

WSN-Based Wildlife Localization Framework in Dense Forests through Optimization Techniques

Mauricio González-Palacio^a, Liliana González-Palacio^b, José Aguilar^{c,d} and Long Bao Le^e

^aUniversidad de Medellín, Carrera 87 #30-65, Medellín, 050026, Antioquia, Colombia

^bUniversidad Eafit, Carrera 49 #7-50, Medellín, 050022, Antioquia, Colombia

^cIMDEA Networks Institute, Av. Mar Mediterráneo, 22 Leganés, Madrid, 28918, Spain

^dCEMISID, Facultad de Ingeniería, Universidad de Los Andes, Mérida, 501, Venezuela

^eInstitut National de la Recherche Scientifique, Rue De la Gauchetière 800, Montreal, H5A 1K6, Quebec, Canada

ARTICLE INFO

Keywords:

Wireless Sensor Networks
Anchor Node Placement
Distance Optimization
Residual Error-Based Localization

ABSTRACT

Wildlife in forests can be threatened by different phenomena like deforestation and changes in land use. It introduces the need to track the diverse terrestrial species of animals to understand their moving patterns and distributions. This allows conservationists to evaluate which species need immediate policies to guarantee survival and how the different land uses impact the ecosystem dynamics. The tracking can be achieved by using collars installed in the animals' necks, namely End Nodes (ENs), which are GPS-based in most cases. However, this technology is energy-consuming and is constrained to the outdoors, with an adequate line of sight with the satellite network. These constraints, which are not always met in forests, motivate the need to develop Wireless Sensor Network (WSN)-based localization solutions where different Low Power Wide Area Network (LPWAN) protocols can be used. These protocols rely on the measurements of the Received Signal Strength Indicator (RSSI), the Time of Flight (ToF), and propagation models to determine the distance between the EN and a set of Anchor Nodes (ANs) with fixed and previously known positions, and then, apply trilateration techniques to estimate the position of the animal. However, the existing approaches might have significant errors due to multipath and shadow fading caused by dense canopies. To address these limitations, this paper proposes a framework to improve the localization accuracy in dense forests using a three-step strategy. First, we provide an optimization setup to adequately select the ANs positions, increasing the redundancy of trilateration and coverage. Then, we propose an optimized method to determine the distance between the EN and the ANs based on bias and variance minimization using the RSSI and ToF. Finally, the optimized redundancy and coverage setup and the optimized distance estimation are used to improve the localization using a scoring method for ANs with the most reliable distance estimations based on residual errors. Numerical studies show that our framework outperforms the state-of-the-art strategies regarding trilateration capacity, distance accuracy, and localization errors.

1. Introduction

Productive human activities for extractive industries and constructions have caused high deforestation rates and changes in land use in dense forest ecosystems, particularly in tropical environments like the Amazonian or Darién jungles [1]. Different initiatives have assessed the threatened wildlife species that habit these areas, like the International Union for Conservation of Nature (IUCN), finding that a significant number of native species are now classified as threatened due to habitat fragmentation [2]. The Amazonian jungle has experienced high and unprecedented land use changes, leading to critical ecosystem variations for countless species [3]. Given this phenomenon, active research and technological advances have arisen to monitor wildlife's position and spatial distribution to measure how environmental alterations impact their survival and behavior [4]. With the collected data, the environmental authorities can identify habitats and common corridors used by endemic species and find policies to mitigate adverse effects caused by land use changes.

The traditional technologies to track the position of animals are primarily based on the Global Positioning System (GPS) [5] that has shown to be effective in open environments but has limited performance for dense forest canopies. The energy consumption demanded by the GPS receivers may lead to rapid battery depletion [6], which can be impractical for long-term monitoring since battery replacements demand recapturing the animals, leading to extra

ORCID(s):

stress [7]. Furthermore, the correct operation of GPS-based hardware requires an adequate line of sight to the satellites. This condition is rarely met in dense canopies with broad areas covered by trees that obstruct signal propagation. These drawbacks can make Wireless Sensor Networks (WSNs) and Low Power Wide Networks (LPWANs) feasible solutions for implementation in the considered scenarios [8]. These technologies have been designed to achieve low energy consumption, which strongly increases the operational lifetime of tracking devices [9, 10]. WSNs and LPWANs are not sky-dependent like GPS, so these localization strategies may exhibit lower errors in dense forests [11]. They work by collecting the Received Signal Strength Indicator (RSSI) or the Time of Flight (ToF) information from an End Node (EN) (usually a collar attached to the animal) at a set of base stations, namely Anchor Nodes (ANs), placed in fixed and known positions. The RSSI and ToF can be used to estimate the position of tagged animals. Nevertheless, these short-term channel metrics are sensitive to the environment and are subject to shadowing, multipath propagation, and delay spread [12]. This may potentially lead to significant errors in distance estimation as signals can be absorbed, redirected, reflected [13], or delayed by vegetative cover [14].

1.1. Related Works

Active research has been conducted to improve the localization accuracy in WSNs and LPWANs. For instance, Islam *et al.* [15] presented a LoRa-based localization system that uses RSSI to calculate EN-to-EN and EN-to-AN distances. They propose using particle filtering for noisy RSSI and machine learning models such as support vector machines and random forests to estimate distances. They perform the localization by trilateration, and the testing over an area of 3.1 km² has shown its efficacy in short-to-medium range distances of 20–170 m, attaining an error about 40 m. However, relying on a single gateway reduces scalability and makes this system vulnerable due to multipath and shadowing effects in denser environments. The iterative search over a large area adds latency and increases computational costs in this system.

Mohar *et al.* [16] considered RSSI as the primary distance metric for node localization to optimize the performance of a 100×100 m WSN. They designed the Jaya algorithm, which operates by minimizing the average localization error through iterative refinement of position estimation of the EN from the information received at the ANs. The results show a reduction in localization error of about 64–80%, with a significant reduction in computation time compared to other optimization algorithms like Particle Swarm Optimization (PSO). However, the proposed design depends on the quality of the RSSI signal and the consistent deployment of the ANs. Their presented results showed that the errors are more significant as the distance increases, so this approach has limitations in large-scale or sparse networks since the variability in RSSI is greater under those conditions.

Wu *et al.* [17] used RSSI as the distance metric, enhancing the localization estimation accuracy using a multipiece log-linear path loss model. The considered area was divided into sub-regions, for each of which the path loss parameters were calibrated. Thus, the algorithm self-adapted to the vegetation and terrain that cause signal attenuation. It was shown the the maximum error was ≈ 3.7 m in a bamboo forest with a validation small area of 0.0012 km². Although this strategy is effective, its deployment demands large measurement campaigns to fit the corresponding models for sub-regions, which can be unsuitable for large areas like those found in forests. Furthermore, the changes in the environment could affect the fitted models since it relies uniquely on the RSSI variations.

Li *et al.* [18] employed the RSSI and ToF to deliver estimated positions of the ENs, proposing a Weighted Cascade Compensation Estimator (WCCE), which can integrate multiple data sources from the channel dynamics. The design achieved low localization errors about 0.5 m when tested in favorable conditions of Signal-to-Noise (SNR) in a small area of 100×100 m². Although the method claims to be effective, it depends on high-quality data, limiting its application to environments with low-noise channels since RSSI and ToF are highly prone to variability in forests [14].

Kaur *et al.* [19] measured the RSSI for distance estimation in an area of 15×15m² and they proposed a modification over the Teaching–Learning-Based Optimization (TLBO) and using virtual ANs to augment the RSSI measurements. However, the scalability of this approach in large areas with dynamic changes in the SNR of the measurements remains untested. Similarly, Chen *et al.* [20] combined RSSI and ToF to create a hybrid metric, namely RSSI-RANGE, for fingerprinting in indoor scenarios. This metric achieved high precision using PSO, Random Forests (RFs), Gaussian and Median Filters, and Kriging interpolation, finding up to 58–63% improvements under experimental conditions. Although this strategy addresses multipath and noise channel issues, it requires a fingerprint database, limiting its scalability to forest-like environments where the collection of measurement campaigns can be time and resource-constrained.

Lyu *et al.* [21] utilized Radio Frequency Identification (RFID) and RSSI measurements for monorail localization in rugged orchard environments. They deployed a testbed combining an STM32 hardware platform with RFID and

Long-Range (LoRa) protocols to enable accurate positioning, reducing errors to 1.27% on flat rails and 21.18% on sloped rails. However, this strategy requires installing RFID readers in multiple locations, which may be unfeasible for forests where animals can move in large areas. Finally, Liu [22] scored reliable ANs based on the distance error distributions. This information was used to create multiple search regions. The method relied on the distance error distributions and actual positions that are unknown a priori. Determining these distributions in a dense environment, like a forest, may be unsuitable without a large-scale measurement campaign. Furthermore, if there are high standard variations in the distance error distributions, obtaining the actual distance may be challenging.

1.2. Contribution of this Paper

The summary of the state-of-the-art strategies for WSN-based localization is presented in Table 1 from which the following can be noticed. First, although some existing approaches consider RSSI variations caused by environmental characteristics, the scenario with high standard variations has not been conducted for complex conditions like those presented in forests. Second, the dense-vegetation dynamics are rarely addressed by the current works. Third, most approaches do not address the selection of the position of ANs to improve EN localization. Fourth, it is common for a database of environmental measurements to be required, which is difficult in forests considering the accessibility issues. Finally, it can be noticed that the current approaches were validated in small areas, but the forests are commonly large areas covered by dense vegetation. To address these limitations, this paper proposes a framework for enhanced localization accuracy in dense forests for terrestrial animals using WSNs based on the ITU-R P.833 [14]. Our approach is composed of three main stages. First, we propose an optimization problem to determine the AN's positions, aiming to maximize coverage and trilateration redundancy and penalize extra overlap between the AN's coverages. We use Genetic Algorithms (GAs) to find optimized solutions depending on the number of available ANs and the area to be covered. Second, we propose an optimized distance estimator based on the channel metrics measured by each AN, namely RSSI and ToF, which are weighted to minimize the bias and variance of the estimator. Third, we propose a residual error-based (REB) metric to score the best ANs to perform the localization, which takes advantage of trilateration's redundancy. Our main contributions can be summarized as follows:

1. We propose a unique optimization problem that helps WSN designers and practitioners choose optimized AN positions to improve localization accuracy. We conduct a mathematical analysis to test the superiority of our approach.
2. We propose a new distance estimator that has been mathematically shown to be optimal regarding bias and variance. We show that our estimator outperforms the average-based distance estimators.
3. We design a REB-based localization strategy that uses the estimated distance and coordinates to calculate the residuals of the predicted positions. Then, we use this metric to score the best estimations to average the final coordinates.
4. Via numerical studies, we test our framework under very tough conditions up to 16 dB of RSSI standard deviation, up to 13 μ s of delay spread with a standard deviation up to 6 dB [14], and a large area up to 100 km². To the best of our knowledge, no previous approaches have been considered for these adverse channel conditions.

The rest of this paper is organized as follows. Section 2 describes the system model we considered for the dense-vegetation environment to account for the variations of RSSI and ToF caused by shadowing, multipath fading, and delay spread and sets our design goal. Section 3 describes in depth the framework for localization of the EN. Section 4 shows the evaluation of the proposed framework under different conditions and network configurations. Finally, we present the conclusion remarks in Section 5. For ease of reference, Table A-1 presents the acronyms, and Table A-2 summarizes the key notations used in this paper.

2. System Model and Design Goal

The system model of our localization setup is depicted in Figure 1(a). First, a terrestrial animal located in a forest-like environment wears an EN, located in the point $p_e = (x_e, y_e)$, sends n upload packets to m different ANs in the set P whose positions are denoted as $p_i = (x_i, y_i) \forall i \in \{1, \dots, m\}$ and $i \in P$, deployed strategically to maximize the coverage and redundancy, as we will study in Section 3.1. Then, all the ANs covering p_e receive the n samples of the ToF and the RSSI from the EN. Subsequently, each AN resends these metrics to the Network Server (NS), including its position p_i . The NS performs the following tasks:

Table 1
Comparison of diverse WSN-based localization designs

| Ref. | Multi-scattering | Vegetation considered | Metric | Technique | AN placement optimization | Database required | Validation area | Max error |
|------------------|------------------|-----------------------|----------------|------------------|---------------------------|-------------------|------------------------|-----------|
| [15] | ✓ | × | RSSI | PSO | × | ✓ | 3.1 km ² | ≈ 40 m |
| [16] | × | × | Not considered | JAYA | × | × | 0.01 km ² | TBD |
| [17] | ✓ | Partially | RSSI | Loglinear models | × | ✓ | 0.0012 km ² | 3.7 m |
| [18] | ✓ | × | RSSI | WCCE | ✓ | ✓ | 0.01 km ² | 0.5 m |
| [19] | × | × | RSSI | TLBO | ✓ | ✓ | 0.0002 km ² | TBD |
| [20] | ✓ | × | RSSI-ToF | Fingerprinting | × | ✓ | 60 m ² | 0.82 m |
| [21] | ✓ | × | Not considered | RFID tagging | × | × | 12 m | 0.5 m |
| [22] | × | × | Not considered | PSO | × | ✓ | 0.01 km ² | 6 m |
| This work | ✓ | ✓ | RSSI-ToF | Optimization | ✓ | × | 100 km ² | < 5% |

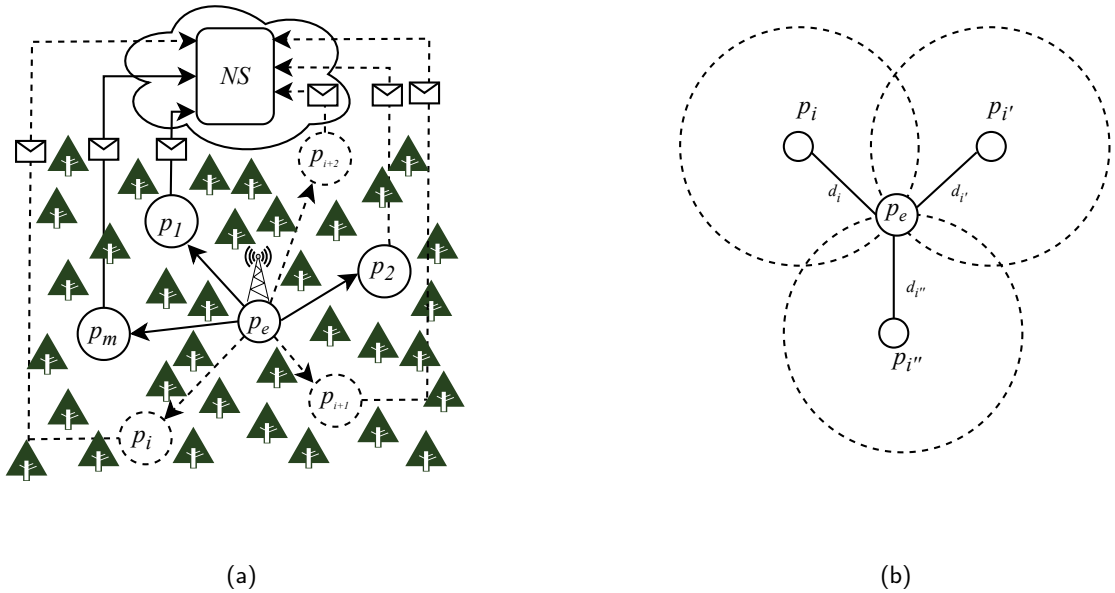


Figure 1: (a) System Model. An EN within a forest-like environment sends n upload packets to a set of ANs where some propagation metrics are registered. (b) Trilateration process. The distances from three ANs in positions p_i, p_i' and p_i'' to p_e are used to determine the localization estimation \hat{p}_e .

1. Estimate the distance \hat{d}_i between p_e and p_i as we will study in Section 3.2, using the n samples of the ToF and the RSSI.
2. Using all the p_i , determine the combinations of possible distances among the EN and the ANs covering p_e . Then, use these distances to obtain $\binom{m}{3}$ estimated positions \hat{p}_e using trilateration, as depicted in Figure 1(b).
3. Finally, use REB estimation to score the combinations of ANs that deliver the most reliable distance estimations to tackle the inaccurate positioning when the ANs can introduce outliers, as we will study in Section 3.3.

To estimate the distances \hat{d}_i , we use a path loss model and the ToF. The following subsections show how we modeled each considered metric.

2.1. RSSI-based Path Loss Model

The RSSI measurements received in any AN $_i$, denoted by RSSI_i , can be modeled as

$$\text{RSSI}_i = A - PL(d_i) + \epsilon_i, \quad (1)$$

where $A = P_{\text{TX}} - L_{\text{TX}} + G_{\text{TX}} - L_{\text{RX}} + G_{\text{RX}}$ is a known constant representing the link budget, which includes the transmit power (P_{TX}), transmitter and receiver antenna gains (G_{TX} and G_{RX}), and transmitter and receiver losses (L_{TX} and L_{RX}). The term $PL(d_i)$ represents the total path loss at a distance d_i , and ϵ_i is the measurement noise, assumed to be independent and identically distributed (i.i.d.) zero-mean Gaussian noise with variance $\sigma_{\text{RSSI}_i}^2$, i.e., $\epsilon_i \sim \mathcal{N}(0, \sigma_{\text{RSSI}_i}^2)$. The total path loss $PL(d_i)$ includes both the Free Space Path Loss (FSPL) denoted by $PL_{\text{FS}}(d_i)$ [12], and the attenuation due to vegetation $PL_{\text{veg}}(d_i)$ [23], given by

$$PL(d_i) = PL_{\text{FS}}(d_i) + PL_{\text{veg}}(d_i). \quad (2)$$

We use the FSPL rather than the lognormal path loss model because we are considering the attenuation caused by the vegetation separately. In that way, the FSPL is calculated as

$$PL_{\text{FS}}(d_i) = 20 \log_{10}(d_i) + 20 \log_{10}(f) + 32.45, \quad (3)$$

with d_i in km and the frequency f in MHz. In addition, we model the vegetation attenuation $PL_{\text{veg}}(d_i)$ using the ITU-R P.833-9 [23] that recommends that the losses caused by trunks, branches, and leaves are given by

$$PL_{\text{veg}}(d_i) = A_{\text{max}} \left(1 - \exp \left(-\frac{\gamma d_i}{A_{\text{max}}} \right) \right), \quad (4)$$

where A_{max} is the maximum attenuation dB and γ is the specific attenuation coefficient in dB/m.

2.2. Time of Flight Model

Since the transmitted signal from the EN propagates throughout the forest-like environment, it experiences the effects of scattering, reflections, and diffractions caused by the multiple trees. Thus, the arrival times at each AN depend on the distance d_i and the delay spread caused by the above-mentioned multipath propagation effects. In that way, the ToF of the arrival wave in the AN $_i$, denoted by T_i , can be modeled by

$$T_i = \frac{d_i}{c} + \Delta T_i, \quad (5)$$

where $\frac{d_i}{c}$ is the propagation time over distance d_i (with c being the speed of light), and $\Delta T_i \sim \mathcal{N}(0, \tau_{\text{rms}}^2)$ is the excess delay that is zero-mean normally distributed with variance τ_{rms}^2 caused by multipath effects, where τ_{rms} is the average Root Mean Square delay spread that has been studied for different environments in [24]. The delay spread also depends on the distance [25] as

$$\tau_{\text{rms}} = T_1 d^\eta u, \quad (6)$$

where T_1 is the mean value of τ_{rms} at 1 km (1000 ns for forest-like environments), d is the distance, η is the loss exponent (1 for forest-like environments), and u is a zero-mean lognormal variate with standard deviation up to 6 dB for the [studied](#) scenario [25]. Considering this challenging environment, where the multipath effects make the measurements of RSSI and ToF inaccurate, our design goal is to find a framework that helps determine the EN position accurately, as we will describe in Section 3.

3. Proposed Localization Scheme

Our proposed localization scheme is based on the following steps. First, we optimize the positions p_i of each AN, maximizing coverage and number of possible trilaterations at each p_e in a particular area and penalizing the excess of coverage overlapping by the ANs. If this penalization is not considered, the optimization algorithm can find solutions such that all the ANs are placed in the center to maximize the number of trilaterations and coverage (Section 3.1). The penalization was also considered to avoid concentric coverage areas, which do not help perform the trilaterations. Second, our framework determines the distance \hat{d}_i between p_e and any AN localized at p_i using an enhanced distance estimator (Section 3.2) based on the RSSI and ToF at the AN side. Third, we use the REB method to take advantage of the multiple trilaterations in the optimization placement problem to find the estimated location \hat{p}_e (Section 3.3).

3.1. Optimized Anchor Node Placement Strategy

We are interested in finding the optimized positions p_i of m ANs within a determined area S . To improve the subsequent localization strategies in Sections 3.2 and 3.3, it is crucial to deal with: *i*) the coverage maximization of S , *ii*) the maximization of the number of trilaterations that can be performed for each point p_e since it will provide redundant localization information that empowers the REB method we will propose, and *iii*) the penalization of extra overlap of coverage areas to avoid that the optimization scheme favors the localization of the ANs in concentric points that may not be useful to determine trilaterations. To find the best tradeoff among these conditions, we propose to tackle the following optimization problem::

$$\begin{aligned} \max_P \quad & \Phi(P) = \alpha \cdot \text{TrilaterationScore}(P) + \beta \cdot \text{CoverageArea}(P) - \lambda \cdot \text{OverlapPenalty}(P) \\ \text{S.t.} \quad & p_i \in S \subseteq \mathbb{R}^2, \quad \forall i = 1, 2, \dots, m \\ & \alpha + \beta + \gamma = 1, \quad \alpha, \beta, \gamma \geq 0, \end{aligned} \quad (7)$$

where $\Phi(P)$ is the objective function to find the optimal localization of the ANs, $\text{TrilaterationScore}(P)$ measures the trilateration capacity of the ANs located in the positions $p_i \in S$, $\text{CoverageArea}(P)$ represents the number of points in S that can be used for trilateration, $\text{OverlapPenalty}(P)$ captures the penalization for extra overlap among the regions covered by different ANs, and α , β , and λ are the weights for the considered metrics, which can be set according to the application characteristics. In the following subsections, we will model these terms and functions. To ease the analysis, we have discretized S in a grid S_{discrete} of $l \times l$ points, where l is the number of points in each axis. We will solve the optimization problem using GAs since they can explore multiple solutions concurrently and find the optimized solutions faster.

3.1.1. Trilateration Score

To consider that a particular AN covers a certain point, we define the coverage indicator as

$$I_i(p) = \begin{cases} 1, & \text{if } \text{RSSI}_i \geq \text{sensitivity}, \\ 0, & \text{if } \text{RSSI}_i < \text{sensitivity}, \end{cases} \quad (8)$$

where $I_i(p)$ is the function to indicate if a point p is covered by the AN_{*i*}, meaning that the RSSI_{*i*} level is over the minimum sensitivity of the corresponding AN, using the link budget in Equation (1). We define the number of ANs covering p by

$$S(p) = \{i \in P \mid I_i(p) = 1\}, \quad (9)$$

where $S(p)$ is a set that contains the ANs that cover p . We can also define the function of the number of trilaterations as

$$T(p) = \begin{cases} 0, & \text{if } |S(p)| < 3, \\ \binom{|S(p)|}{3}, & \text{if } |S(p)| \geq 3, \end{cases} \quad (10)$$

where $T(p)$ is the number of combinations of trilateration for the point p , and $|S(p)|$ represents the number of elements of $S(p)$, that is, the number of ANs covering p . Finally, we define the TrilaterationScore(p) by

$$\text{TrilaterationScore}(p) = \frac{\sum_{p \in S_{\text{discrete}}} T(p)}{\binom{m}{3} \times l \times l} \quad (11)$$

That is, from Equation (11), we count all the possible trilaterations in each point of S_{discrete} and normalize the result over the maximum number of trilaterations, which can be obtained by finding $\binom{m}{3}$ combinations of the ANs times the grid size $l \times l$.

3.1.2. Coverage Area Score

To maximize the coverage over the area S , we evaluate all the discretized points to determine whether they are covered by at least three ANs, guaranteeing the ability of the system to perform trilaterations. The function to assess this condition can be defined as

$$\text{CoverageArea}(P) = \frac{\sum_{p \in S_{\text{discrete}}} \delta(|S(p)| \geq 3)}{l^2}, \quad (12)$$

$$\delta(\Omega) = \begin{cases} 1, & \text{if } \Omega \text{ is true,} \\ 0, & \text{if } \Omega \text{ is false,} \end{cases}$$

where $\delta(\Omega)$ is a function that evaluates if the condition Ω is true or false. In this case, Ω evaluates if the assessed point p is covered by more than three ANs. To obtain a score in the interval $[0, 1]$, we divide the points covered for more than three ANs by the total number of points in the discrete area ($l \times l$).

3.1.3. Overlap Penalty Score

If we only consider the scores of trilateration and coverage, the optimization can favor the ANs' positions excessively close among them and probably near the center of S . This distribution causes the optimized ANs' localizations not to help perform the trilaterations. Because of that, we introduce a score to penalize the excess of overlap among the coverage of the ANs' positions. The coverage area set C_i of the AN $_i$ is defined by

$$C_i = \{p \in S \mid I_i(p) = 1\}. \quad (13)$$

That is, for each point p , we define the coverage area using the coverage indicator in Equation (8) such that we have an $l \times l$ matrix per each AN that stores boolean values indicating if the AN $_i$ covers the point. Using all the coverage sets, we define the fractional overlap by

$$F_{i,j} = \frac{|C_i \cap C_j|}{\min(|C_i|, |C_j|)}, \quad (14)$$

where $F_{i,j}$ is a proportion of overlapping between the coverage of the AN $_i$ and the coverage of the AN $_j \forall i, j \in \{1, \dots, m\}$ which is scaled by the smaller coverage area of the ANs being compared resulting in a score in the interval $[0,1]$. We define a function to penalize the excess of overlap by

$$\psi_{i,j} = \begin{cases} F_{i,j}, & \text{if } F_{i,j} > F_{\text{threshold}}, \\ 0, & \text{if } F_{i,j} \leq F_{\text{threshold}}, \end{cases} \quad (15)$$

where $\psi_{i,j}$ represents a penalty when the hyperparameter $F_{\text{threshold}}$ is surpassed by the anchor nodes AN $_i$ and AN $_j$. Finally, we define the OverlapPenalty(P) as

$$\text{OverlapPenalty}(P) = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^m \psi_{i,j}}{\binom{m}{2}}. \quad (16)$$

From Equation (16), we sum all the penalties $\psi_{i,j}$ between each pair of ANs i and j and scale the total summation with all the combinations of m ANs taken 2.

3.2. Proposed Estimator for the Distance between the EN and the AN

Once we have optimized the ANs' positions, we must determine the distance from p_e to all the ANs that cover it. Thus, a novel scheme to estimate the distance \hat{d}_i from p_e to p_i is proposed in this section, based on the RSSI and ToF reported in the corresponding AN $_i$. Subsequently, we perform a theoretical analysis to show that our estimator is unbiased and its variance is less than the state-of-the-art estimators based on central tendency statistics using the n RSSI upload packets.

3.2.1. Combined-ToF-RSSI Distance Estimator

As previously mentioned, we utilize the RSSI and the ToF to estimate the distance \hat{d}_i . If the value of $n \rightarrow \infty$, the use of the mean of the RSSI samples measured at the AN $_i$, denoted by $\overline{\text{RSSI}}_i$, may be enough to obtain \hat{d}_i from the path loss model in Equation (1). Using the mean guarantees that the estimator is unbiased with variance $\sigma_{\text{RSSI}_i}^2/n \approx 0$. However, since the terrestrial animal wears the EN, it might not be easily recharged. For this reason, the EN should operate as long as possible without battery replacement. Due to these energy constraints, the number of upload packets n should be small, reducing the transmissions required to perform the localization. Thus, we propose an enhanced distance estimator, showing that it exhibits an imperceptible bias and lower variance of d_i than using the sample mean of RSSI measurements alone. We consider two estimators as follows. First, we use the sample mean of RSSI measurements $\overline{\text{RSSI}}_i$. Second, we use the ToF, which can be employed to determine a ToF-estimated distance \hat{d}_{it} (the subindex t denotes that the distance was estimated using the ToF). Then, we can use \hat{d}_{it} to determine an estimated RSSI, represented by $\widehat{\text{RSSI}}_{it}$, using the path loss model in Equation (1). In that way, we can mix both RSSI estimations as a weighted sum as

$$\widehat{\text{RSSI}}_i = a\overline{\text{RSSI}}_i + b\widehat{\text{RSSI}}_{it}, \quad (17)$$

where a and b are the weights ($a + b = 1$), which will be treated later in this section. Regarding $\overline{\text{RSSI}}_i$, it can be determined by

$$\overline{\text{RSSI}}_i = \frac{1}{n} \sum_{j=1}^n \text{RSSI}_{ij}, \quad (18)$$

where RSSI_{ij} is the j^{th} measurement of the i^{th} AN. On the other hand, we use the ToF measured at the AN $_i$ to estimate \hat{d}_{it} by using the speed of light c . However, we need to guarantee that $\widehat{\text{RSSI}}_{it}$ is unbiased, so it is required to adjust the ToF measured by the AN $_i$, denoted by \hat{T}_i , by subtracting the mean delay spread τ_{rms} as

$$\hat{T}_i = \frac{1}{n} \sum_{j=1}^n T_{ij} - \tau_{\text{rms}}, \quad (19)$$

where T_{ij} is the j^{th} measurement of the AN $_i$. This adjusted time is used to estimate $\hat{d}_{it} = c \cdot \hat{T}_i$. Subsequently, we use the estimated distance in the path loss model of Equation (1) to obtain $\widehat{\text{RSSI}}_{it}$. It can be shown that the expected value of \hat{d}_{it} is $\mathbb{E}[\hat{d}_{it}] = c \cdot (\mathbb{E}[\hat{T}_i]) = c \cdot (d_i/c + \mathbb{E}[\Delta T_i] - \tau_{\text{rms}}) = d_i + c \cdot (\tau_{\text{rms}} - \tau_{\text{rms}}) = d_i$, making it an unbiased estimator of d_i . Since $\text{Var}[T_{ij}] = \tau_{\text{rms}}^2$, the variance of \hat{d}_{it} can be expressed as

$$\text{Var}[\hat{d}_{it}] = c^2 \cdot \text{Var}[\hat{T}_i] = c^2 \cdot \left(\frac{\text{Var}[T_{ij}]}{n} \right) = \frac{c^2 \cdot \tau_{\text{rms}}^2}{n} \quad (20)$$

Our goal is to minimize the variance $\text{Var}[\widehat{\text{RSSI}}_i]$ while ensuring the estimator remains unbiased. By guaranteeing these conditions, we may significantly improve the traditional distance estimators based on path loss models. Thus, we must find the optimal values for a and b that make our RSSI estimator in Equation (17) optimal. To achieve this, we propose the following optimization problem that minimizes the variance:

$$\begin{aligned} \min_{a,b} \quad & a^2 \text{Var}[\overline{\text{RSSI}}_i] + b^2 \text{Var}[\widehat{\text{RSSI}}_{it}] \\ \text{S.t.} \quad & a + b = 1 \end{aligned} \quad (21)$$

We use the Lagrangian to obtain the optimal values for a and b as

$$\mathcal{L}(a, b, \lambda) = a^2 \text{Var}[\overline{\text{RSSI}}_i] + b^2 \text{Var}[\widehat{\text{RSSI}}_{it}] - \lambda(a + b - 1), \quad (22)$$

where λ measures the increase in the objective function when there is a marginal relaxation in the constraint $a + b = 1$. Taking partial derivatives $\partial \mathcal{L} / \partial a = \partial \mathcal{L} / \partial b = \partial \mathcal{L} / \partial \lambda = 0$ and solving for a and b , it is possible to show that

$$a = \frac{\text{Var}[\widehat{\text{RSSI}}_{it}]}{\text{Var}[\overline{\text{RSSI}}_i] + \text{Var}[\widehat{\text{RSSI}}_{it}]}, \quad b = \frac{\text{Var}[\overline{\text{RSSI}}_i]}{\text{Var}[\overline{\text{RSSI}}_i] + \text{Var}[\widehat{\text{RSSI}}_{it}]} \quad (23)$$

At this point, we have an RSSI estimation with a lower variance than the mean of the measurements $\text{RSSI}_{ij} \forall j \in \{1, \dots, n\}$, traditionally used to obtain the distance estimation from a path loss model. In our case, since the combination of Equations (1)-(4) produce an expression that is not possible to solve analytically, we can use a numerical method like Newton Raphson to obtain the estimation \hat{d}_i .

3.2.2. Bias Analysis of the Distance Estimator

As part of the performance analysis of the proposed distance estimator in Equation (17), we analyze the bias of $\widehat{\text{RSSI}}_i$ in Equation (17). From this equation, we analyze the terms for $\overline{\text{RSSI}}_i$ and $\widehat{\text{RSSI}}_{it}$ separately. In the first case, it can be noticed that $\mathbb{E}[\overline{\text{RSSI}}_i] = \mu_{\text{RSSI}_i}$, that is, when n is large enough, the sample mean of the RSSI values received at AN_i converges to the population mean μ_{RSSI_i} . In the second case, the bias of $\widehat{\text{RSSI}}_{it}$ can be obtained by calculating the second-order Taylor expansion around d_i as

$$b_{\widehat{\text{RSSI}}_{it}} = \mathbb{E}[\widehat{\text{RSSI}}_{it}] - \mu_{\text{RSSI}_i} = \mathbb{E}[A - PL(d_i) - PL'(d_i)(\hat{d}_i - d_i) - \frac{1}{2} PL''(d_i)(\hat{d}_i - d_i)^2] - \mu_{\text{RSSI}_i}, \quad (24)$$

where $b_{\widehat{\text{RSSI}}_{it}}$ is the bias of $\widehat{\text{RSSI}}_{it}$, $PL'(d_i)$ is the first derivative of $PL(d_i)$, $PL''(d_i)$ is the second derivative of $PL(d_i)$, and μ_{RSSI_i} is the population mean of RSSI_i . Notice that since \hat{d}_i is unbiased, $\mathbb{E}[(\hat{d}_i - d_i)] = 0$. Furthermore, since $\mathbb{E}[A - PL(d_i)]$ is indeed the mean μ_{RSSI_i} , these terms are also cancelled out. Thus, $b_{\widehat{\text{RSSI}}_{it}}$ is

$$b_{\widehat{\text{RSSI}}_{it}} = -\frac{1}{2} PL''(d_i) \mathbb{E}[(\hat{d}_i - d_i)^2] = -\frac{1}{2} PL''(d_i) \text{Var}[\hat{d}_{it}]. \quad (25)$$

The total bias $b_{\widehat{\text{RSSI}}_i}$ is the expected value of the weighted sum in Equation (17) such that

$$b_{\widehat{\text{RSSI}}_i} = a \mathbb{E}[\overline{\text{RSSI}}_i] + b \mathbb{E}[\widehat{\text{RSSI}}_{it}] - \mu_{\text{RSSI}_i} = a \mu_{\text{RSSI}_i} + b \mu_{\text{RSSI}_i} + b \cdot b_{\widehat{\text{RSSI}}_{it}} - \mu_{\text{RSSI}_i} = b \cdot b_{\widehat{\text{RSSI}}_{it}}. \quad (26)$$

From Equation (25), it can be noticed that the bias depends on the second derivative of the path loss and the variance $\text{Var}[\hat{d}_{it}]$. Remarkably, the ToF variance in forest-like scenarios is small [25], the term $b \cdot b_{\widehat{\text{RSSI}}_{it}} \approx 0$, so the bias $b_{\widehat{\text{RSSI}}_i} \approx 0$, that is, our proposed estimator is approximately unbiased.

3.2.3. Variance Analysis of the Distance Estimator

We want to show that the variance of the estimator in Equation (17) is less or equal than the mean $\overline{\text{RSSI}}_i$, guaranteeing more concentrated distance measurements. Thus, we can calculate the variance of $\widehat{\text{RSSI}}_i$ as

$$\text{Var}[\widehat{\text{RSSI}}_i] = a^2 \text{Var}[\overline{\text{RSSI}}_i] + b^2 \text{Var}[\widehat{\text{RSSI}}_{it}]. \quad (27)$$

We use the weights in Equation (23) to obtain the estimator variance as:

$$\text{Var}[\widehat{\text{RSSI}}_i] = \text{Var}[\overline{\text{RSSI}}_i] \cdot \underbrace{\frac{\text{Var}[\widehat{\text{RSSI}}_{it}]}{\text{Var}[\overline{\text{RSSI}}_i] + \text{Var}[\widehat{\text{RSSI}}_{it}]}_{\gamma}. \quad (28)$$

Since $0 < \gamma \leq 1$, we have shown that $\text{Var}[\widehat{\text{RSSI}}_i] \leq \text{Var}[\overline{\text{RSSI}}_i]$.

3.3. REB Localization Strategy

The last step of our localization framework is to accurately estimate the actual EN position $p_e = (x_e, y_e)$ within a forest-like environment. Once each AN_i delivers the distance \hat{d}_i , we use $\binom{m}{3}$ combinations to estimate a set of coordinates $\hat{p}_e^{(i)} = (\hat{x}_e^{(i)}, \hat{y}_e^{(i)})$. To this end, $\hat{p}_e^{(i)}$ can be inferred using the previously-known coordinates of three ANs $\{p_i, p_{i'}, p_{i''}\}$ solving the trilateration equations by

$$(\hat{x}_e^{(i)} - x_i)^2 + (\hat{y}_e^{(i)} - y_i)^2 = \hat{d}_i^2, \quad (\hat{x}_e^{(i)} - x_{i'})^2 + (\hat{y}_e^{(i)} - y_{i'})^2 = \hat{d}_{i'}^2, \quad (\hat{x}_e^{(i)} - x_{i''})^2 + (\hat{y}_e^{(i)} - y_{i''})^2 = \hat{d}_{i''}^2, \quad (29)$$

where (x_i, y_i) , $(x_{i'}, y_{i'})$, and $(x_{i''}, y_{i''})$ are the coordinates for the ANs located in p_i , $p_{i'}$ and $p_{i''}$. As we commented previously, the localization errors can be inaccurate due to shadowing and multipath effects present in dense forests. To improve these errors, we introduce a REB scoring method to select the most reliable estimated positions as follows. For each estimated position $\hat{p}_e^{(i)}$, we compute the residual error r_{ij} using all the ANs that cover the corresponding point. The residual error for the AN_j and the i^{th} position is calculated as

$$r_{ij} = \hat{d}_j - \sqrt{(x_j - \hat{x}_e^{(i)})^2 + (y_j - \hat{y}_e^{(i)})^2}, \quad (30)$$

where \hat{d}_j is the estimated distance from the EN to the AN_j and the square root in the second term represents the Euclidean norm of the estimation $(\hat{x}_e^{(i)}, \hat{y}_e^{(i)})$ and the position of the AN_j in (x_j, y_j) . The residual error r_{ij} in Equation (30) measures the difference between the observed distance \hat{d}_j and the distance obtained with the estimation of the position $\hat{p}_e^{(i)}$. To evaluate the overall consistency of the estimated position $\hat{p}_e^{(i)}$, we accumulate all the square residual errors by

$$R_i = \sum_{j \in C} r_{ij}^2, \quad (31)$$

where R_i is the accumulated residual error for the estimated position $\hat{p}_e^{(i)}$, and C is the set of ANs covering the EN. Low values of R_i indicate better goodness of fit between the estimated position and the observed distance, which suggests that $\hat{p}_e^{(i)}$ is likely to be near the true position p_e . Using R_i , we can rank all the estimated positions and discard those that offer poor performance. Let us define $K \in (0, 1]$ as the proportion of the positions set to obtain the final localization and $m' = \left\lceil K \cdot \binom{m}{3} \right\rceil$ as the number of selected positions. We define the estimated position as

$$\hat{p}_e = \frac{1}{m'} \sum_{i=1}^{m'} \hat{p}_e^{(i)}. \quad (32)$$

4. Numerical Results

This section evaluates the proposed strategies to improve the localization of wildlife in dense forests. First, we describe the simulation scenarios for each case. Then, we evaluate the optimization for the AN's placement and compare its performance versus other relevant strategies in the literature. Subsequently, we evaluate the proposed distance estimator for diverse standard deviations to check its behavior in adverse scenarios and contrast the results with ToF-based and RSSI-based distance estimation strategies. Finally, we test the REB redundant localization framework with different numbers of ANs.

4.1. Database Generation

We used the system model in Section 2 to generate a comprehensive database that reflects the complexity of the propagation metrics in forests, namely RSSI and ToF. We considered the settings in Table 2. First, it is essential to state that our simulation area was inspired by the characteristics of tropical, flat, and wet forests with dense vegetation like the National Park Los Katíos in the Chocó/Darien jungle (Colombia), located in the coordinates $7^{\circ}48'0''$ N, $77^{\circ}9'0''$ W, which is cataloged as a UNESCO World Heritage Site. This national park has different endemic terrestrial species like the chochoan spider monkey (*Ateles Hybidus*), the Central American tapir (*Tapirus Bairdii*), and the jaguar (*Panthera Onca*) that are species threatened by deforestation [26]. In this context, it is crucial to continuously monitor the wildlife's position to identify critical areas for survival and assess the ecosystem's health, among others [27]. Under the previous conditions, we considered an area of 100 km^2 and divided it into 100.000 discrete points. Regarding the ANs, we considered 10, 15, 20, and 25 receivers to evaluate our approach to redundancy of trilateration. Regarding the link budget parameters (like powers and losses, among others), we considered typical values for Low Power Wide Area Networks like LoRaWAN [28, 29], a protocol designed for energy efficiency. For instance, the receiver sensitivity was set to -136 dBm , a typical value for LoRaWAN gateways with a Spreading Factor of 12, guaranteeing the maximum coverage [23]. For the standard deviation of the received signal σ_{RSSI} , we evaluated different values from 5 to 16 dB to check our approach under very challenging conditions [13, 30, 31]. We used the frequency $f = 900\text{ MHz}$ since it is the approved Industrial, Scientific, and Medical (ISM) band for America [32]. To accurately model the vegetation attenuation, we used the specification ITU-R P.833, using parameters derived from the empirical measurements performed in the Mulhouse forest in France, which are similar to the characteristics of the forest Los Katíos regarding distance, frequency, and vegetation density [14]. Once we ran the optimization for the localization of the ANs, we used these positions to calculate the Euclidean distance from each AN and each point in the grid, simulating the movement of the terrestrial animal in 10.000 possible points, following a uniformly distributed random position, since the animal can be located in any point of the grid. For each point and each AN, we simulated $n = 50$ subsequent measurements of the RSSI and ToF, which are needed for the distance and localization estimations. It was performed by using the models in Equations (2) - (4) iterating for different standard deviations and distances.

4.2. Evaluation of the Anchor Nodes Placement Optimization

We evaluated the redundancy in the coverage of the ANs using a different number of them, varying it in our numerical studies with values of 10, 15, 20, and 25. To solve the optimization problem in (7), we used a GA through the `differential_evolution` class of the library `scipy.optimize` of Python. According to the structure of this algorithm, we started with ten candidates for the first generation, which represent the positions of the ANs in the grid. Each configuration was evaluated by balancing the objectives previously described in Section 3.1 as follows: *i*) maximize trilateration combinations, *ii*) enhance area coverage, and *iii*) minimize the overlap between the zones of coverage. We set $\alpha = \beta = \lambda = 1/3$ to represent the same importance for each requirement, guaranteeing optimized trilateration combinations in the area without favoring the placement of the ANs in the center since these positions may not be useful for solving the trilateration equations. We used 100 generations to find the global optimized value progressively. To empower the diversity exploration, we introduced high mutation and recombination rates with values of 0.7 that allowed the GA to avoid local optima.

To show the superior performance of our optimization scheme, we compared the obtained results with other common AN placement strategies, such as equispaced [33] and randomly [34] located AN positioning. In the first strategy, the ANs are placed following an organized grid of points. In the second strategy, the ANs are located without following a specific criterion, and the placement is arbitrary. An example of the three strategies for 25 anchor nodes is depicted in Figure 2, where it can be noticed that our strategy finds AN positions where the coverage, trilateration ability, and overlap are adequately balanced. The performance of our strategy is depicted in Figure 3, which shows a heatmap of each scheme considering the proportion of the total area covered by a particular number of ANs. From this

Table 2

Database generation parameters. The transmitter and receiver are equivalent to the EN and the ANs, respectively.

| Parameter | Symbol | Value | Unit |
|------------------------------------|-----------------|--------------------|-----------------|
| Area | S | 100 | km ² |
| Number of grid points | l^2 | 10.000 | N/A |
| Transmit power | P_{TX} | 0 | dBm |
| Transmitter losses | L_{TX} | 1 | dB |
| Transmitter antenna gain | G_{TX} | 3 | dBi |
| Receiver losses | L_{RX} | 1 | dB |
| Receiver antenna gain | G_{RX} | 3 | dBi |
| Receiver sensitivity | sensitivity | -136 | dBm |
| Standard deviation of path loss | σ_{RSSI} | 5, 10, and 16 | dB |
| Frequency | f | 900 | MHz |
| Number of ANs | m | 10, 15, 20, and 25 | N/A |
| Maximum vegetation attenuation | A_{max} | 26.5 | dB |
| Specific attenuation coefficient | γ | 0.17 | m ⁻¹ |
| Distance loss exponent | η | 1 | N/A |
| Number of measurements per node | n | 50 | N/A |
| Mean value of delay spread | T_1 | 1000 | ns |
| Standard deviation of delay spread | u | 6 | dB |

Table 3

Distribution for the areas covered by 25 anchor nodes according to quartiles Q1 (0-3 ANs), Q2 (4-7 ANs), Q3 (8-11 ANs), and Q4 (12-15 ANs)

| Strategy | Q1 | Q2 | Q3 | Q4 |
|------------|------|--------|--------|-------|
| Random | 3.5% | 33.93% | 59.19% | 3.38% |
| Equispaced | 0% | 57.16% | 42.8% | 0.04% |
| Optimized | 0.1% | 26.2% | 64.7% | 9% |

figure, the following can be noticed. First, in Figure 3(a) where 10 ANs are considered, the percentage of the area that is not covered by at least three ANs is 31.1%, 40%, and 15% for random, equispaced, and optimized approaches, meaning that our scheme outperformed the other strategies in coverage and trilateration abilities. Furthermore, notice that our scheme also outperformed the other strategies when the area covered by more than six ANs is analyzed for the same case (ten ANs), delivering 22.8%, 3.9%, and 25.9% for random, equispaced, and optimized strategies. This enhanced performance indicates that we can deliver more estimated positions and improve localization accuracy. Second, we can analyze the scenario where there are more ANs to cover the area, e.g., 25 ANs, as depicted in Figure 3(d). To conduct the performance analysis, we divided the number of ANs covering a percentage of the area in quartiles, i.e., quartile 1 (Q1) for 0 to 3 ANs, quartile 2 (Q2) for 4 to 7 ANs, quartile 3 (Q3) for 8 to 11 ANs, and quartile 4 (Q4) for 11 to 15 ANs. The summary of this analysis is depicted in Table 3. Note that even with a high number of ANs, the random strategy still has areas in the first quartile of 3.75%, which means that the ability of trilateration is limited in the corresponding zones. In contrast, the equispaced and optimized strategies are close to zero. On the other hand, the last quartile (Q4) shows that our optimized strategy covers 9% of the area, while random and equispaced strategies cover 3.38% and 0.04%, respectively. This means that our scheme has a superior performance in terms of redundancy in trilateration.

4.3. Evaluation of the Optimized Distance Estimation

We assessed the framework's performance proposed in Section 3.2, comparing its accuracy with other state-of-the-art methods such as the ToF, mean, median, moving-average, and Wiener-based estimators. In all the cases, the aim is to estimate the RSSI using the n upload packets and obtain the distance from the path loss model in Equation (2), using a numerical approach. We chose Newton Raphson, a classical numerical method that is computationally efficient and highly accurate. The ToF-based estimator uses the average of the ToF measurements to get the distance directly using the speed of light by $d = c \cdot \overline{\text{ToF}}$ [35]. The mean-based estimator calculates the sample average of the RSSI [36]. This unbiased estimator is a simple and cost-effective solution broadly used for different applications,

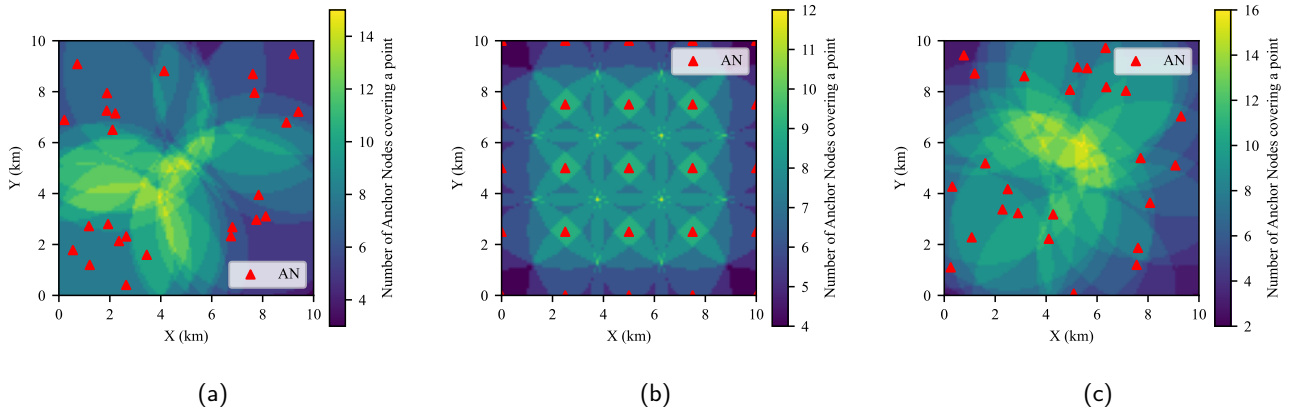


Figure 2: Strategies for placement for 25 ANs using (a) our optimization approach, (b) the equispaced strategy, and (c) random placement.

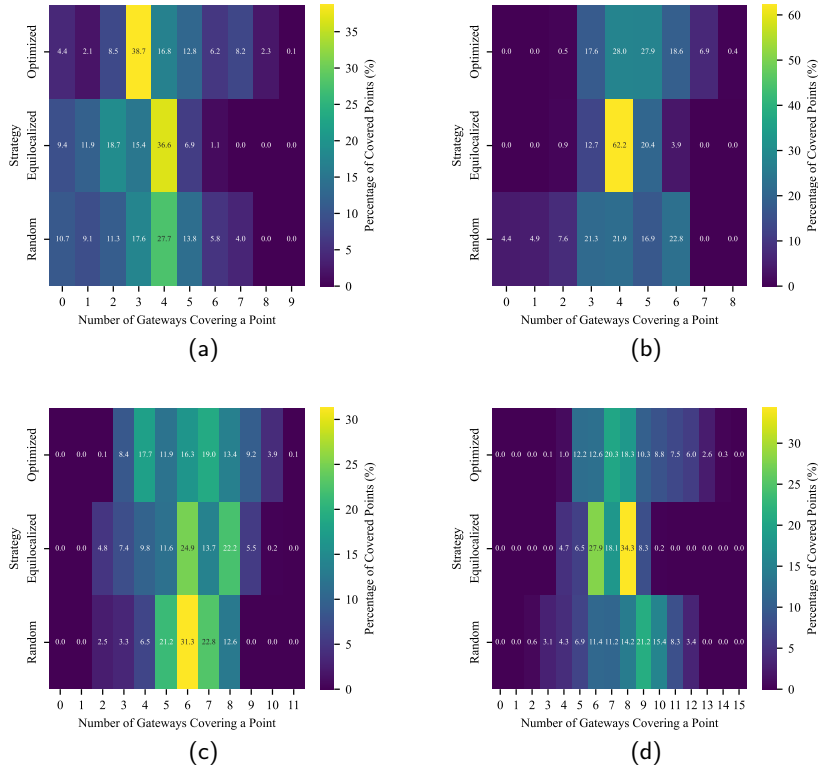


Figure 3: Percentage of covered areas for different ANs when there are a total of deployed ANs of (a) 10, (b) 15, (c) 20, and (d) 25.

like in Transmission Power Control for LPWAN deployments [13]. Since the mean-based estimator is sensitive to outliers likely occurring in the considered environment, we also evaluated the median-based approach [37], which is more resilient to this phenomenon. We also used the moving-average(MA)-based estimator that filters the RSSI values to smooth short-term variations and emphasize the longer-term behavior, helping diminish the noise caused by shadowing and multipath phenomena [38]. Finally, we applied the Wiener filter, which has been extensively used for

Table 4

Performance results of the MAPEs obtained by various estimators. The quartiles, median, and mean are given in percentage.

| Std. Dev (dB) | Estimator | Q1 | Median | Q3 | Mean | MAPE Std. Dev. |
|---------------|--------------|-------|--------|-------|-------|----------------|
| 5 | Proposed | 1.71 | 3.74 | 6.69 | 4.65 | 3.81 |
| | ToF-based | 2.82 | 6.6 | 12.07 | 9.5 | 10.84 |
| | Mean-based | 2.66 | 5.37 | 9.52 | 6.56 | 5.07 |
| | Wiener-based | 2.61 | 5.7 | 9.84 | 6.74 | 5.03 |
| | MA-based | 2.59 | 5.41 | 9.58 | 6.65 | 5.22 |
| | Median-based | 3.09 | 6.38 | 10.72 | 7.44 | 5.43 |
| 10 | Proposed | 2.54 | 5.55 | 9.56 | 6.83 | 5.63 |
| | ToF-based | 3.39 | 6.8 | 12.28 | 9.61 | 11.22 |
| | Mean-based | 5.42 | 10.95 | 18.8 | 13.25 | 10.28 |
| | Wiener-based | 5.61 | 11.19 | 18.18 | 12.98 | 9.61 |
| | MA-based | 5.03 | 11.33 | 19.14 | 13.52 | 10.86 |
| | Median-based | 6.1 | 12.61 | 20.93 | 14.43 | 10.56 |
| 16 | Proposed | 2.48 | 5.61 | 9.85 | 7.28 | 6.45 |
| | ToF-based | 2.8 | 6.43 | 11.92 | 9.0 | 9.33 |
| | Mean-based | 8.27 | 16.71 | 28.47 | 20.69 | 16.98 |
| | Wiener-based | 8.13 | 17.39 | 28.17 | 20.08 | 15.84 |
| | MA-based | 9.01 | 17.77 | 29.71 | 21.36 | 17.04 |
| | Median-based | 10.14 | 18.71 | 31.23 | 22.05 | 16.26 |

RSSI estimation since it optimally estimates this variable by minimizing the mean square error between the predicted and measured values, incorporating the noise statistics [39].

The results of the application of different distance-estimation methods are depicted in Figure 4 and Table 4 where we analyzed the Mean Average Percentage Error (MAPE) for different values of σ_{RSSI} . The MAPE is calculated by

$$\text{MAPE} = \frac{100}{n} \cdot \sum_{i=1}^n \frac{|d_i - \hat{d}_i|}{d_i}. \quad (33)$$

The results can be interpreted as follows. First, at low $\sigma_{\text{RSSI}} = 5$ dB, our distance estimator achieved the lowest MAPE median, mean, and standard deviation, with values of 3.74%, 4.65%, and 3.81% respectively, validating the lower bias and variance of our framework. In contrast, the ToF-based, mean-based, MA-based, and Wiener-based estimators exhibited similar performance of about 5 to 6% MAPE, indicating less accuracy under low noise conditions. Contrasting these results with those obtained at harsh noise conditions, i.e., $\sigma_{\text{RSSI}} = 16$ dB, it can be observed a similar performance where our approach outperforms the baseline strategies analyzed. Thus, our strategy preserves its low MAPE in the median, mean, and standard deviation of 5.61%, 7.28%, and 6.45%, respectively, compared with the other baselines, which were about 16% to 22%, exhibiting a considerable error for localization. The ToF-based estimator preserved a similar performance for different noise conditions but did not outperform our optimized distance estimator for the considered scenarios.

4.4. Evaluation of the REB Localization Strategy

Via numerical studies, we evaluated the approach previously described in Section 3.3 and contrasted the results versus central tendency proposals [40] since the estimated positions $\hat{p}_e^{(i)}$ of the EN converge to the mean according to the law of large numbers [41]. It is essential to remark that due to the effects of shadowing, multipath fading, and delay spread, some ANs combinations for trilateration can deliver inaccurate positions even when we proposed an enhanced distance estimator in Section 3.2. Thus, using all the combinations of ANs to average the estimated positions may be influenced by outliers. In that way, using range-based localization estimators that are less sensitive to outliers has been broadly implemented [42]. To perform the comparison, we delivered two strategies under the following scenarios:

1. Obtain the distance from the EN and each AN that covers the EN using our estimator and the strategy that performs the best according to the results in Section 4.3, that is, the ToF-based estimator, iterating the actual position of the EN for each point in the grid of 10.000 points of the validation area.

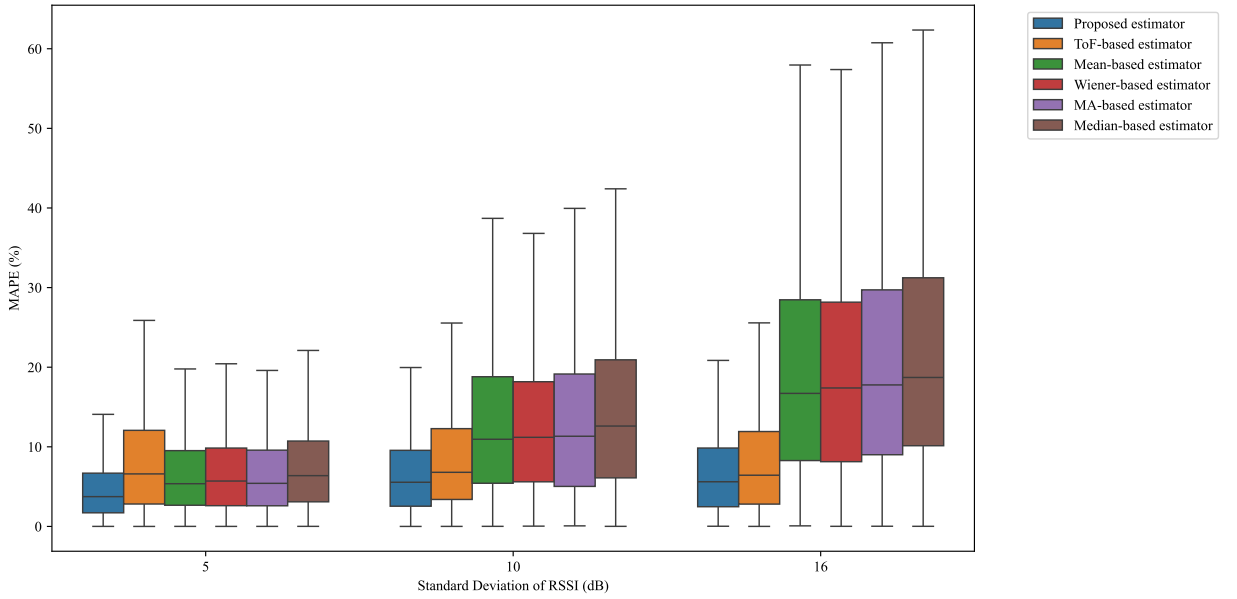


Figure 4: Boxplots of the obtained MAPEs for the proposed estimator versus other estimators in the literature.

2. Find all the $\hat{p}_e^{(i)}$ using trilateration techniques according to Equation (29).
3. Find the median values for the estimated coordinates \hat{x}_e and \hat{y}_e .
4. Use the method proposed in Section 3.3 to score the quality of the set of $\hat{p}_e^{(i)}$, and select the best ones with $K = 0.15$.
5. Contrast the obtained results.

We present the results in Table 5 and Figure 5, from the following can be noticed. Specifically, Table 5 presents the MAPE and the standard deviation of methods, which were evaluated for different numbers of ANs ($m \in \{10, 15, 20, 25\}$) and values of σ_{RSSI} (5, 10, and 15 dB). Overall, the REB method outperforms the median-based localization method in terms of accuracy and precision (lower standard deviation of errors) in the tested case uses, as we will detail. For instance, when $m = 10$ and $\sigma_{\text{RSSI}} = 5$ dB, the REB and median-based methods delivered MAPEs of 3.75% and 5.91% with standard deviations of 5.76% and 10.96%, indicating that our approach achieves better performance under scenarios with few ANs and low channel variability. As this variability increases in scenarios with greater effects of shadowing and multipath, i.e., $\sigma_{\text{RSSI}} = 16$ dB, the performance of both methods degrades, which is expected due to the uncertainty in the RSSI measurements at each AN. However, it can be noticed that our approach is more resilient regarding this variability. For example, in the scenario where $n = 10$ and $\sigma_{\text{RSSI}} = 16$ dB only degraded its MAPE to 5.53% and standard deviation of 7.15%, while the median-based strategy increased the MAPE up to 9.66% and a significant standard deviation of 15.87%. These results confirm that our REB method is more robust in handling higher channel variability. On the other hand, the evaluated scenarios also highlighted the performance of both methods when the number of ANs increased. Thus, when m is increased from 10 to 25 with $\sigma_{\text{RSSI}} = 5$ dB, the REB method decreased its MAPE and standard deviation from 3.75% and 5.76% to 2.57% and 3.76%, respectively. The same analysis can be performed for the median-based localization method. This improvement can be attributed to the extra data given by more ANs, which improved the spatial information.

5. Conclusions

In this paper, we presented a localization framework based on WSNs for dense forests, where the GPS-based localization tasks may not be performed due to limitations in energy consumption and direct communications with the satellites. We presented a three-step strategy as follows. First, we proposed an optimization framework to select the best

Table 5

MAPE and standard deviation of MAPE of the REB localization strategy and the median-based strategy for different number of ANs and different values of σ_{RSSI}

| m | σ_{RSSI} (dB) | REB MAPE (%) | REB Std.Dev (%) | Median MAPE (%) | Median Std.Dev.(%) |
|----|-----------------------------|--------------|-----------------|-----------------|--------------------|
| 10 | 5 | 3.75 | 5.76 | 5.91 | 10.96 |
| | 10 | 5.49 | 7.53 | 10.70 | 20.43 |
| | 16 | 5.53 | 7.15 | 9.66 | 15.87 |
| 15 | 5 | 3.88 | 5.22 | 4.75 | 6.65 |
| | 10 | 4.72 | 7.45 | 6.87 | 11.23 |
| | 16 | 5.28 | 8.35 | 7.63 | 14.44 |
| 20 | 5 | 2.85 | 4.90 | 3.43 | 5.70 |
| | 10 | 3.01 | 3.71 | 5.38 | 6.54 |
| | 16 | 4.37 | 5.91 | 6.11 | 9.01 |
| 25 | 5 | 2.57 | 3.45 | 3.46 | 4.34 |
| | 10 | 3.49 | 5.42 | 5.00 | 7.60 |
| | 16 | 4.89 | 6.31 | 6.54 | 8.83 |

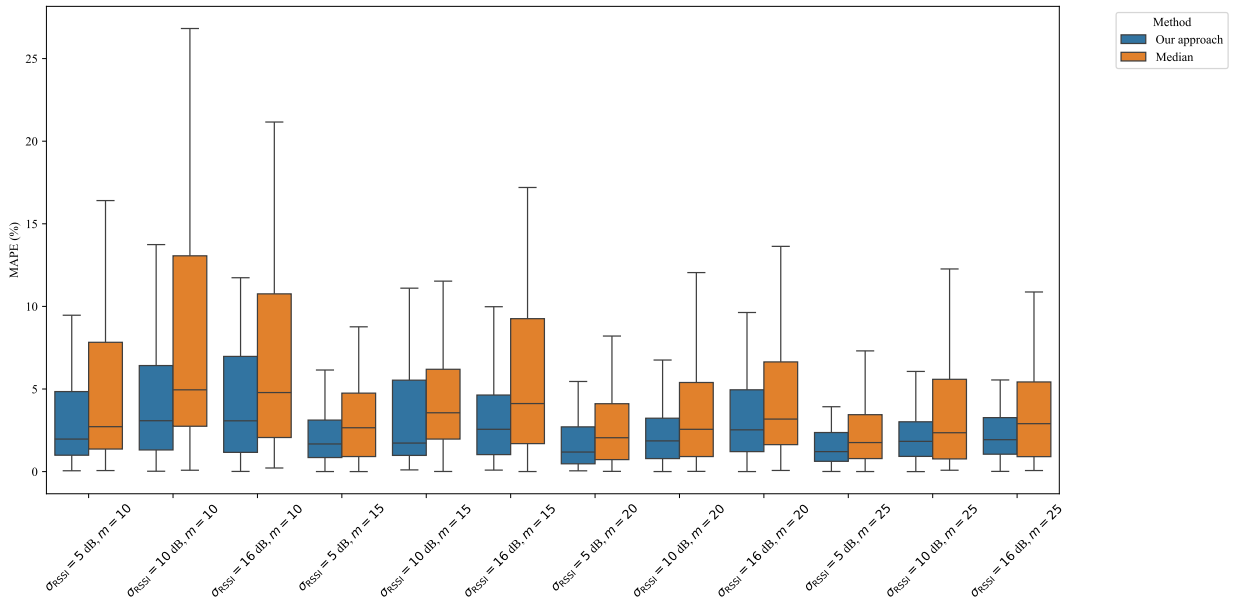


Figure 5: Trilateration boxplots for different standard deviations and number of ANs. We compare our approach to the median-based localization strategy.

positions for a set of ANs, maximizing the coverage and trilateration redundancy and penalizing the extra overlap of the coverage of ANs. This is to tackle the cases where an AN is contained within another AN coverage, being useless for trilateration purposes. To solve this optimization problem, we used GAs to evaluate many different solutions in parallel where it can converge more rapidly to the globally optimized positions of the ANs. We compared the results with state-of-the-art techniques like equispaced-based and randomly deployed ANs localization techniques. Second, we proposed a hybrid RSSI-ToF-based estimator to determine the distance between the EN and the AN, showing imperceptible bias and lower variance than the mean-based estimators. We compared the performance of our approach with different strategies based on central tendency, like the mean, median, moving average, and the Wiener filter, finding that our method outperformed all the considered strategies. Finally, we proposed a score-based method to leverage the quality in the trilateration estimations of $\binom{m}{3}$ AN combinations based on the residual errors between the estimated distance in the last step and the distance after trilateration. We compared the results with other estimation methods

Table A-1

List of acronyms

| Acronym | Meaning | Acronym | Meaning |
|---------|--------------------------------------|---------|---|
| AN | Anchor Node | GA | Genetic Algorithm |
| EN | End Node | GPS | Global Positioning System |
| FSPL | Free Space Path Loss | IUCN | International Union for Conservation of Nature |
| ISM | Industrial, Scientific, and Medical | ITU-R | International Telecommunication Union - Radiocommunication Sector |
| KDE | Kernel Density Estimation | LPWAN | Low Power Wide Area Network |
| MAPE | Mean Average Percentage Error | MA | Moving Average |
| NS | Network Server | P.833 | ITU-R Recommendation P.833 |
| PSO | Particle Swarm Optimization | REB | Residual Error-Based |
| RF | Random Forest | RFID | Radio Frequency Identification |
| REB | Residual Error-Based | RSSI | Received Signal Strength Indicator |
| SNR | Signal-to-Noise Ratio | ToF | Time of Flight |
| TLBO | Teaching–Learning-Based Optimization | WCCE | Weighted Cascade Compensation Estimator |
| WSN | Wireless Sensor Network | - | - |

less sensitive to outliers, like the median-based strategies, finding that our REB method significantly outperformed this baseline method. In general, we conclude that our framework provides a reliable solution for wildlife localization in dense forests, from the election of the ANs localization to the accurate positioning of animals. In future work, we plan to deploy a testbed with our localization framework and perform the energy consumption analysis based on this deployment.

A. Acronyms and Symbols

To ease the readability of this manuscript, we present the acronyms used in this paper in Table A-1, and the list of key notations in Table A-2.

References

- [1] J. S. Albert, A. C. Carnaval, S. G. Flantua, L. G. Lohmann, C. C. Ribas, D. Riff, J. D. Carrillo, Y. Fan, J. J. Figueiredo, J. M. Guayasamin, Human impacts outpace natural processes in the amazon, *Science* 379 (6630) (2023) eabo5003.
- [2] IUCN, The iucn red list of threatened species, [Accessed 15 November 2024] (2024).
URL <https://www.iucnredlist.org/>
- [3] T. E. Lovejoy, C. Nobre, Amazon tipping point: Last chance for action (2019).
- [4] A. M. Manville, B. B. Levitt, H. C. Lai, Health and environmental effects to wildlife from radio telemetry and tracking devices—state of the science and best management practices, *Frontiers in Veterinary Science* 11 (2024) 1283709.
- [5] L. A. Gould, A. D. Manning, H. M. McGinness, B. D. Hansen, A review of electronic devices for tracking small and medium migratory shorebirds, *Animal Biotelemetry* 12 (1) (2024) 11.
- [6] J. Cabezas, R. Yubero, B. Visitación, J. Navarro-García, M. J. Algar, E. L. Cano, F. Ortega, Analysis of accelerometer and gps data for cattle behaviour identification and anomalous events detection, *Entropy* 24 (3) (2022) 336.
- [7] N. L. Babic, C. P. Johnstone, S. Reljić, A. Sergiel, D. Huber, R. D. Reina, Evaluation of physiological stress in free-ranging bears: current knowledge and future directions, *Biological Reviews* 98 (1) (2023) 168–190.
- [8] D. Zorbas, A. Sabyrbek, Supporting critical downlink traffic in lorawan, *Computer Communications* 228 (2024) 107981.
- [9] N. Khidani, F. Hendessi, An energy efficient prediction based protocol for target tracking in wireless sensor networks, *Ad Hoc Networks* 167 (2025) 103688. doi:<https://doi.org/10.1016/j.adhoc.2024.103688>.
URL <https://www.sciencedirect.com/science/article/pii/S1570870524002993>
- [10] E. Ossongo, M. Esseghir, L. Merghem-Boulahia, A multi-agent federated reinforcement learning-based optimization of quality of service in various lora network slices, *Computer Communications* 213 (2024) 320–330.
- [11] L. Zhang, X. Zhou, D. Li, Z. Yang, Hcnnet: Hybrid coupled cooperative network for robust indoor localization, *ACM Transactions on Sensor Networks* 20 (4) (2024) 1–22.
- [12] A. Goldsmith, *Wireless Communications*, Cambridge University Press, Standford, CA, 2005.
- [13] M. González-Palacio, D. Tobón-Vallejo, L. M. Sepúlveda-Cano, M. Luna-delRisco, C. Roehrig, L. B. Le, Machine-learning-assisted transmission power control for lorawan considering environments with high signal-to-noise variation, *IEEE Access* 12 (2024) 54466–54487.

- [14] Recommendation itu-r p.833-10 - attenuation in vegetation, Tech. Rep. P.833-10, International Telecommunication Union, Geneva, Switzerland (Sep. 2021).
URL https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.833-10-202109-I!!PDF-E.pdf
- [15] K. Z. Islam, D. Murray, D. Diepeveen, M. G. Jones, F. Soheli, Lora localisation using single mobile gateway, *Computer Communications* 219 (2024) 182–193.
- [16] S. S. Mohar, S. Goyal, R. Kaur, Jaya nl-wsn: Jaya algorithm for node localization issue in wireless sensor network, *Wireless Personal Communications* 137 (1) (2024) 287–324.
- [17] P. Wu, L. Yu, X. Yi, L. Xu, L. Liu, Y. Yi, T. Jiang, C. Tao, Research on wsn reliable ranging and positioning algorithm for forest environment, *Scientific Reports* 14 (1) (2024) 5417.
- [18] J. Li, P. Li, P. Li, L. Tang, X. Zhang, Q. Wu, Self-position awareness based on cascade direct localization over multiple source data, *IEEE Transactions on Intelligent Transportation Systems* 25 (1) (2022) 796–804.
- [19] G. Kaur, K. Jyoti, N. Mittal, V. Mittal, R. Salgotra, Optimized approach for localization of sensor nodes in 2d wireless sensor networks using modified learning enthusiasm-based teaching–learning-based optimization algorithm, *Algorithms* 16 (1) (2022) 11.
- [20] H. Chen, J. Yang, Z. Hao, M. Ga, X. Han, X. Zhang, Z. Chen, Research on indoor positioning method based on lora-improved fingerprint localization algorithm, *Scientific Reports* 13 (1) (2023) 13981.
- [21] S. Lyu, Q. Li, Z. Li, H. Liang, J. Chen, Y. Liu, H. Huang, Precision location-aware and intelligent scheduling system for monorail transporters in mountain orchards, *Agriculture* 13 (11) (2023) 2094.
- [22] J. Liu, L. Wei, X. Du, L. Jin, A localization algorithm based on coevolutionary noise-suppressing newton method for wireless sensor network, *IEEE Sensors Journal* (2024).
- [23] W. G. Lima, A. V. Lopes, C. M. Cardoso, J. P. Araújo, M. C. Neto, M. E. Tostes, A. A. Nascimento, M. Rodriguez, F. J. Barros, Lora technology propagation models for iot network planning in the amazon regions, *Sensors* 24 (5) (2024) 1621.
- [24] L. J. Greenstein, V. Erceg, Y. S. Yeh, M. V. Clark, A new path-gain/delay-spread propagation model for digital cellular channels, *IEEE Transactions on Vehicular Technology* 46 (2) (1997) 477–485.
- [25] H. R. Anderson, *Fixed broadband wireless system design*, John Wiley & Sons, Hoboken, NJ, 2003.
- [26] U. W. H. Centre, Los katiós national park — whc.unesco.org, <https://whc.unesco.org/en/list/711/>, [Accessed 14 November 2024] (2024).
- [27] E. Arnon, S. Cain, A. Uzan, R. Nathan, O. Spiegel, S. Toledo, Robust time-of-arrival location estimation algorithms for wildlife tracking, *Sensors* 23 (23) (2023) 1–23. doi:10.3390/s23239460.
URL <https://www.mdpi.com/1424-8220/23/23/9460>
- [28] M. González-Palacio, L. M. Sepúlveda-Cano, J. Quiza-Montealegre, J. P. D’amato, Mejoramiento del algoritmo adr en una red de internet de las cosas lorawan usando aprendizaje de máquina, *Revista Ibérica de Sistemas y Tecnologías de Información* 2020 (5) (2020) 164–188.
- [29] M. González-Palacio, L. M. Sepúlveda-Cano, D. P. Tobón-Vallejo, C. Azurdia-Meza, Characterisation of path loss models in wireless sensor networks: scenarios, variables, and open problems, *International Journal of Ultra Wideband Communications and Systems* 5 (3) (2022) 164–188.
- [30] M. González-Palacio, D. Tobón-Vallejo, L. M. Sepúlveda-Cano, S. Rúa, G. Pau, L. B. Le, Lorawan path loss measurements in an urban scenario including environmental effects, *Data* 8 (1) (2022) 4.
- [31] M. González-Palacio, D. Tobón-Vallejo, L. M. Sepúlveda-Cano, S. Rúa, L. B. Le, Machine-learning-based combined path loss and shadowing model in lorawan for energy efficiency enhancement, *IEEE Internet of Things Journal* 10 (12) (2023) 10725–10739.
- [32] G. Cuzzo, R. Marini, C. Buratti, K. Mikhaylov, On the support of the 2.4 ghz band in the lorawan standard, *IEEE Internet of Things Magazine* (2024) 66–71.
- [33] H. Pan, X. Qi, M. Liu, L. Liu, An uwb-based indoor coplanar localization and anchor placement optimization method, *Neural Computing and Applications* 34 (19) (2022) 16845–16860.
- [34] K. Shilpi, Sensor node localization using nature-inspired algorithms with fuzzy logic in wsns, *The Journal of Supercomputing* 80 (19) (2024) 26776–26804.
- [35] B. Li, Y. Xu, Y. Liu, Z. Shi, Lorawaps: A wide-area positioning system based on lora mesh, *Applied Sciences* 13 (17) (2023) 9501.
- [36] H. Vo, V. L. Tran, F. Ferrero, F.-Y. Lee, M.-H. Tsai, Advance path loss model for distance estimation using lorawan network’s received signal strength indicator (rss), *IEEE Access* 12 (12) (2024) 83205–83216.
- [37] Z. Wei, Z. Zhou, A combined filtering method for zigbee indoor distance measurement, *Sensors* 24 (10) (2024) 3164.
- [38] L. Bouse, S. A. King, T. Chu, Simplified indoor localization using bluetooth beacons and received signal strength fingerprinting with smartwatch, *Sensors* 24 (7) (2024) 2088.
- [39] P. Savazzi, E. Goldoni, A. Vizziello, L. Favalli, P. Gamba, A wiener-based rssi localization algorithm exploiting modulation diversity in lora networks, *IEEE Sensors Journal* 19 (24) (2019) 12381–12388.
- [40] Z. Wu, Y. Li, X. Zhang, X. Meng, X. Lv, Y. Wu, Multiple anchors and ris-aided localization method in complex nlos environments, *IEEE Internet of Things Journal* 24 (2024) 13578 – 13588.
- [41] K. L. Judd, The law of large numbers with a continuum of iid random variables, *Journal of Economic theory* 35 (1) (1985) 19–25.
- [42] G. Pettorru, V. Pilloni, M. Martalò, Trustworthy localization in iot networks: A survey of localization techniques, threats, and mitigation, *Sensors* 24 (7) (2024) 2214.

Table A-2
 Mathematical symbols

| Symbol | Meaning | Symbol | Meaning |
|----------------------------------|--|------------------------------|--|
| A | Link budget constant | A_{\max} | Maximum vegetation attenuation |
| a, b | Optimal weights for RSSI estimator | α, β, λ | Weights in optimization problem |
| $b_{\widehat{\text{RSSI}}_{it}}$ | Bias of estimated RSSI from ToF | c | Speed of light |
| C_i | Coverage area of AN_i | d_i | Distance between EN and AN_i |
| $\delta(\Omega)$ | Indicator function for condition Ω | ΔT_i | Excess delay due to multipath |
| ϵ_i | Measurement noise in RSSI | η | Distance loss exponent |
| f | Frequency of operation | $F_{i,j}$ | Fractional overlap between AN_i and AN_j |
| G_{RX} | Receiver antenna gain | G_{TX} | Transmitter antenna gain |
| γ | Specific attenuation coefficient | \hat{d}_i | Estimated distance between EN and AN_i |
| \hat{d}_{it} | Estimated distance from ToF | \hat{p}_e | Estimated position of EN |
| $\hat{p}_e^{(i)}$ | Estimated position from i^{th} trilateration | \hat{T}_i | Adjusted Time of Flight at AN_i |
| $I_i(p)$ | Coverage indicator function | K | Proportion of positions used in localization |
| L_{RX} | Receiver losses | L_{TX} | Transmitter losses |
| l | Number of points per axis in discretized grid | m | Number of Anchor Nodes |
| n | Number of measurements per node | P | Set of Anchor Node positions |
| p_e | Position of the End Node | p_i | Position of Anchor Node i |
| $PL(d_i)$ | Total path loss at distance d_i | $PL'(d_i)$ | First derivative of path loss function |
| $PL''(d_i)$ | Second derivative of path loss function | $PL_{\text{FS}}(d_i)$ | Free space path loss at d_i |
| $PL_{\text{veg}}(d_i)$ | Vegetation attenuation at d_i | p_i, p_i', p_i'' | Positions of ANs used in trilateration |
| $\Phi(P)$ | Objective function for AN placement | $\psi_{i,j}$ | Overlap penalty between AN_i and AN_j |
| Q | Quartile | R_i | Accumulated residual error for position $\hat{p}_e^{(i)}$ |
| r_{ij} | Residual error between $\hat{p}_e^{(i)}$ and AN_j | σ_{RSSI_i} | Standard deviation of RSSI noise at AN_i |
| S | Area under consideration | $S(p)$ | Set of ANs covering point p |
| RSSI_i | Received Signal Strength Indicator at AN_i | sensitivity | Receiver sensitivity threshold |
| τ_{rms} | Root mean square delay spread | T_i | Time of Flight to AN_i |
| T_1 | Mean τ_{rms} at 1 km | $T(p)$ | Number of trilaterations at point p |
| u | Zero-mean lognormal variate | $\text{Var}[\cdot]$ | Variance operator |
| $\widehat{\text{RSSI}}_i$ | Weighted RSSI estimator at AN_i | $\widehat{\text{RSSI}}_{it}$ | Estimated RSSI from ToF at AN_i |
| $\overline{\text{RSSI}}_i$ | Mean RSSI value at AN_i | $\mathbb{E}[\cdot]$ | Expected value operator |
| $\text{CoverageArea}(P)$ | Coverage area function in optimization | $\text{OverlapPenalty}(P)$ | Overlap penalty function in optimization |
| $\text{TrilaterationScore}(P)$ | Trilateration score function | | |