



Dealing with Outliers in Wireless Sensor Networks Localization: An Iterative and Selection-Minimization Strategy

Boris Mederos-Madrazo¹ · Jose Diaz-Roman² · Francisco Enriquez-Aguilera² · Juan Cota-Ruiz²

Received: 8 December 2023 / Accepted: 4 April 2024 / Published online: 26 April 2024
© The Author(s) 2024

Abstract

In recent years, there has been considerable interest in robust range-based Wireless Sensor Network (WSN) localization due to the increasing importance of accurately locating sensors in various WSN applications. However, achieving precise localization is often hampered by the presence of outliers or underestimations in range measurements, particularly when employing the RSS technique. To tackle these issues, we introduce a Two-Step Localization (namely, the SelMin approach). In the initial phase, the approach utilizes Second-Order Cone Programming (SOCP) to minimize distance discrepancies. It does this by comparing a reference Euclidean Distance Matrix (EDM) with a weighted one derived from imprecise distances between sensor nodes. In the subsequent phase, a heuristic method is employed to identify a specific number of imprecise distances, referred to as outliers, that will be disregarded in the first phase, and this two-phase process continues iteratively. The experimental results demonstrate that the SelMin strategy performs better than the DSCL method when evaluated using the Root Mean Square Error (RMSE) metric. This superior performance is maintained even in challenging conditions, such as when there are many outliers (i.e., around 30%) in the network. This indicates that SelMin is a reliable and robust choice for these environments.

Keywords Localization · Outliers-handling · SOCP

1 Introduction

A wireless sensor network (WSN) is comprised of several sensors (nodes) that are interconnected and can process and exchange data with each other. These nodes are deployed to

collect data for a variety of purposes, including exploration, monitoring, surveillance, and maintenance, to mention a few.

The necessity to know the position of the nodes must be handled since in many applications, the data must be linked to a location to be relevant such as with sensors used to monitor forest fires, military surveillance, etc. Placing GPS devices on each individual node and manually configuring their positions can be challenging, especially for large networks with thousands of sensors. This is mainly due to the high expenses associated with the sensors. To tackle this problem, a solution is to outfit a subset of sensors with GPS (referred to as anchors), and then estimate the locations of the other nodes by making use of these known anchor nodes and wireless communications. This approach helps overcome the impracticality caused by cost and scale.

Range-based methods in Wireless Sensor Networks (WSNs) are characterized by their reliance on establishing the distances between nodes within the network. Typically, these approaches involve the computation or measurement of the physical separation between sensor nodes. This aspect holds significance in various WSN applications, such as tracking, localization, and monitoring. Within range-based strategies,

Jose Diaz-Roman, Francisco Enriquez-Aguilera and Juan Cota-Ruiz authors have contributed equally to this work.

✉ Juan Cota-Ruiz
jcota@uacj.mx

Boris Mederos-Madrazo
boris.mederos@uacj.mx

Jose Diaz-Roman
david.roman@uacj.mx

Francisco Enriquez-Aguilera
fenrique@uacj.mx

¹ Department of Mathematics, UACJ, Ave. del Charro 450 Nte., Cd. Juarez 32320, Chihuahua, Mexico

² Department of Electrical and Computer Engineering, UACJ, Ave. del Charro 450 Nte., Cd. Juarez 32320, Chihuahua, Mexico

the task of establishing the location can be achieved by employing methods like analyzing Received Signal Strength (RSS), evaluating Time of Arrival (TOA), measuring the round trip duration, or even utilizing a fusion of these techniques [1]. These methodologies rely on estimating sensor distances to conclude the process of positioning [2].

Once the distances between sensor nodes are acquired, the localization procedure can be carried out through either a central processing unit or a distributed processing approach. The former proves to be less effective in extensive networks due to constraints in information transfer, insufficient processing capabilities, and unequal energy usage. On the contrary, the latter method is more adaptable and exhibits great scalability. Nonetheless, distributed processing comes with its own disadvantages, primarily its iterative nature. This characteristic can lead to heightened overall power consumption due to the need for a significant number of iterations for rapid convergence. Additionally, this approach can be susceptible to the propagation of errors, adding to its limitations [3].

It's commonly understood that in range-based localization algorithms, the process of determining the positions of unknown sensors is closely tied to estimating the distances between them. This connection means that an unusual or non-typical distance estimation between sensors has the potential to lead to undesirable outcomes in the localization process [4–9]. This is why numerous studies have been conducted to address this scenario, with the goal of developing methods to filter, remove, or mitigate the impact of these undesired distances (outliers) on the localization procedure. For example, The study in [10] explores the use of the weighted three minimum distances method (WTM), which considers three minimal distances to enhance atypical RSSI measurements for localization. By utilizing an experimental channel model and a new location optimization formula, grounded in multilateration and distance estimation, this technique strives for more accurate positioning. Despite requiring more time for computation compared to other established algorithms like Semidefinite Programming (SDP), Levenberg-Marquardt (LM), and nonlinear least squares (NLS), it achieves more precise estimations. Also, in [4], the authors proposed an approach for removing atypical distance measurements. They achieve this by applying geometric constraints from the Cayley-Menger determinant and utilizing Euclidean norm, the 1-norm, minimization decoding algorithm. Large errors are initially identified with the assistance of neighboring anchor nodes connected to regular nodes. Subsequently, these significant errors are filtered out using a predefined threshold, resulting in the determination of the best-estimated positions for the unknown nodes. In the same way, in [7], initially, the Frobenius norm and the 1-norm (i.e., L1) are used to establish a framework for reconstructing an Euclidean Distance Matrix (EDM) that contains noise, outliers, and missing

values. This process is referred to as Normalized Matrix Completion (NRMC). Subsequently, an efficient algorithm using an alternative address multiplier approach is developed to address the NRMC problem. Lastly, a multidimensional scaling method is utilized to determine the positions of unidentified nodes based on a complete EDM. Also, in [5], a novel approach utilizing the least squares (LS) method is introduced for Wireless Sensor Network (WSN) localization. This method involves determining the location by assessing a condition number derived from a coordinate matrix, which effectively prevents the presence of outliers. Also, the research introduces the concept of a condition number threshold to enhance location accuracy while mitigating the impact of outliers.

In the majority of the analyzed approaches, it's crucial to keep the occurrence of outliers relative to the total number of measurements in the network as low as possible to ensure favorable outcomes [11]. For instance, the study development in [12] introduces a method named the greedy search-based random sample consensus (GS-RSC) algorithm for identifying non-line-of-sight (NLOS) measurements. This algorithm, through its dual components—the localization and identification modules—along with comprehensive time-difference-of-arrival (TDOA) data, effectively reduces errors from outliers and NLOS paths. By iteratively working in tandem, these modules accurately weed out irregular data, thus ensuring highly accurate localization solutions that are less affected by both outliers and NLOS-path errors. Furthermore, it's important to take into account scenarios where outliers are caused by either a multiplying factor or a mitigating factor affecting the actual measurement value. For example, when estimating distances in wireless sensor networks using RSS readings, there are notable disadvantages. Ambient conditions such as obstructions and multipath propagation can affect signal strength, leading to inaccurate and atypical distance estimations. Variations in hardware can alter RSS signal strength, complicating the standardization of measurements across devices. Additionally, this method's accuracy and reliability are constrained by its dependence on a predetermined model that correlates signal strength with distance, which might not accurately represent every deployment environment. To address this challenge, the proposed method, comprising two sequential phases, is an iterative process that progressively selects (or eliminates) anticipated outlier distances during each iteration. In the initial stage, an optimization method is employed to minimize the discrepancy between two EDMs, one serving as the reference and the other as the estimated EDM. In the second phase of the proposed approach, a heuristic method is utilized to eliminate the estimated distances between sensors that exhibit significant errors. It should be noted that at each iteration, the EDM is converted to positions using the MDS algorithm.

The remainder of this paper unfolds as follows: Sect. 2 describes the mathematical representation of the problem.

Section 3 provides an overview of the proposed algorithm, covering both phases of the procedure: the minimization and selection of outliers. In Sect. 4, the experiment’s outcomes are presented, where the quantity of outlier measurements is systematically raised. Ultimately, Sect. 5 offers concluding remarks.

2 WSN Localization: Mathematical formulation problem

This research limits the discussion only to two dimensions (2D) for node localization in WSNs. Consider n randomly deployed sensor nodes over a certain area denoted as $\mathbf{S} = \{s_1, s_2, s_3, \dots, s_m, s_{m+1}, \dots, s_n\}$ with true locations

$$\begin{aligned} \mathbf{P} &= [\mathbf{p}_1 = [x_1, y_1]^T, \mathbf{p}_2 = [x_2, y_2]^T, \\ \mathbf{p}_3 &= [x_3, y_3]^T, \dots, \mathbf{p}_m = [x_m, y_m]^T, \\ \mathbf{p}_{m+1} &= [x_{m+1}, y_{m+1}]^T, \dots, \mathbf{p}_n = [x_n, y_n]^T] \end{aligned}$$

respectively; the first m sensor nodes have unknown positions whose position estimates are represented as

$$\begin{aligned} \bar{\mathbf{P}} &= [\bar{\mathbf{p}}_1 = [x_1, y_1]^T, \bar{\mathbf{p}}_2 = [x_2, y_2]^T, \\ \bar{\mathbf{p}}_3 &= [x_3, y_3]^T, \dots, \bar{\mathbf{p}}_m = [x_m, y_m]^T] \end{aligned}$$

and the remainder $n - m$ sensor nodes have known positions, obtained by GPS or a similar scheme (i.e., Anchors), with $mn - m$. Let all sensor nodes have connectivity with any other sensor nodes in network, and the noisy distance r_{ij} between sensor nodes i and j can be estimated by methods like RSS, ToA, AoA or combination of them [13–16], where $(i, j) \in \mathbf{S} \times \mathbf{S}$. The sets of undirected neighbors between sensors-sensors, sensors-anchors, and anchors-anchors can be described as

$$\begin{aligned} \mathbf{C} &:= \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq m, i \neq j\}, \\ \mathbf{K} &:= \{(k, l) : 1 \leq l \leq m, m + 1 \leq k \leq n, l \neq k\}, \end{aligned}$$

and

$$\mathbf{A} := \{(i, j) : m + 1 \leq i \leq n, m + 1 \leq j \leq n, i \neq j\},$$

respectively. Here,

$$\|\mathbf{p}_i - \mathbf{p}_j\| = d_{ij}, \forall (i, j) \in \mathbf{C},$$

where d_{ij} represents the true distance between sensors i and j , and $\|\cdot\|$ denotes the 2-norm. Similarly, $\|\mathbf{p}_l - \mathbf{p}_k\| = d_{lk}, \forall (l, k) \in \mathbf{K}$ where d_{lk} represents the true

distance between the sensor l and the anchor k . Then, the position of unknown sensors (i.e., the first m nodes) can be estimated as the following non-convex optimization function [17–21].

$$\min_{\bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_m} \sum_{(i,j) \in \mathbf{C}} (\|\bar{\mathbf{p}}_i - \bar{\mathbf{p}}_j\| - r_{ij}\hat{o}_{ij})^2 + \sum_{(k,l) \in \mathbf{K}} (\|\bar{\mathbf{p}}_k - \mathbf{p}_l\| - r_{kl}\hat{o}_{kl})^2 \tag{1}$$

where the noisy estimated distance between sensors i and j (or sensor l and anchor k) is composed as

$$r_{ij} = d_{ij} + e_{ij}, \tag{2}$$

$$r_{kl} = d_{kl} + e_{kl}, \tag{3}$$

respectively, and e_{ij} and e_{kl} have normal distribution $\mathcal{N}(0, sd^2)$, which represent the noise introduced by the environment and the ranging measurement techniques, and sd is the standard deviation. Also, in many cases, a range measurement between sensor s_i and sensor s_j lies in an abnormal distance (i.e., an outliers) introduced as \hat{o}_{ij} that represents a random variable. The last problem can be established as a geometric undirected graph $\mathbf{G} = (\mathbf{V}, \mathbf{L})$ where \mathbf{V} and \mathbf{L} represent the set of vertices (i.e., sensor nodes \mathbf{S}) and the set of edges (i.e., links among sensor nodes) $\mathbf{L} = \{(i, j) | (i, j) \in \mathbf{V} \times \mathbf{V}, i \neq j\}$, respectively. Then, the adjacency matrix (i.e., symmetric with zeros on its diagonal) of size $N \times N$ can be represented as $\mathbf{MC} = (\mathbf{MC}_{ij})$, where

$$\mathbf{MC}_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in \mathbf{L} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and the matrix distance, derived as weights of \mathbf{MC} , can be denoted as $\mathbf{D} = (\mathbf{D}_{ij})$ with

$$\mathbf{D}_{ij} = \begin{cases} r_{ij}\hat{o}_{ij} & \text{if } \{i, j\} \in \mathbf{L} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Equation (1) can be redeveloped in a convex form using relaxation as follows:

$$\begin{aligned} \min_{\mathbf{Y}} & \|\mathbf{D}^2 - K(\mathbf{Y})\|^2 \\ \text{st} & : \mathbf{Y} \geq 0 \end{aligned} \tag{6}$$

where the symmetric matrix $\mathbf{Y} = \bar{\mathbf{P}}\bar{\mathbf{P}}^T$ is constrained to be positive semidefinite, which is denoted by the expression ≥ 0 , and the operator $K(\mathbf{Y}) = \text{diag}(\mathbf{Y})\mathbf{1}^T - 2\mathbf{Y} + \mathbf{1}\text{diag}(\mathbf{Y})^T$ [22]. Once obtained an estimation of \mathbf{D} by a SDP or SOCP programming [23–26], a transformation from distances to positions using the MDS method must be employed [27, 28].

3 The Selection and Minimization Algorithm

As aforementioned in Sect. 2, range based localization algorithms are strongly dependent of accuracy on range estimates, so atypical range measurements can provide large errors on position estimates. In order to mitigate the effect of outliers, a centralized localization approach is proposed, and it is divided into two schemes: the minimization and outliers-selection processes. Here, the minimization stage is derived from Eq. (6) and can be rewritten as

$$\mathbf{Y}^{i+1} = \arg \min_{\mathbf{Y}^i} \left\| (\mathbf{D}^2 - K(\mathbf{Y}^i)) \odot \mathbf{O}^i \right\|^2 \quad (7)$$

st : $\mathbf{Y}^i \geq 0$

where the symbol \odot denotes the Hadamard product between the matrices $\mathbf{D}^2 - K(\mathbf{Y}^i)$ and \mathbf{O}^i .

In the optimization process (7), $\mathbf{O}^i = (o_{rs}^i)$ represents a symmetric matrix of ones and zeros whose indexes of 1's indicate those range measurements r_{rs} with $\{r, s\} \in \mathbf{L}$ with non potential outliers. Ideally, without *outliers* in range measurements $\mathbf{O}^i \equiv \mathbf{MC}$ and with *outliers* in range measurements for instance $(s, t) \in \mathbf{L}$ leads to $o_{rs}^i = 0$, then $\#\{\{r, s\} \in \mathbf{L} : o_{rs}^i = 1\} < \#\mathbf{L}$. Initially, at the iteration $i = 0$, $\mathbf{O}^i \equiv \mathbf{MC}$, and through iterations, the selection process (i.e., the second stage) modifies the matrix \mathbf{O}^i with zeros in those indexes $\{i, j\} \in \mathbf{L}$ with potential outliers. In (7), the Hadamard product between the matrix \mathbf{O}^i and the term $\mathbf{D}^2 - K(\mathbf{Y}^i)$ allows the minimization process focuses only on those range measurements considered as non outliers. This kind of optimization problems with semidefinite constraints can be solved using a software like SEDUMI [29]. Once the optimization process (7) has finished, a new square distance matrix \mathbf{Y}^{i+1} is generated, and it will be used in the selection process.

In the second stage (i.e., the selection process), the distance errors between \mathbf{Y}^{i+1} and \mathbf{D}^2 are weighted by \mathbf{O}^i as shown in the Eq. (8). Then, the distance errors between sensors obtained in (8) are normalized as indicated by Eq. (9)

$$\mathbf{W}_{ed} = \left(\left| \mathbf{Y}^{i+1} - \mathbf{D}^2 \right|^{1/2} \right) \odot \mathbf{O}^i, \quad (8)$$

$$\bar{\mathbf{W}}_{ed} := \frac{\mathbf{W}_{ed}}{\max(\mathbf{W}_{ed})} = \begin{bmatrix} \frac{\mathbf{W}_{ed_1}}{m_1} \\ \frac{\mathbf{W}_{ed_2}}{m_2} \\ \vdots \\ \frac{\mathbf{W}_{ed_n}}{m_n} \end{bmatrix}, \quad (9)$$

$$m_i = \max(\mathbf{W}_{ed_i}), \quad i = 1, \dots, n,$$

where \mathbf{W}_{ed_i} , $i = 1, \dots, n$ denotes the i^{th} row of \mathbf{W}_{ed} and the term $\max(\mathbf{W}_{ed_i})$ refers to its maximum component m_i which in turn corresponds to the maximum distance error of it. Each row of the $\bar{\mathbf{W}}_{ed}$ matrix is ultimately weighted with a maximum value of 1, which corresponds to the maximum distance error.

Utilizing the weight matrix $\bar{\mathbf{W}}_{ed}$ and the procedure denoted as $[\cdot]_x$ in (10), the x higher values that display notable distance errors, in each row of the matrix, are isolated. Here, $x = \text{round}(\rho * n)$ where $\rho \in [0, 1]$ represents a percentage of the total n sensors to be rule out. To elaborate, for each sensor, the process identifies its x neighboring sensors that possess the highest errors and convert them to 0s within a binary connectivity matrix \mathbf{B} . Meanwhile, the remaining sensors maintain a value of 1, indicating their role as non outlier measures within the localization process.

$$\mathbf{B} = [\bar{\mathbf{W}}_{ed}]_x. \quad (10)$$

Lastly, the outlier matrix \mathbf{O}^i is modified for the subsequent iteration, as depicted in Equation (11). Thus, after finishing the selection process, a new refined binary matrix \mathbf{O}^{i+1} is provided to the minimization process to start the iterative process again. It is important to note that within this approach, once the estimated distance with a neighboring sensor is recognized as an outlier, it ceases to contribute in the localization process. Also, for each sensor, having at least a minimum of α measurements with nearby sensors is necessary; otherwise, this condition prevents the removal of additional neighboring sensors. Algorithm 1 summarizes this selection-minimization scheme in detail.

$$\mathbf{O}^{i+1} = \mathbf{O}^i \odot \mathbf{B}. \quad (11)$$

Algorithm 1 The SelMin algorithm that minimizes error distances among sensor nodes using an alternation process.

Require: $\mathbf{D}, \mathbf{Y}^0, \mathbf{O}^0, \rho = .15, \mathbf{A}, \alpha = 6, \text{Thr} = e - 10, e_0 = e_{-1} = 0.$

Ensure: \mathbf{Y}^{i+1}

- 1: **Initialize:** $i = 0$
- 2: **while** $i \leq 15$ or $|e_i - e_{i-1}| \leq \text{Thr}$ **do**
- 3: $\mathbf{Y}^{i+1} = \arg \min_{\mathbf{Y}^i} \left\| (\mathbf{D}^2 - K(\mathbf{Y}^i)) \odot \mathbf{O}^i \right\|^2$
- 4: **subject to :** $\mathbf{Y}^i \succeq 0$
- 5: $\mathbf{W}_{ed} = \left(|\mathbf{Y}^{i+1} - \mathbf{D}^2|^{1/2} \right) \odot \mathbf{O}^i$
- 6: $\bar{\mathbf{W}}_{ed} = \frac{\mathbf{W}_{ed}}{\max(\mathbf{W}_{ed})}$
- 7: $\mathbf{B} = [\bar{\mathbf{W}}_{ed}]_x$
- 8: $\mathbf{O}^{i+1} = \mathbf{O}^i \odot \mathbf{B}$
- 9: $\bar{\mathbf{P}}^{i+1} = \text{MDS} \left((\mathbf{Y}^{i+1})^{1/2}, \mathbf{A} \right)$
- 10: $e_{i+1} = \left\| K(\mathbf{Y}^{i+1}) - K(\mathbf{Y}^i) \right\|_F$
- 11: $i = i + 1$
- 12: **end while**

It is worth mentioning that in row 9 of algorithm 1, the multi-dimensional scaling (MDS) procedure is used to convert the new estimated distance matrix into estimated sensor positions useful to compute the RMSE during the simulation process. Here, \mathbf{A} represents the set of anchor positions in the network, and ρ is heuristically selected with a value of 0.15.

4 Analysis and Results

To evaluate the performance of the proposed algorithm, a test methodology has been created where 30 independent sensor networks are used. Each network contains 50 nodes, all interconnected in one hop, where only three nodes are

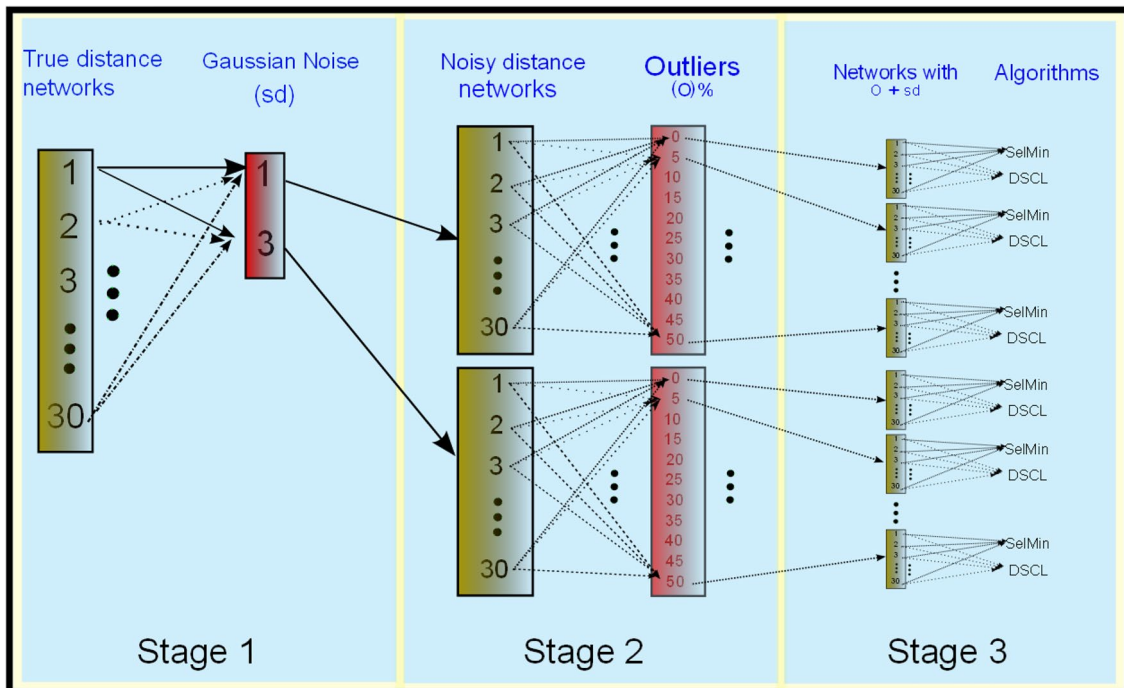


Fig. 1 Test methodology with 30 reference networks, each with 50 nodes and three reference anchors. The estimated distances between each sensor are affected by Gaussian noise and atypical errors (outliers)

non-collinear anchors. These sensor nodes are distributed randomly and independently in an area of [250 m x 250 m]. Taking into account the previous description, the process to generate the test networks is divided into three stages as shown in Fig 1.

In a first stage, without loss of generality, all the true distances between each of the sensors in network 1 are affected by a Gaussian noise level (with zero mean) and two standard deviations ($sd = 1$ and $sd = 3$) determined by

equations (2-3). This last process generates two independent test networks of network 1. This methodology is repeated for each of the remaining networks, generating two sets of 30 independent networks that will be used in stage two. In stage 2, outliers are randomly added to the estimated distances between sensors. These added outliers range from 0% to 50% in steps of 5. That is, each percentage level of outliers affects all 30 benchmark networks (a test set), so there will be 22 independent test sets for stage 3. It is important to mention

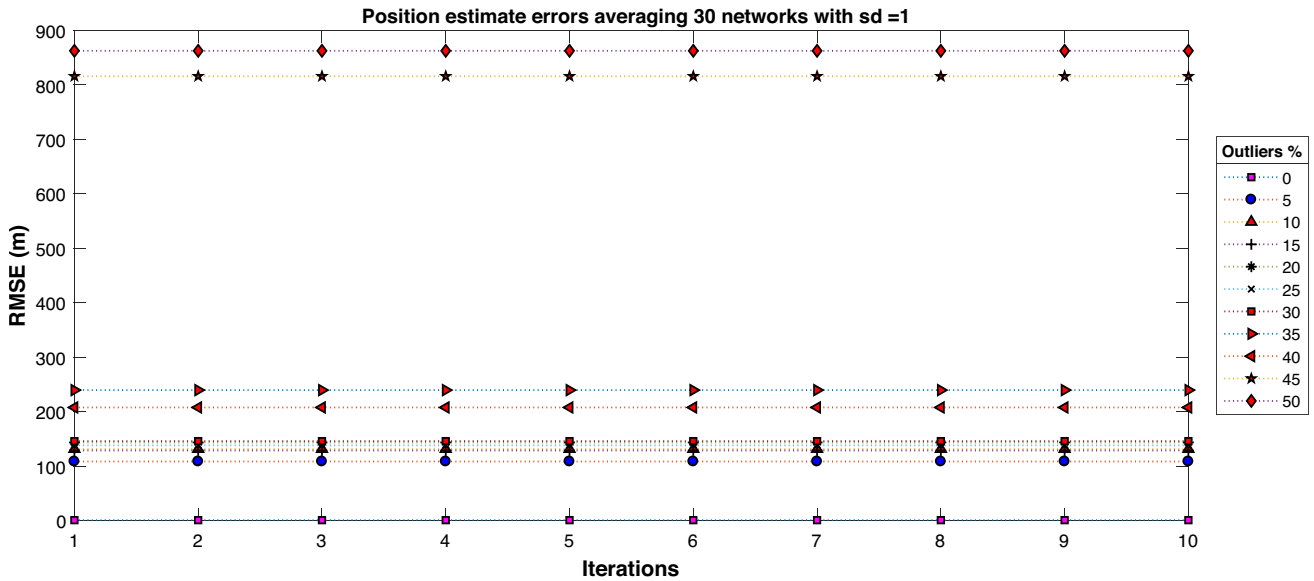


Fig. 2 Accuracy performance of the minimization scheme at different levels of outliers on range estimates and a standard deviation of one

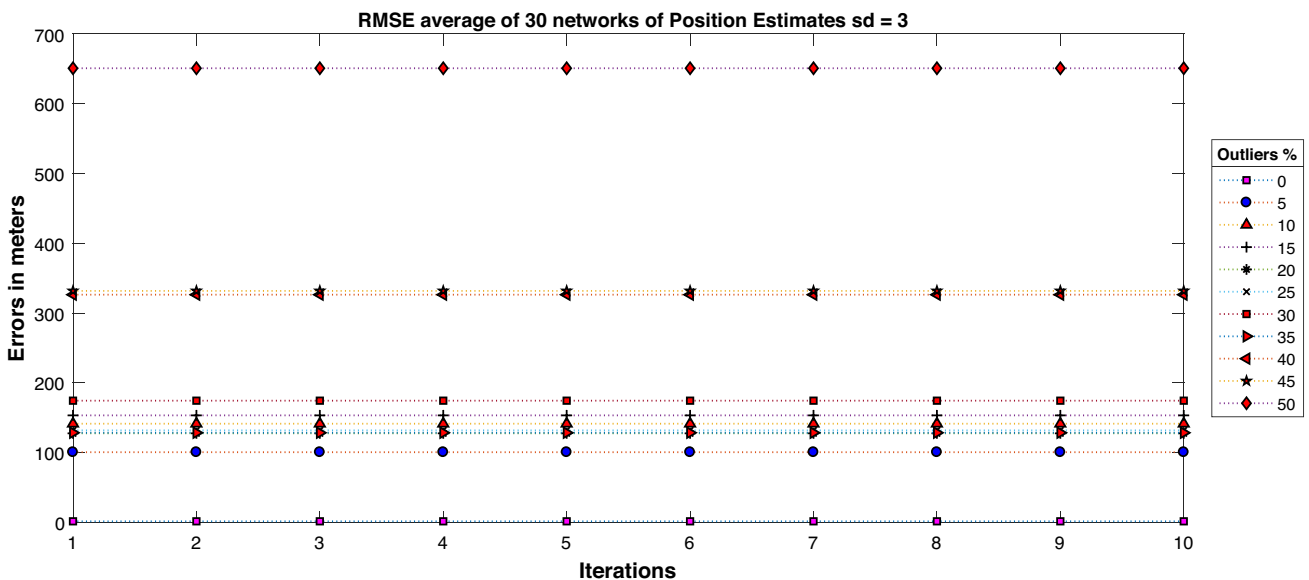


Fig. 3 Accuracy performance of the minimization scheme at different levels of outliers on range estimates and a standard deviation of three

Table 1 Accuracy performance of the minimization approach at different percentage of outliers in the network when $\mathbf{O}^i \equiv \mathbf{MC}$

Outlier Percentage (%)	0	5	10	15	20	25	30	35	40	45	50
Number of outliers	0	122.9	245.1	367	490.3	613.1	734.8	857.1	980.2	1103	1225
$sd = 1$											
RMSE (m)	0.53	108.3	131.1	128.6	143.3	137.9	145.7	239.5	207.6	816	862.8
$sd = 3$											
RMSE (m)	1.63	100.7	141.5	153.5	127.7	132	174.5	128.6	326.4	331.8	650.9

that outliers between sensor-sensor and sensor-anchors are affected by the random variable \hat{o}_{ij} , \hat{o}_{kl} respectively, described in equation (1), that takes only three possible values 0.1, 1 and 10. Obviously, when \hat{o}_{ij} or \hat{o}_{kl} is 1 there is not an outlier. For example, considering a connected network of 50 sensor nodes, so there are 2450 distance estimates between sensors. If only 10% of the distance measurements are affected by outliers, 2205 distance estimates will have \hat{o}_{ij} and $\hat{o}_{kl} = 1$, and the remaining 245 distance estimates will be randomly affected either by \hat{o}_{ij} and $\hat{o}_{kl} = 0.1$ or 10.

In the stage 3, every test set, composed by 30 networks, is used to evaluate the performance of the proposed algorithm (i.e., SelMin). Position estimates are evaluated using the RMSE metric as shown in Equation (12).

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m \|\mathbf{p}_i - \bar{\mathbf{p}}_i\|^2} \tag{12}$$

where $\|\cdot\|$ represents the 2-Norm, \mathbf{p}_i and $\bar{\mathbf{p}}_i$ are the true and estimated position of the sensor s_i , respectively; and m is the number unknown sensors in the network. Also, the DSCL algorithm [30], which uses the Levenberg-Marquardt (LM)

approach to obtain initial estimates [31], is used to compare the performance of the proposed method.

As described in Sect. 3, the proposed approach is composed of two processes: the minimization (7) and the selection (8-11) schemes working alternately. It is significant to notice that the selection procedure modifies \mathbf{O}^i by attempting to omit outliers so that they are not taken into account in the minimization stage. If the term \mathbf{O}^i remains constant, the proposed approach will only depend of the minimization stage. Thus, the outliers will significantly affect the minimization process. For example, figures 2 and 3 show the accuracy performance of the proposed scheme (with $sd = 1$ and $sd = 3$) if only the minimization approach is used when $\mathbf{O}^i \equiv \mathbf{MC}$ for $i = \{1, 2, 3, \dots, 10\}$. It is clear that all iterations have the same pattern (or remains constant). As more outliers are added to the network, the RMSE rises as well. Clearly, the only findings that are helpful are those where the estimated distances between the sensors do not include any outliers.

Table 1 summarizes the performance of the optimization algorithm (7), which minimizes the error of distances between sensors when considering different levels of Gaussian noise and percentage of outliers in the measurements.

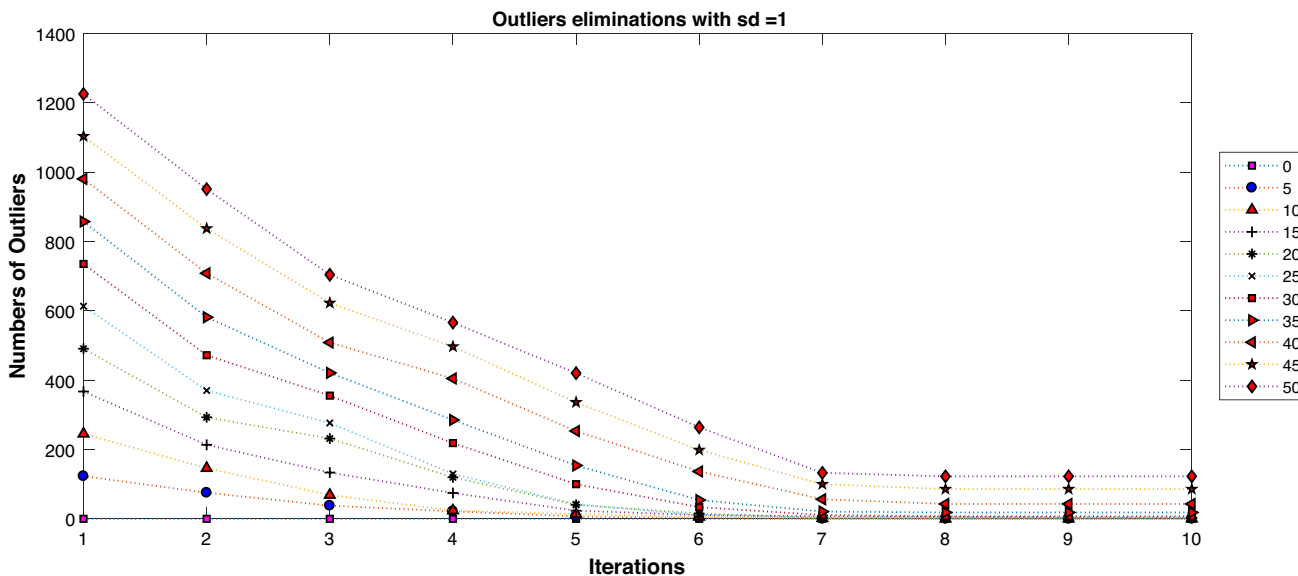


Fig. 4 The selection and the minimization process (SelMin approach) working alternately to reduce outliers considering an $sd = 1$

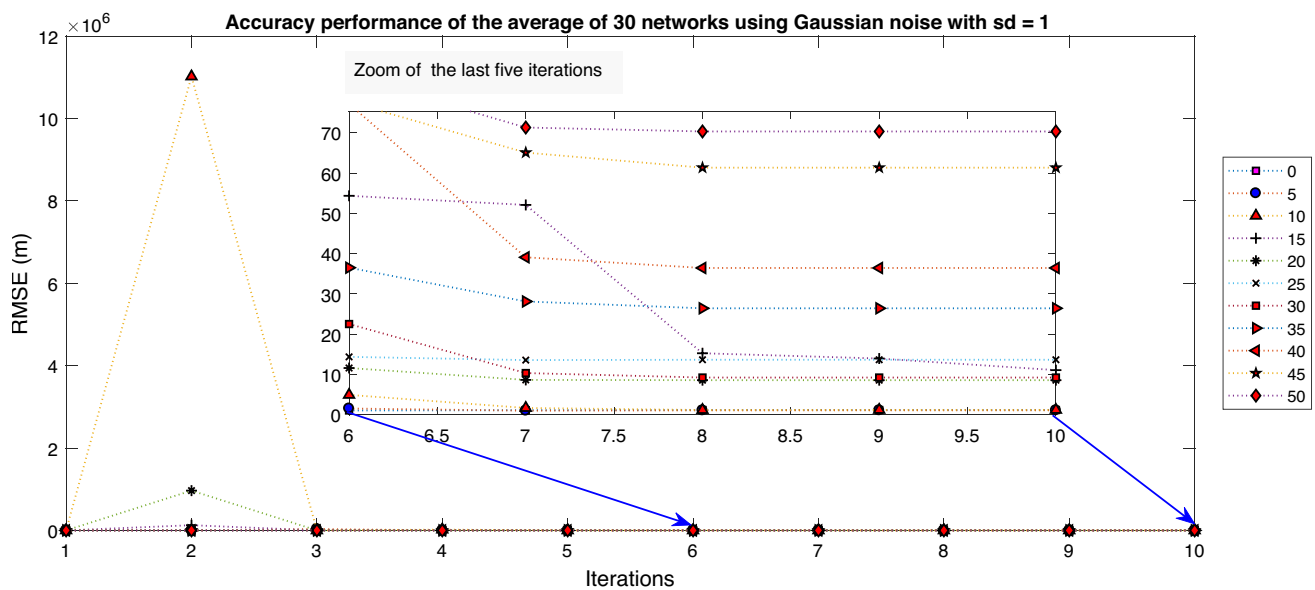


Fig. 5 The selection and the minimization process (SelMin approach) working alternatively to reduce errors on position estimates considering an $sd = 1$ with a zoom in the last five iterations

Table 2 Accuracy performance and outliers reduction summary of the SelMin approach at different percentages of outliers and noise in the networks versus the DSCL approach

Outlier Percentage (%)	0	5	10	15	20	25	30	35	40	45	50
Number of outliers	0	122.9	245.1	367	490.3	613.1	734.8	857.1	980.2	1103	1225
SelMin with $sd = 1$											
Final Outliers	0	0	0	6	1.8	3.6	7.2	18.2	42.7	85.6	122.3
RMSE (m)	1.10	1.07	1.17	15.2	8.57	13.64	9.22	26.42	36.4	61.3	70.32
SelMin with $sd = 3$											
Final Outliers	0	0	0.06	1.4	2.4	3.4	14.9	28	66.3	96	144.8
RMSE (m)	3.08	3.2	3.36	5.78	12.5	9.7	24.4	42.4	54.1	83.4	104.1
DSCL with $sd = 1$											
RMSE (m)	0.62	4.9	9.8	15.64	28.11	29.58	46	61.49	82.74	160.9	149.7
DSCL with $sd = 3$											
RMSE (m)	1.66	7.37	12.21	17.27	23.49	34.71	47.97	63.94	88.42	132.6	166.6

Conversely, when the minimization scheme (7) is employed in conjunction with the outliers-selection approach (8-11), the estimation of distances between sensors shows notable improvement within a few iterations, as will be explained later. Fig. 4 shows the reduction in the number of outliers through iterations of the SelMin algorithm 1 with a noise level of $sd = 1$. As can be observed, the algorithm completes the minimal outlier reduction for each contamination level in 8 iterations (averaged over the 30 networks), and the final RMSE for each contamination level is depicted in Fig. 5. As can be seen, while the removal of outliers is in process (8-11), the minimization approach (7) which indirectly affects the estimation of the sensor positions presents

variations. However, once the outlier removal stabilizes (at iteration 8), so does the estimation of sensor positions. Table 2 displays the outcomes achieved of the SelMin and DSCL schemes after 10 iterations, considering various percentages of outliers and levels of noise.

Regarding the SelMin approach, it is obvious that the precision of the estimated positions is reduced in direct proportion to how well outliers in the estimated distances between sensors are minimized. Regarding the last point, it can be considered that the RMSE position estimates, with $sd = 1$, between 0% and 30% of outliers present outstanding results as shown in Table 2. For instance, with 30% of

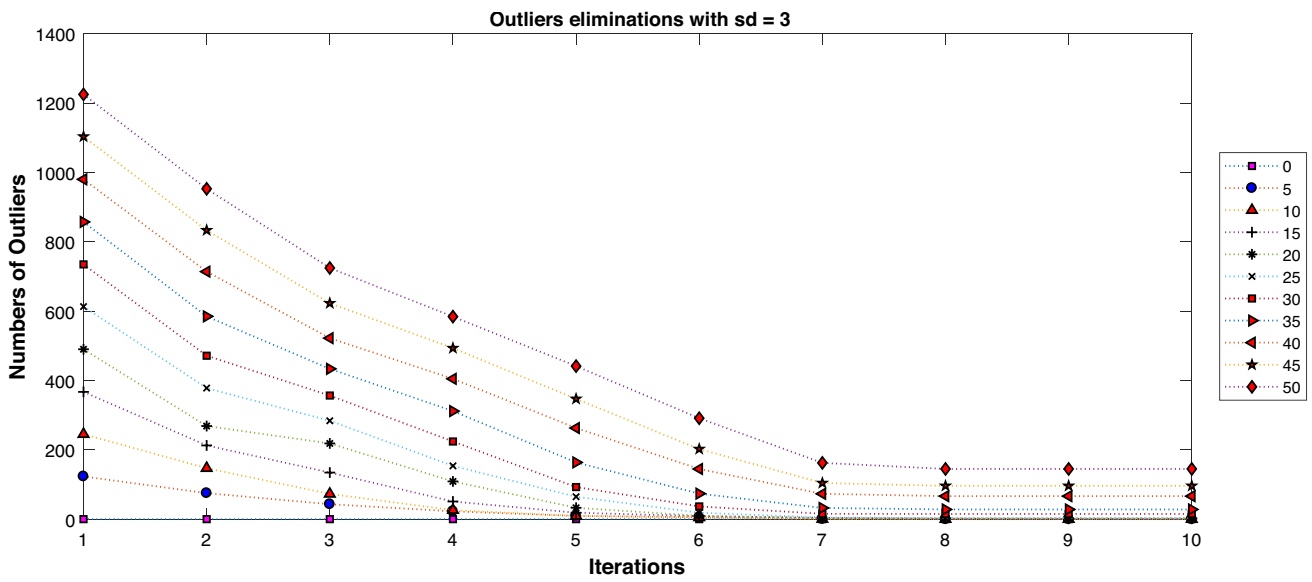


Fig. 6 The selection and the minimization process (SelMin approach) working alternatively to reduce outliers considering an $sd = 3$

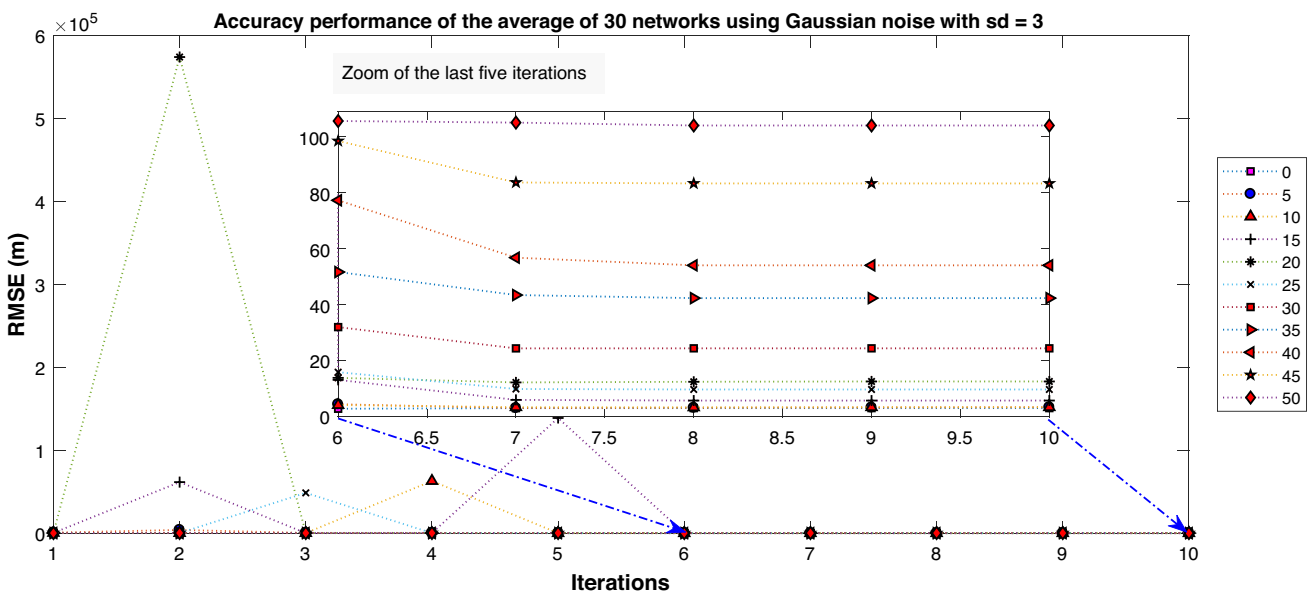


Fig. 7 The selection and the minimization process (SelMin approach) working alternatively to reduce errors on position estimates considering an $sd = 3$ with a zoom in the last five iterations

outliers (i.e., 734.8), these are reduced to 0.29% (i.e., 7.2) giving a final RMSE of 9.22m of error in average.

On the other hand, Fig. 6 depicts the outliers reduction considering a Gaussian noise level of $sd = 3$ among distance measurements.

This procedure follows a similar behavior to that of the previous analysis. That is, all the levels of outliers suffer a considerable reduction, and these are minimized at 8 iterations. However, for this analysis, the best results are obtained in the range of 0% to 25% of outliers as shown Table 2. This

means, assuming a 25% of outliers contamination in the measurements, the reduction is quite significant (from 613.1 to 3.4), producing an RMSE of 9.7m of inaccuracy on position estimates. As expected, the position estimations generally become less accurate with increasing errors from $sd = 1$ to $sd = 3$ in the distance estimates between sensors in the 30 reference networks. Finally, after reducing the impact of outliers (after 8 iterations), the distance error is correspondingly minimized, as illustrated in Figure 7.

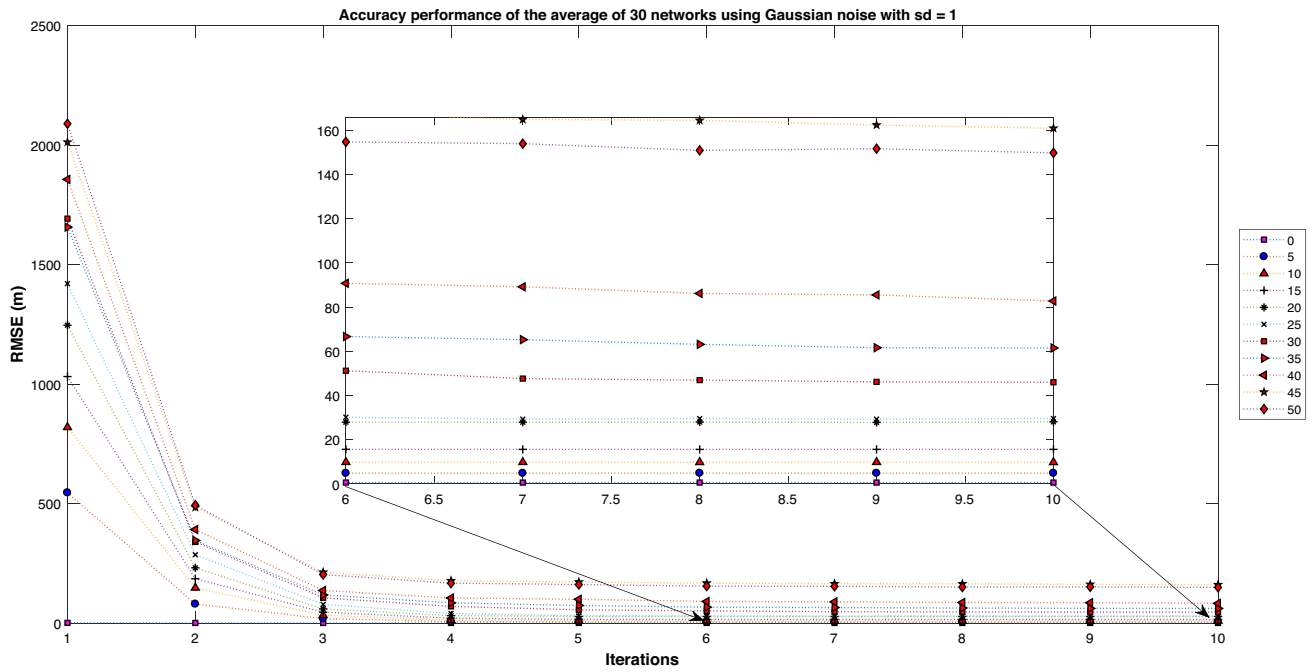


Fig. 8 The Root Mean Square Error (RMSE) for position estimates provided by the DSCL algorithm is being assessed under a standard deviation (sd) of 1. This evaluation involves 10 iterations, and specific attention is given to the final five iterations

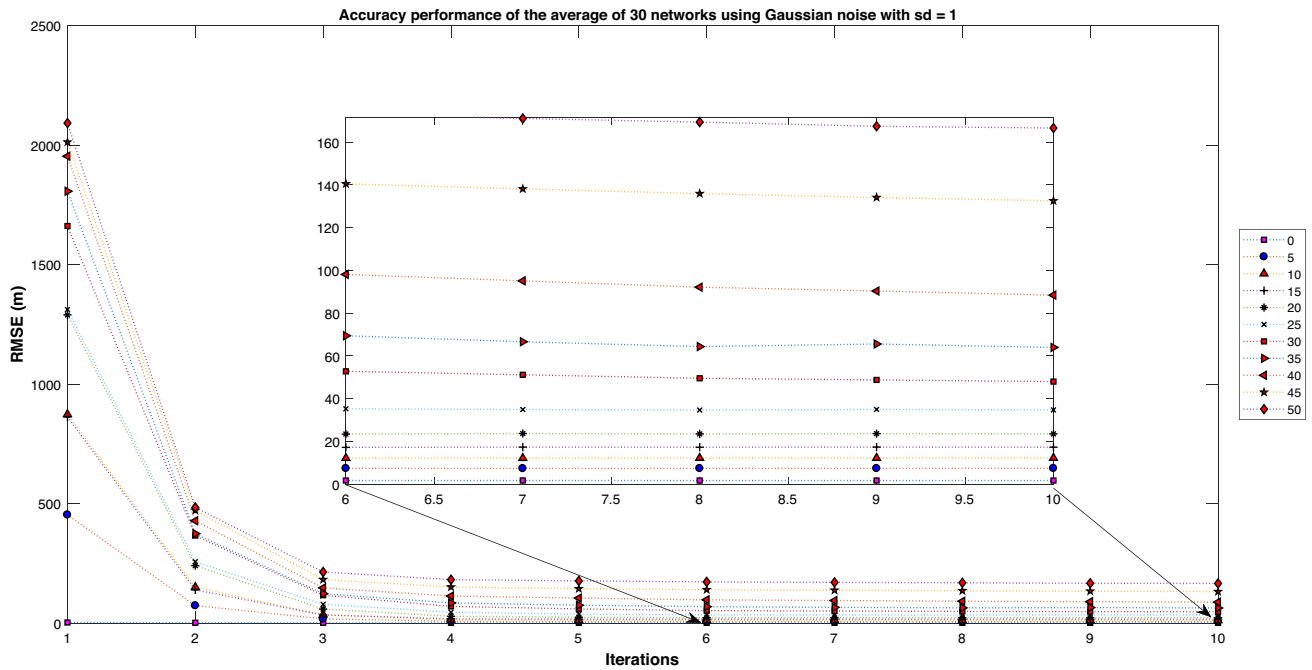


Fig. 9 The Root Mean Square Error (RMSE) for position estimates provided by the DSCL algorithm is being assessed under a standard deviation (sd) of 3. This evaluation involves 10 iterations, and specific attention is given to the final five iterations

To evaluate the performance of the proposed algorithm, the same 30 reference networks and procedures, used previously, were tested in the DSCL algorithm and results are also summarized in the bottom of Table 2. Figures 8 and 9 show the error in the estimated positions after 10 iterations considering $sd = 1$ and $sd = 3$, respectively.

Contrasting the findings in Table 2, it is observed that when outliers are absent, the DSCL algorithm provides a slightly more accurate estimation of the sensor positions compared to the proposed algorithm. However, in the presence of outliers, the proposed scheme demonstrates remarkable performance, surpassing the DSCL algorithm.

5 Conclusions and Future Works

This paper introduces a method for enhancing centralized Wireless Sensor Network (WSN) localization based on range measurements. This method utilizes a two-step process: one step aims to reduce errors in distance measurements between sensor nodes, while the other step focuses on eliminating inconsistencies in distance measurements. This iterative dual-step approach enhances the system's resilience to anomalies in distance estimates due to environmental factors and network configurations. Experimental results using simulated data sets confirm the effectiveness of this approach, particularly when a significant number of outliers distance estimates are present. For instance, results demonstrate that the proposed scheme achieves total or partial elimination of outliers in less than seven iterations, indicating not only its suitability for integration in hostile environments, particularly those relying on RSS measurements with hardware constraints, but also its potential for saving energy. On the other hand, when encountering outliers amounting to less than 10%, the SelMin algorithm maintained a similar behavior regarding the minimization of the error in the position estimates, which indicates its robustness against a moderate presence of atypical measurements. Additionally, the algorithm maintains a satisfactory RMSE even at a 25% outlier rate, exhibiting less than half the error compared to the DSCL algorithm in identical circumstances. In practical applications, the suggested method can be executed on constrained embedded systems, like Raspberry Pi and LattePanda, among others.

Shortly, there are plans to develop and test the proposed method in a distributed setting, taking advantage of matrix completion and multi-hop networks. An additional approach will be to employ real testbeds to evaluate the model under different constraints, such as noise ratio, anisotropic networks, and the number of anchor nodes. Additionally, future research will investigate the application of

artificial intelligence techniques to identify “outliers” more thoroughly.

Author Contributions Mederos-Madrazo and Cota-Ruiz introduced the primary concept, and alongside Diaz-Roman and Enriquez-Aguilera, they conducted experiments and scrutinized the data. Furthermore, all members interpreted the outcomes and drafted the manuscript.

Funding Not applicable

Availability of data and materials Not applicable

Declarations

Conflict of interest The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Code availability Not applicable

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Rappaport TS (2002) Wireless communications-principles and practice, (the book end). Microw J 45(12):128–129
2. Mesmoudi A, Feham M, Labraoui N (2013). Wireless sensor networks localization algorithms: a comprehensive survey. arXiv preprint [arXiv:1312.4082](https://arxiv.org/abs/1312.4082)
3. Sivasakthiselvan S, Nagarajan V (2020). Localization techniques of wireless sensor networks: A review. In: 2020 international conference on communication and signal processing (ICCSPP), pp. 1643–1648 . IEEE
4. Wang D, Zhang Q, Wan J (2016). A novel secure localization algorithm against distance outliers in wireless sensor networks. In: 2016 8th international conference on intelligent human-machine systems and cybernetics (IHMSC), 1, 343–346. IEEE
5. Ge Y, Zheng Z, Yan B, Yang J, Yang Y, Meng H (2016). An rssi-based localization method with outlier suppress for wireless sensor networks. In: 2016 2nd IEEE international conference on computer and communications (ICCC), 2235–2239. IEEE
6. Chen Y.C, Sun W.C, Juang J.C (2010). Outlier detection technique for rssi-based localization problems in wireless sensor networks. In: Proceedings of SICE Annual Conference 2010, 657–662. IEEE

7. Xiao F, Chen L, Sha C, Sun L, Wang R, Liu AX, Ahmed F (2018) Noise tolerant localization for sensor networks. *IEEE/ACM Trans Netw* 26(4):1701–1714
8. Dwivedi R.K, Rai A.K, Kumar R(2020). A study on machine learning based anomaly detection approaches in wireless sensor network. In: 2020 10th international conference on cloud computing, data science & engineering (confluence), 194–199. IEEE
9. Almuzaini KK, Gulliver A et al (2010) Range-based localization in wireless networks using density-based outlier detection. *Wirel. Sensor Netw.* 2(11):807
10. Du J, Yuan C, Yue M, Ma T (2022) A novel localization algorithm based on rssi and multilateration for indoor environments. *Electronics* 11(2):289
11. Yang Z, Jian L, Wu C, Liu Y (2013) Beyond triangle inequality: Sifting noisy and outlier distance measurements for localization. *ACM Transactions on Sensor Networks (TOSN)* 9(2):1–20
12. Wang Y, Ho K, Wang Z (2023) Robust localization under nlos environment in the presence of isolated outliers by full-set tdoa measurements. *Sig Process* 212:109159
13. Aspnes J, Eren T, Goldenberg DK, Morse AS, Whiteley W, Yang YR, Anderson BD, Belhumeur PN (2006) A theory of network localization. *IEEE Trans Mob Comput* 5(12):1663–1678
14. Wang Q, Li X, et al (2023) Source localization using rssi measurements with sensor position uncertainty. *Int J Distrib Sensor Netw*
15. Wang Q-h, Lu T-t, Liu M-l, Wei L-f, et al (2013) Research on the wsn node localization based on toa. *J Appl Math*
16. Zuo P, Peng T, Wu H, You K, Jing H, Guo W, Wang W (2020) Directional source localization based on rssi-aoa combined measurements. *China Commun* 17(11):181–193
17. Hendrickson B (1995) The molecule problem: Exploiting structure in global optimization. *SIAM J Optim* 5(4):835–857
18. Carter MW, Jin HH, Saunders MA, Ye Y (2007) Spaseloc: an adaptive subproblem algorithm for scalable wireless sensor network localization. *SIAM J Optim* 17(4):1102–1128
19. Nie J (2009) Sum of squares method for sensor network localization. *Comput Optim Appl* 43(2):151–179
20. Aspnes J, Goldenberg D, Yang YR (2004) On the computational complexity of sensor network localization. In: Algorithmic aspects of wireless sensor networks: first international workshop, ALGOSENSORS 2004, Turku, Finland, July 16, 2004. Proceedings 1, 32–44. Springer
21. Eren T, Goldenberg O, Whiteley W, Yang YR, Morse AS, Anderson BD, Belhumeur PN (2004) Rigidity, computation, and randomization in network localization. In: IEEE INFOCOM 2004, 4, 2673–2684. IEEE
22. Dokmanic I, Parhizkar R, Ranieri J, Vetterli M (2015) Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Process Mag* 32(6):12–30
23. Biswas P, Liang T-C, Toh K-C, Ye Y, Wang T-C (2006) Semidefinite programming approaches for sensor network localization with noisy distance measurements. *IEEE Trans Autom Sci Eng* 3(4):360–371
24. Srirangarajan S, Tewfik AH, Luo Z-Q (2008) Distributed sensor network localization using socp relaxation. *IEEE Trans Wirel Commun* 7(12):4886–4895
25. Ji S, Sze K.-F, Zhou Z, So A.M.-C, Ye Y (2013). Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In: 2013 Proceedings IEEE INFOCOM, 2499–2507. IEEE
26. Wang Z, Zheng S, Ye Y, Boyd S (2008) Further relaxations of the semidefinite programming approach to sensor network localization. *SIAM J Optim* 19(2):655–673
27. Torgerson WS (1952) Multidimensional scaling: I. theory and method. *Psychometrika* 17(4):401–419
28. Shang Y, Ruml W (2004). Improved mds-based localization. In: IEEE INFOCOM 2004, vol. 4, pp. 2640–2651. IEEE
29. Tomic S, Beko M, Dinis R (2014) Distributed rssi-based localization in wireless sensor networks based on second-order cone programming. *Sensors* 14(10):18410–18432
30. Cota-Ruiz J, Rosiles J-G, Rivas-Perea P, Sifuentes E (2013) A distributed localization algorithm for wireless sensor networks based on the solutions of spatially-constrained local problems. *IEEE Sens J* 13(6):2181–2191
31. Cota-Ruiz J, Rosiles J-G, Sifuentes E, Rivas-Perea P (2012) A low-complexity geometric bilateration method for localization in wireless sensor networks and its comparison with least-squares methods. *Sensors* 12(1):839–862

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.