UDC 004.8 https://doi.org/10.20535/2786-8729.6.2025/326821

METHOD FOR DETECTING ANOMALOUS BEARING MEASUREMENTS BASED ON THE ANALYSIS OF DISTRIBUTION DENSITY HISTOGRAM

Illia Starodubtsev *
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
https://orcid.org/0009-0007-0361-6870

Oleksii Pysarchuk
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
https://orcid.org/0000-0001-5271-0248

*Corresponding author: i.starodubtsev.io21mp@kpi.ua

Subject of the research. Statistical models and algorithmic techniques for processing bearing-measurement datasets with the goal of detecting and filtering anomalous values. Object of the research. Methods for handling outliers in bearing-measurement datasets. Purpose of the work. To develop a method that improves the reliability of determining the direction to a radio-emission source in a single-station radio direction-finding system.

The article proposes a method for detecting anomalous measurements by analysing the histogram of their probability-density distribution. Propagation characteristics of very-high-frequency radio waves complicate radio direction finder operation, producing both random and systematic anomalies. The approach examines bearing-density histogram intervals rather than individual measurements, enabling an integrated assessment of sample structure and anomaly identification.

The method's performance was assessed on bearing datasets of 100, 500 and 1000 measurements at confidence levels of 80–95 %. The evaluation of the method's robustness was made for small and large arrays. The study analysed the influence of systematic and random anomalies on histogram construction and confidence-interval determination. Key metrics were bearing-estimation accuracy, root-mean-square error and method stability across varying confidence levels.

Results show the method minimises the impact of anomalous measurements at the tails and within the main cluster. Experiments demonstrate a reduced root-mean-square error and a consistent rise in direction-finding accuracy after anomaly removal. These findings confirm the technique's potential for further development and integration into synthesised radio direction-finding systems, where bearing reliability and precision are critical.

Keywords: radio direction finding, VHF range, statistical learning, anomalies, histogram.

1. Introduction

Current circumstances clearly demonstrate the importance of utilizing the radio-frequency spectrum to address a broad range of radio monitoring tasks. Notably, one of the principal objectives is to determine the coordinates of sources of radio-frequency emission (SRFE). One approach to achieving this goal involves the use of radio-direction-finding systems (RDFS) operating in the very-high-frequency (VHF) band (30–300 MHz). Owing to its short wavelength (1–10 m), this band enables high-precision bearing measurements to SRFE.

At the same time, despite its widespread application in modern radio monitoring systems, the VHF band presents numerous challenges arising from its physical characteristics. For example, multipath propagation, limited line-of-sight conditions, elevated background-noise levels, and pronounced sensitivity to weather effects lead to various types of anomalous measurements (AM) that reduce the accuracy of SRFE bearing determination. By contrast, such AM are not as prevalent in the high-frequency band.

Therefore, the propagation characteristics of the VHF band significantly complicate the operation of RDFS and promote the occurrence of AM. Improving the accuracy of SRFE bearing determination underscores the need to develop reliable methods for detecting these anomalies. Ongoing investigation into statistically grounded, sample-efficient techniques for detecting and filtering anomalous VHF bearing measurements is indispensable for ensuring dependable radio-direction finding in today's increasingly congested spectrum and will remain a critical research priority in the foreseeable future.

The relevance of this study stems from the need to improve the reliability of estimating directions to SRFE amid an increasingly congested spectrum and rising accuracy demands in radio monitoring. The problem addressed is the development of a statistically grounded and sample-efficient method for the automatic detection and filtering of anomalous bearing measurements in small- to medium-sized datasets. Solving this problem will reduce coordinate-estimation errors and enhance the robustness of single-station RDFS. Accordingly, the proposed approach aims to eliminate a critical limitation of existing methods, which either require large training datasets or lack sufficient sensitivity to strong in-sample anomalies.

2. Literature review and problem statement

The problem of detecting AM belongs to the general class of anomaly detection problems. Classical statistical anomaly detection algorithms encompass various approaches, including anomaly filtering, Bayesian statistics, cluster analysis, and central tendency methods [1–5]. In general, these traditional algorithms operate as follows: construct a model for the data distribution, compute the distance of the observations from the center of the distribution or the primary cluster, and label as anomalous any values that exceed a specified threshold distance. The main drawbacks of these approaches are their dependence on prior distribution assumptions, limited applicability to small sample sizes, and poor detection of anomalies in both the tails and the center of the data distribution.

Traditional statistical filters, Kalman-type estimators and machine-learning outlier detectors only partially address this challenge: many rely on large training datasets, assume specific noise models or struggle with small, rapidly acquired bearing batches.

Consequently, devising robust, data-economical techniques for detecting and compensating AM, especially in small samples, remains an open scientific and engineering problem that directly affects operational reliability, regulatory compliance and public safety.

Recent research proposes algorithms [6–8] that rely on constructing data distribution histograms and analysing deviations within the resulting intervals such as Histogram-based Outlier Score, The Adjusted Histogram-Based Outlier Score, Relative Kernel Density-Based Outlier Score. The general idea involves selecting a predefined number of intervals for the histogram, normalizing their values using a specific criterion, and computing an integral anomaly score. Although this approach allows for simultaneously capturing various segments of the distribution, it is primarily aimed at evaluating individual data points.

To enhance the reliability of AM analysis, it is advisable to focus not on individual values but on entire intervals within the distribution histogram. This generalized approach evaluates density and deviations within each interval instead of examining every single measurement in detail. As a result, it reduces the influence of both random and systematic anomalies, which may emerge in the distribution tails as well as at its center.

Accordingly, employing the distribution histogram to assess entire intervals – rather than individual measurements – can significantly enhance the method's consistency with real-world observation conditions. This approach ensures robustness against random errors and facilitates reliable detection of systematic disturbances. Ultimately, such modelling offers the prospect of more accurate detection of AM and, consequently, improved overall detection system performance.

This allows us to argue that it is advisable to conduct a study dedicated to the development of a statistically grounded, sample-efficient method for detecting and mitigating AM in VHF RDFS, integrating histogram-based analysis.

3. The aim and objectives of the study

The purpose of the study is to develop a method for detecting AM by analysing the distribution histogram. This will address the shortcomings identified above and enhance analytical robustness against various types of errors without compromising the efficiency of the detection process.

To achieve this goal, the following tasks were defined:

– develop and validate a histogram-based method for assessing measurement anomalies at the histogram interval level.

– experimentally verify the effectiveness of the proposed method using model data.

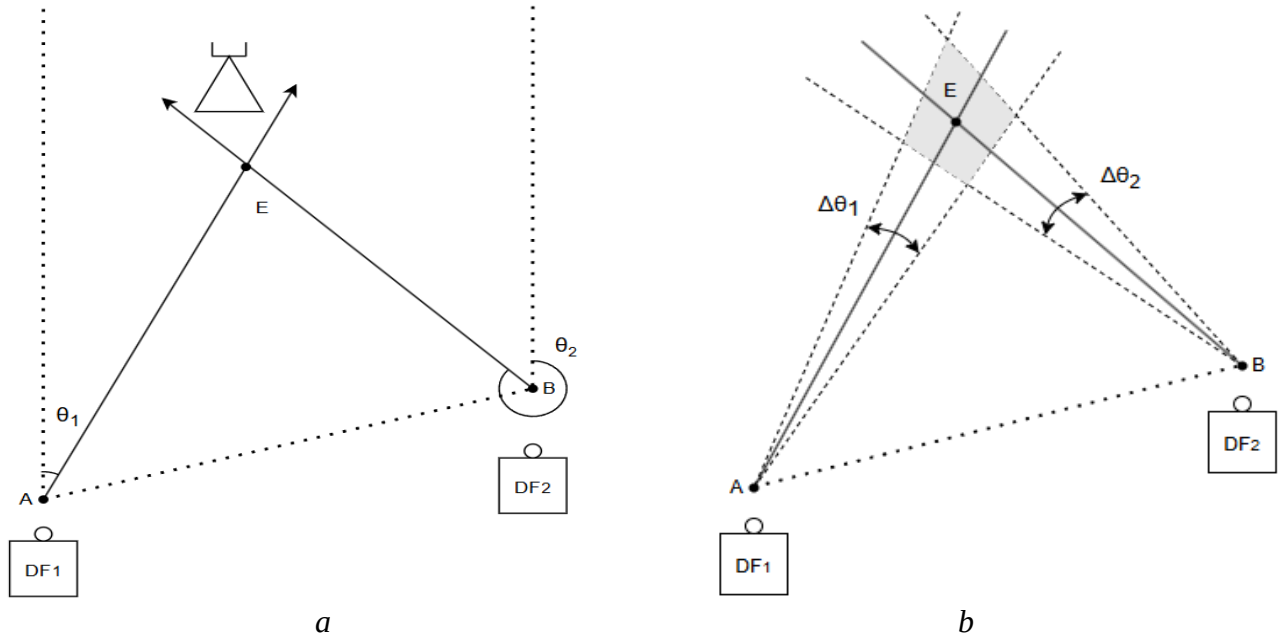
4. The study materials and methods

4.1. The study materials and methods of anomaly detection

As part of this study, the processes related to bearing signal processing are not considered. The input information consists of bearing data for SRFE, which have already been generated using primary processing algorithms. It is assumed that, over a short stationary interval – during which spatial orientations remain unchanged – RDFS can collect between 100 and 1000 bearing measurements, forming a statistically representative sample. This volume of data is sufficient for the subsequent application of methods for bearing measurement and assessment, including determining the specific SRFE to which the bearing belongs.

A dual RDFS uses triangulation to determine the coordinates of the SRFE. RDFS positioned at points A (DF1) and B (DF2) measure the bearing angles θ1 and θ2, which are referenced to true north as shown in Figure 1, *a*. The intersection of these directions determines the position of the SRFE at point E. The accuracy of determining the coordinates at point E is heavily dependent on the precision of the bearing measurements. In an ideal scenario, where bearing measurements are error-free, the coordinates of SRFE can be calculated with high accuracy. However, under real-world conditions, errors arise due to both random and systematic AM.

The operation of RDFS in the presence of anomalous bearings is illustrated in Figure 1, *b*. RDFS located at points A (DF1) and B (DF2) determine the directions to SRFE at point E. However, due to measurement errors Δθ1 and Δθ2, the actual position of SRFE falls within an uncertainty region shown in grey. Bearing lines are depicted with dashed lines to indicate possible deviations, while the grey region represents the potential location of the SRFE based on the measurement accuracy of each RDFS.



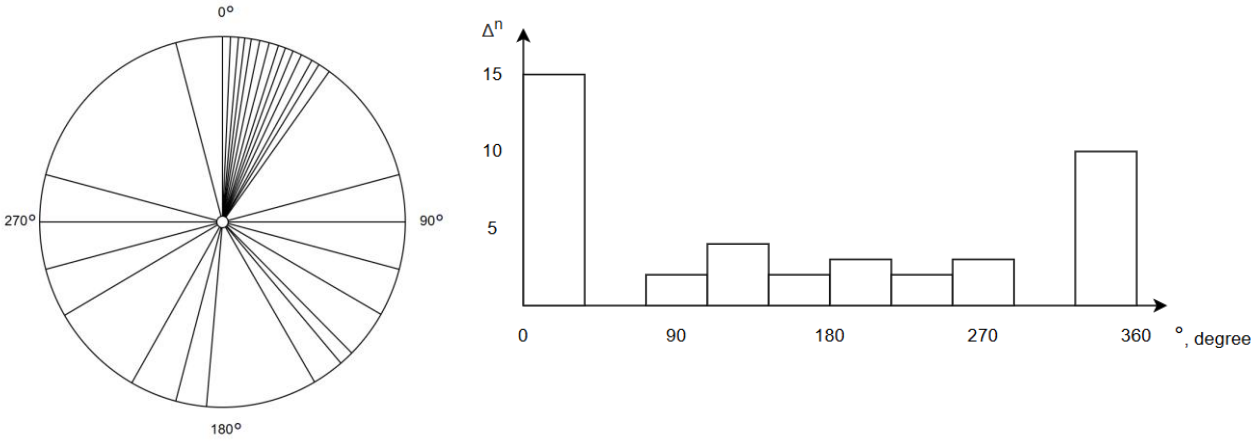
a *b*

Fig. 1. Principle of RDFS operation under different conditions: *a* – error-free environment; *b* – real-world environment.

A sample of bearing measurements obtained from the observation process can be represented either as a polar plot or as a histogram of their distribution, thereby allowing assessment of how the bearing angles are distributed. Under ideal conditions, the bearings exhibit a normal distribution centered around the true angle. However, the presence of anomalous values can significantly affect the outcome.

The polar plot of the collected bearings is illustrated in Figure 2*a*, where RDFS is positioned at the center, and the lines represent the measured bearing values. Although the true bearing angle is set at 15, the observed measurements show considerable variability. The bearing angles are unevenly distributed, displaying a higher concentration around the true value, while also exhibiting substantial deviations attributable to AM.

A histogram of the obtained bearings is depicted in Figure 2*b*, allowing an evaluation of their frequency distribution. The horizontal axis marks bearing angles ranging from 0° to 360°, while the vertical axis indicates the number of measurements in each interval.



a *b*

Fig. 2. Bearings distribution: *a* – polar plot; *b* – histogram.

The main aim and objectives are addressed at three levels: model, algorithm, and technology.

4.2. Adaptive model for detecting and filtering anomalous bearing measurements

A histogram of the data distribution enables shifting the focus from individual values to the frequency characteristics of the sample, thus streamlining the analysis and enhancing its statistical stability. This approach is particularly crucial for small samples, in which individual anomalous values can have a significant impact on the distribution.

Assume that during the RDFS observation process targeting an SRFE, a dataset in the form of a bearing distribution is obtained:

$$E = \{e_1, e_2, \dots, e_n\}, \quad (1)$$

where, n – number of measurements.

In general, it is assumed that the obtained sample is primarily affected by anomalies that are normally distributed, which can be classified into two main types: AM1 and AM2. Anomalies of the first type (AM1) typically arise from random noise or interference of natural or man-made origin, leading to minor deviations from the expected bearing values. Anomalies of the second type (AM2) are caused by malfunctions or errors in the measurement equipment and usually produce more substantial deviations, which can critically affect the accuracy of subsequent estimates.

Next, the bearing values are represented as a data distribution histogram, where each interval is characterized by a frequency f_m and a center c_m. In general form, such a histogram can be expressed as:

$$hist = \{(c_1, f_1), (c_2, f_2), \dots, (c_m, f_m)\}, \quad (2)$$

where, m – number of intervals.

Selecting the optimal number of intervals is challenging. Many intervals can lead to a noisy distribution, where anomalous values stand out as significant elements. Conversely, an excessively small number of intervals results in over-smoothing, causing the loss of critical features, such as the main cluster of measurements.

Various heuristic methods are employed to achieve an appropriate balance between detail and generalization in statistical analysis. These methods consider sample characteristics such as sample size, data bounds (min/max), standard deviation, or interquartile range. They help determine the width of each interval in a way that balances sensitivity to anomalous values with the retention of the most significant distribution features.

In this article, the Freedman-Diaconis rule is used to determine the histogram bin width (number of intervals). The choice of this method is guided by the Shannon entropy measure, which quantifies the degree of disorder in a distribution. Maximizing the entropy helps select several intervals that best capture the data's structure, preserving both the bulk of the sample's distribution and its local features.

The next step is to determine confidence intervals for the histogram values, considering the sample size. This is accomplished via the cumulative distribution function (CDF), which helps form intervals that encompass a specified proportion of the dataset at a chosen confidence level. The method retains an interval around the central values of the distribution, recognizing that AM are scattered within the main body of data; this reduces their influence on the interval boundaries. Consequently, the selected interval covers the central portion of the data distribution while excluding extreme values likely arising from AM1 or AM2 anomalies. By adopting this approach, one can establish a statistically justified range that captures the primary data cluster while minimizing the impact of individual outliers.

A confidence interval for a random variable represented by a histogram can be established by computing the CDF and identifying the bounds corresponding to a specified confidence level.

$$h = \{h_1, h_2, \dots, h_n\}, \tag{3}$$

where, h_n – histogram frequency values.

$$e = \{e_1, e_2, \dots, e_k\}, \tag{4}$$

where, e_k – histogram interval boundaries.

First, the total number of observations is calculated:

$$N = \sum_{i=0}^{n} h_i, \tag{5}$$

where, h_i – number of observations in the i–th interval.

The probability of a random variable falling into the i interval is defined as:

$$p_i = \frac{h_i}{N}, \tag{6}$$

CDF determines the cumulative probability that the value of a random variable will be less than or equal to the limit e_i:

$$F(e_i) = \sum_{j=0}^{n} p_j, \tag{7}$$

For a given confidence level γ, the parameter α is calculated as:

$$\alpha = 1 - \gamma, \tag{8}$$

Then the lower bound of the confidence interval corresponds to the quantile of the level a over two, and the upper limit is the quantile of the level one minus a over two.

$$i_{low} = min\left\{i: F(e_i) \geq \frac{\alpha}{2}\right\}, \tag{9}$$

$$i_{high} = min\left\{i: F(e_i) \geq 1 - \frac{\alpha}{2}\right\}, \tag{10}$$

The boundaries of the confidence interval are determined by the corresponding values of the interval limits:

$$\left[e_{i_{low}}, e_{i_{high}}\right], \tag{11}$$

The final step in computing the probabilistic bearing is determining the weighted average of the histogram interval centers that lie within the chosen confidence interval. First, each interval center is calculated as the midpoint of its bin, providing a representative value for that interval. Next, only those centers falling within the previously defined confidence interval are retained. Each selected center is assigned a weight corresponding to the frequency of values in its interval. The weighted average is then computed by summing the products of these centers and their frequencies and dividing by the total frequency of all data within the confidence interval.

The central value of the interval is determined by the formula:

$$c_i = \frac{e_{i-1}+e_i}{2}, \tag{12}$$

The weighted average value is calculated by the formula:

$$\underline{c} = \frac{\sum_{i=0}^{n} c_i \times h_i}{\sum_{i=0}^{n} h_i}, \tag{13}$$

This approach accounts not only for the positions of values in the distribution but also for their density, thereby ensuring that the computed average remains representative. By giving greater weight to intervals with higher data density (which most likely encompass the true bearing), the method ensures these dense intervals have a stronger influence on the result.

Using a certain confidence level allows the method to incorporate a larger portion of the data distribution, which is crucial for robust results in the presence of anomalies. However, increasing the confidence level to very high values (95% or above) can lead to the inclusion of extreme values. Conversely, lowering the confidence level to around 80–85% may exclude too much representative data, adversely affecting accuracy. Therefore, controlling the confidence level is important not only for achieving an optimal root mean square error (RMSE) and high accuracy. It is also crucial for ensuring the stability and consistency of the computed bearing in the presence of AM1 and AM2 anomalies. Considering this, it is important to investigate how RMSE and accuracy change with the confidence level and to determine an optimal range that balances the inclusion of representative data against the exclusion of anomalies.

Controlling the confidence level is important not only to achieve optimal root mean square error (RMSE) and high method accuracy, but also to ensure the stability and consistency of the computed bearing in the presence of AM1 and AM2 anomalies. Thus, it is important to investigate how RMSE and accuracy change with confidence level and to determine the optimal range that balances inclusion of representative data and exclusion of anomalies.

An adaptive model for detecting and removing AM has been developed. Its main differences, advantages, and defining features are as follows. First, the model can dynamically determine the number of histogram intervals based on an entropy criterion, thereby balancing the retention of the central measurement cluster against the isolation of anomalies. Second, the algorithm involves constructing confidence intervals via CDF, which facilitates the formation of the main dataset, while providing the ability to exclude tail intervals. Third, the method readily accommodates modifications: by adjusting the confidence level and optimizing the number of intervals, it can be tailored to different sample sizes. Because the model does not require ongoing prior knowledge about the nature of anomalies, it can be easily integrated into RDFS to enhance bearing accuracy. Consequently, the proposed approach preserves statistical stability and analytical flexibility, while substantially mitigating the influence of measurement errors on determining the location of SRFE.

4.3. Algorithm for detecting and filtering anomalous bearing measurements

Summarizing the model described above, the algorithm can be outlined in the following steps:
– Obtain a sample of bearing measurements E.
– Generate a distribution histogram from this sample.
– Compute the CDF for the sample's histogram.
– Determine the boundaries of the desired confidence interval.
– Calculate the centers of the histogram bins that lie within this confidence interval.
– Compute the weighted average of these selected bin centers.
– Output the results of the weighted average calculation.

4.4. Technology for detecting and filtering anomalous bearing measurements

The technology for detecting and filtering AM refers to the practical implementation of the synthesized model and its algorithm within a software system architecture. This implementation is illustrated by the structural diagram of the algorithm in Figure 3.

The input file (dataset.csv) contains a single column labelled "azimuth" and comprises a total of 1000 bearing values, as shown in Figure 4. The algorithm is implemented in Python, utilizing the NumPy and Pandas libraries.

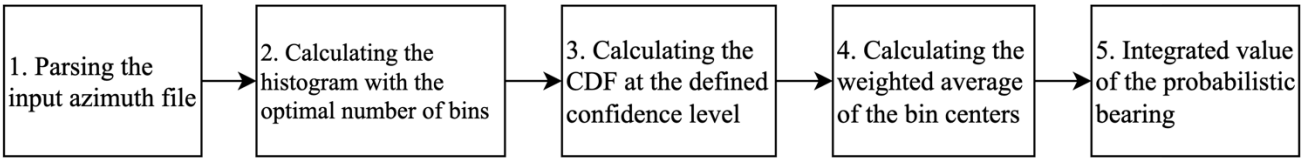


Fig. 3. Structural diagram of the algorithm

	azimuth
1	43
2	46
3	44
4	47
5	53
6	44
7	45
8	48
9	53
10	43
11	46
12	45
13	45
14	49
15	48
16	44
17	46
18	47
19	49
20	46

Fig. 4. Input azimuth sample *dataset.csv*

The algorithm illustrated in Figure 3 carries out the detection and filtering of AM sequentially, following the developed model and algorithm. At the first stage (block 1), the CSV file is parsed and the sample E (bearing measurements from the RDFS) is loaded into a DataFrame. An example of a histogram of the input data's distribution is presented in Figure 5*a*. At the second stage (block 2), a histogram of the bearing data is generated using the chosen number of intervals. An example histogram obtained at this stage is shown in Figure 5*b*. During the third stage (block 3), CDF is computed according to the specified confidence level. This CDF accumulates the frequency in each histogram bin, reflecting the probability that a random bearing value does not exceed that interval boundary. Next, the lower and upper limits of the confidence interval are determined. These limits exclude the most extreme measurements, retaining the central portion of the sample that most likely represents the true bearing. An example of the data distribution histogram after the calculations in block 3 is displayed in Figure 5*c*.

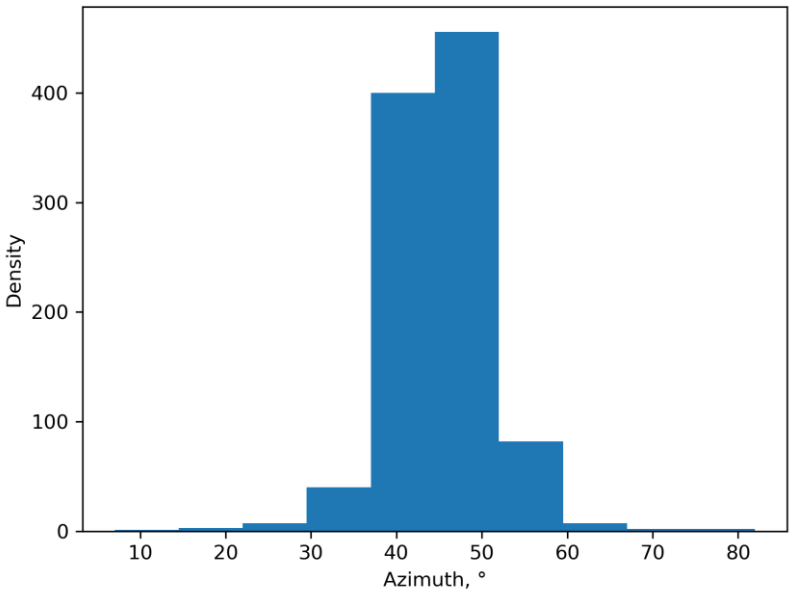


Fig. 5 (a). Distribution of azimuth measurements: of input dataset

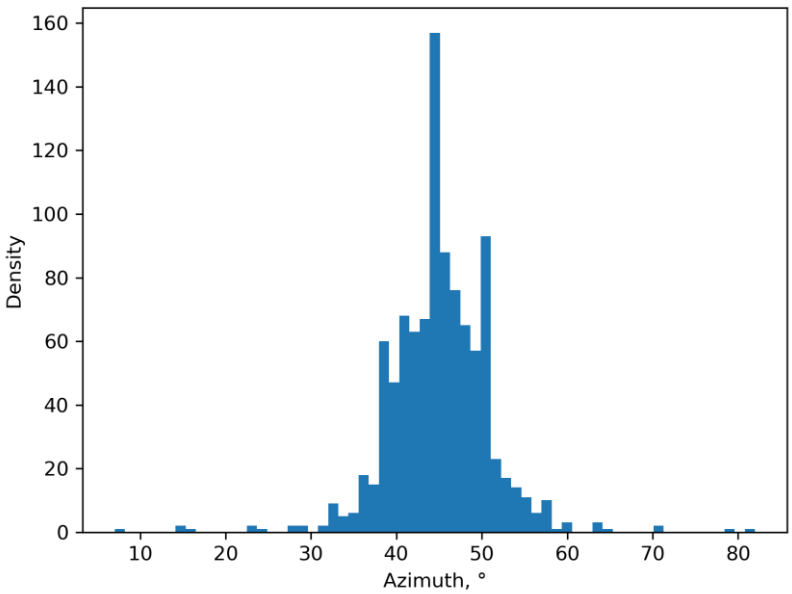


Fig. 5 (*b*). Distribution of azimuth measurements: with calculated optimal number of intervals

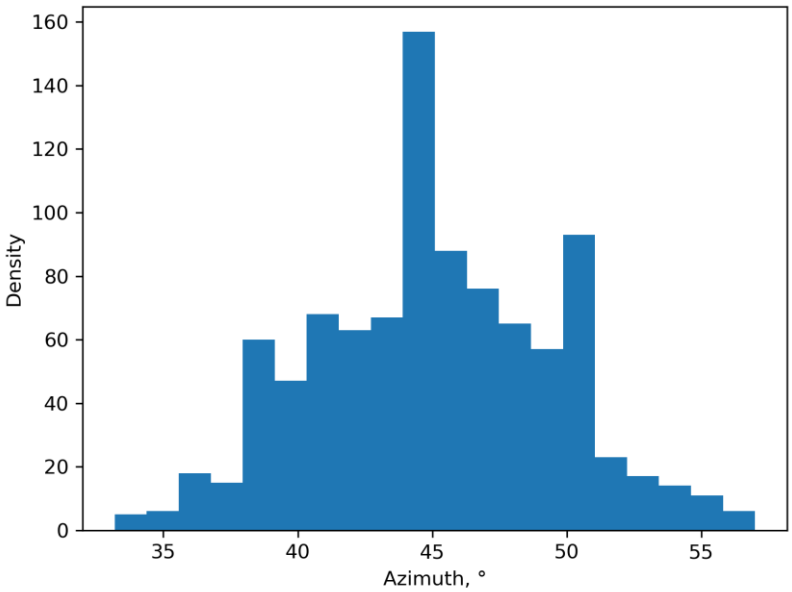


Fig. 5 (*c*). Distribution of azimuth measurements: after applying CDF with defined confidence level

Next (block 4), the centers of the intervals within the defined confidence range are computed, followed by the calculation of their weighted average value. According to the step-by-step algorithm described above, this weighted average constitutes a probabilistic bearing estimate that maximizes consideration of the data's frequency density while minimizing the impact of AM. The final step (block 5) is to output the final probabilistic bearing value, which represents the direct result of the algorithm and can be integrated into RDFS to enhance the accuracy of locating SRFE.

5. Results of investigating the anomaly evaluation algorithm at histogram intervals

5.1. Overview of experimental setup and evaluation metrics

To verify the method, samples of sizes $n = 100$, $n = 500$, and $n = 1000$ were used. The confidence level varied from 80% to 95% in 1% increments. The main analysis metrics were accuracy (Δ) and RMSE, which were employed to assess how the confidence level and sample size affect the algorithm's stability and accuracy.

5.2. Method for anomaly assessment in measurements

Within the scope of the first objective, an algorithm was developed for evaluating measurement anomalies at the histogram interval level. The core of this approach lies in establishing a certain confidence level for the measured data sample. Following this, anomalous intervals (e.g., those with abnormally high or low frequencies) are either excluded or assigned a reduced weight.

The method relies on two primary parameters:

– Confidence level, which defines cut-off intervals for anomalous values.

– Variance-based assessment, used to identify intervals with anomalous values.

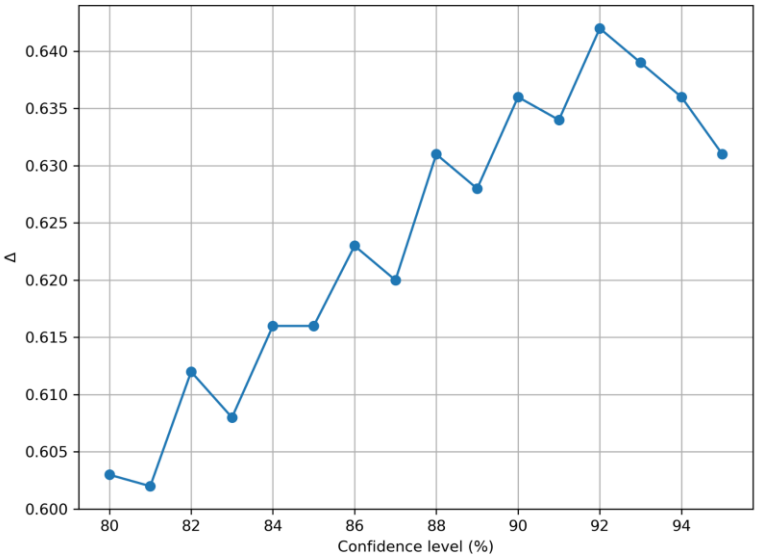
For each confidence level accuracy and RMSE are calculated for the respective samples. This approach enables control over the volume of excluded data, striking a balance between two opposing effects: excessive removal of valid measurements (reducing the completeness of the analysis) and the inclusion of anomalies (reducing accuracy).

5.3. Experimental Results

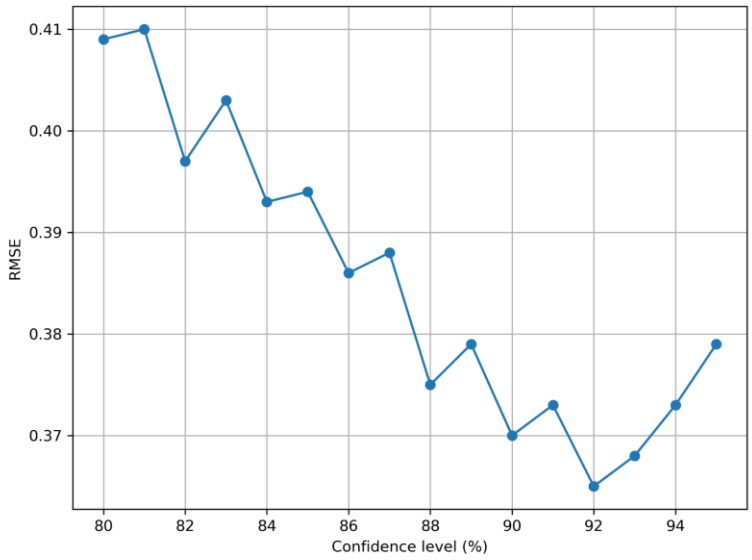
To achieve the second objective, we conducted a series of experiments using model datasets of various sizes. The confidence level varied from 80% to 95% in 1% increments. Based on the

experimental results, six plots were created to illustrate the relationships between accuracy and RMSE in respective confidence levels as shown in Figures 6–8. Each data point on the graphs represents the computed results for a particular confidence level, while the solid lines show the overall trend of how the metrics change.

For the sample of size n = 100, Figure 6*a*. shows how accuracy depends on the confidence level. The accuracy ranges from 0.60 to 0.64, indicating fewer stable results for smaller datasets. The dependence of RMSE on the confidence level is presented in Figure 6*b*. The minimum value (about 0.36) occurs at a 92% confidence level. These graphs exhibit some fluctuations, reflecting the algorithm's sensitivity to small sample sizes and the impact of anomalies.



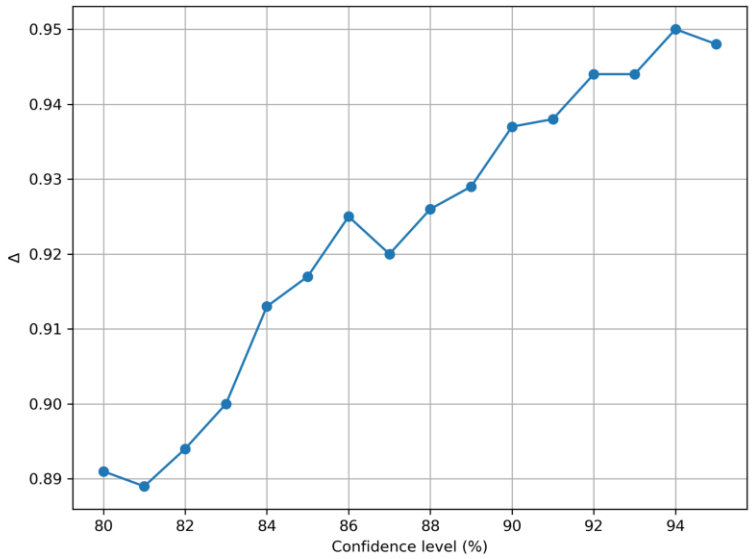
a



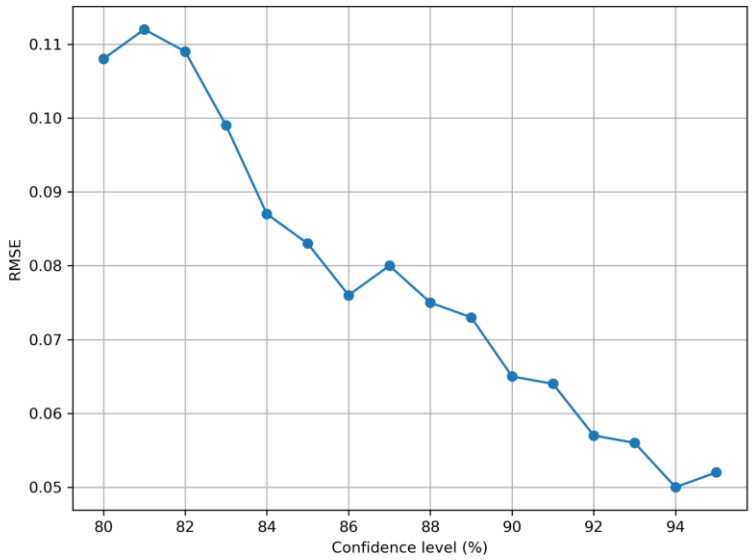
b

Fig. 6. Dependence of accuracy and RMSE on the confidence level for n = *100*: *a* – accuracy; *b* – RMSE.

The results for the sample of size n = 500 are shown in Figure 7. Accuracy grows more steadily from 80% to 94%, and RMSE decreases significantly from 0.11 to 0.05. The lines on the graphs display fewer fluctuations, demonstrating greater algorithmic stability as the number of data points increases.



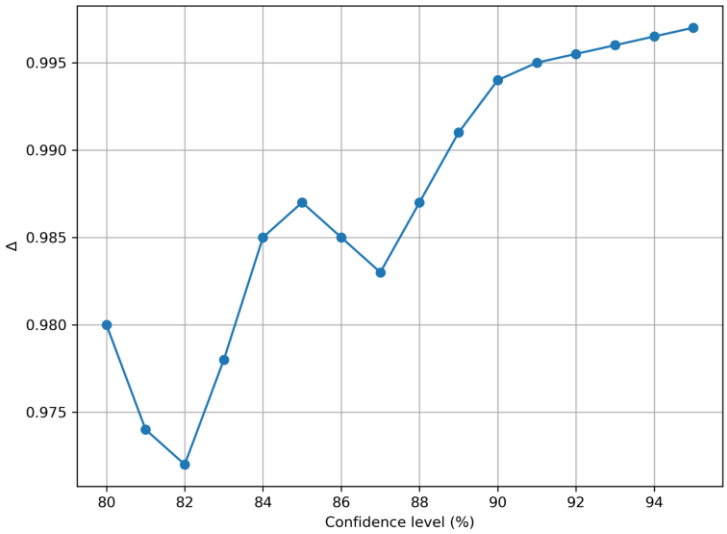
a



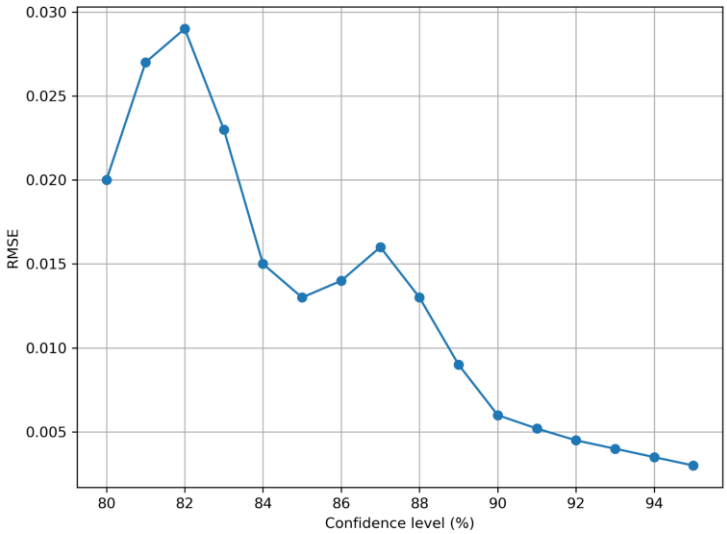
b

Fig. 7. Dependence of accuracy and RMSE on the confidence level for n = *500*: *a* – accuracy; *b* – RMSE.

The results for the sample of size n = 1000 are shown in Figure 8. In terms of accuracy, there is a slight instability between 80% and 82% confidence levels; however, starting at 88%, the accuracy continues to improve and reaches about 99.5% at 95% confidence. Meanwhile, RMSE is substantially reduced as the confidence level increases. From 90% onward, the curve trends almost linearly toward zero. This outcome highlights the algorithm's capacity to mitigate the effect of anomalies and sustain high accuracy, even when the data exhibit variability.



a



b

Fig. 8. Dependence of accuracy and RMSE on the confidence level for $n = 1000$: *a* – accuracy; *b* – RMSE.

Overall, the increasing stability of the method with larger sample sizes confirms the appropriateness of applying the proposed approach to more extensive datasets. Results indicate that a 92–95% confidence level is optimal for minimizing RMSE and achieving high accuracy across different sample sizes. Nonetheless, for very small samples, excessively high confidence levels can result in including extreme values, which lead to increasing RMSE. This finding underscores the need to select a confidence level based on the amount of available data.

6. Discussion of results of the anomaly evaluation algorithm

The findings confirm the effectiveness of the proposed anomaly assessment algorithm operating at the histogram interval level. For the smaller sample ($n = 100$), results exhibit greater fluctuations, because the algorithm is more vulnerable to individual anomalies that can substantially affect the overall evaluation.

When dealing with medium and large samples ($n = 500$ and $n = 1000$), the algorithm

demonstrates relatively steady accuracy gains and a marked reduction in RMSE as the confidence level increases – an outcome attributable to enhanced data representativeness and a smaller proportion of anomalous values. According to the experimental data, a 92–95% confidence interval proves optimal for minimizing the effect of anomalies while maintaining maximum achievable accuracy. However, with small samples, exceedingly high confidence levels can result in including extreme values, thus inflating RMSE.

Overall, the experimental investigations affirm both the novelty and the practical significance of the developed algorithm. Its application not only allows for monitoring anomalous intervals in a sample but also supports adaptive confidence level adjustment in accordance with data size and statistical characteristics. Future work will focus on making the method less sensitive to anomalous values in small samples and testing its performance on real-world datasets.

Conclusions

The first research objective was to develop an adaptive method for detecting and filtering AM at the histogram interval level. The method assesses frequency characteristics of the sample rather than individual data points, thereby enhancing statistical stability for small sample sizes. It uses the Freedman-Diaconis heuristic to select the number of histogram intervals, and CDF to form confidence intervals that exclude AM1 and AM2 anomalies. This ensures minimal distortion of the core measurement cluster while preserving essential density distribution features. The approach is novel in its ability to automatically adapt to different sample sizes and mitigating the effects of random and systematic measurement errors. Its practical value is evident in increased bearing accuracy and improved reliability of RDFS.

The second objective was to experimentally verify the effectiveness of this method on model data. Tests on samples of different sizes showed that both accuracy and RMSE depend on the chosen confidence level. An optimal 92–95% confidence interval emerges, ensuring reduced RMSE and robust anomaly filtering without discarding too many representative values. Larger samples exhibit more stable performance, underscoring the algorithm's ability to isolate random or systematic anomalies while retaining most valid measurements. By dynamically balancing interval boundaries, the method demonstrates significant applicability for bearing analysis in radio monitoring tasks.

References

[1] W. W. Hsieh, Introduction to Environmental Data Science. Cambridge, U.K.: Cambridge Univ. Press, 2023, https://doi.org/10.1017/9781107588493.

[2] D. Samariya and A. Thakkar, "A comprehensive survey of anomaly detection algorithms," Annals of Data Science, vol. 10, no. 3, pp. 829–850, Jun. 2023, https://doi.org/10.1007/s40745-021-00362-9.

[3] S. Baccari, M. Hadded, H. Ghazzai, H. Touati, and M. Elhadef, "Anomaly detection in connected and autonomous vehicles: A survey, analysis, and research challenges," IEEE Access, vol. 12, pp. 19250–19276, 2024, https://doi.org/10.1109/ACCESS.2024.3361829.

[4] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," Decision Analytics Journal, vol. 6, Art. no. 100164, Mar. 2023, https://doi.org/10.1016/j.dajour.2023.100164.

[5] H. Xu, L. Zhang, P. Li, and F. Zhu, "Outlier detection algorithm based on k-nearest neighbors–local outlier factor," Journal of Algorithms & Computational Technology, vol. 16, Art. no. 17483026221078111, Mar. 2022, https://doi.org/10.1177/17483026221078111

[6] U. Binzat and E. Yıldıztepe, "The adjusted histogram-based outlier score – AHBOS," Mugla Journal of Science and Technology, vol. 9, no. 1, pp. 92–100, Jun. 2023, https://doi.org/10.22531/muglajsci.1252876

[7] N. Banić, N. Elezović, "TVOR: Finding discrete total variation outliers among histograms," IEEE Access, vol. 9, pp. 1807–1832, 2021, https://doi.org/10.1109/ACCESS.2020.3047342

[8] A. Wahid and A. C. S. Rao, "RKDOS: A relative kernel density-based outlier score," IETE Technical Review, vol. 37, no. 5, pp. 441–452, Sep. 2020, https://doi.org/10.1080/02564602.2019.1647804

УДК 004.8

МЕТОД ВИЯВЛЕННЯ АНОМАЛЬНИХ ВИМІРІВ ПЕЛЕНГУ ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ГІСТОГРАМИ ЩІЛЬНОСТІ ЇХ РОЗПОДІЛУ

Ілля Стародубцев
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
https://orcid.org/0009-0007-0361-6870

Олексій Писарчук
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
https://orcid.org/0000-0001-5271-0248

Предметом дослідження є статистичні моделі та методи обробки вибірки вимірювань пеленгів з метою виявлення й фільтрації аномальних значень. Об’єктом дослідження є методи обробки аномальних значень у вибірках вимірювань пеленгів. Метою роботи є розробка методу для підвищення достовірності визначення напрямку на джерело радіовипромінювання у однопозиційній радіопеленгаторній системі.

Стаття присвячена розробці методу виявлення аномальних вимірів за аналізом гістограми щільності їх розподілу. Природа поширення радіохвиль у діапазоні УКХ формує чинники, що ускладнюють роботу радіопеленгаторів і зумовлюють випадкові та систематичні аномалії. Підхід базується на аналізі інтервалів гістограми щільності пеленгів замість ізольованої обробки окремих вимірювань, що дозволяє комплексно оцінювати структуру вибірки і виявляти аномалії.

Ефективність методу перевірено на вибірках пеленгів 100, 500 і 1000 вимірів у діапазоні довіри 80–95 %. Такий набір експериментальних умов дав змогу оцінити стійкість методу для малих та великих масивів. Проаналізовано вплив систематичних і випадкових аномалій на побудову гістограм і визначення довірчого інтервалу. Ключові метрики: точність оцінювання пеленгу, середньоквадратична похибка й точність методу при різних рівнях довіри.

Результати показали, що метод мінімізує вплив аномальних вимірів на кінцях і всередині основного кластера. Експерименти демонструють зниження середньоквадратичної похибки та стабільне зростання точності визначення напряму на джерело радіовипромінювання після виключення аномалій. Отримані результати підтверджують перспективність подальшого розвитку методики й її інтеграцію у синтезовані системи радіопеленгації, де вирішальними є надійність і точність визначення пеленгу.

Ключові слова: радіопеленгація, УКХ діапазон, статистичне навчання, аномалії, гістограма.