Leveraging RSS Data For An Improved Radiolocation Estimation Algorithm Realization In LoRaWAN Using A Two-Tier Normalization

Colette E. Agbo, Udora N. Nwawelu*, and Mamilus A. Ahaneku

Department of Electronic Engineering, University of Nigeria, Nsukka, Enugu State, Nigeria

**Corresponding author. E-mail: nwabuoku.nwawelu@unn.edu.ng*

Received: November 26, 2022; Accepted: March 20, 2023

An attempt to further enhance the accuracy and reliability of M-iWMLR localization algorithm using a new weight matrix that was formulated with two-tier RSS data normalization is presented. The two-tier normalization: data clipping and z-score normalization were applied to form a new weight matrix in this work. Data clipping was first applied to reduce significantly the effects of outliers on the RSS data while z-score normalization provides data consistency. The new localization algorithm herein, referred to as Ext.M-iWMLR algorithm is carefully evaluated by the use of location accuracy (location error), root mean square error (RMSE), range of error, and R^2 score metrics. This algorithm is validated with the Modified Improved Weighted Multiple Linear Regression (M-iWMLR). The simulation results generated with MATLAB show that the Ext.M-iWMLR algorithm, at 95 percentile reduced the mean location error by 19.45%. The range of error and RMSE are reduced by 11.08% and 17.95%, respectively. Furthermore, the respective R^2 scores were increased by 5.71% and 17.17% for the latitude and longitude coordinates. It was established that the new weight matrix formulated through two-step normalization enhanced all the considered metrics.

Keywords: Data clipping; Z-score normalization; LoRaWAN; Location accuracy; RSS

http://dx.doi.org/10.6180/jase.202312_26(12).0014

1. Introduction

Number of Internet of Things (IoT) devices with myriad of applications are rapidly increasing. The applications of these devices can be seen in just a few fields such as environment monitoring and control, smart cities, agriculture, navigation system, fleet management, Medical diagnostics, among others [1–6]. These devices can communicate through IoT supportive network such as Long Range Wide Area Network (LoRaWAN), Sigfox, and NarrowBand IoT (NB-IoT) [7]. The LoRaWAN, Sigfox, and NB-IoT are popular Low Power Wide Area Networks (LPWAN) technologies. The usage of these technologies to estimate the position of IoT end terminals are growing in unprecedented pattern. Reasons being that such network standards are scalable, consumes low power and can be used to locate radio terminal in an environment with poor visibility of the sky [6–8].

Various location dependent parameters (LDP) have been used to formulate various localization algorithms that can be used to estimate the position of a transmitting IoT terminal in LPWAN. The LDP are Received Signal Strength (RSS), Angle of Arrival (AoA), Time of Arrival (ToA), and ToA variants such as Time Different of Arrival (TDoA), among others [2, 4–6, 9–14]. Localization algorithms formulated using RSS information are the most common algorithms to estimate radio terminal location. This is due to the fact that RSS information are found in the network and the backend can estimate the terminal's location, resulting to both device cost and system complexity reduction. Also, it does not require any form of synchronization either at the transmitting or receiving end [4, 6]. However, RSS data is affected by physical environment as distance varies [13, 15–18]. On the contrary, AoA-based or ToA-based algorithms (i.e., localization algorithm formulated with AoA or ToA parameters) can record lower location errors. Nevertheless, such algorithms require extra hardware at both terminal ends for synchronization in order to increase location accuracy. Consequently, extra hardware grafting at both terminal ends with hope of achieving proper synchronization would not only increase the device cost but also increase the system complexity [4, 6].

It is not an over statement to say fact that most, if not all the LDP for formulating location algorithms are obtained through measurements. Environments where such data are collected are uncontrolled environments. The measured data from such environments are expected to have outliers. If such data is not managed properly while formulating a localization algorithm, it could result to an algorithm that records appreciable location error. Thus, in formulating localization algorithm, it would be a good practice to incorporate a method to handle outliers. Outliers in this context is defined as data that deviate so much from the remaining data. They lie at a particular extreme of the scale.

This work tries to extend the previous contributions of the authors on the M-iWMLR algorithm [19]. In M-iWMLR algorithm, mean zero scaling (z-score normalization) was used on the RSS data to realize a modified weight matrix that was incorporated to the iWMLR algorithm [14]. In this current research, a two-tier normalization for realizing a new weight matrix is presented. Using the weight matrix formulated through two-tier normalization instead of that formed by z-score normalization only will certainly reduce significantly the location estimation error of a radio device. Enumerated below are the strengths of this work:

1. The total of 130,430 data sample points which comprised of RSS, latitude and longitude coordinates information were used. Relying on large data sample in order to realize a reliable localization estimation algorithm cannot be overemphasized.

2. The effects of data clipping on outliers and consequently on the terminal location accuracy were investigated by varying the clipping points. The aim is to obtain an optimum clipping point for clipping normalization.

3. The influential new weight matrix was used to extend the M-iWMLR algorithm.

2. Related work

There are many radio terminal localization algorithms in existence and more are on the way due to the relevance of these algorithms to location-based services. The existing location algorithms can be categorized as RSS-based, AoA-based, or ToA-based algorithm. This article reviewed a good number of RSS-based algorithms due to the fact that RSS information are available on the network [6].

RSS data from GSM network was used to formulate a trained algorithm for radio terminal location using multilayered perceptron neural network [12]. The location accuracy achieved complied with the Federal Communication Commission (FCC) requirement for emergency location based services [20]. Nevertheless, environmental phenomenon affects the RSS data used in the formation of the artificial neural network based location approaches. Also, much effort and time are required to train and maintain the trained data.

Multi linear regression (MLR) approach was used to formulate a location estimation algorithm. Data from GSM network which consisted of RSS and geographical coordinate information of the target terminal were used to develop the MLR model [13]. The MLR algorithm though simple, recorded a noticeable location error which can be due to the enviromental effect on the RSS values [15].

Recently, it was demonstrated that by incorporating a weight factor with MLR, location accuracy would be improved [11]. The performance of the algorithm that was formulated from such combination, WMLR algorithm was evaluated with the RSS data from GSM network infrastructure. Results show an improvement in terminal location accuracy. However, it was observed that the approach used to apportion weights to the RSS values during weight formation still resulted to coarse location estimation.

Meanwhile, min-max scaling was used to form a refined weight matrix that improved the WMLR algorithm, resulting to iWMLR algorithm [14]. The iWMLR algorithm, when evaluated with the RSS data from the LoRaWAN infrastructure recorded an improved location accuracy. The iWMLR algorithm achieved 8.6m and 21.6m location accuracy at 67 and 95 percentiles, respectively. However, it was observed that min-max scaling used to achieve the refined weight function is most suitable when there are few or no outliers or when all the data lies within the same range of values. Since most of the data lies within the smallest part of the scale, application of min-max scaling is not encouraged.

The mean zero scaling also known as z-score normalization was used recently to form a weight matrix that further improved the iWMLR algorithm, resulting to M-iWMLR algorithm [19]. The M-iWMLR algorithm, when evaluated with the RSS data from the outdoor LoRaWAN infrastructure recorded an improved performance metric values. The M-iWMLR algorithm achieved 8.4m and 21.4m location accuracy at 67 and 95 percentiles, respectively. However, it was observed that z-score normalization can be used when data with minimum number but not so extreme outliers are involved.

From the review done in this work and the close observance of the empirical RSS data from LoRaWAN, it is evident that the effects of extreme outliers can reduce the location accuracy of any location estimation algorithm. In the collected RSS data, most of the values are at the extreme (i.e., the smallest part of the scale) [19]. The action sought is to leverage the RSS data through incorporating a method to handle outliers during location algorithm formulation. This action would certainly increase significantly the location accuracy of a localization estimation algorithm.

3. Methodology

The method of data preparation down to the formulation of a new position estimation algorithm in LoRaWAN are discussed subsequently.

3.1. Data Preparation

Data used for the formulation and subsequently evaluation of our algorithm are LoRaWAN data set. The data set has been collected and used to formulate and evaluate localization algorithms [5, 6]. The collected set of LoRaWAN data was made available in the data repository provided in [21]. The data was deposited in the aforementioned repository to allow researchers of interest to compare formulated algorithms with the same data set. Total of 130,430 messages were collected by 72 LoRaWAN gateways which are sited around the Antwerp city of Belgium [6, 21]. Each of the messages comprised of RSS values in dBm together with the GPS latitude and longitude value in km. The latitude and longitude values served as true values. During the measurement campaign, RSS value for any gateway that could not receive a message from the transmitting device (i.e., non-hearable gateways) was set to -200 dBm [5, 6].

The major goal of this work is to formulate a better localization algorithm from the provided LoRaWAN messages. To put the LoRaWAN data set in proper perspective, the data set is divided into two subsets namely major subset and minor subset. The major subset will comprise of the 70% of the data set and it will be used in formulating the new localization algorithm. The minor subset will comprise of the 30% of the data set and it will be used to evaluate and validate the formulated algorithm.

3.2. Adoption of the Concept of Multiple Linear Regression (MLR) Algorithm

The knowledge from the MLR algorithm and its variants namely WMLR, iWMLR, and M-iWMLR algorithms are vital in formulating the new localization algorithm. Thus, the reader is referred to the scholarly works on the respective aforementioned algorithms [11, 13, 14, 19]. Only the vital aspect from the MLR algorithms as it relates to the M-iWMLR algorithm are discussed. The reason been that the proposed algorithm is a modified version of the M-iWMLR algorithm.

Suppose that the dependent location coordinate variables (x_i, y_i) are linear functions of two or more independent location variables, $R_{i,1}, R_{i,2}, \ldots, R_{i,n}$ as shown in Equations (1) and (2):

$$x_i = \beta_0 + \beta_1 R_{i,1} + \beta_2 R_{i,2} + \ldots + \beta_n R_{i,n} \tag{1}$$

$$y_i = \alpha_0 + \alpha_1 R_{i,1} + \alpha_2 R_{i,2} + \ldots + \alpha_n R_{i,n} \tag{2}$$

The coefficient values $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ and $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n$ to fit a set of data (x, $y, R_1, R_2, \ldots, R_n$) can be found using the least square method on Equations (1) and (2). The MLR comprised of more than one linear equations. If n is the number of data points and $i \in (1, 2, \ldots, m)$ is the size of the sample, the MLR algorithm for estimating a radio terminal location was formulated using the set of linear equations that was presented in matrix form in Equations (3a) and (4a) and in compact matrix as in Equations (3b) and (4b):

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ x_m \end{bmatrix} = \begin{bmatrix} 1 & R_{1,1} & R_{1,2} & \cdots & R_{1,n} \\ 1 & R_{2,1} & R_{2,2} & \cdots & R_{2,n} \\ 1 & R_{3,1} & R_{3,2} & \cdots & R_{3,n} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ 1 & R_{m,1} & R_{m,2} & \cdots & R_{m,n} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} \tag{3a}$$

$$\mathbf{x} = \mathbf{R}\boldsymbol{\beta} \tag{3b}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & R_{1,1} & R_{1,2} & \cdots & R_{1,n} \\ 1 & R_{2,1} & R_{2,2} & \cdots & R_{2,n} \\ 1 & R_{3,1} & R_{3,2} & \cdots & R_{3,n} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ 1 & R_{m,1} & R_{m,2} & \cdots & R_{m,n} \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \alpha_n \end{bmatrix} \tag{4a}$$

$$\mathbf{y} = \mathbf{R}\boldsymbol{\alpha} \tag{4b}$$

The vectors, β and α are obtained from Equations (3b) and (4b) and presented as in Equations (5a) and (5b), respectively.

$$\boldsymbol{\beta} = \left(\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{x} \tag{5a}$$

$$\boldsymbol{\alpha} = \left(\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1}\mathbf{R}^{\mathrm{T}}\mathbf{y} \tag{5b}$$

where $R^{\top}$ denotes the transpose of the R matrix.

With the known vectors, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as presented in Equations (5a) and (5b), the estimated terminal location was obtained with Equations (6a) and (6b).

$$x_{\mathrm{Est}} = [\mathbf{1}\ \mathbf{R}] \cdot \boldsymbol{\beta} \tag{6a}$$

$$\mathbf{y}_{\mathrm{Est}} = [\mathbf{1}\ \mathbf{R}] \cdot \boldsymbol{\alpha} \tag{6b}$$

Where $\mathrm{x}_{\mathrm{Est}}$ and y_{Est} are the target estimated location for the respective latitude and longitude coordinates.

The M-iWMLR is the recent improvement to MLR series. It is realisead by incorporating weight matrix into the MLR as in Equation (7) [11].

$$\mathrm{L}_{\text{modified}} = \mathrm{R} + \left[\widehat{\mathrm{Y}}_{\mathrm{i,j}}\right] \times \varnothing \tag{7}$$

The matrix, R as in Equation (3) through (6) was replaced with $\mathrm{L}_{\text{modified}}$ of Equation (7) for the coordinates of x and y. The $\widehat{Y}_{\mathrm{i,j}}$, which denotes weighted matrix is created from R matrix. The scaler multiplier, $\varnothing$ was used to vary $\widehat{Y}_{i,j}$ until an optimum value is realized [11]. Meanwhile, 10 is used for $\varnothing$ in [11].

The weighted matrix used to form the M-iWLMR algorithm was realized with z-score normalization [19]. The z-score normalization uses the mean and standard deviation of the observed data to scale each data point. The z-score normalization for two-dimensional data vector is expressed as in Equation (8);

$$\widehat{Y}_{i,j} = \frac{(\mathrm{Y}_{\mathrm{i,j}} - \mu)}{\sigma} \tag{8}$$

Where, Y denotes the RSS value for the ith and j th row and column, respectively. The μ represents the mean of the RSS value per row while σ represents the standard deviation of the RSS value per row.

Using z-score normalization to realize a weight matrix has certainly improved the M-iWMLR in terms of localization accuracy. However, as we have observed, z-score normalization can be used when data with minimum but not so extreme outliers are involved. Examining carefully the LoRaWAN data set that was deposited in the repository [21], over 80% of the RSS data values are at the extreme with the minimum RSS value of -200 dBm. Thus, exploring methods that will increase the location estimation accuracy of radio devices become germane. One of the possible way to achieve a better location estimation accuracy is to incorporate a method to handle outliers in our localization algorithm formulation process.

3.3. The M-iWMLR Algorithm with Weight Modification

The weight matrix in the M-iWMLR algorithm is modified through two-level normalization. On the first level, clipping normalization is used on the RSS data, which helps to cap the features (RSS values) above or below a certain value. Once clipping is realized, other forms of normalization can then be carried out and this will take us to the second level normalization. On the second level, z-score normalization is applied. Reason being that our working data after a careful clipping process have minimum number of outliers that are not extreme. Thus, the z-score normalization is best for this proposal as deduced from our previous results [19].

3.3.1. Clipping Normalization

Clipping in this context involves forming a set of new RSS values from the original RSS values by trimming the original RSS values. This process can be employed when there is need to scale or shift from the present scale to a scale closer to a value within a working data. It can be used to reduce the effect of outliers by reducing the range of the RSS values. As we observed, the RSS data in the repository have numerous outliers and more of the data lie within the lower scale [19].

In order to carry out data clipping on the original RSS values, there is need to set thresholds for the minimum (clip sub-peak) and maximum (clip peak) values. We first observe the range of the RSS values. The original RSS values ranged from -65 to -200 dBm [21]. We strongly assumed in this work that -200 dBm constituted outliers. This value was set for the gateways that could not receive a message from the transmitting device [6]. Before the -200 dBm that constituted the outliers, there are RSS values of -122 dBm that are not constituent of the outliers. Without the outliers, the RSS values range from -65 to -122 dBm. Thus, our minimum value can take up any value from -123 to -200 dBm and this can be achieved by adjusting the minimum value. We clipped the maximum at -65 dBm and started re-adjusting the minimum value to find the optimum value for better accuracy. The RSS clipping ($\mathrm{RSS}_{\mathrm{Clip}}$) is achieved using an expression presented in Equation (9):

$$\mathrm{RSS}_{\mathrm{Clip}} = \begin{cases} \text{max} & \text{if} \quad \text{RSS} > \text{max} \\ \text{RSS} & \text{if} \quad \text{min} \leq \text{RSS} \geq \text{max} \\ \text{min} & \text{if} \quad \text{RSS} < \text{min} \end{cases} \tag{9}$$

The clipping normalization process is presented in flowchart in Figure 1.

3.3.2. Z-score Normalization

The concept of z-score normalization has been used to form a weight matrix that was used to form the M-iWMLR location algorithm [19]. This concept is adopted for the nor-

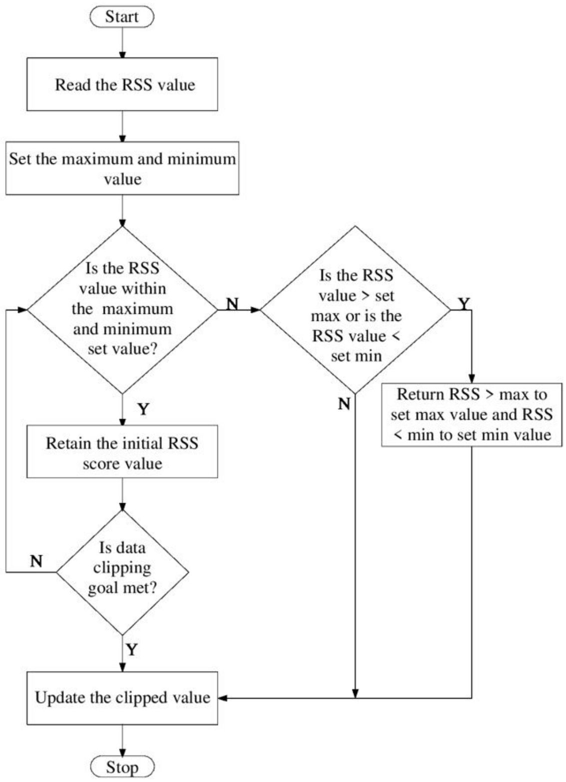

Fig. 1. Flowchart Illustrating a Program to Solve Clipping Normalization.

malization of the clipped RSS value. Using z-score normalization, the RSS value with high magnitude is allotted high value whereas that with low magnitude is allotted low value [19]. For the two-dimensional RSS data vector, the z-score normalization expression has been presented in Equation (8). The z-score normalization process to realize the second level normalization is presented in Figure 2.

The RSS data clipping followed by z-score normalization process results to a new matrix weight (W_{nom}). By substituting W_{nom} for $\widehat{Y}_{i,j}$ in Equation (7) lead to Equation (10).

$$\mathrm{L}_{\text{modified}} = \mathrm{R} + [W_{\mathrm{nom}}] \times \varnothing \tag{10}$$

When matrix, R as in Equation (3) through (6) is replaced with the matrix, $\mathrm{L}_{\text{modified}}$ of Equation (10) for the coordinates of x and y, the proposed location estimation algorithm in its compact form is presented as in Equation (11).

$$\left(\mathrm{Lat}_{\mathrm{Est}(x)}, \mathrm{Lon}_{\mathrm{Est}(y)}\right) = ([1\ \mathrm{L}_{\text{modified}}] \cdot \alpha, [\mathrm{L}_{\text{modified}}] \cdot \beta) \tag{11}$$

Where, $\mathrm{Lat}_{\mathrm{Est}(x)}$ and $\mathrm{Lon}_{\mathrm{Est}(y)}$ represents the respective estimated latitude and longitude of the radio terminal.

The flowchart for realizing the Ext.M-iWMLR terminal location algorithm is shown in Figure 3.

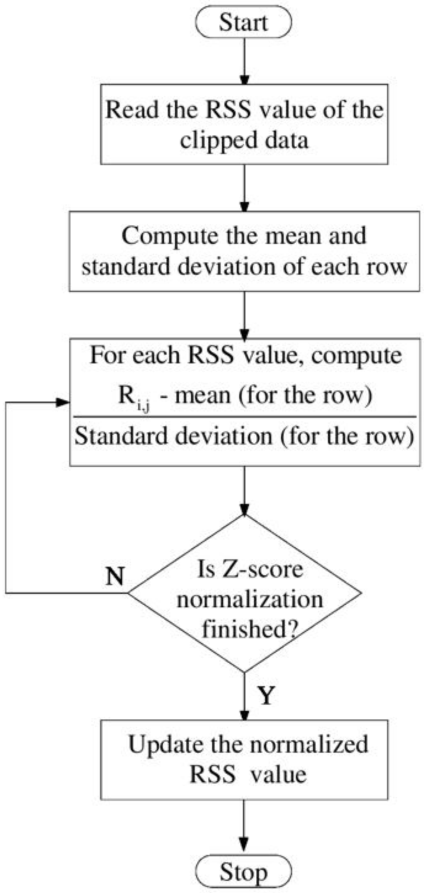

Fig. 2. Flowchart Illustrating a Program to Solve Z-score Normalization.

3.4. Simulations

As earlier explained in subsection 3.1, the empirical RSS data is divided into two; 70% of the data is for the algorithm formation while the 30% of the rest of the data is used to evaluate the formulated algorithm. The performance indices used for evaluating the proposed Ext.M-iWMLR algorithm include;

- Location Error: The distance between the measured coordinate and estimated coordinate is the location error. In this work, the mean location error is used.
- RMSE: This is the mean of the square of all the errors.
- Range of error: The difference between the maximum and minimum error values is referred to as range of error.
- R^2 score: This metric is used to investigate how closely the measured terminal position is to the fitted line of regression.

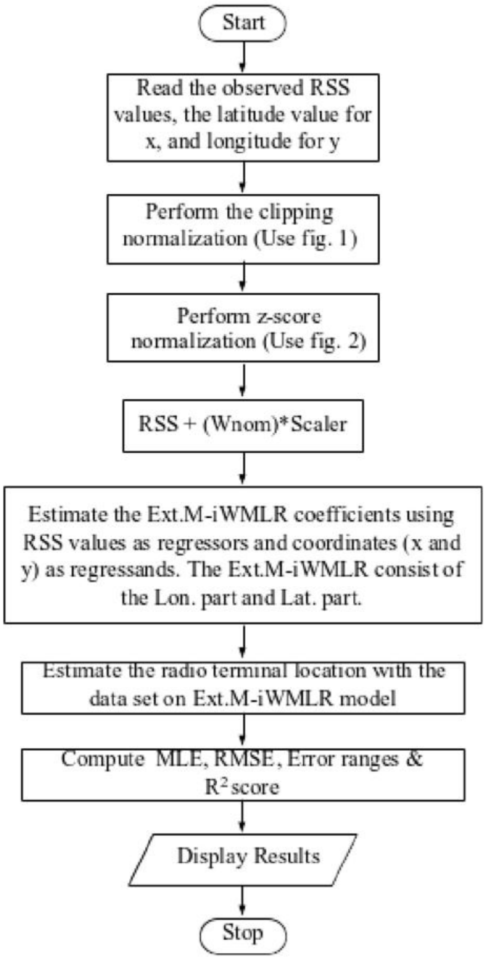

Fig. 3. Flowchart for Ext.M-iWMLR Terminal Location Algorithm.

The implementation of the Ext.M-iWMLR localization algorithm is done in MATLAB. The simulation is carried out under two scenarios. The first scenario covers the Ext.M-iWMLR algorithm simulation while the second scenario handles its validation.

In scenario one, only the Ext.M-iWMLR localization algorithm was simulated. The Ext.M-iWMLR algorithm is evaluated using the RSS information from LoRaWAN infrastructure. The location error, RMSE, range of error, R^2 score indicators are used to evaluate the Ext.M-iWMLR algorithm. For demonstrating the reliability of the Ext.M-iWMLR algorithm, 95% is the accuracy level used as benchmark. In the second scenario, the Ext.M-iWMLR localization algorithm is validated with the M-iWMLR algorithm. Both algorithms are allowed to run concurrently during validation. This is to ascertain the level of improvement the Ext.M-iWMLR algorithm has achieved and to further carry out a load test on the Ext.M-iWMLR algorithm.

4. Results discussions

In scenario one, we first present and discuss the results obtained while choosing the optimum clipped value. Without the outliers, the RSS values range from -65 to -122 dBm. Consequently, the minimum value can take up any value from -123 to -200 dBm and was achieved by adjusting the minimum value. The maximum was clipped at -65 dBm and started re-adjusting the minimum value in order to find an optimum value for better accuracy. It was started by increasing the -200 dBm by 5 steps (i.e., -200+5, -200+10, . . . , -200+75), resulting to -195, -190 ,. . . , -125 dBm. This preceding action was taken to scale the RSS value from its present scale to a scale closer to a value within a working data. Figure 4 is the plot of the clipped values against the mean location error (MLE). As the minimum value increased up to -130 dBm (i.e., -200+70), the MLE and the clipped values showed an inverse relationship. This results to a considerable decrease in the MLE leading to better accuracy. Exactly at point -125 dBm (i.e., -200+75), there was a very large increase in the MLE (5.14 x 10^4 m) that remained constant up to -123 dBm (i.e., -200+77).

To observe what happened from -130 to -125 dBm (i.e, -200+70 to -200+75), the RSS value was varied at -129, -128, -127, -126, -125 all in dBm and noticed that RSS values maintained normal relationship across -129 to -126 dBm and became abnormal at the exact point of -125 dBm. Figure 5 shows the plot of the clipped values against the MLE before the anomaly. To prevent the clipped value from having an adverse effect on the RSS values ranging from -65 to -122 dBm, -130 dBm (-200+70) was selected to be the optimum value and the minimum value is thus clipped at -130 dBm.

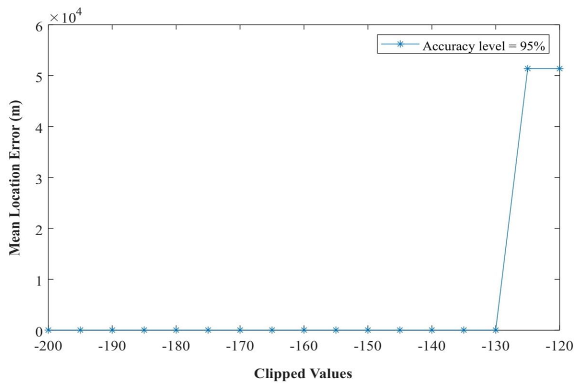

Fig. 4. Plot of the Clipped Values Against the Mean Location Error.

The simulation results from the Ext.M-iWMLR algorithm are now presented for analysis. Figure 6 is the plot

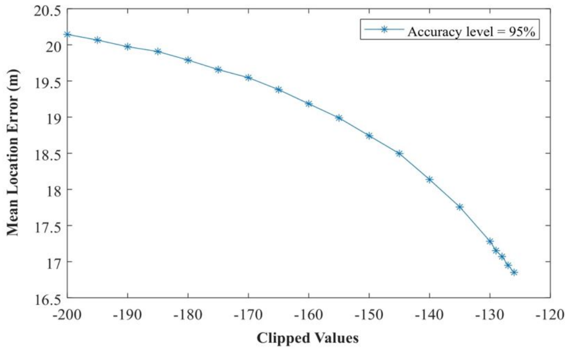

Fig. 5. Plot of the Mean Location Error against the Clipped Values before the Anomaly.

of the MLE for the Ext.M-iWMLR algorithm.

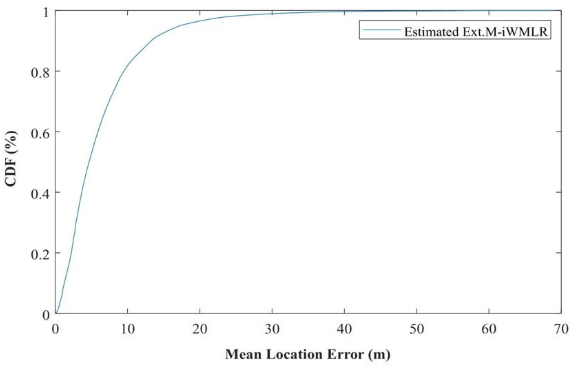

Fig. 6. Plot of MLE for extended M-iWMLR Algorithm.

Figure 6 shows the cumulative distribution frequency (CDF) against the MLE of the Ext.M-iWMLR algorithm. As observed, there is an exponential increase in the Ext.M-iWMLR algorithm MLE. At 95 percentile, the Ext.M-iWMLR algorithm recorded 17.2813m of MLE, RMSE of 6.4207m and error range of 68.5968m. Furthermore, the Ext.M-iWMLR algorithm recorded R^2 scores of 0.8958 and 0.7888 for the latitude and longitude, respectively.

In scenario 2, an insight on the level of improvement achieved by the Ext.M-iWMLR algorithm is demonstrated by validating the Ext.M-iWMLR algorithm with M-iWMLR algorithm. Figure 7 shows the plots of the CDF against the MLE of the M-iWMLR and the Ext.M-iWMLR algorithms. Table 1 summarized the result of the performance metrics used to compare the Ext.M-iWMLR algorithm and the M-iWMLR algorithm.

The plots in Figure 7 clearly show that the Ext.M-iWMLR algorithm outperformed significantly than the M-iWMLR algorithm. For MLE, the Ext.M-iWMLR algorithm and the M-iWMLR algorithm recorded 17.2813 m

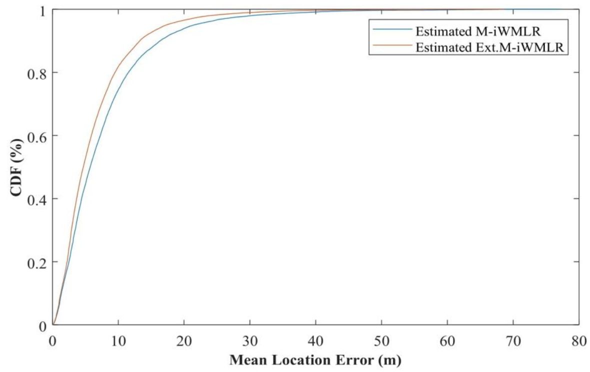

Fig. 7. Plots of MLE for Ext.M-iWMLR and M-iWMLR Algorithm.

and 21.4534 m, respectively. This implies that the Ext.M-iWMLR localization algorithm reduced the MLE of the device location by 19.45%. For the RMSE, the Ext.M-iWMLR algorithm and the M-iWMLR algorithm achieved 6.4207 m and 7.8249 m, correspondingly. This indicates that the proposed Ext.M-iWMLR algorithm has reduced the RMSE of the device location by 17.95%. Also, the Ext.M-iWMLR algorithm and the M-iWMLR localization algorithm recorded 68.5968 m and 77.1483 m range of errors, accordingly. It means that the Ext.M-iWMLR algorithm has enhanced the range of error of the device location by 11.08%. The R^2 scores for the latitude $\left(R_x^2\right)$ and longitude $\left(R_y^2\right)$ further established the effectiveness of the Ext.M-iWMLR algorithm over the M-iWMLR algorithm. In the same Table 1 , the achievable $\left(\mathrm{R}_\mathrm{x}^2\right)$ scores for the Ext.M-iWMLR and M-iWMLR algorithms are 0.8958 and 0.8474 , respectively. At the same time, the achievable $\left(\mathrm{R}_\mathrm{y}^2\right)$ scores for the Ext.M-iWMLR and M-iWMHR algorithms are 0.7888 and 0.6732 , correspondingly. Usually, the algorithm with the highest R^2 score performs better than the other algorithm. It indicates that the Ext.M-iWMLR algorithms achieved 5.71% and 17.17% improvement in terms of R_x^2 and R_y^2, respectively. It is worthy to note that at 0 percentile as shown in Figure 7, the respective metric values of MLE for the Ext.M-iWMLR and M-iWMLR algorithms are the same. Reason being that at the clipped value of (-200+0), the Ext.M-iWMLR algorithm behaves like M-iWMLR algorithm. Also, there is no effect of clipping on the Ext.M-iWMLR algorithm at that point.

In addition, the range of error values achieved by the Ext.M-iWMLR algorithm was used to evaluate its reliability. Conventionally, the smaller the range of error the better the reliability. As observed from Table 1, the Ext.M-iWMLR algorithm achieved smaller range of error than M-iWMLR algorithm. Thus, the Ext.M-iWMLR algorithm

Table 1. MLE, RMSE, Range of Error, and R^2 scores for the Ext.M-iWMLR and M-iWMLR Algorithms with LoRaWAN at 95 Percentile.

				$\mathbf{R}^2$ scores	
Algorithms	MLE (m)	RMSE (m)	Range of Error (m)	$\mathbf{R}_x^2$	$\mathbf{R}_y^2$
M-iWMLR	21.4534	7.8249	77.1483	0.8474	0.6732
Ext.M-iWMLR	17.2813	6.4207	68.5968	0.8958	0.7888

is more reliable for use in estimation of a terminal device in LoRaWAN.

Load test was also carried out on the Ext.M-iWMLR algorithm to further drive home the enhancement of the proposed Ext.M-iWMLR algorithm over the M-iWMLR algorithm. The load test is a type of performance test carried out to observe the behavior of the Ext.M-iWMLR algorithm under expected load. Here we investigated the performance of Ext.M-iWMLR algorithm and the M-iWMLR algorithm at 100 percentile using MLE metric. At 100 percentile, the plots in Figure 7 recorded the MLE values of 68.6201m and 77.1557m from the Ext.M-iWMLR algorithm and the M-iWMLR algorithm, respectively. For clarity, the result of the load test is presented in plot of Figure 8. It indicates that at 100 percentile, the Ext.M-iWMLR algorithm also reduced the MLE of the device location by 11.06%.

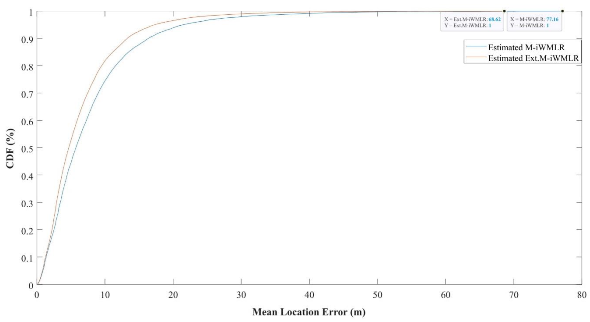

Fig. 8. Plot of Load test on the Ext.M-iWMLR and M-iWMLR Algorithm at 100 percentile.

5. Conclusions

In this paper, the Ext.M-iWMLR algorithm for radiolocation estimation with LoRaWAN is presented. The goal centered on achieving a radiolocation estimation algorithm with better location accuracy by reducing location error. This was achieved by; thorough investigation of a very large RSS LoRaWAN data set, using data clipping to reduce the effects of outliers on the location accuracy, normalizing the clipped data with z-score method, forming a new weight matrix and the M-iWMLR algorithm modification with the influential new weight matrix. The new weight matrix was formed with the data obtained after z-score normalization. The Ext.M-iWMLR algorithm implementation is achieved in MATLAB. Also, the M-iWMLR algorithm was used to validated the Ext.M-iWMLR algorithm. The validation was successfully done by allowing the M-iWMLR and Ext.M-iWMLR algorithms to run concurrently. Four key performance indicators namely location errors, RMSE, range of error, and R^2 scores are used to assess the Ext.M-iWMLR algorithm. The accuracy level was benchmarked at 95% when evaluating the Ext.M-iWMLR algorithm perfomance. Results shows that the proposed Ext.M-iWMLR algorithm reduced the MLE by 19.45%. The RMSE and the range of error are reduced by 17.95% and 11.08%, respectively. Furthermore, the R^2 scores are increased by 5.71% and 17.17% at latitude and longitude coordinates, correspondingly. The obtained range of error value in addition divulge the reliability of the Ext.M-iWMLR algorithm. Finally, load test was carried out on the Ext.M-iWMLR and M-iWMLR algorithms at 100 percentile. It was observed that the proposed Ext.M-iWMLR algorithm reduced the MLE of the device location by 11.06%.

One major significant advantage of Ext.M-iWMLR algorithm is that it considered a method of handling outliers during the algorithm formation process. Certainly, the Ext.M-iWMLR algorithm improved greatly the location accuracy of radio terminals. However, improving radio localization accuracy should be an incessant process since many IoT and other smart devices with location capabilities are now used for various location based applications and services. Future work will try to investigate the time complexity of the Ext.M-iWMLR algorithm with other variant of MLR based location algorithms.

Acknowledgments

For making availabile the data used in this work, the authours thank M. Aernouts et al., The authors also appreciate the reviewers for the constructive criticism done in this work.

References

[1] W. Meng, W. Xiao, and L. Xie, (2011) *"An efficient EM algorithm for energy-based multisource localization in* *wireless sensor networks"* **IEEE Trans. Instrum. Meas** ***60***(3): 1017–1027. DOI: 10.1109/TIM.2010.2047035.

[2] A. Yassin, Y. Nasser, M. Awad, and A. Al-dubai, (2016) *"Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications"* **IEEE communications surveys & tutorials** ***19***(2): 1327–1346. DOI: 10.1109/COMST.2016.2632427.

[3] L. Lee, M. Jones, G. S. Ridenour, S. J. Bennett, A. C. Majors, B. L. Melito, and M. J. Wilson. "Comparison of accuracy and precision of GPS-enabled mobile devices". In: *Proc. - 2016 16th IEEE Int*. 2016 6th Int. Symp. Cloud Serv. Comput. IEEE SC2 2016 2016 Int. Symp. Secur. Priv. Soc. Netwo, 36: Conf. Comput. Inf. Technol. CIT 2016, 2017, 73–82.

[4] M. K. Ardakani, S. Mohammad, and H. Reza, (2020) *"A hybrid adaptive approach to improve position tracking measurements"* **ICT Express** ***6***(4): 273–279. DOI: 10.1016/j.icte.2020.05.012.

[5] M. Aernouts, R. Berkvens, V. K. Van, and M. Weyn, (2020) *"Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas"* **MDPI Data**: 1–15. DOI: 10.3390/data3020013.

[6] T. Janssen, R. Berkvens, M. Weyn, T. Janssen, R. Berkvens, and M. Weyn, (2020) *"Benchmarking RSS-based Localization Algorithms with LoRaWAN"* **Internet of Things**: 1–29. DOI: 10.1016/j.iot.2020.100235.

[7] B. E. N. Buurman, J. Kamruzzaman, and S. Member, (2020) *"Low-Power Wide-Area Networks: Design Goals, Architecture, Suitability to Use Cases and Research Challenges"* **IEEE Access** ***8***(2020): 17179–17220.

[8] Y. Roth, J. Doré, L. Ros, and V. Berg, (2018) *"The Physical Layer of Low Power Wide Area Networks: Strategies, Information Theory's Limit and Existing Solutions"* **Advances in Signal Processing: Reviews** ***1***:

[9] Y. H. S. Xie and Y. Wang, (2014) *"An Improved E-Min-Max Localization Algorithm in Wireless Sensor Networks"* **IEEE**: 1–4.

[10] L. S. Ezema and C. I. Ani, (2020) *"Weighted Multiple Linear Regression Model for Mobile Location Estimation in GSM Network"* **International Journal of Interdisciplinary Telecommunications and Networking** ***12***(1): 57–69. DOI: 10.4018/IJITN.2020010105.

[11] L. S. Ezema and C. I. Ani, (2017) *"Artificial Neural Network Approach to Mobile Location Estimation in GSM Network"* **International Journal of Electronics and Telecommunications** ***63***(1): 39–44. DOI: 10.1515/eletel-2017-0006.

[12] L. S. Ezema and C. Ani, (2018) *"Multi Linear Regression Model for Mobile Location Estimation in GSM Network"* **Indian Journal of Science and Technology** ***9***(6): 1–6. DOI: 10.17485/ijst/2016/v9i6/75195.

[13] U. N. Nwawelu, M. A. Ahaneku, and B. O. Ezurike, (2022) *"Improving Weighted Multiple Linear Regression Algorithm for Radiolocation Estimation in LoRaWAN"* **International Journal of Interdisciplinary Telecommunications and Networking** ***14***(1): 1–12. DOI: 10.4018/IJITN.299369.

[14] L. S. Ezema and C. Ani, (2019) *"Mobile Location Estimation in GSM/UMTS"* **International Journal of Emerging Technology & Research** ***1***(3): 63–70.

[15] U. N. Nwawelu, A. N. Nzeako, and M. A. Ahaneku, (2012) *"The Limitations of Campus Wireless Networks: A Case Study of University of Nigeria, Nsukka [Lionet]"* **International Journal of Networks and Communications** ***2***(5): 112–122. DOI: 10.5923/j.ijnc.20120205.04.

[16] U. N. Nwawelu, A. N. Nzeako, M. A. Ahaneku, and V. C. Chijindu, (2014) *"Performance Evaluation of the Lionet Quality of Service in Nsukka Campus of the University of Nigeria"* **Int. J. Commun. Netw. Syst. Sci** ***7***(4): 122–129. DOI: 10.4236/ijcns.2014.74014.

[17] M. A. Ahaneku, A. N. Nzeako, and U. N. Nwawelu, (2014) *"Assessment of Radiation Variations with Distance in the Vicinity of GSM Base Stations Antenna"* **International Journal of Scientific & Engineering Research** ***5***(4): 639–645.

[18] M. A. Ahaneku, A. N. Nzeako, and U. N. Nwawelu, (2015) *"Investigation of Electromagnetic Radiations by GSM Base Stations in Nigeria for Compliance Testing"* **Advances in Physics Theories and Applications** ***47***: 10–18.

[19] U. N. Nwawelu and M. A. Ahaneku, (2022) *"A Holistic Approach of Achieving Accurate Radio Location Estimation in Long Range Wide Area Network"* **International Journal of Interdisciplinary Telecommunications and Networking** ***14***(1): 1–14. DOI: 10.4018/IJITN.312256.

[20] F. C. C. (FCC). *Wireless E911 Location Accuracy Requirements*. Federal Register. location-accuracy-requirements, 2014.

[21] M. Aernouts, R. Berkvens, V. K. Van, and M. Weyn. "Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas (1.3): [Data Set].Zenodo". In: 2019.