to Mohamed [13] described the G_a *Statistic* steps as follows: At first, the $_{(1)}, \theta_{(2)},\ldots, \theta_{(n)}$ of observations need to be arranged in ascending order. Next, calculate G_{ai}, i = 1,…, n as in (7) and (8) for a a-step spacing. Then, define $G_i = \min$ $(G_{ai}, G_{a,-a})$ for i = 1, 2,…,n which is the smaller of a-step spacing on both side of θ_i. Lastly, define $G_a = \max_{i=1,2,\ldots,n} (G_i)$ and the ith observations similar to $\max_{i=1,2,\ldots,n} (G_i)$ will be identified as outlier if the G_a value is larger than the cut-off points, C_g. Since this paper only detected a single outlier, the a-step spacing that will be used is 1-step spacing. Therefore, 1-step spacing in G_a *Statistic* will be known as G_1 *Statistic*. Thus, the G_1 *Statistic* will be used as discordancy test to detect outlier in VM sample.

2.5. Cut-off points for the *RCDu Statistic* and *G_1 Statistic*

The two discordancy tests of *RCDu Statistic* and G_1 *Statistic* are used to detect the outlier in VM sample. The generated samples are the observations of *(n-1)* from VM (α, κ) and another observation from VM $(\alpha + \lambda\pi, \kappa)$, where λ is degree of contamination and the λ value is between 0 and 1. Then, the sample is generated based on different sample sizes of n between 10 and 100 and the value of concentration parameter, κ lies in range 2 to 10. The values of sample size, n and concentration parameter, κ are based on the previous research in [13]. This is due to ensure that the cut-off point values for *RCDu Statistic* and G_1 *Statistic* are significant and tally with the combination of sample size, n and concentration parameter, κ.

Besides, the Monte Carlo simulation studies is used to obtain the cut-off points for *RCDu Statistic* and G_1 *Statistic*. The cut-off points for G_1 *Statistic* is referred [13] and the new cut-off points for *RCDu Statistic* is obtained. The SPlus statistical package has been used in processing the estimation of the percentage points for both statistics. The process is repeated 3000 times at 10%, 5% and 1% upper percentile.

The results of cut-off point for *RCDu Statistic* is tabulated in Table 1 respectively. This tabulated values can be used as cut-off points for *RCDu Statistic*. The cut-off points of *RCDu Statistic* showed that the cut-off points increases, as sample sizes increases for all combinations of concentration parameter and upper percentile. This is due to the large distance between the observations and circular median. However, the cut-off points decreases, as concentration parameter increases for all combination of sample sizes and upper percentile. This indicated where there is a small distance between the observations and circular median.

The cut-off points for G_1 *Statistic* is referred to [13]. Based on [13], Monte Carlo simulation studies is used and the process is replicated 4,000 times to obtain the cut-off points. The combinations of sample size and concentration parameter that have been used for G_1 *Statistic* are same as [12] for *RCDu Statistic*. As a result, [13] find out that the cut-off points decreases as the sample sizes increases for all combination of concentration parameter and upper percentile. This is due to the large spacing between each observations. However, the cut-off points decreases as the concentration parameter increases for all sample sizes and upper percentile. This can be indicated that there is a small spacing between each observation.

Table 1. The cut-off points for the *RCDu Statistic*

		κ							
n	level	2.01	2.86	5.29	5.84	6.53	7.42	8.61	10.27
10	10%	2.57	1.96	1.25	1.17	1.10	1.01	0.94	0.87
	5%	2.79	2.28	1.38	1.30	1.22	1.11	1.03	0.95
	1%	3.01	2.83	1.69	1.63	1.45	1.33	1.27	1.13
20	10%	2.85	2.23	1.38	1.29	1.20	1.11	1.01	0.93
	5%	2.98	2.57	1.51	1.42	1.31	1.22	1.11	1.01
	1%	3.09	3.00	1.86	1.70	1.57	1.45	1.28	1.18
30	10%	2.91	2.46	1.46	1.37	1.28	1.16	1.06	0.97
	5%	3.01	2.76	1.60	1.52	1.38	1.26	1.14	1.05
	1%	3.10	3.02	1.85	1.76	1.63	1.47	1.35	1.22
50	10%	3.01	2.68	1.54	1.43	1.34	1.23	1.15	1.02
	5%	3.07	2.91	1.69	1.55	1.44	1.32	1.24	1.09
	1%	3.12	3.08	2.02	1.81	1.69	1.53	1.45	1.28
100	10%	3.08	2.90	1.69	1.55	1.46	1.33	1.21	1.09
	5%	3.11	3.02	1.82	1.67	1.58	1.42	1.29	1.16
	1%	3.13	3.11	2.17	1.96	1.86	1.60	1.46	1.33

2.6. Performance tests for the *RCDu Statistic* and G_1 *Statistic*

A good test should have three important things which are high power function, high probability of identifying a contaminating value and low probability of wrongly identifying a good observation as discordance, [24, 25]. Therefore, the performance tests of *RCDu Statistic* and G_1 *Statistic* are tested in proportion of correct detection, masking and swamping rate. The performance test for both statistics will be considered good if *RCDu Statistic* and G_1 *Statistic* follow these properties i) proportion of correct detection is equal to one, ii) masking and iii) swamping rate are equal to zero. It can be indicated that both statistics are expected to perform well in detecting the outlier if the proportion of correct detection equal to one. Moreover, there will be no outlier is undetected if the masking rate is equal to zero. In addition, there is no "non-outlier" is classified as an outlier for both statistics of swamping rate equal to zero.

Next, the performance tests of *RCDu Statistic* and G_1 *Statistic* will be conducted with the combination of concentration parameter, κ and sample size, *n* at 5% upper percentile. The 5% upper percentile is used because it is based on the previous research conducted by [12, 13]. Besides, 5% upper percentile also is commonly used by the other researchers in order to test the statistic performance. Those performance tests of *RCDu Statistic* and G_1 *Statistic* are illustrated in Figure 1-6.

The *RCDu Statistic* and G_1 *Statistic* have a better performance in proportion of correct detection to detect outlier at all contamination level when sample size, n = 10 and fixed concentration parameter, $\kappa = 5.29$. In proportion of correct detection to detect outlier, the *RCDu Statistic* and G_1 *Statistic* are performed well with large value of concentration parameter, $\kappa = 10.27$ and fixed sample size, n = 50 at all level of contamination. For the combination of fixed concentration parameter, $\kappa = 5.29$ and all sample sizes, n = 10, 50 and 100, the *RCDu Statistic* is more powerful than G_1 *Statistic* in detecting a single outlier. For the case of combination between fixed sample size, n = 10 and 50 and all concentration parameter, $\kappa = 2.84$, 5.29 and 10.27, the *RCDu Statistic* is performed better than G_1 *Statistic*. As compared to G_1 *Statistic*, the *RCDu Statistic* outperform masking effect. This is because the masking rate of *RCDu Statistic* is close to zero and can be indicated that there is no an outlier undetected. The findings found that the G_1 *Statistic* is performed well than *RCDu Statistic* in swamping effect. This is due to the swamping rate of G_1 *Statistic* is lower than *RCDu Statistic*.
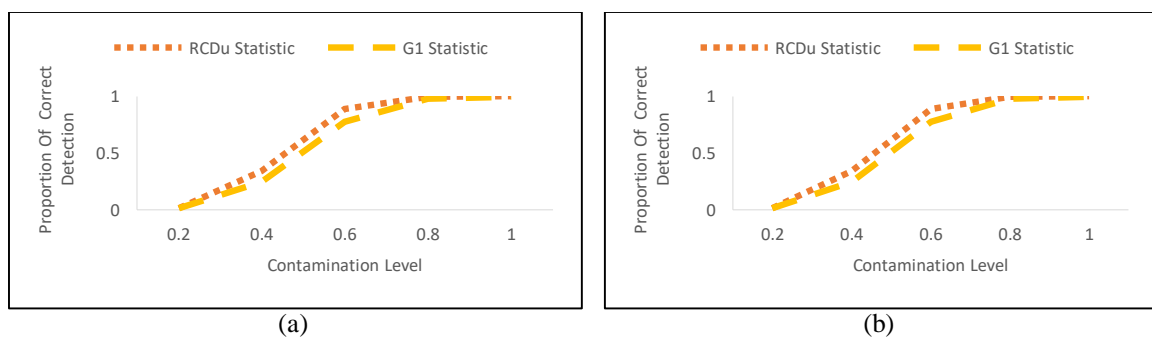

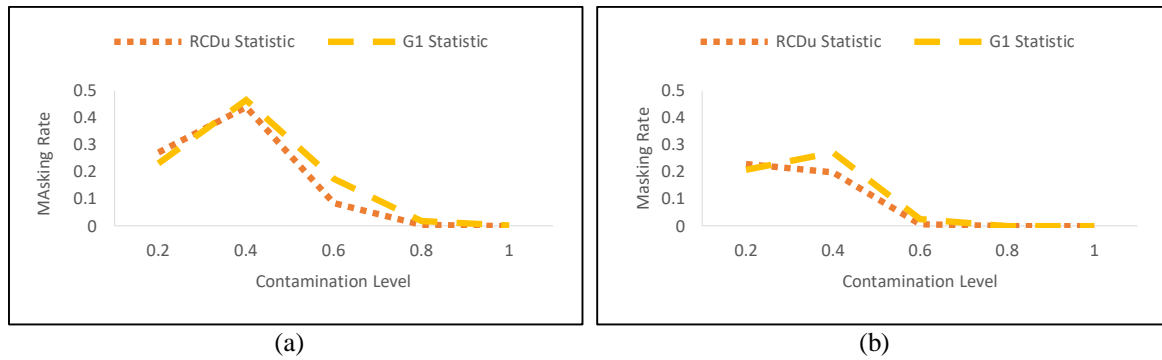Figure 1. Comparison performance of *RCDu* and G_1 statistics in proportion of correct detection, (a) κ = 5.29 and n = 10, (b) κ = 10.27 and n = 50

Figure 2. Comparison performance of *RCDu* and G_1 statistics in masking rate, (a) κ = 5.29 and n = 10, (b) κ = 10.27 and n = 50
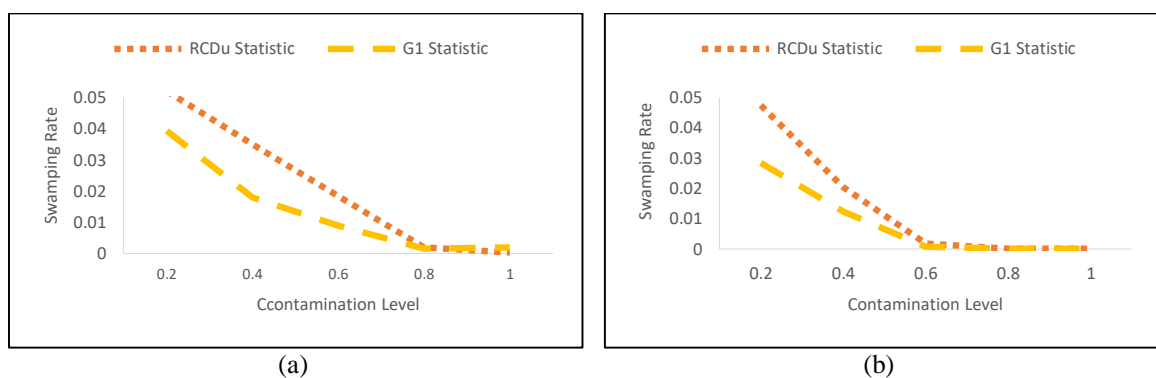

Figure 3. Comparison performance of *RCDu* and G_1 statistics in swamping rate, (a) κ = 5.29 and n = 10, (b) κ = 10.27 and n = 50

3.   RESULTS AND ANALYSIS

This paper used secondary data from a series experiment of homing ability of the northern cricket frogs defined by [26]. The data represent measurement of 14 frogs' directions that have been released from a dark environmental chamber. Several plots have been plotted to illustrate the distribution of data set. The direction of mean and median frog data are 145.974° and 140.5°. Both direction values seem close to each other due to the direction values are between 140° and 145°. The circular variance and standard deviation of frog data are 0.275 and 45.931°. The estimated concentration parameter is around 2 where the data seem to be not highly concentrated. The linear plot and circular plot are shown in the Figure 4 and Figure 5. Both plot clearly show that the observation 14th is a suspected to be an outlier due to this observation is not consistent with the rest of the observations. The circular histogram of frog data is displayed in Figure 6. The direction of mean is showed on the solid straight line and the arched line is the mean direction of 95% confidence interval. The data that close to the mean direction is the mode data while the observation that suspected to be an outlier located opposite the direction of mean bar.

Next, further test is conducted where the frog data need to be applied in the *RCDu Statistic* and G_1 *Statistic* to test either the observation 14th is classified as an outlier or not. The frog data contained 14 sample sizes and new estimated concentration parameter, κ from von Mises (VM) distribution is obtained. The new estimated concentration parameter, κ is 2.17 and it is depends on the sample size. Therefore, due to the sample size is less than 20, so the concentration parameter is greater than 2. Then, the new cut-off points with new estimated concentration parameter, κ = 2.17 is measured in order to test the performance in proportion of correct detection to detect outlier for *RCDu Statistic* and G_1 *Statistic*. The new 5% significance level of cut-off points with estimated concentration parameter, κ = 2.17 for both statistics are 2.861934 and 1.677554. And then, the estimated proportion of correct detection for *RCDu Statistic* and G_1 *Statistic* are 3.063053 and 2.024582 which those values exceed the cut-off points. Thus, the observation 14th is classified as an outlier. In the frog data set, the two discordancy tests of *RCDu Statistic* and G_1 *Statistic* have identified the observation 14th is

classified as an outlier. Therefore, by this findings, we recommend to the biologist to apply further test in order to know whether the frog of observation 360º is happened due to the certain possible factors. The possible factors that might be happened are the human or measurement error. Sometimes the presence of outlier is due to the environment factor. For instance, flood, climate change and ecological disaster. The propose method in handling the outliers that can be used by the biologist is the robust method. The special characteristic of robust method will not affect the analysis and estimation values when the outliers exist in the data set.
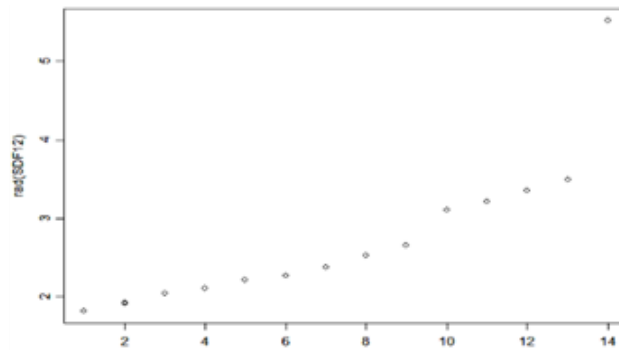

Figure 4. Linear plot of frog data
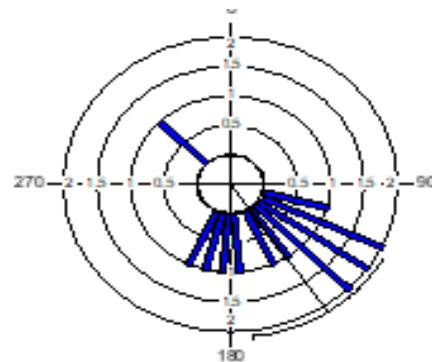

Figure 5. Circular plot of frog data

Figure 6. Circular histogram of frog data

4. CONCLUSION

This paper focused on comparing two discordancy tests between *RCDu Statistic* and *G_1 Statistic* to detect a single outlier in univariate circular data that follow von Mises (VM) distribution. The *RCDu Statistic* and *G_1 Statistic* are based on the circular median and spacing theory. The new cut-off points for *RCDu Statistic* is obtained through Monte Carlo simulation studies and cut-off points for *G_1 Statistic*. The process to obtain the cut-off points is replicated by 3000 times. Each simulation processes have the combination of concentration parameter, κ = 2.84, 5.84, 10.27 and sample size, n = 10, 50, 100. Those combination are used in order to compare the performance of outlier proportion of correct detection, masking and swamping rate between tests of *RCDu Statistic* and *G_1 Statistic*. The tests statistic of *RCDu Statistic* and *G_1 Statistic* will be indicated as a good test if the proportion of correct detection is 1 and rate of masking and swamping are 0. The higher the outlier proportion of correct detection, the better the test statistic. The lower the masking and swamping rate, the better the test statistic. As a result, via simulation studies, the generated sample from von Mises (VM) distribution is depending on the sample size, n and concentration parameter, κ. Next, for the performance of *RCDu Statistic* and *G_1 Statistic*, both statistics have the similar performance in proportion of correct detection to detect outlier when the value of contamination level is more than 0.8. Despite, they are not performing well to detect outlier in proportion of correct detection for large sample size and low concentration parameter. The *RCDu Statistic* and *G_1 Statistic* are performed better when sample size, n is small and concentration parameter, κ is high. In addition, the results of comparing the performance of *RCDu Statistic* and *G_1 Statistic* are discussed.

The *RCDu Statistic* is better than G_1 *Statistic* for von Mises (VM) sample for all value of sample sizes, *n* and concentration parameter, κ. This is due to the *RCDu Statistic* perform well in detecting a correct single outlier. Moreover, the *RCDu Statistic* has a lowest masking rate compared to G_1 *Statistic.* However, the G_1 *Statistic* is better than *RCDu Statistic* for swamping effect due to a lowest swamping rate. Thus, this can be indicated that *RCDu Statistic* is better than G_1 *Statistic* due to *RCDu Statistic* meet two properties of a good performance to detect outlier in von Mises (VM) sample. Finally, throughout this study, the *RCDu Statistic* and G_1 *Statistic* are applied in the real data set of frog to detect any possible outlier. The distribution of frog data is measured by using Watson's U^2 Test. It is indicated that the frog data follow von Mises (VM) distribution. As a result, the observation 14th is declared as an outlier by using the *RCDu Statistic* and G_1 *Statistic*. Hence, it can be concluded that *RCDu Statistic* and G_1 *Statistic* were able to detect a single outlier in univariate von Mises (VM) sample.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Mardia, K. V. and Jupp, P. E., "Directional Statistics," *John Wiley & Sons*, London, 2000.
[2]   Mulder, K., & Klugkist, I, "Bayesian estimation and hypothesis tests for a circular Generalized Linear Model.,"*Journal of Mathematical Psychology*, pp. 4-14, 2017.
[3]   Bluman, Allan, "Elementary Statistics," *brief version, New York*: McGraw-Hil, 2000.
[4]   F. Wilcoxon, "Personal Communication to J. K. Brewer in Class Presentation for Statistics 405," *Florida State University*, Tallahassee, 1962.
[5]   Tukey, J. W., "Exploratory Data Analysis," *Addison-Wesley*, Reading, MA, 1977.
[6]   Grubbs, F. E., "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1-21, 1969
[7]   P. C. Mahalanobis, "On tests and measures of group divergence," *Journal of the Asiatic Society of Bengal*, vol. 26, pp. 541–588, 1930.
[8]   Mulder, K., & Klugkist, I., "Bayesian estimation and hypothesis tests for a circular Generalized Linear Model," *Journal of Mathematical Psychology*, pp. 4-14, 2017.
[9]   Mardia, K. V., "Statistics of directional data," *Journal of the Royal Statistical Society*, Series B, vol. 37, pp. 349-393, 1975.
[10] Collett, D., "Outliers in circular data," *Applied Statistics*, vol. 29, no. 1, pp. 50-57, 1980.
[11] Abuzaid, A. H., "Some problems of outliers in circular data. PhD thesis: Institute of Mathematical Science," University of Malaya, 2010.
[12] Mahmood, A. E., Rana, S., Midi, H., & Hussin, A. G., "Detection of Outliers in Univariate Circular Data using Robust Circular Distance," *Journal of Modern Applied Statistical Methods*, vol. 16, no. 2, pp. 418-438, 2017.
[13] Mohamed, I. B., Rambli, A., Khaliddin, N., & Ibrahim, A. I. N., "A new discordancy test in circular data using spacings theory," *Communications in Statistics – Simulation and Computation*, vol. 45, no. 8, pp. 2904-2916, 2015.
[14] Von Mises, R., "Uber die "Ganzzahligkeit" der atomgewichte und vewandte fragen. *Physikal. Z.*, vol. 19, pp. 490-500, 1918.
[15] Jammalamadaka, S. R. and SenGupta, A., "Topics in Circular Statistics," World Scientific Press, Singapore, 2001.
[16] Ducharme, G. R., & Milasevic, P., "Some asymptotic properties of the circular median," *Communications in Statistics-Theory and Methods*, vol. 16, no. 3, pp. 659- 664, 1987.
[17] He, X., & Simpson, D. G., "Robust direction estimation," *The Annals of Statistics*, vol. 20, no. 1, pp. 351-369, 1992.
[18] Barnett, V & Lewis, T, "Outliers in Statistical Data," *3rd Edition, Wiley,* New York, 1994.
[19] Wang, S., & Serfling, R., "General Foundations for Studying Masking and Swamping Robustness of Outlier Identifiers," *Department of Mathematics*, University of Texas at Dallas, Richardson, Texas, 2013.
[20] Chiang, J. T., "The Masking and Swamping Effects Using the Planted Mean-Shift Outliers Models," *Int. J. Contemp. Math. Sciences*, vol. 7, no. 2, pp. 297-307, 2007.
[21] Maimon, O., & Rockach, L., "Outlier Detection," Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, 2005.
[22] Pyke, R., "Spacings," *Journal of the Royal Statistical Society*: Series B, vol. 27, no. 3, pp. 395– 449, 1965.
[23] Rao, J. S., "Some Contributions to the Analysis of Circular Data," Ph.D. thesis, Indian Statistical Institute, Calcutta, India, 1969.
[24] David, H. A., "Order Statistics," *New York and London*: Wiley, 1970.
[25] Barnett, V & Lewis, T., "Outliers in Statistical Data," *New York: John Wiley & Sons*, 1984.
[26] Ferguson, D. E., Landreth, H. F. and Mckeown, J. P., "Sun compass orientation of the northern cricket frog, Acris crepitans," *Animal Behaviour*, vol. 15, no. 1, pp. 45-53, 1967.