ISSN 1996-3343

Asian Journal of Applied Sciences

http://knowledgiascientific.com

Knowledgia SCIENTIFIC
A Place to Publish Outstanding Research

OPEN ACCESS Asian Journal of Applied Sciences

ISSN 1996-3343
DOI: 10.3923/ajaps.2017.126.133

Research Article
Robust Circular Distance and its Application in the Identification of Outliers in the Simple Circular Regression Model

[1]Ehab A. Mahmood, [1,2]Habshah Midi, [3]Sohel Rana and [4]Abdul Ghapor Hussin

[1]Department of Mathematics, University Putra Malaysia, Jalan Upm, 43400 Serdang, Selangor,Malaysia
[2]Institute for Mathematical Research, University Putra Malaysia, Jalan Upm, 43400 Serdang, Selangor, Malaysia
[3]Department of Applied Science, East West University, Dhaka, Bangladesh
[4]Faculty of Defence Science and Technology, National Defence University of Malaysia, Jalan Upm, 43400 Serdang, Selangor, Malaysia

Abstract

Background and Objective: The existence of outliers in any type of data influences the efficiency of an estimator. Few methods for detecting outliers in a simple circular regression model have been proposed in the study but it suspected that they are not very successful in the presence of multiple outliers in a data set. This study aimed to investigate new statistic to identify multiple outliers in the response variable in a simple circular regression model. **Materials and Methods:** The proposed statistic is based on calculating robust circular distance between circular residuals and circular location parameter. The performance of the proposed statistic is evaluated by the proportion of detected outliers and the rate of masking and swamping. The simulation study is applied for different sample sizes at 10 and 20% ratios of contamination. **Results:** The results from simulated data showed that the proposed statistic has the highest proportion of outliers and the lowest rate of masking comparing with some existing methods. **Conclusion:** The proposed statistic is very successful in detecting outliers with negligible amount of masking and swamping rates.

Key words: Circular regression, outliers, masking, swamping

Received: January 30, 2017 Accepted: March 27, 2017 Published: June 15, 2017

Citation: Ehab A. Mahmood, Habshah Midi, Sohel Rana and Abdul Ghapor Hussin, 2017. Robust circular distance and its application in the identification of outliers in the simple circular regression model. Asian J. Applied Sci., 10: 126-133.

Corresponding Author: Ehab A. Mahmood, Department of Mathematics, University Putra Malaysia, Jalan Upm, 43400 Serdang, Selangor, Malaysia

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The simple circular regression model is one of the circular models that is proposed to represent the relationship between two circular variables. This model can be used in many scientific fields. For example, in studying bird migration, the study is interested to observe wind direction and flight direction of the birds or in medical studies, some circular variables are recorded for the study of vector cardiograms. Sometimes, it is interested to predict one variable given other. This can be done by simple circular regression model. However, the existence of outliers can cause a huge effect of the statistical analysis and the final outcomes. In real-life applications, samples from any field might include noise or outliers. Outlier is an observation which appears inconsistent (extreme) with the other observations in the statistical data and effect on the results. There are often two problems with methods of detecting outliers: 'Masking' and 'Swamping' problems. Masking is the inability of the procedure to detect correct outliers and swamping is the identification of inliers as outliers[1]. It is now evident that the presence of outliers causes misleading conclusions to be drawn from the results. Thus, researchers are interested in improving the ways of detecting outliers in statistical data. Many researchers have proposed methods to identify outliers in linear regression model. However, there are only few methods in the study that develop methods of detecting outliers in a simple circular regression model.

Jammalamadaka and Sarma[2] proposed a regression model when both the response and the explanatory variables are circular. Downs and Mardia[3] suggested a regression model in which both the response and the explanatory variables are circular with means β and α, respectively. Hussin *et al.*[4] extended the previous models and suggested a simple circular regression model when both the response and the explanatory variables are circular variables. Kato *et al.*[5] suggested another regression model in the case when both the response and the explanatory variables are circular and assumed that the angular error follows a wrapped Cauchy distribution. Hussin *et al.*[6] extended the COVRATIO statistic, which is used to detect outliers in the linear regression model to the detection of outliers in the functional relationship model. Abuzaid *et al.*[7] suggested using the COVRATIO statistic to detect outliers in the response variable in a simple circular regression model by using a row deletion approach. Rambli[8] achieved many objectives in his studies by developing procedures for identifying outliers in circular regression models by using COVRATIO and the mean circular error statistic DMCEs that was proposed by Abuzaid[9]. Abuzaid *et al.*[10] also proposed the mean circular error statistic DMCEc to identify outliers in the response variable of a simple circular regression model by using a row deletion approach. Later, Abuzaid[11] compared the performance of COVRATIO statistic for a Simple Circular (SC) regression model and a Complex Linear (CL) regression model. It was found that the COVRATIO statistic performs better for the SC model than for the CL model. Hussin *et al.*[12] proposed a complex linear regression model to fit the circular data by using the complex residuals to detect any possible outliers.

In this study, a new approach is proposed to identify outliers in the response variable in a simple circular regression model.

MATERIALS AND METHODS

Simple circular regression model: Hussin *et al.*[4] proposed a simple circular regression model when both the response variable y and the explanatory variable x are circular variables and there is a linear relationship between them; their model is given in Eq. 1:

$$y_i = \alpha + \beta x_i + \varepsilon_i \,(\mathrm{mod}\ 2\pi) \tag{1}$$

where, α and β are the parameters and ε is the circular random error, which follows the von Mises distribution with a circular mean μ and concentration parameter k. It is obvious that the angles ϑ and ϑ+2π give the same point on the circle. The von Mises distribution is called the normal distribution for the circular data. Let $\vartheta_1, \vartheta_2, \ldots\ldots, \vartheta_n$ follow the von Mises distribution with mean direction μ and concentration parameter k, which can be denoted by [vM(μ,k)], then the probability density function of the von Mises distribution is given in Eq. 2[13]:

$$g(\vartheta,\mu,k) = \frac{1}{2\pi I_0(k)} e^{k\cos(\vartheta-\mu)} \tag{2}$$

where, I_0 denotes the modified Bessel function of the first kind and order zero, which can be defined as:

$$I_0(k) = \frac{1}{2\pi}\int_0^{2\pi} e^{k\cos(\vartheta)}\, d\vartheta$$

The maximum likelihood estimates of the model parameters are given in following Eq. 3-6[12]:

$$\hat{\alpha} = \begin{cases} \tan^{-1}(s/c) & \text{if } s>0, c>0 \\ \tan^{-1}(s/c) + \pi & \text{if } c<0 \\ \tan^{-1}(s/c) + 2\pi & \text{if } s<0, c>0 \end{cases} \tag{3}$$

where, $s = \sum \sin(y_i - \hat{\beta}x_i);\ c = \sum \cos(y_i - \hat{\beta}x_i)$

$$\hat{\beta}_1 \approx \hat{\beta}_0 + \frac{\sum x_i \sin(y_i - \hat{\alpha} - \hat{\beta}_0 x_i)}{\sum x_i^2 \cos(y_i - \hat{\alpha} - \hat{\beta}_0 x_i)} \quad (4)$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i (\mathrm{mod}\, 2\pi) \quad (5)$$

$$\hat{k} = A^{-1}\left(\frac{\sum \cos\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)}{n}\right) \quad (6)$$

where, $A^{-1}(\omega) \approx \frac{9 - 8\omega + 3\omega^2}{8(1-\omega)}$, with A(.) being the ratio of the modified Bessel function of the first kind and of order one to that of the first kind and of order zero.

Conventional methods for the detection of outliers in a simple circular regression model

COVRATIO statistic: Abuzaid *et al.*[7] proposed COVRATIO statistic to detect outliers in the response variable of a simple circular regression model. This statistic is based on the covariance matrix of the simple circular regression model. The COVRATIO statistic is given in Eq. 7:

$$\mathrm{COVRATIO}_{(-i)} = \frac{\left|\mathrm{COV}_{(-i)}\right|}{\left|\mathrm{COV}\right|} \quad (7)$$

where, |COV| is the determinant covariance matrix of coefficients for the full data set:

$$\left|\mathrm{COV}\right| = \frac{1}{\hat{k}A(\hat{k})}$$

and $|COV_{(-i)}|$ is the determinant covariance matrix of coefficients for the reduced data set formed by excluding the i-th row:

$$\left|\mathrm{COV}_{(-i)}\right| = \frac{1}{\hat{k}_{(-i)}A(\hat{k}_{(-i)})}$$

The i-th observation is identified as an outlier if $|COVRATIO_{(-i)}-1|$ exceeds the cut-off point.

Mean circular error statistic: Abuzaid[9] and Abuzaid *et al.*[10] suggested two statistics, the DMCEs and DMCEc statistics, to identify outliers in the response variable y in a simple circular regression model.

DMCEs statistic: Abuzaid *et al.*[10] proposed to use sine function as a measure of mean circular error to identify outliers, where sin is an increasing function on the interval [0, π/2]. The mean circular error is given as follows:

$$\mathrm{MCEs} = \frac{1}{n}\sum \sin\left(\frac{d_i}{2}\right)$$

where, $d_i = \pi - \left|\pi - \left|y_i - \hat{y}_i\right|\right|$ is the circular distance between y_i and $\hat{y}_i$, MCEs∈[0,1]. The existence of outliers is expected to increase the value of MCEs and the removal of outlier decreases the value of MCEs. Thus, the statistic to detect outliers is given in Eq. 8:

$$\mathrm{DMCEs}_{(i)} = |\ \mathrm{MCEs} - \mathrm{MCEs}_{(-i)}| \quad (8)$$

where, $MCEs_{(-i)}$ is MCEs with the i-th observation removed. The cut-off point represents the maximum absolute difference between the value of the statistic for the full data and the reduced data set (formed by excluding the i-th observation) which shown in Eq. 9:

$$\mathrm{cut\ DMCEs} = \max|\mathrm{MCEs} - \mathrm{MCEs}_{(-i)}| \quad (9)$$

The i-th observation is identified as an influential observation if $DMCEs_{(i)}$ is greater than the cut-off point.

DMCEc statistic: Abuzaid[9] proposed to use cosine function as an alternative measure of mean circular error. This statistic is given by:

$$\mathrm{MCEc} = 1 - \frac{1}{n}\sum \cos(y_i - \hat{y}_i)$$

where, MCEc ∈[0,2]. If y_i is an outlier then the circular distance between y_i and $\hat{y}_i$ is expected to be relatively large. Hence, the existence of outlier in a data set will increase value of MCEc. Consequently, the removal outlier will decrease the value of the statistic. The statistic to identify outlier is given in Eq. 10:

$$\mathrm{DMCEc}_{(i)} = |\ \mathrm{MCEc} - \mathrm{MCEc}_{(-i)}| \quad (10)$$

where, $MCEc_{(-i)}$ is MCEc with the i-th observation removed. The cut-off point is the maximum absolute difference between the value of the statistics for the full data set and the reduced data sets as obtained in Eq. 11:

$$\text{cut DMCEc} = \max|\text{MCEc}-\text{MCEc}_{(-i)}| \quad (11)$$

If $DMCEc_{(i)}$ is greater than the cut-off point, the i-th observation is detected as an outlier.

Proposed robust circular distance for Y, RCD_y: The distance between two circular observations is completely different from the distance between two linear observations as circular data are in angular form. The maximum distance between two circular observations cannot be more than π in a circle. For example, if $\vartheta i = 350^\circ$ and $\vartheta_j = 10^\circ$ the difference between them is equal to 340°, but this is not the true circular distance. It can be seen from the hypothetical Fig. 1 that $\vartheta_i = 350^\circ$ and $\vartheta_j = 10^\circ$ are not far from each other. It is obvious that their distance is equal to 20°.

It is important to mention that the von Mises distribution is symmetric around the circular mean. Jammalamadaka and SenGupta[14] defined circular distance (cd) in Eq. 12:

$$cd = \pi - \left|\pi - \left|\vartheta_i - \vartheta_j\right|\right| \quad (12)$$

where, ϑ_i and ϑ_j are circular observation. It is anticipated that any observations in which their circular residuals lie far away from their circular mean can be considered as outliers. However, the circular mean is not robust against outliers. Hence, it is proposed using circular median in this regard. This issue has motivated us to formulate a new statistic to detect outliers in simple circular regression model, by employing circular median instead of circular mean. It is called this statistic robust circular distance, denoted as RCD_y.

The following steps are applied to compute the proposed robust circular distance (RCD_y): First, calculate the absolute value of the estimated circular residuals:

$$\widehat{cr}\left(\widehat{cr}_i = \pi - \left|\pi - \left|y_i - \hat{y}_i\right|\right|\right)$$

Note that it cannot calculate the difference between y_i and $\hat{y}_i$ directly as it is done with the linear method, because of the circular geometry theory. Second, compute the robust circular distance $(RCD_{(i)})_y$ between $\widehat{cr}_i$ and circular median of:

$$\widehat{cr}\left[RCD_i\right]_y = \pi - \left|\pi - \left|\widehat{cr} - \text{med}\left(\widehat{cr}\right)\right|\right|$$

Then any y_i is suspected to be an outlier if its corresponding $[RCD_i]_y$ is relatively large. This is due to the fact that if $y_{(i)}$ is an outlier, it affects on the value of

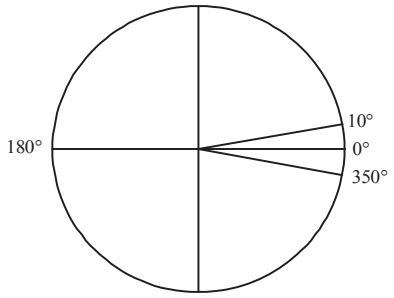

Fig. 1: A hypothetical figure for the distance between two circular observations

$\widehat{cr}$. Subsequently, the circular distance between $\widehat{cr}_i$ and circular median of $\widehat{cr}$ is relatively large. Hence, the cut-off point should be the maximum value of the $[RCD]_y$ defined in Eq. 13:

$$\text{Cut } RCD_y = \max\,[RCD]_y \quad (13)$$

Any circular data that $[RCD_i]_y$ greater than the cut-off point is declare as an outlier.

RESULTS AND DISCUSSION

Simulated cut-off point of the RCD_y statistic: A simulation study is designed to determine the cut-off points (percentage points) of the null hypothesis for the distribution of no outliers in circular data of the proposed statistic. The same procedure has been used by Pearson and Hartley[15] and Collett[16] to determine cut off points in their studies. For any data set of sample size n and concentration parameter k, it will be rejected the null hypothesis if the value of computed statistic is larger than the cut-off points, which suggested that there is an outlier in the data set. In this simulation study, first, it is considered 20 different sample sizes of n = 10, 20, 30, ..., 200 and 9 values of concentration parameter k = 2, 3, 5, 6, 8, 10, 12, 15, 20. Second, generate a set of explanatory variables X of size n, such that:

$$\left[x \sim \text{vM}\left(\frac{\pi}{4}, 10\right)\right]$$

Third, generate a set of circular random errors, such that [e~vM(0, k)] for each sample size n and concentration parameter k. Next, fix the initial values of the parameters in the model at $\alpha = 0$, $\beta = 1$. Then, calculate the values of the response variable Y and fit the generated circular data. The RCD_y statistic is then computed and its maximum value is determined. These processes are replicated 5000 times for

each combination of sample size n and concentration parameter k. The 10 and 5% upper percentile values of the maximum RCD_y are determined as shown in Table 1 and 2, respectively. The priority is to calculating these values that they can be used as cut-off points of the RCD_y statistic to identify outliers of the simple circular regression model according to the sample sizes and values of the concentration parameter.

It is noticeable from Table 1 and 2 that the cut-off point is an increasing function of the sample size n for any value of the concentration parameter k. This is reasonable, because when sample size is increased, the data will be more spread out. Consequently, the circular distance between them and the circular mean increases. Moreover, the cut-off point is a decreasing function of the concentration parameter k for any sample size n. This is because increasing the concentration parameter causes the concentration of the circular data around the circular mean to increase.

Performance of the RCD_y statistic: The performance of the RCD_y, COVRATIO, DMCEs and DMCEc statistics are examined

Table 1: Cut-off points of RCD_y with 10% upper percentile

	k								
n	2	3	5	6	8	10	12	15	20
10	2.02	1.35	0.843	0.797	0.637	0.586	0.495	0.434	0.396
20	2.25	1.62	0.968	0.890	0.725	0.664	0.576	0.485	0.460
30	2.42	1.90	1.15	1.00	0.850	0.746	0.662	0.582	0.511
40	2.48	1.96	1.20	1.06	0.876	0.772	0.686	0.611	0.535
50	2.50	2.11	1.26	1.11	0.918	0.800	0.722	0.631	0.552
60	2.52	2.23	1.32	1.14	0.940	0.826	0.736	0.655	0.564
70	2.54	2.30	1.36	1.16	0.966	0.842	0.755	0.668	0.573
80	2.55	2.36	1.39	1.19	0.990	0.857	0.773	0.680	0.584
90	2.55	2.41	1.40	1.22	0.995	0.868	0.798	0.695	0.591
100	2.56	2.43	1.42	1.24	1.00	0.888	0.797	0.707	0.594
110	2.56	2.44	1.44	1.25	1.02	0.895	0.797	0.711	0.603
120	2.57	2.46	1.49	1.27	1.04	0.900	0.814	0.714	0.615
130	2.57	2.48	1.52	1.27	1.04	0.910	0.819	0.720	0.622
140	2.57	2.50	1.52	1.29	1.05	0.916	0.828	0.731	0.624
150	2.58	2.52	1.54	1.32	1.06	0.927	0.836	0.740	0.629
160	2.58	2.54	1.55	1.33	1.07	0.933	0.840	0.743	0.634
170	2.58	2.55	1.56	1.33	1.08	0.943	0.845	0.745	0.640
180	2.58	2.56	1.57	1.34	1.08	0.950	0.853	0.747	0.644
190	2.59	2.57	1.59	1.34	1.09	0.957	0.857	0.753	0.646
200	2.59	2.58	1.61	1.36	1.11	0.969	0.861	0.763	0.650

Table 2: Cut-off points of RCD_y with 5% upper percentile

	k								
n	2	3	5	6	8	10	12	15	20
10	2.31	1.64	0.983	0.919	0.726	0.667	0.573	0.498	0.450
20	2.43	2.02	1.18	1.01	0.878	0.754	0.662	0.546	0.505
30	2.53	2.22	1.31	1.12	0.943	0.826	0.738	0.639	0.557
40	2.55	2.32	1.37	1.16	0.977	0.842	0.755	0.657	0.574
50	2.57	2.40	1.42	1.23	1.01	0.883	0.794	0.690	0.600
60	2.58	2.46	1.50	1.25	1.03	0.900	0.804	0.713	0.611
70	2.59	2.48	1.53	1.28	1.05	0.923	0.833	0.725	0.623
80	2.60	2.55	1.55	1.33	1.07	0.941	0.845	0.740	0.635
90	2.61	2.56	1.57	1.34	1.08	0.946	0.872	0.745	0.642
100	2.62	2.58	1.60	1.35	1.09	0.957	0.867	0.769	0.649
110	2.62	2.58	1.63	1.37	1.12	0.966	0.861	0.769	0.654
120	2.61	2.60	1.66	1.40	1.14	0.974	0.878	0.769	0.664
130	2.61	2.61	1.68	1.41	1.14	0.982	0.878	0.774	0.669
140	2.61	2.61	1.70	1.43	1.15	0.987	0.888	0.779	0.675
150	2.61	2.62	1.72	1.45	1.15	0.995	0.902	0.788	0.679
160	2.61	2.62	1.73	1.44	1.16	1.00	0.910	0.805	0.682
170	2.61	2.63	1.74	1.43	1.17	1.02	0.912	0.802	0.686
180	2.61	2.64	1.76	1.46	1.17	1.02	0.915	0.805	0.694
190	2.62	2.64	1.78	1.47	1.18	1.03	0.916	0.806	0.694
200	2.62	2.65	1.81	1.50	1.19	1.04	0.920	0.806	0.700

by using Monte Carlo simulations. Four different sample sizes are used, namely n =10, 50, 100 and 150 and six concentration parameters, k = 2, 3, 5, 6, 8 and 10. The data are contaminated in the response variable Y according to the following formula in Eq. 14:

$$y_{cont} = y_{clean} + \lambda\pi \bmod(2\pi) \quad (14)$$

where, λ is the degree of contamination, such that ($0 \leq \lambda \leq 1$). If $\lambda = 0$, there is no contamination. If $\lambda = 1$, the circular observation is located at the anti-mode of its initial location.

For all combinations of sample sizes and concentration parameters, it is generated 10 and 20% contaminated data with $\lambda = 0.8$. To evaluate the performance of all the statistics, three measures are considered namely, the proportion of outliers detected, the masking and the swamping rates. The processes are replicated 5000 times for each combination of sample size n and concentration parameter k. In each time of replication, it is observed that the number of detected true outliers (generated). Then the proportion of outliers are calculated as follows:

$$\text{Proportion of outliers} = \frac{\text{Sum of detected true outliers}}{P \times n \times 5000}$$

where, P is percentage of contamination. Similarly, to calculate the rate of masking and the rate of swamping, it is observed that the number of generated outliers detected as inlier (clean observation) and the number of inlier detected as outlier, respectively as follows:

$$\text{Rate of masking} = \frac{\text{Sum of detected outliers as inliers}}{P \times n \times 5000}$$

$$\text{Rate of swamping} = \frac{\text{Sum of detected inliers as outliers}}{(n - (P \times n)) \times 5000}$$

A good method is one that has the highest detection rate for the outliers and low masking and swamping rates. Figure 2 and 3 exhibit the proportion of outliers detected and the rate of masking and swamping with 5% upper percentile for n = 50 and 100. The results of Fig. 2 and 3 showed that the rates of swamping are zero or close to zero for all statistics. However, the rates of masking of COVRATIO statistic are very high and the proportions of outliers detected are very low, for all combinations of sample sizes, concentration parameters and percentage of outliers. It was noticed that the proportion of outliers detected of the MCEs statistic is low when the concentration parameter is less than 6 with 10%

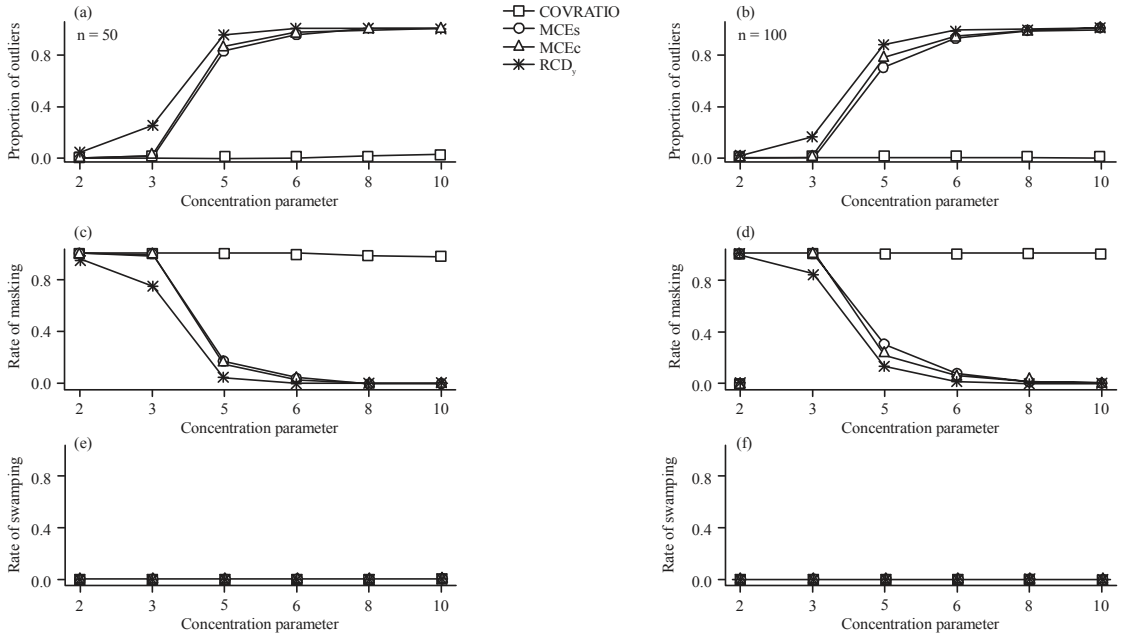

Fig. 2(a-f): Proportion of outliers detected and rate of masking and swamping with 5% cut-off points and 10% contamination

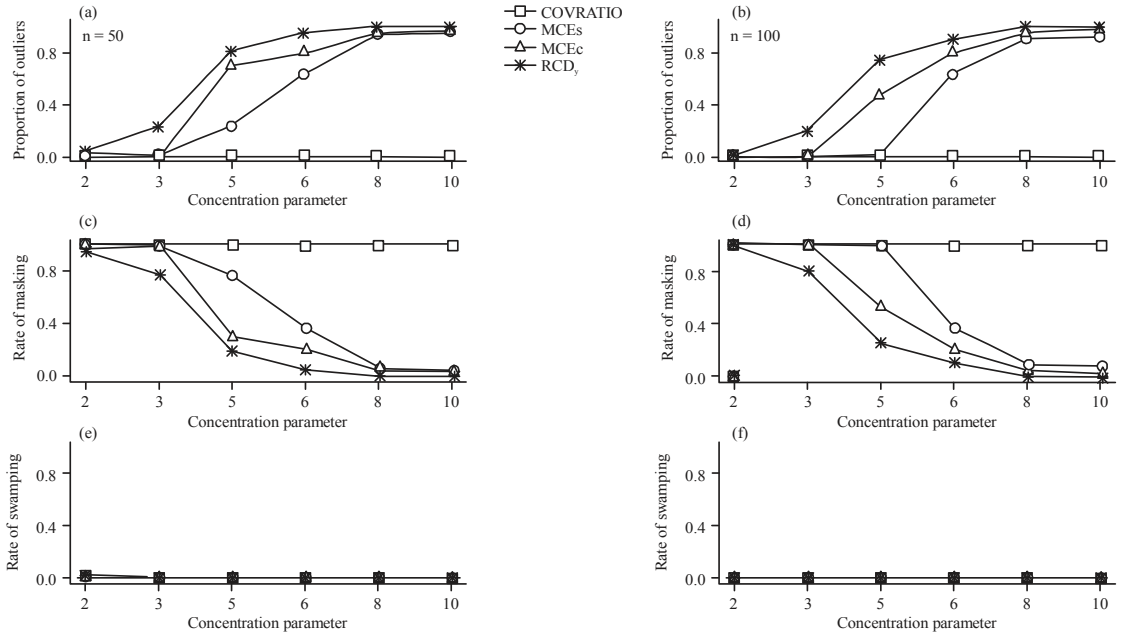

Fig. 3(a-f): Proportion of outliers detected and rate of masking and swamping with 5% cut-off points and 20% contamination

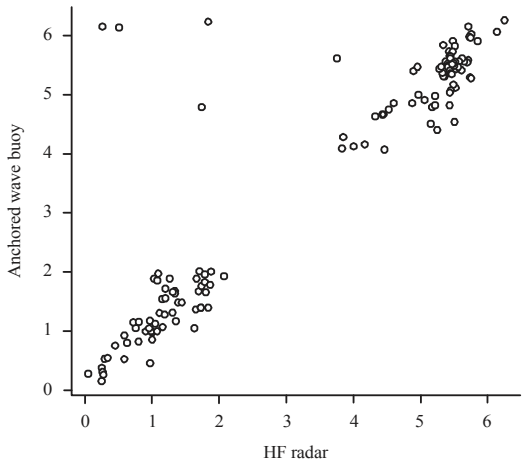

Fig. 4: Scatter plot of the wind direction data

contamination and this proportion significantly decrease with 20 % contamination. Consequently, it has high rate of masking. The MCEc statistic relatively has a higher proportion of outliers detected than the MCEs statistic and the proportion increases with the concentration parameter but it is low at 20% contaminated. The proportion of outliers detected of the proposed RCD_y statistic is relatively low for small value of k. This is acceptable because the circular data will be more spread around the circumference of the circle when the concentration parameter is low. Consequently, it is very difficult to identify outliers in this case[16]. As expected, the RCD_y statistic gives a greater proportion of outliers detected than the other statistics. The proportion is an increasing function of the concentration parameter and increases to 100% for values of the concentration parameter greater than 5. Therefore, the rate of masking is very low and is a decreasing function of the concentration parameter, decreasing down to 0%.

In general, the proposed RCD_y statistic is very successful in the detection of outliers because the circular median is one of the robust location parameter in a circular data. For these reasons, the RCD_y statistic is the best when compared to the other three measures. It has the highest proportion of outliers detected and the lowest rates of both masking and swamping. Due to space constraints, the results for n = 10 and 50 are not shown. However, the results are consistent.

Practical example: The wind direction data that is studied by Abuzaid *et al.*[7] is considered. The data represent measurements by using an HF radar system and an anchored wave buoy with sample size (n = 129). Figure 4 showed the scatter plot of the wind direction data, where X are the circular observations measured by the HF radar and Y are the circular observations measured by the anchored wave buoy. The estimated concentration parameter is $\hat{k} = 7.34$. By referring to Table 2 with 5% upper percentile, the cut-off point is equal to 1.28. The RCD_y statistic is calculated and the results are

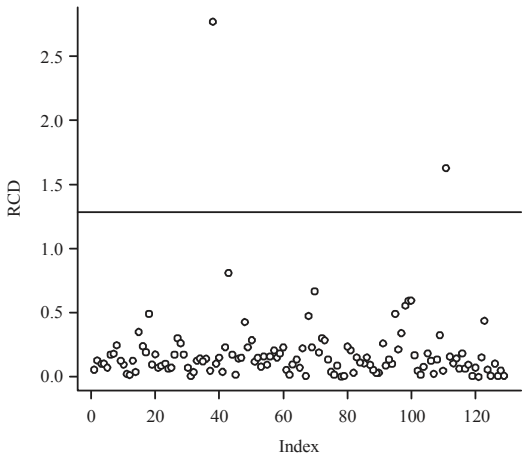

Fig. 5: $[RCD_i]_y$ statistic of the wind direction data

plotted in Fig. 5. It can be seen that the observations numbered 38 and 111 exceed the cut-off point. Therefore, they are considered as outliers. These results are in agreement with the results of Abuzaid *et al.*[7]. Abuzaid *et al.*[7] pointed out that the two points at the left top of the plot in Fig. 4 are not outliers. They are consistent with the rest of the observations at the right top or left bottom because of the closed range property of the circular variable.

CONCLUSION

A new statistic is proposed to detect outliers in Y in a simple circular regression model. Proportion of detection outliers and rate of masking and swamping are used to evaluate performance of the proposed method. The results showed that the proposed RCD_y statistic is very successful in identifying genuine outliers for different sample sizes and with very low rates of masking and swamping.

REFERENCES

1. Maronna, R.A., R.D. Martin and V.J. Yohai, 2006. Robust Statistics, Theory and Methods. John Wiley and Sons Ltd., Hobokon, New Jersey, USA.
2. Jammalamadaka, S.R. and Y.R. Sarma, 1993. Circular Regression. In: Statistical Sciences and Data Analysis, Matsusita, K. (Ed.)., VSP., Utrecht, pp: 109-128.
3. Down, T.D. and K.V. Mardia, 2002. Circular regression. Biometrika, 89: 683-698.
4. Hussin, A.G., N.R.J. Fieller and E.C. Stillman, 2004. Linear regression model for circular variables with application to directional data. J. Applied Sci. Technol., 9: 1-6.
5. Kato, S., K. Shimizu and G.S. Shieh, 2008. A circular-circular regression model. Statistica Sinica, 18: 633-645.
6. Hussin, A.G., A. Abu Zaid and I. Mohamed, 2009. Detection of outliers in the unreplicated linear circular functional relationship model via functional form. Proceedings of the International Conference on Nonparametric Methods for Measurement Error Models and Related Topics, May 3-5, 2009, Ottawa, Canada.
7. Abuzaid, A., I. Mohamed, A.G. Hussin and A. Rambli, 2011. COVRATIO statistic for simple circular regression model. Chiang Mai J. Sci., 38: 321-330.
8. Rambli, A., 2011. Outlier detection in circular data and circular-circular regression model. Master's Thesis, Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur.
9. Abuzaid, A.H., 2010. Some problems of outliers in circular data. Ph.D. Thesis, Faculty of Science, University Malaya, Kuala Lumpur.
10. Abuzaid, A.H., A.G. Hussin and I.B. Mohamed, 2013. Detection of outliers in simple circular regression models using the mean circular error statistic. J. Stat. Comput. Simul., 83: 269-277.
11. Abuzaid, A.H., 2013. On the influential points in the functional circular relationship models. Pak. J. Stat. Operat. Res., 9: 333-342.
12. Hussin, A.G., A.H. Abuzaid, A.I.N. Ibrahim and A. Rambli, 2013. Detection of outliers in the complex linear regression model. Sains Malaysiana, 42: 869-874.
13. Pewsey, A., M. Neuhauser and G.D. Ruxton, 2013. Circular Statistics in R. Oxford University Press, Oxford, UK.
14. Jammalamadaka, S.R. and A. SenGupta, 2001. Topics in Circular Statistics. World Scientific Publishing Company, Singapore, ISBN-13: 978-9810237783, Pages: 350.
15. Pearson, E.S. and H.O. Hartley, 1966. Biometrika Tables for Statisticians, Volume 1. 3rd Edn., Cambridge University Press, New York.
16. Collett, D., 1980. Outliers in circular data. J. Royal Stat. Soc. Ser. C (Applied Stat.), 1: 50-57.