STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, July 2025, pp 454–468.
Published online in International Academic Press (www.IAPress.org)

Enhanced Outlier Detection in Linear-Circular Regression Using Circular Distance and Mean Resultant Length

Thunchanok Chaitongdee [1], Wuttichai Srisodaphol [1,*], Oktsa Dwika Rahmashari [1], Benjawan Rattanawong [2], Khanuengnij Prakhammin [2]

[1] *Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand*
[2] *Department of Applied Statistics, Faculty of Engineering, Rajamangala University of Technology Isan Khon Kaen Campus, Khon Kaen 40000, Thailand*

Abstract In the study of outlier identification in linear-circular regression, two new methods are proposed. By calculating the circular distance of each erroneous value and using the mean resultant length for outlier identification, these methods aim to enhance the precision and reliability of outlier detection. Their effectiveness will be assessed through comprehensive simulations on datasets with and without contamination, comparing them with the previous method. Additionally, the methods were tested on real-world data, specifically wind speed and wind direction data, to further validate their practical applicability. Three metrics are used to evaluate their performance: the probability of correctly identifying all outliers, the masking effect, and the swamping effect. While occasional misclassification of inliers as outliers is possible, the results indicate that both proposed methods demonstrate strong overall performance.

Keywords Circular Distance, Linear-Circular Regression, Mean Resultant Length, Masking Effect, Swamping Effect

AMS 2010 subject classifications 62J99, 62P99

DOI: 10.19139/soic-2310-5070-2459

1. Introduction

Linear-circular regression is a model used to investigate the relationship between a linear independent variable and a circular dependent variable. Gould [1] applied this approach to study the relationship between the time elapsed from the start of the QRS cycle (as the circular dependent variable) and heart rate (as the linear independent variable). Moreover, research by Kim and SenGupta [2] and Sikaroudi and Park [3] on a mixture of linear-linear regression models for linear-circular regression has also contributed to the development of linear-circular regression, building upon the foundational work of Gould [1]. Circular data, which serves as the dependent variable in this model, is defined as data measured along the circumference of a unit circle. A key characteristic of circular data is its cyclical nature, where 0 degrees equals 360 degrees, or 0 radians equals 2π radians. Due to these unique properties, the analysis of circular data has gained popularity across various fields. In medical research, circular data has been utilized to study the curvature of the posterior cornea [4]. Additionally, research on wind direction behavior in Malaysia during monsoon seasons has been conducted using a replicated functional relationship in the von Mises distribution [5]. Circular data plays a significant role in biology, particularly in understanding animal navigation behaviors, such as pigeon homing [6], and is often used to compare two independent samples of circular data in biological studies. Linear-circular regression has also been applied in various studies. For instance, Johnson

*Correspondence to: Wuttichai Srisodaphol (Email: wuttsr@kku.ac.th). Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

and Wehrly [7] explored the application of linear-circular regression in analyzing angular-linear distributions, including the relationship between temperature and wind direction. Kouassi et al. [8] examined the relationship between satellite elevation angle (as the dependent variable) and GNSS carrier phase (as the independent variable). This application is crucial for enhancing the accuracy of location and height estimations, reducing fluctuations, and increasing the reliability of GNSS systems. Further studies have expanded the application of linear-circular regression. For example, Beyene et al. [9] analyzed the timing of ice-out events in lakes. Additionally, Rueda et al. [10] applied circular-circular regression to study the relationship between gene expression during the cell cycle in two species with different cycle lengths. Their analysis employed maximum likelihood estimation and model selection using generalized degrees of freedom, applied to gene expression data from two datasets. In ecology, Rivest et al. [11] used circular-circular regression to examine animal movement, such as tracking bison in Prince Albert National Park. In a related study, Polsen [12] explored the use of circular-circular regression to analyze wind direction data from weather stations in Texas, focusing on the relationship between wind direction at different stations and predicting wind direction distribution across time and locations. Lastly, Nazer and Ali [13] analyzed real data from 50 patients at Al-Kindi Teaching Hospital in Baghdad to study the relationship between circular time data and peak systolic blood pressure. They employed two estimation methods: Simple Circular Regression (SCR) and Nadaraya-Watson (NW).

Outliers in circular data are identified based on their position in the circle, which plays a crucial role in their detection. Unlike other data types, circular outliers are determined by magnitude and their angular location relative to the circular distribution. Several studies have explored outlier detection in circular data. For instance, Mahmood et al. [14] introduced a method for detecting outliers in univariate circular data using circular distance, with practical applications to frog movement and wind direction data. Similarly, Romroeng et al. [15] proposed new cut-off points for circular distance to enhance the accuracy of outlier detection in univariate circular data.

Outliers in circular regression models can significantly compromise prediction accuracy. Several studies have proposed methods for detecting outliers in circular regression. For example, Di et al. [16] proposed an outlier detection method using the minimum spanning tree in circular regression models. Another study applied the resultant length principle to detect outliers in circular regression models. Regarding linear-circular regression models, Sert and Kardiyen [17] introduced an outlier detection technique for linear-circular regression models using circular medians. Their method assumes that errors follow a wrapped Cauchy distribution and calculates circular distances between actual and predicted values to derive errors, which are then evaluated based on their deviation from the median of all errors. Outliers are identified using cut-off points corresponding to the 90th, 95th, and 99th quantiles. To demonstrate the practical application of their method, they analyzed real-world data collected every 10 minutes from a wind turbine's SCADA system in Turkey. The study focused on the relationship between wind speed (a linear independent variable) and wind direction (a circular dependent variable), highlighting the importance of accurate outlier detection in assessing and predicting wind turbine performance in power generation. The performance of their proposed method was evaluated using three metrics: the rate of detached outliers misclassified as inliers, the rate of detached inliers misclassified as outliers, and the rate of true detached outliers correctly identified. Results from both real and simulated datasets demonstrated that the method is highly effective, particularly when the degree of contamination is moderate to high. Furthermore, increasing the sample size and ensuring greater homogeneity in the data further improved the method's accuracy and reliability. However, its performance may be limited in small sample sizes, as the statistical measures used in the method become less reliable under such conditions.

Given the limitations of previous work, particularly its reduced reliability in small sample sizes, there is a clear need for more effective methods to detect outliers in linear-circular regression models. To address this challenge, the present study begins by examining the concepts of circular distance and mean resultant length, which are fundamental measures for assessing the dispersion of circular data. Building on these principles, we propose two new outlier detection methods for linear-circular regression: the Iterative Circular Distance of Residuals (ICDR) method and the Mean Resultant Length of Residuals (MRLR) method. Both methods leverage circular distance-based logic to improve the accuracy and robustness of outlier detection in linear-circular settings. Furthermore, this research conducts a comparative analysis of the proposed methods against the method by Sert and Kardiyen [17], evaluating their effectiveness in identifying outliers within the specific context of linear-circular regression.

The remainder of this study is structured as follows: Section 2 discusses the materials, including the linear-circular regression model, circular distance, and performance criteria. Section 3 introduces the proposed methods. Section 4 presents the results of a simulation study for both uncontaminated and contaminated data, along with findings from real data analysis. Section 5 provides a discussion of the findings, and Section 6 concludes the study.

2. Materials

2.1. Linear-circular regression

Linear-circular regression analyzes the relationship between a linear independent variable and a circular dependent variable. This approach was first introduced by Gould in 1969 [1]. The model for linear-circular regression is defined as follows:

$$y = \mu_0 + \sum_{j=1}^{k} \beta_j x_j \text{mod}(2\pi). \tag{1}$$

The circular dependent variable is represented by y, where μ_0 denotes the intercept, indicating the initial mean direction. The regression coefficient is β_j, while x_j represents the linear independent variables, and k is the number of parameters. Both μ_0 and β_j are unknown parameters that need to be estimated through the regression analysis [18].

2.2. Circular distance

Circular distance can be measured using various methods, each determining the shortest arc length between two points, a and b, on a circle. One common approach utilizes the cosine function to define the distance,

$$d(a, b) = 1 - \cos(a - b) \tag{2}$$

this approach, the distance increases monotonically as the angle between the points grows from 0 to π, taking values from 0 to 2. Another method directly calculates the arc length, given by

$$d_0(a, b) = \pi - |\pi - |a - b|| \tag{3}$$

ensuring that the maximum possible distance between any two points does not exceed π [19].

2.3. Mean resultant length

The resultant length is computed by summing the individual vectors in an angular dataset. This process involves calculating the sum of the cosine and sine components for each angle and then determining the length of the resultant vector from these summed components. Figure 1 illustrates this process: the upper image shows the circular data, while the lower image displays the resultant length. Solid lines represent vectors for each data point, while the dotted vector indicates the overall direction, with its length corresponding to the resultant length. The calculation begins by converting each angle into a vector with cosine and sine components. These components are then summed to form a composite vector, from which the resultant length is derived [20].

The calculated length serves as the basis for determining the mean resultant length, which quantifies the dispersion of angular data around the circle. The mean resultant length is a measure of dispersion, obtained by averaging all vectors in the dataset using the following formula:

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \tag{4}$$

where $\bar{C} = \frac{1}{n}\sum_{j=1}^{n} cos(\theta_j)$, $\bar{S} = \frac{1}{n}\sum_{j=1}^{n} sin(\theta_j)$ and θ_j represents the angle of each vector. The mean resultant length ranges from 0 to 1. If $\bar{R}$ is close to 1, it indicates that the angular data are highly concentrated or point in the same direction. Conversely, if $\bar{R}$ is close to 0, it suggests that the data are widely dispersed around the circle. Additionally, $\bar{R}$ is used to compute the mean direction, which represents the average orientation of the angular data in a circular system [21].

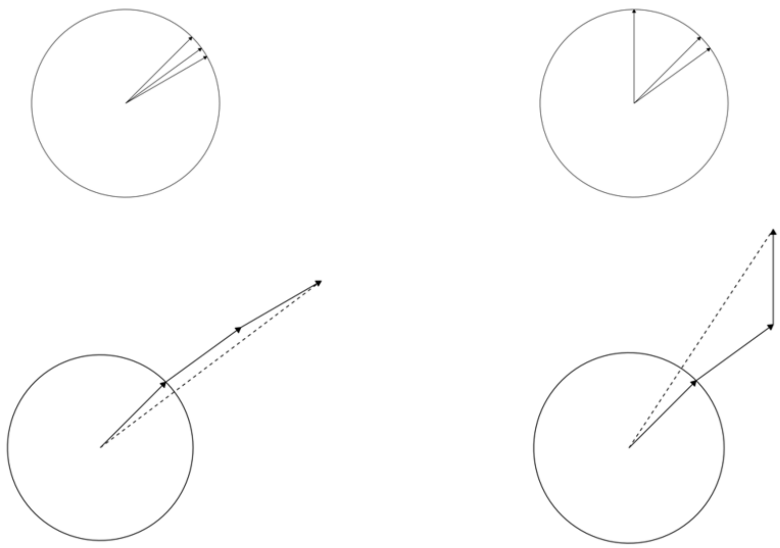

Figure 1. Visualization of the resultant length calculation for circular data.

2.4. Performance criterion

Performance is evaluated using the following three metrics [22].

1. Probability of successfully detecting all outliers ($pout$):
$$pout = \frac{n(\text{success})}{n(\text{out})} \tag{5}$$
where n(success) is the number of outliers correctly identified, and n(out) is the total number of outliers. A pout value close to 1 indicates that the method effectively detects all outliers.
2. Probability of outliers is falsely detected as inliers (masking effect, $pmask$):
$$pmask = \frac{n(\text{failure})}{n(\text{out})} \tag{6}$$
where n(failure) is the number of outliers misclassified as inliers and n(out) is the total number of outliers. A $pmaskp$ value close to 0 suggests that no outliers were mistakenly classified as inliers.
3. Probability of inliers is falsely detected as outliers (swamping effect, $pswamp$):
$$pswamp = \frac{n(\text{false})}{n - n(\text{out})} \tag{7}$$
where n(false) is the number of inliers misclassified as outliers, n is the total number of observations, and n(out) is the total number of outliers. A $pswamp$ value close to 0 indicates that no inliers were mistakenly classified as outliers.

3. Methods

3.1. Iterative Circular Distance of Residuals (ICDR)

The ICDR method uses circular distance to partition the data for outlier detection. The steps of the ICDR method are as follows:

Algorithm: ICDR Method

Inputs: y and $\hat{y}$
Outputs: y_{inlier} and $y_{outlier}$

for each i=1 to n do
  Calculate circular distance $e_i = \pi - |\pi - |y_i - \hat{y}_i||$
  Calculate $dist_{min(i)} = 1 - cos(e_i - min(e_i))$
  Calculate $dist_{max(i)} = 1 - cos(e_i - max(e_i))$
  Create matrix Group 1 with a sample size n_1 and Group 2 with a sample size n_2
    if $dist_{min(i)} < dist_{max(i)}$
      then send e_i to Group 1
      else send e_i to Group 2
    end if
end for
repeat
  for each j = 1 to n_1 do
    Fit a linear-circular regression model for Group 1
  end for
  Calculate $C = max(e_{j(group1)})$
  for each k = 1 to n_2 do
    Predict values for Group 2 and calculate residuals
    if $e_k < C$
      then send e_k to Group 1
      else send e_k to Group 2
    end if
  end for
Until Group 1 and Group 2 no longer change
for j=1 to n_1 do
  Label all remaining data points in Group 1 as y_{inlier}
end for
for k=1 to n_2 do
  Label all remaining data points in Group 2 as $y_{outlier}$
end for
Return y_{inlier} and $y_{outlier}$

3.2. *Mean Resultant Length of Residual (MRLR)*

The MRLR method uses circular distance for data partitioning and applies the mean resultant length to effectively detect outliers. It operates by iteratively refining group assignments and identifying outliers based on predefined cut-off points: 0.001, 0.005, and 0.01. The steps of the MRLR method are as follows:

Algorithm: MRLR Method

Inputs: y and $\hat{y}$
Outputs: y_{inlier} and $y_{outlier}$

for each i=1 to n do
  Calculate circular distance $e_i = \pi - |\pi - |y_i - \hat{y}_i||$

Calculate $dist_{min(i)} = 1 - cos(e_i - min(e_i))$
Calculate $dist_{max(i)} = 1 - cos(e_i - max(e_i))$
Create matrix Group 1 with a sample size n_1 and Group 2 with a sample size n_2
  if $dist_{min(i)} < dist_{max(i)}$
    then send e_i to Group 1
    else send e_i to Group 2
  end if
end for
repeat
  for each j = 1 to n_1 do
    Calculate $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$ where $\bar{C} = \frac{1}{n_1} \sum_{j=1}^{n_1} cos(e_j)$ and $\bar{S} = \frac{1}{n_1} \sum_{j=1}^{n_1} sin(e_j)$
  end for
  for each k = 1 to n_2 do
    $e_{(1)group2} = min(e_k)_{G_2}$ do
    Calculate $\bar{R}_{new} = \sqrt{\bar{C}^2 + \bar{S}^2}$ where $\bar{C} = \frac{1}{n_2} \sum_{k=1}^{n_2} cos(e_k)$ and $\bar{S} = \frac{1}{n_2} \sum_{k=1}^{n_2} sin(e_k)$
    if $\bar{R} - \bar{R}_{new} = C_R$
      then send $e_{(1)group2}$ to Group 1
      else send $e_{(1)group2}$ to Group 2
    end if
  end for
Until Group 1 and Group 2 no longer change
for j=1 to n_1 do
  Label all remaining data points in Group 1 as y_{inlier}
end for
for k=1 to n_2 do
  Label all remaining data points in Group 2 as $y_{outlier}$
end for
Return y_{inlier} and $y_{outlier}$

The algorithms for the ICDR and MRLR methods are illustrated in Figure 2 – 3, respectively.

4. Results

This research introduces the ICDR and MRLR methods for outlier detection in linear-circular regression and provides a comparative analysis with the Sert and Kardiyen method. To evaluate their effectiveness, simulation studies were conducted using both contaminated and uncontaminated datasets. The evaluation focuses on key metrics such as the masking effect, the swamping effect, and accuracy in identifying all outliers. Furthermore, the proposed methods were applied to real-world data, specifically wind speed and wind direction data, to assess their practical applicability under various conditions.

4.1. Simulation study

Simulation studies were conducted using Monte Carlo methods with 1,000 repetitions on uncontaminated and contaminated data with 1%, 5%, and 10% contamination. The independent variable was drawn from a standard normal distribution, while the dependent variable consisted of circular data generated with specified parameters and errors from the von Mises distribution (μ_0=1, β_1=3). The study focused on outlier detection in linear-circular regression models under scenarios with κ = 0.5, 1, 5, 10, and contamination degrees γ= 0.5, 0.9. Sample sizes of n= 20, 60, 100, and 200 were included. By varying these parameters, we evaluated the robustness and efficacy of outlier detection methods. Details of the dataset generation process are provided below.

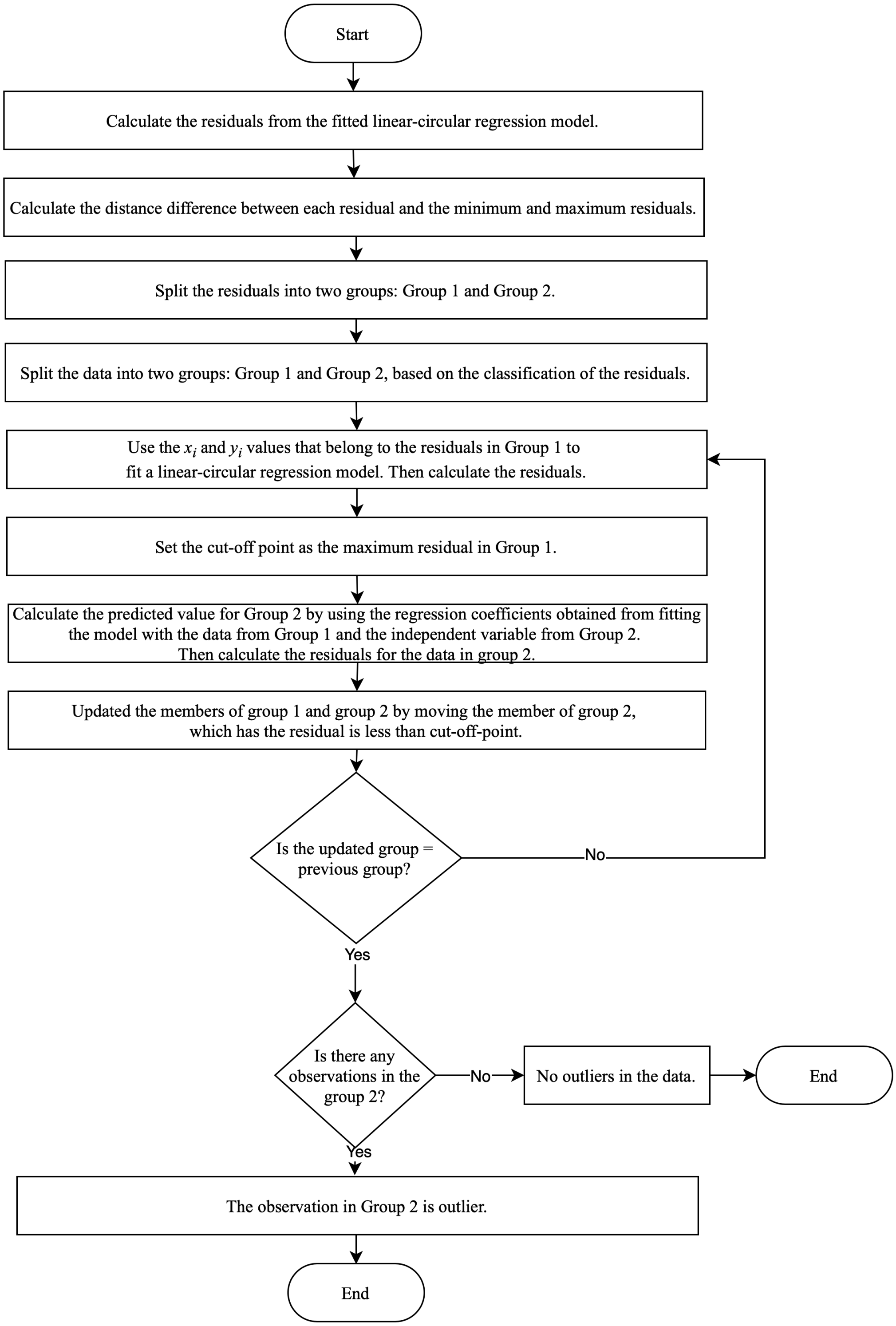

Figure 2. **T**he workflow of the ICDR method.

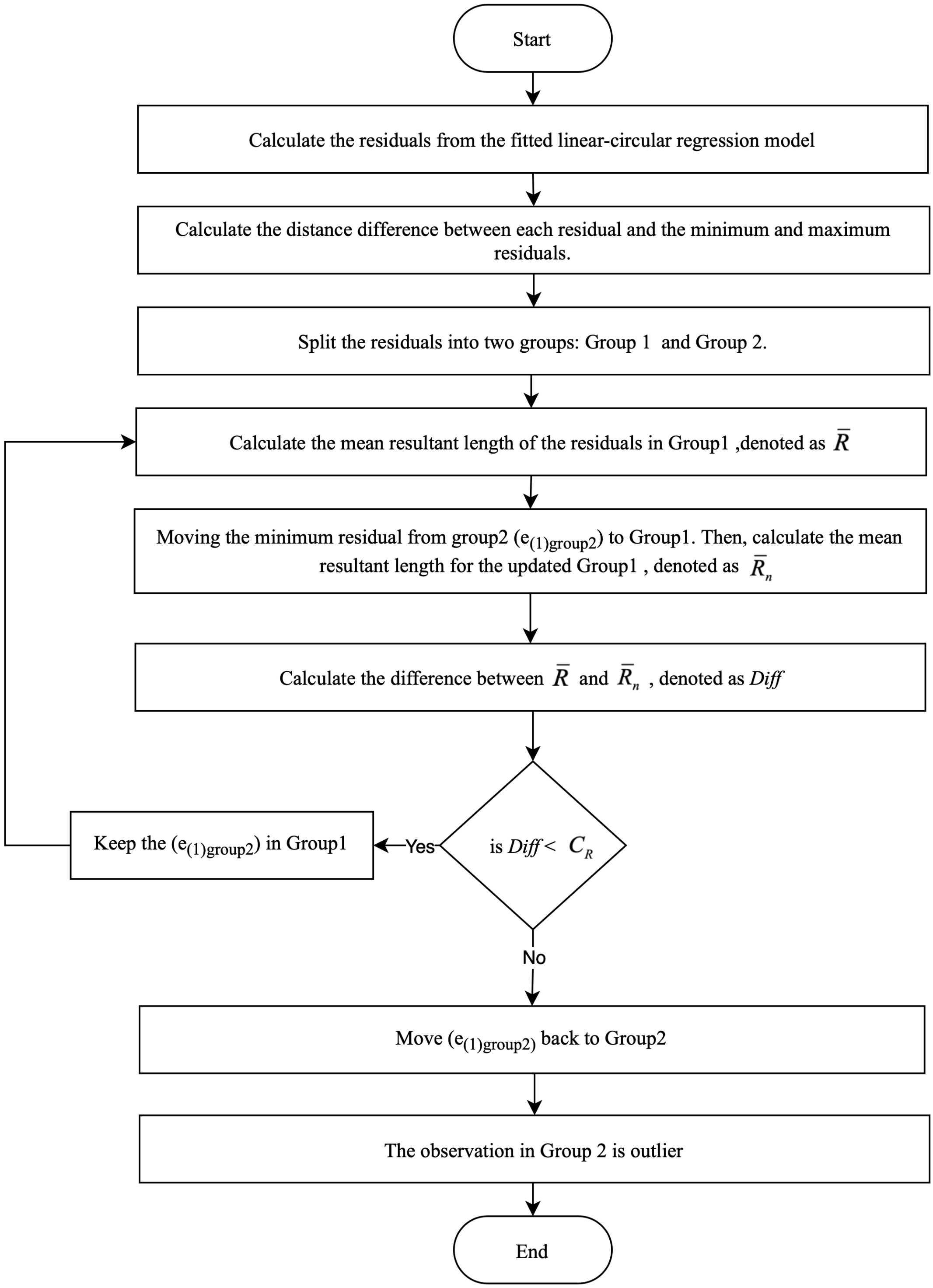

Figure 3. **T**he workflow of the MRLR method.

4.1.1. Uncontaminated data i: Simulate data for the independent variable (x_i) for $i = 1, 2, ..., n$ with sample size n from normal distribution, $x_i \sim N(0, 1)$.

ii: Simulate the residual (e_i) for $i = 1, 2, ..., n$ with sample size n from von Mises distribution, $e_i \sim vM(0, \kappa)$.

iii: Calculate the dependent variable (y_i) for $i = 1, 2, ..., n$ with parameters μ_0 and β_1, and sample size n from Eq. (8),

$$y_i = \mu_0 + \beta_1 x_i + e_i mod(2\pi) \tag{8}$$

iv: Fit the linear-circular regression model to the data, and calculate the residuals (e_i) for $i = 1, 2, ..., n$.

v: Detect outliers using the ICDR method, MRLR method, and Sert and Kardiyen method. Then, count the number of inlier samples that are incorrectly classified as outliers.

vi: Repeat Steps i.-v. 1,000 times in each value of n and κ.

vii: Calculate the $pswamp$ in each method and compare the effectiveness of the different methods.

The probability of inliers being falsely detected as outliers ($pswamp$) under uncontaminated data is presented in Table 1.

Table 1. The $pswamp$ for uncontaminated data.

κ	n	Sert & Kardiyen	ICDR	MRLR (C_R= 0.001)	MRLR (C_R= 0.005)	MRLR (C_R= 0.01)
0.5	20	0.061	0.283	0.317	0.312	0.207
0.5	60	0.079	0.248	0.280	0.279	0.270
0.5	100	0.053	0.294	0.333	0.261	0.048
0.5	200	0.057	0.304	0.340	0.036	0.000
1	20	0.081	0.224	0.224	0.243	0.227
1	60	0.062	0.211	0.224	0.170	0.058
1	100	0.057	0.212	0.223	0.169	0.058
1	200	0.053	0.208	0.221	0.058	0.000
5	20	0.068	0.242	0.248	0.068	0.027
5	60	0.056	0.163	0.144	0.014	0.002
5	100	0.054	0.134	0.087	0.004	0.000
5	200	0.053	0.102	0.033	0.000	0.000
10	20	0.068	0.255	0.200	0.022	0.003
10	60	0.056	0.174	0.063	0.001	0.000
10	100	0.056	0.148	0.027	0.000	0.000
10	200	0.053	0.118	0.005	0.000	0.000

The results in Table 1 highlight the performance of various outlier detection methods on uncontaminated data in linear-circular regression models. The method proposed by Sert and Kardiyen [17] exhibits consistently low and stable $pswamp$ values, ranging from 0.053 to 0.081. These values slightly decrease as the sample size increases, indicating the method's robustness to uncontaminated data across different concentration parameter (κ) levels.

The ICDR method, while exhibiting higher $pswamp$ values compared to Sert and Kardiyen's method, shows a gradual decline in these values as the sample size increases, indicating improved reliability with larger datasets. The MRLR method, evaluated at three different cut-off thresholds (C_R = 0.001, 0.005 and 0.01), demonstrates varying levels of sensitivity to inliers. Specifically, C_R= 0.01 achieves the lowest $pswamp$ values (ranging from 0.000 to 0.270), suggesting a reduced likelihood of misclassifying inliers as outliers. In comparison, C_R= 0.005 yields $pswamp$ values between 0.000 and 0.312, while C_R= 0.001 produces the highest $pswamp$ values (0.005 to 0.340), indicating increased sensitivity and a greater risk of false positive detections.

Overall, the MRLR method with C_R= 0.01 demonstrates the best performance for uncontaminated data. Nonetheless, the selection of detection method and parameter settings should be guided by the sample size to achieve optimal outlier detection accuracy.

4.1.2. Contaminated data i: Simulate data for the independent variable (x_i) for $i = 1, 2, ..., n$ with sample size n from normal distribution, $x_i \sim N(0, 1)$.

ii: Simulate the residual (e_i) for $i = 1, 2, ..., n$ with sample size n from von Mises distribution, $e_i \sim vM(0, \kappa)$.

iii: Calculate the dependent variable (y_i) for $i = 1, 2, ..., n$ with parameters μ_0 and β_1, and sample size n from Eq. (8).

iv: Fit the linear-circular regression model to the data and calculate the residuals (e_i) for $i = 1, 2, ..., n$.

v: Find the difference between any e_i and the $target$ when the $target = \pi - (\gamma \times \pi)$, then sort the differences between e_i and the $target$ in ascending order.

vi: Define the sorted e_i in Step v, positions 1 to the number of outliers, as outlier positions.

vii: Contaminated value of the dependent variable (y_i) by adjusting the y_i value with the contamination value from Eq. (9),

$$y_{i(adjusted)} = y_i + \gamma\pi mod(2\pi) \tag{9}$$

viii: Detect outliers using the ICDR method, MRLR method, and Sert and Kardiyen method. Then, count the n(success), n(failure), and n(false).

ix: Repeat Steps i.-viii. 1,000 times in each each scenario.

x: Calculate the $pout$, $pmask$, and $pswamp$ in each method and compare the effectiveness of the different methods.

The probability of successfully detecting all outliers $(pout)$, the probability of outliers being falsely detected as inliers $(pmask)$, and the probability of inliers being falsely detected as outliers $(pswamp)$ with 1%, 5%, and 10% contamination are presented in Tables 2–4, respectively.

Table 2. The $pout$, $pmask$, and $pswamp$ for 1% contamination.

κ	γ	n	Sert & Kardiyen			ICDR			MRLR (C_R= 0.001)			MRLR (C_R= 0.005)			MRLR (C_R= 0.01)		
			$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$
0.5	0.5	20	0.887	0.113	0.040	0.981	0.019	0.186	1.000	0.000	0.197	1.000	0.000	0.196	1.000	0.000	0.195
0.5	0.5	60	0.965	0.035	0.045	0.984	0.016	0.264	1.000	0.000	0.283	1.000	0.000	0.280	1.000	0.000	0.195
0.5	0.5	100	0.982	0.018	0.045	0.991	0.009	0.283	1.000	0.000	0.309	1.000	0.000	0.294	0.855	0.145	0.048
0.5	0.5	200	0.998	0.002	0.047	0.997	0.003	0.297	1.000	0.000	0.334	0.803	0.197	0.034	0.000	1.000	0.000
0.5	0.9	20	0.950	0.050	0.048	0.971	0.029	0.202	1.000	0.000	0.216	1.000	0.000	0.216	1.000	0.000	0.212
0.5	0.9	60	0.984	0.016	0.049	0.980	0.020	0.268	1.000	0.000	0.296	1.000	0.000	0.186	1.000	0.000	0.186
0.5	0.9	100	0.995	0.005	0.049	0.985	0.015	0.281	1.000	0.000	0.317	1.000	0.000	0.246	0.846	0.154	0.046
0.5	0.9	200	0.997	0.003	0.043	0.996	0.004	0.301	1.000	0.000	0.328	0.843	0.158	0.038	0.000	1.000	0.000
1	0.5	20	0.985	0.015	0.008	0.975	0.025	0.026	1.000	0.000	0.028	1.000	0.000	0.028	1.000	0.000	0.027
1	0.5	60	0.995	0.005	0.046	1.000	0.000	0.190	1.000	0.000	0.195	1.000	0.000	0.194	1.000	0.000	0.138
1	0.5	100	1.000	0.000	0.046	1.000	0.000	0.197	1.000	0.000	0.205	1.000	0.000	0.169	1.000	0.000	0.056
1	0.5	200	1.000	0.000	0.044	1.000	0.000	0.201	1.000	0.000	0.209	1.000	0.000	0.057	0.000	1.000	0.000
1	0.9	20	0.970	0.030	0.004	1.000	0.000	0.017	1.000	0.000	0.018	1.000	0.000	0.017	1.000	0.000	0.017
1	0.9	60	1.000	0.000	0.048	0.999	0.001	0.196	1.000	0.000	0.205	1.000	0.000	0.205	1.000	0.000	0.135
1	0.9	100	0.998	0.002	0.048	1.000	0.000	0.201	1.000	0.000	0.211	1.000	0.000	0.167	0.998	0.002	0.057
1	0.9	200	1.000	0.000	0.047	1.000	0.000	0.020	1.000	0.000	0.216	1.000	0.000	0.057	0.000	1.000	0.000
5	0.5	20	0.935	0.065	0.025	0.999	0.001	0.026	1.000	0.000	0.024	0.937	0.063	0.000	0.892	0.108	0.000
5	0.5	60	0.997	0.003	0.051	1.000	0.000	0.001	1.000	0.000	0.001	0.997	0.003	0.001	0.997	0.003	0.001
5	0.5	100	1.000	0.000	0.050	1.000	0.000	0.001	1.000	0.000	0.001	1.000	0.000	0.001	1.000	0.000	0.000
5	0.5	200	1.000	0.000	0.047	1.000	0.000	0.001	1.000	0.000	0.001	1.000	0.000	0.000	0.000	1.000	0.000
5	0.9	20	1.000	0.000	0.021	0.999	0.001	0.011	1.000	0.000	0.011	1.000	0.000	0.011	1.000	0.000	0.009
5	0.9	60	1.000	0.000	0.044	1.000	0.000	0.007	1.000	0.000	0.007	1.000	0.000	0.007	1.000	0.000	0.002
5	0.9	100	1.000	0.000	0.047	1.000	0.000	0.008	1.000	0.000	0.008	1.000	0.000	0.003	1.000	0.000	0.000
5	0.9	200	1.000	0.000	0.043	1.000	0.000	0.003	1.000	0.000	0.005	1.000	0.000	0.005	0.000	1.000	0.000
10	0.5	20	1.000	0.000	0.019	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	60	1.000	0.000	0.046	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	100	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	200	1.000	0.000	0.045	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
10	0.9	20	1.000	0.000	0.012	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	60	1.000	0.000	0.041	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	100	1.000	0.000	0.045	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	200	1.000	0.000	0.044	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000

Table 3. The $pout$, $pmask$, and $pswamp$ for 5% contamination.

κ	γ	n	Sert & Kardiyen			ICDR			MRLR (C_R= 0.001)			MRLR (C_R= 0.005)			MRLR (C_R= 0.01)		
			$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$
0.5	0.5	20	0.921	0.079	0.045	0.972	0.028	0.181	1.000	0.000	0.196	1.000	0.000	0.196	1.000	0.000	0.194
0.5	0.5	60	0.884	0.116	0.024	0.985	0.015	0.251	1.000	0.000	0.264	1.000	0.000	0.263	1.000	0.000	0.212
0.5	0.5	100	0.923	0.077	0.021	0.994	0.006	0.276	1.000	0.000	0.286	1.000	0.000	0.265	0.936	0.064	0.068
0.5	0.5	200	0.909	0.091	0.013	1.000	0.000	0.297	1.000	0.000	0.305	0.928	0.072	0.054	0.000	1.000	0.000
0.5	0.9	20	0.932	0.068	0.044	0.970	0.030	0.200	1.000	0.000	0.217	1.000	0.000	0.217	1.000	0.000	0.215
0.5	0.9	60	0.961	0.039	0.037	0.976	0.024	0.257	1.000	0.000	0.286	1.000	0.000	0.281	1.000	0.000	0.182
0.5	0.9	100	0.979	0.021	0.036	0.994	0.006	0.279	1.000	0.000	0.311	1.000	0.000	0.273	0.861	0.139	0.046
0.5	0.9	200	0.993	0.007	0.032	1.000	0.000	0.291	1.000	0.000	0.330	0.794	0.206	0.033	0.000	1.000	0.000
1	0.5	20	0.985	0.015	0.008	0.975	0.025	0.026	1.000	0.000	0.028	1.000	0.000	0.028	1.000	0.000	0.027
1	0.5	60	0.972	0.028	0.022	0.999	0.001	0.173	1.000	0.000	0.175	1.000	0.000	0.175	1.000	0.000	0.144
1	0.5	100	0.987	0.013	0.018	1.000	0.000	0.180	1.000	0.000	0.182	1.000	0.000	0.171	1.000	0.000	0.062
1	0.5	200	1.000	0.000	0.027	1.000	0.000	0.206	1.000	0.000	0.220	1.000	0.000	0.057	0.000	1.000	0.000
1	0.9	20	0.970	0.030	0.004	1.000	0.000	0.017	1.000	0.000	0.018	1.000	0.000	0.017	1.000	0.000	0.017
1	0.9	60	0.994	0.006	0.031	0.999	0.001	0.162	1.000	0.000	0.204	1.000	0.000	0.202	1.000	0.000	0.133
1	0.9	100	0.987	0.013	0.027	1.000	0.000	0.199	1.000	0.000	0.211	1.000	0.000	0.165	0.992	0.008	0.054
1	0.9	200	1.000	0.000	0.027	1.000	0.000	0.206	1.000	0.000	0.220	1.000	0.000	0.057	0.000	1.000	0.000
5	0.5	20	0.983	0.017	0.087	0.999	0.001	0.029	1.000	0.000	0.028	0.932	0.068	0.001	0.882	0.118	0.000
5	0.5	60	0.962	0.038	0.026	0.936	0.064	0.001	0.935	0.065	0.000	0.933	0.067	0.000	0.933	0.067	0.000
5	0.5	100	0.972	0.028	0.024	0.959	0.041	0.000	0.959	0.041	0.000	0.959	0.041	0.000	0.959	0.041	0.000
5	0.5	200	1.000	0.000	0.028	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
5	0.9	20	1.000	0.000	0.020	1.000	0.000	0.009	1.000	0.000	0.009	1.000	0.000	0.009	1.000	0.000	0.008
5	0.9	60	1.000	0.000	0.007	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.003
5	0.9	100	1.000	0.000	0.005	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.000
5	0.9	200	1.000	0.000	0.002	1.000	0.000	0.002	1.000	0.000	0.002	1.000	0.000	0.000	0.002	0.980	0.000
10	0.5	20	1.000	0.000	0.019	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	60	1.000	0.000	0.009	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	100	1.000	0.000	0.009	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	200	1.000	0.000	0.005	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
10	0.9	20	1.000	0.000	0.014	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	60	1.000	0.000	0.005	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	100	1.000	0.000	0.003	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	200	1.000	0.000	0.001	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000

In Tables 2-4, the simulation study assess the effectiveness of five outlier detection methods, Sert and Kardiyen, ICDR, and MRLR with three different cut-off thresholds (C_R= 0.001, 0.005, and 0.01), across varying scenarios defined by concentration parameter (κ), contamination level (γ), and sample size (n). The Sert and Kardiyen method demonstrates a conservative approach to outlier detection, particularly evident in scenarios with low concentration parameters ($\kappa = 0.5$ or $\kappa = 1$). While it achieves relatively low swamping rates, meaning it rarely misclassifies inliers as outliers, it often fails to detect all true outliers, as reflected in its lower pout and higher pmask values. As the concentration increases (κ = 5 and 10), the performance of the Sert and Kardiyen method improves, achieving perfect or near-perfect detection ($pout$=1.000) with minimal to zero false classifications.

The ICDR method achieves very high pout values (often reaching 1.000) even under difficult conditions such as low concentration parameters (e.g., $\kappa = 0.5$) and small sample sizes, highlighting its excellent ability to detect all outliers. At the same time, its low $pmask$ values indicate a minimal tendency to misclassify outliers as inliers, reflecting its high sensitivity and accuracy. However, the method does show moderately elevated $pswamp$ rates in certain settings, particularly when κ is low and n is large (e.g., $\kappa = 0.5, n = 200, pswamp$ = 0.277-0.301), implying a tendency to identify some inliers as outliers mistakenly. Even so, these swamping rates remain within reasonable bounds and generally decrease as the data becomes more concentrated.

The MRLR method, evaluated with three cut-off thresholds (C_R = 0.001, 0.005, and 0.01), demonstrates exceptionally high sensitivity in detecting outliers, as evidenced by its consistently perfect or near-perfect pout values (often 1.000) across all simulated conditions, including small sample sizes and low concentration parameters (κ). This indicates that MRLR is highly effective at identifying all true outliers regardless of data complexity. Additionally, $pmask$ values remain at or near zero, reflecting a strong capability in avoiding the misclassification of outliers as inliers, especially for C_R = 0.001 and C_R = 0.005. However, the $pswamp$ values, indicating the false classification of inliers as outliers, vary depending on the chosen cut-off. The most conservative threshold (C_R = 0.001) tends to produce higher swamping rates in scenarios with low κ and larger n, suggesting reduced specificity under such conditions. The C_R = 0.005 threshold often strikes a better balance, with slightly lower $pswamp$ rates than C_R = 0.001 while maintaining high pout and low $pmask$. Meanwhile, C_R = 0.01 exhibits more variability; in some cases (e.g., $\kappa = 0.5, n = 200$), it results in extreme outcomes such as a complete failure to detect outliers ($pout$ = 0.000, $pmask$ = 1.000), indicating its potential unreliability in complex scenarios. Therefore, MRLR is an effective method for outlier detection; however, the selection of the threshold significantly impacts its balance between sensitivity and specificity.

Table 4. The $pout$, $pmask$, and $pswamp$ for 10% contamination

κ	γ	n	Sert & Kardiyen			ICDR			MRLR (C_R= 0.001)			MRLR (C_R= 0.005)			MRLR (C_R= 0.01)		
			$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$	$pout$	$pmask$	$pswamp$
0.5	0.5	20	0.851	0.149	0.026	0.977	0.023	0.153	1.000	0.000	0.161	1.000	0.000	0.161	1.000	0.000	0.160
0.5	0.5	60	0.714	0.286	0.068	0.986	0.014	0.229	1.000	0.000	0.238	1.000	0.000	0.238	1.000	0.000	0.214
0.5	0.5	100	0.610	0.390	0.003	0.995	0.005	0.257	1.000	0.000	0.260	1.000	0.000	0.255	0.979	0.021	0.092
0.5	0.5	200	0.697	0.303	0.019	1.000	0.000	0.277	1.000	0.000	0.289	0.954	0.046	0.073	0.000	1.000	0.000
0.5	0.9	20	0.912	0.088	0.032	0.961	0.039	0.175	1.000	0.000	0.190	1.000	0.000	0.190	1.000	0.000	1.875
0.5	0.9	60	0.954	0.046	0.030	0.990	0.011	0.245	1.000	0.000	0.269	1.000	0.000	0.263	1.000	0.000	0.165
0.5	0.9	100	0.982	0.018	0.030	0.998	0.002	0.266	1.000	0.000	0.298	1.000	0.000	0.223	0.897	0.103	0.046
0.5	0.9	200	0.998	0.002	0.031	1.000	0.000	0.276	1.000	0.000	0.319	0.865	0.135	0.036	0.000	1.000	0.000
1	0.5	20	0.925	0.075	0.002	1.000	0.000	0.007	1.000	0.000	0.007	1.000	0.000	0.007	1.000	0.000	0.007
1	0.5	60	0.748	0.252	0.002	1.000	0.000	0.073	1.000	0.000	0.074	1.000	0.000	0.074	1.000	0.000	0.067
1	0.5	100	0.684	0.316	0.001	1.000	0.000	0.159	1.000	0.000	0.160	1.000	0.000	0.157	1.000	0.000	0.074
1	0.5	200	0.530	0.470	0.000	1.000	0.000	0.170	1.000	0.000	0.160	1.000	0.000	0.081	0.000	1.000	0.000
1	0.9	20	0.920	0.080	0.003	0.982	0.018	0.017	1.000	0.000	0.018	1.000	0.000	0.018	1.000	0.000	0.017
1	0.9	60	0.949	0.051	0.010	0.998	0.002	0.096	1.000	0.000	0.101	1.000	0.000	0.100	1.000	0.000	0.068
1	0.9	100	0.860	0.141	0.007	1.000	0.000	0.199	1.000	0.000	0.212	1.000	0.000	0.166	1.000	0.000	0.054
1	0.9	200	0.981	0.019	0.013	1.000	0.000	0.206	1.000	0.000	0.223	1.000	0.000	0.056	0.000	1.000	0.000
5	0.5	20	1.000	0.000	0.000	0.998	0.002	0.000	0.998	0.002	0.000	0.998	0.002	0.000	0.998	0.002	0.000
5	0.5	60	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
5	0.5	100	0.675	0.325	0.000	0.675	0.325	0.000	0.675	0.325	0.000	0.675	0.325	0.000	0.675	0.325	0.000
5	0.5	200	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
5	0.9	20	0.973	0.027	0.000	0.998	0.002	0.002	0.998	0.002	0.002	0.998	0.002	0.002	0.998	0.002	0.002
5	0.9	60	0.998	0.002	0.000	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.003	1.000	0.000	0.003
5	0.9	100	1.000	0.000	0.000	1.000	0.000	0.002	1.000	0.000	0.002	1.000	0.000	0.002	1.000	0.000	0.000
5	0.9	200	1.000	0.000	0.000	1.000	0.000	0.002	1.000	0.000	0.003	1.000	0.000	0.001	0.997	0.003	0.001
10	0.5	20	0.995	0.001	0.000	0.998	0.002	0.000	0.998	0.000	0.000	0.998	0.002	0.000	0.998	0.002	0.000
10	0.5	60	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	100	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.5	200	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
10	0.9	20	0.980	0.020	0.000	0.999	0.001	0.000	0.999	0.001	0.000	0.999	0.001	0.000	0.999	0.001	0.000
10	0.9	60	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	100	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
10	0.9	200	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000

4.2. Real Data Results

The effectiveness of the proposed methods is evaluated using real-world data from the Wind Turbine SCADA Dataset [23], which contains operational data from wind turbines in Turkey collected in 2018. This dataset includes parameters such as wind speed, wind direction, and turbine power output, recorded at 10-minute intervals. The study focuses on analyzing the relationship between wind speed (a linear variable) and wind direction (a circular variable) to assess the performance of the proposed outlier detection methods. Data collected between 6:00 and 7:00 AM on July 1–8, representing the summer season in Turkey, are selected for analysis. Since the dataset does not explicitly label outliers, artificial outliers are introduced by incorporating three records from the winter season (January 1, 2018, between 6:00 and 6:30 AM). These records were assigned to observations 57, 58, and 59. Using winter-season records as artificially introduced outliers in a summer-season dataset provides a realistic test scenario that reflects seasonal shifts in wind patterns. This setup facilitates the evaluation of the proposed methods' sensitivity to angular deviations that naturally occur over time.

The proposed ICDR and MRLR with three cut-off thresholds (C_R = 0.001, 0.005, and 0.01) methods are applied to detect outliers in the dataset. Both methods successfully identify the artificially labeled outliers (observations 57, 58, and 59) and also flag observations 1, 2, and 3 as potential outliers, demonstrating their sensitivity to deviations within the data. As illustrated in Figure 3, inlier data points are represented by yellow dots, while detected outliers appear as red dots at positions 1, 2, 3, 57, 58, and 59. Notably, only observations 57, 58, and 59 are true outliers, as they are deliberately introduced from a different season. This analysis highlights the robustness and adaptability of the proposed methods in real-world applications, particularly in scenarios where actual outliers are not explicitly identified.

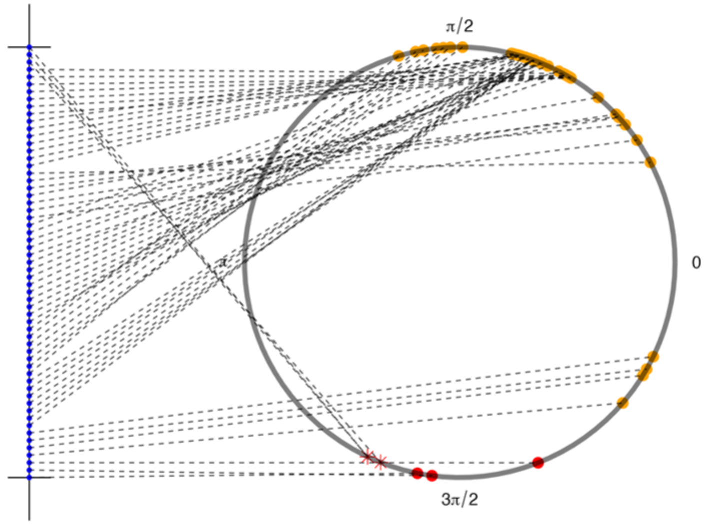

Figure 4. Wind direction data plot: yellow dots as correctly identified inliers, red dots as misidentified inliers, red stars as correctly detected outliers.

5. Discussion

The proposed methods, ICDR and MRLR, demonstrate notable improvements in detecting outliers in linear-circular regression models compared to the method proposed by Sert and Kardiyen [17]. The proposed methods provide more flexibility and accuracy, especially under difficult circumstances, however, Sert and Kardiyen's method consistently demonstrates robustness with constant low $pswamp$ values in uncontaminated data and high $pout$ with low $pmask$ in contaminated data. MRLR, which leverages the principle of mean resultant length, focuses on the consistency and concentration of circular data around the mean direction. The MRLR method excels at quantifying the degree of clustering in circular data. This allows it to effectively distinguish outliers that deviate significantly from the mean direction and inliers that align with the overall data structure. With a cut-off point of C_R = 0.01, MRLR achieves near-zero $pswamp$ values across both uncontaminated and contaminated data, outperforming Sert and Kardiyen's method in terms of minimizing the misclassification of inliers as outliers. Additionally, the MRLR method provides the ability to adjust sensitivity through the cut-off threshold (C_R), making it more adaptable to varying contamination levels and data structures. This flexibility is especially useful in situations with high contamination, because MRLR reliably and accurately detects actual outliers. However, when dealing with high sample sizes, MRLR has a disadvantage.

The ICDR method incorporates circular distance measures, which effectively capture the unique characteristics of circular data. This method ensures accurate data point differentiation in circular regression, especially for datasets with high concentrations or large sample sizes. Even while ICDR's $pswamp$ values are marginally greater than the MRLR method's, it still shows excellent accuracy in differentiating outliers from inliers. While ICDR is more effective in high data concentration conditions than Sert and Kardiyen's method, it might not be as reliable in datasets with smaller sample sizes. Both ICDR and MRLR outperform Sert and Kardiyen's method in scenarios with high contamination and complex data structures. While MRLR is highly effective in detecting outliers under most conditions, it fails to identify outliers in larger sample size, which limits its robustness in such cases. ICDR, on the other hand, demonstrates consistent performance but is slightly less effective than MRLR overall in reducing misclassification rates.

The real data application results align with the simulation study, demonstrating that the proposed method can detect all true outliers while exhibiting some swamping effect, affirming its practical applicability and effectiveness.

However, this study focuses on the von Mises distribution, which is widely used for symmetric circular errors due to its analytical tractability. Future work may explore heavy-tailed or asymmetric error models, such as the wrapped Cauchy distribution, to evaluate the robustness of the proposed methods under more challenging conditions. Moreover, hybrid strategies that combine the strengths of ICDR's circular distance partitioning with MRLR's resultant length-based refinement may also be investigated. Such approaches could improve performance across a broader range of data conditions. Additionally, future studies might compare the proposed methods with other outlier detection techniques, like MST-based methods and strong distance measures designed for circular data, to better understand how widely these methods can be used.

6. Conclusion

This study presents two new ways to find outliers in linear-circular regression models, using circular distance and mean resultant length (MRL). We evaluated and compared the proposed methods, ICDR and MRLR, to Sert and Kardiyen's [17] method. The evaluation was conducted using simulated datasets under both uncontaminated and contaminated conditions. The results demonstrated that ICDR and MRLR significantly improved outlier detection, particularly in challenging scenarios. However, some issues were found: MRLR became less reliable when there were many data points and high levels of contamination because of masking effects, while ICDR did not work as well with datasets that had fewer data points or smaller sample sizes. So, the results show that ICDR and MRLR are good at finding outliers in difficult situations, but they also have some weaknesses. Thus, the findings point out that ICDR and MRLR are strong methods for improving outlier detection in linear-circular regression models, but there is still a need to improve them to fix their weaknesses.

Acknowledgement

The authors would like to thank the Department of Statistics, Faculty of Science, Khon Kaen University for financial support.

REFERENCES

1. A.L.Gould, *A regression technique for angular variates*, Biometrics, vol. 25, no. 4, pp. 683–700, 1969.
2. S. Kim, and A. SenGupta, *Inverse circular-linear/linear-circular regression*, Communications in Statistics - Theory and Methods, vol. 44, no. 32, pp. 4772–4782, 2015.
3. A.E. Sikaroudi, and C. Park, *A mixture of linear-linear regression models for a linear-circular regression*, Statistical Modelling, vol. 21, no. 3, pp. 220–243, 2021.
4. A. Rambli, I. Mohamed, K. Shimizu, and N.M. Ramli, *A half-circular distribution on a circle*, Sains Malaysiana, vol. 48, no. 4, pp. 887–892, 2019.
5. N.H. Moslim, N.A. Mokhtar, Y.Z. Zubairi, and G. Hussin, *Understanding the behaviour of wind direction in Malaysia during monsoon seasons using replicated functional relationship in von Mises distribution*, Sains Malaysiana, vol. 50, no. 7, pp. 2035–2045, 2021.
6. L. Landler, G.D. Ruxton, and E.P. Malkemper, *Advice on comparing two independent samples of circular data in biology*, Scientific Reports, vol. 11, 20337, 2021.
7. R.A. Johnson, and T.E. Wehrly, *Some angular-linear distributions and related regression models*, Journal of the American Statistical Association, vol. 73, no. 363, pp. 602–606, 1978.
8. H.K.W. Kouassi, H. Issa, G. Stienne, and S. Rebou, *A linear-circular regression estimate for data fusion: application to GNSS carrier-phase signal processing*, Digital Signal Processing, vol. 117, 103172, 2021.
9. M.T. Beyene, S. Jain, and R.C. Gupta, *Linear-circular statistical modeling of lake ice-out dates*, Water Resources Research, vol. 54, no. 10, pp. 7841–7858, 2018.
10. C. Rueda, M.A. Fernández, S. Barragán, K.V. Mardia, and S.D. Peddada, *Circular piecewise regression with applications to cell-cycle data*, Biometrics, vol. 72, no. 4, pp. 1266–1274, 2016.

11. L.P. Rivest, T. Duchesne, A. Nicosia, and D. Fortin, *A general angular regression model for the analysis of data on animal movement in ecology*, Journal of the Royal Statistical Society, Series C: Applied Statistics, vol. 65, no. 3, pp. 445–463, 2016.
12. O. Polsen, *Circular–circular regression model, with application to meteorological data*, The Journal of KMUTNB, vol. 34, no. 2, pp. 1–11, 2024.
13. R.S. Nazer, and O.A. Ali, *Nadaraya-Watson estimation of a circular regression model on peak systolic blood pressure data*, Journal of Economics and Administrative Sciences, vol. 30, no. 144, pp. 473–484, 2024.
14. E.A. Mahmood, S. Rana, H. Midi, and A.G. Hussin, *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.
15. N. Romroeng, W. Srisodaphol, and B. Rattanawong, *Procedure for outlier detection in a circular regression model using resultant length*, AIP Conference Proceedings, 3123, 020022, 2024.
16. N.F.M. Di, S.Z. Satari, and R. Zakaria, *Outlier detection in circular regression model using minimum spanning tree method*, Journal of Physics: Conference Series, 1366, 012102, 2019.
17. S. Sert, and F. Kardiyen, *A method for detecting outliers in linear-circular non-parametric regression*, PLoS ONE, vol. 18, no. 6, e0286448, 2023.
18. A.H. Abuzaid, *Some problems of outliers in circular data*, Doctoral Dissertation, University of Malaya, 2010.
19. S.R. Jammalamadaka, and A. SenGupta, *Topics in Circular Statistics*, World Scientific Press, Singapore, 2001.
20. J. Cremers, and I. Klugkist, *One direction? A tutorial for circular data analysis using R with examples in cognitive psychology*, Frontiers in Psychology, 9, 2040, 2018.
21. K.V. Mardia, and P.E. Jupp, *Directional Statistics*, John Wiley & Sons, Chichester, 2000.
22. N.S. Zulkipli, S.Z. Satari, and W.N.S.W. Yusoff, *The effect of different similarity distance measures in detecting outliers using single-linkage clustering algorithm for univariate circular biological data*, Pakistan Journal of Statistics and Operation Research, vol. 18, no. 3, pp. 561–573, 2022.
23. B. Erisen, *Wind Turbine Scada Dataset*, Kaggle, https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset (Accessed 10 January 2025), 2018.