Article

A Probabilistic Approach for Optimal Design of Indoor Positioning Systems

Bráulio Henrique O. U. V. Pinto [1],*, Horácio A. B. F. de Oliveira [1], Eduardo J. P. Souto [1]

[1]　Institute of Computing, Federal University of Amazonas, Manaus, Amazonas, Brazil

*　Correspondence: braulio@icomp.ufam.edu.br

Abstract: Indoor Positioning Systems (IPSs) are designed to provide solutions for location-based services. Wireless local area network (WLAN)-based positioning systems are the most widespread around the globe and are commonly found to have a ready-to-use infrastructure composed mostly of access points (APs). They provide useful information on signal strength to be processed by adequate location algorithms, which are not always capable of achieving the desired localization error only by themselves. In this sense, this paper proposes a new method to improve the accuracy of IPSs by optimizing some of their most relevant infrastructure components. Included are the arrangement of APs over the environment, the number of reference points (RPs), and the number of samples per location estimation test. A simulation environment is also proposed, in which the impact of key influencing factors on system accuracy is analyzed. Finally, a case study is simulated to validate an optimal combination of design parameters and its compliance with the requirements of localization error and the limited number of access points. Our simulation results clearly show that the desired localization accuracy, which is set as a goal, can be achieved while maintaining the factors already mentioned at minimal levels, which decreases both system deployment costs and computational effort.

Keywords: indoor positioning; access point placement; path loss model; optimization

---

1. Introduction

Indoor positioning systems (IPSs) are a reality and provide location information of devices and persons for different applications in the real world. With the appropriate technology, it is possible to locate products in a warehouse, firefighters in a burning building, medicines in a hospital, maintenance tools spread over a plant, and so forth [1].

Applications already well established as Google Maps, Waze, and Uber are also location-based services, except that they are used outdoors. In this case, the most widespread technology is the *Global Positioning System* (GPS). Unfortunately, GPS does not perform well indoors, as it needs, among other factors, direct line of sight to the satellites and the device whose location one wants to know [2].

An indoor positioning system must take into account some factors whose effects compromise the accuracy when estimating the location. Lack of line of sight, the influence of obstacles and obstructions such as walls and human movement, multipath propagation, and interference noises are examples of factors that result in the low performance of the most commonly deployed solutions [3].

Most of these systems use wireless technologies such as WiFi or Bluetooth due to a wide available and accessible infrastructure, which saves time and related costs of deployment [4]. A common architecture consists of mobile devices, access points (APs), and a central server. The main goal is to obtain location information of the mobile devices. To do this, the devices should transmit signals, whose power levels are captured by the access points which are spread over the environment. The power levels, well-known in the literature as *Received Signal Strength* (RSS), are passed on to the central server. After that, these data are processed using techniques and appropriate algorithms to determine the location of the devices. Figure 1 illustrates better this situation.
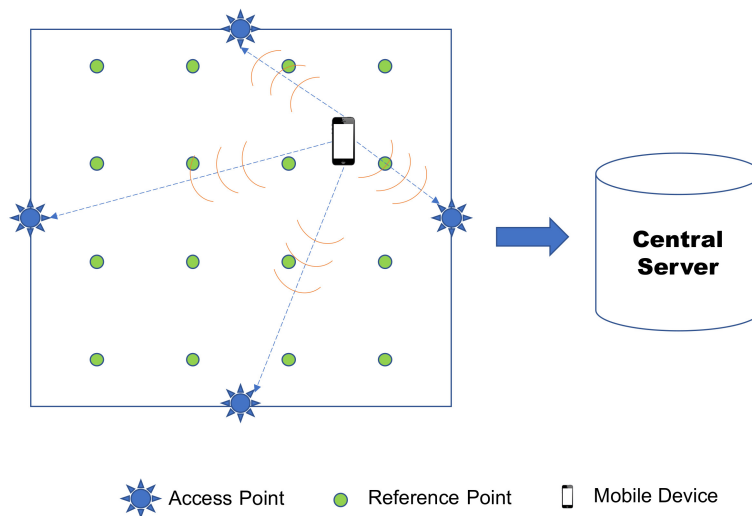
Figure 1. An example of a typical WLAN IPS topology.

There is a vast literature of positioning algorithms used for IPSs, which include deterministic and probabilistic methods. The first ones are quite common in fingerprinting-based localization, which basically consists of two main steps: an offline phase, in which RSS measurements (fingerprints) are previously collected in the environment; and an online phase, in which machine learning techniques and algorithms are used for the location estimation by a comparison between the offline database and the RSS data collected in real time. One of the first and most traditional system is the RADAR [5], which achieved an accuracy of 2 – 3 m. The second, probabilistic methods, are much more common in propagation model-based systems, when the model is able to describe the environment reasonably. One advantage of this method is a better computational efficiency. A well-known probabilistic-based solution is the HORUS [6], which achieved an accuracy of approximately 2 m. Besides that, many systems provide hybrid solutions taking into account specificities of the indoor environment, seeking in general to improve accuracy. In this sense, IPS optimal design is also a hot research topic, since high localization performance can be achieved by means of a few infrastructure modifications [7].

In this work, we propose a model and simulation based approach to optimize the deployment of the most relevant design factors that influence the accuracy of an IPS. In our work, they are restricted to the number of reference points (RPs), the number of samples collected per test, and the arrangement of the access points (APs) over the environment. Starting from the model parameters which describe the environment, we address the influencing factors and analyze their impact on the positioning error. Then, we propose a method to improve the system accuracy while keeping each correspondent factor value at a minimum level. This way, the desired accuracy can be achieved by simply adjusting the values of the factors which compound the positioning system infrastructure.

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 details the probabilistic model approach. Section 4 discuss the impact of relevant design factors on system accuracy, and a method to find an optimal combination of factors to achieve a required accuracy is proposed. Section 5 presents a case study to validate the method proposed. Section 6 concludes the paper.

2. Related Work

One of the first discussions about optimizing IPS design factors to improve accuracy was posed by the work of Kaemarungsi and Krishnamurthy [8]. They used a probabilistic model to represent the RSS variation over the environment in fingerprinting-based systems. A framework was developed to

analyze the influence of the number of APs, grid spacing between training points, and environment parameters on localization error. Although the fundamental theory and intuitions behind the design problem are carefully described, the authors neither propose nor apply a specific method to improve accuracy for real systems. Hara and Fukumura [9] proposed an efficient method to achieve a required localization accuracy with the deployment of a minimum number of access points. They use the maximum likelihood estimation to determine the location of targets and develop mathematical formulas that relate the variables involved in the optimization. Although they validate the proposed system experimentally, they do not take into account the impact of the arrangement of APs over the environment.

The previous consideration of finding an optimal AP placement is explored by Zhao et al. [10] in which, given a fixed number of APs, a Differential Evolution algorithm is used to find the APs placement. An interesting result was that the best arrangements had the APs distributed in a zigzag pattern instead of the hitherto classical AP distribution at the borders. Also, the results were validated with model-based simulations and testbed experiments. Nevertheless, the number of possible places for the APs to be allocated was considered fixed, which did not allow its associated impact – on both system accuracy and computational effort – could be addressed. He et al. [11] used a genetic algorithm to determine both the minimum number of APs and the best arrangement of APs over an area to achieve the desired localization error. They simulate a fingerprinting-based IPS using the Nearest-Neighbors (NN) method for the location estimations. Like the preceding work, they also found a similar behavior for the distribution of APs – the authors called it a "serrated" pattern. Although more complete from a simulation point of view, the work considered a fixed number for the RPs (or training points in this case) throughout all the simulations performed. Consequently, the impact of RPs on the system accuracy was not addressed.

From the 2010's on, most of the works have tried to apply efficient algorithms to optimize the placement of APs in an indoor area. Farkas et al. [12] used an algorithm based on simulated annealing to find a minimum number of APs to achieve a required criterion of AP perceivability. They discretize the area in a reasonable number of points and signalize the problem of optimization as NP-complete, suggesting a method to approximate the global optimum solution. With n possible AP location possibilities they compare the time complexity results obtained by their solution – $\mathcal{O}(n)$ – with the brute force algorithm – $\mathcal{O}(2^n)$. Despite the detailed approach, the work is geared towards localization with triangulation techniques. Aomumpai et al. [13], on the other hand, worked with a path loss model considering obstructions and used the Binary Integer Linear Programming (BILP) method to optimize the number of APs. They compare their results with other approaches based on the average localization error achieved and do not mention the algorithm complexity issues. The scenarios and results were all obtained by simulation.

Rajagopal et al. [14] proposed a toolchain to optimize the number of beacons (APs) while keeping a sufficient signal coverage over the indoor plan. The metrics used to compare different configurations were an enhanced Geometric Dilution of Precision (GDOP) and the cumulative distribution function of the localization error. Several floor plans were used as scenarios and they demonstrated the improvement made by their method. Despite the promising results, the simulations were based on an ideal ray-tracing model, which perhaps might not be extended to more complex environments. Also, there is not any mention relating to the time complexity of the proposed method. Similar work is presented by Sharma and Badarla [15], except for the use of a Mixed Integer Linear Programming (MILP) approach and the extension for three dimensional (3D) indoor localization. In spite of the improvements obtained, neither a generalizing propagation model nor a testbed experiment for validation was addressed. Jia et al. [16] combined the previous works by proposing a technique to reduce the possible AP configurations for fingerprinting-based IPSs. They use a lognormal shadowing path loss model that includes walls and people attenuation, which brings more reality to the results obtained. Still, the work lacks considerations about the influence of the number of RPs (or training points simulated) and does not mention the time complexity of the proposed method.

As we can see, many authors have tried to address the problem of optimizing the influencing factors that surround IPSs, mainly for improving localization accuracy. The majority of the research presented before relies on the construction of path loss models. However, a more complete performance analysis including the most relevant factors which might influence the localization error was not apparently done yet. In contrast, besides presenting a generalizing path loss model, we address the influence of the arrangement of APs, the number of RPs, and the number of samples collected per test on the system accuracy. Moreover, we consider the influence of the number of tests performed, and the use of the 95% interval of cumulative error distribution metric (instead of the classical average error) for more representative results. Finally, we analyze the time complexity of the method proposed, which can give researchers some insights for improvements and future work. Hereupon, to the best of our knowledge, no work has either combined or addressed these listed design factors and optimized their values for achieving a required localization error. Thus, we seek to point out the nuances involving optimization for IPSs with both analytical and simulation approaches.

3. Modeling of the Probabilistic-based Positioning System

3.1. Mathematical Principles

Although geometrical approaches for indoor positioning are vastly used, the high RSS variability is still a big issue to handle. One reasonable choice is to use statistical models to deal with uncertainty [17]. More specifically, the log-distance path loss model is a well established one to represent indoor signal propagation [18], as shown in Equation (1):

$$PL(dB) = PL(d_0) + 10\alpha \log\left(\frac{d}{d_0}\right) + X_\sigma \tag{1}$$

where $PL(d_0)$ is a constant which represents the path loss in dB at a distance d_0 used as a reference, α is the path loss exponent, and X_σ is a normal random variable with zero mean and standard deviation σ in dB, that is, $X \sim \mathcal{N}(0, \sigma^2)$. All these parameters are specific for each place and describe the distribution of RSS at a point distant d from a transmitter. They are often obtained by collecting and processing RSS measurements with linear regression techniques or maximum likelihood estimation [17]. The use of this simulation model is convenient due to its generalizing capability for indoor systems and the very efficient usage of computational effort in the process of localization.

For simplicity, we consider an IPS topology consisting of transmitters fixed over the room and receivers which one wants to locate. The first ones are known as APs, which are often WiFi or Bluetooth-based [1,4]. The last ones are usually smart devices that can receive RSS information provided by the APs. This way, the goal is to simply collect and process RSS data to estimate the device location. The model presented before can be slightly modified to describe the distribution of RSS at each point over the area:

$$r = P_t - PL(dB) = \{P_t - PL(d_0)\} - \left\{10\alpha \log\left(\frac{d}{d_0}\right) + X_\sigma\right\} \tag{2}$$

where r is the perceived power in the receiver device and P_t is the AP transmission power. It is important to notice that r is also a random variable, which can be represented by $r \sim \mathcal{N}(\mu_r, \sigma^2)$, in which μ_r is the RSS expected value for a point in the environment:

$$\mu_r = P_t - PL(d_0) - 10\alpha \log\left(\frac{d}{d_0}\right) \tag{3}$$

The equations listed above describe the distribution of RSS given a point distant d from an AP, which is known in the literature as the *likelihood function*, whose probability density function (p.d.f) is given by:

$$p(r|l) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2}\left(\frac{r - \mu_r}{\sigma}\right)^2\right\} \tag{4}$$

where l is such that $d = \|l - l_{AP}\|$, with l and l_{AP} being the test point and the AP coordinates, respectively.

On the other hand, the main interest is to know the distribution of l given the RSS information, which is obtained by the collected data. In this case, the *posterior function* contains the necessary information to estimate the location coordinate l. According to Bayes' rule:

$$p(l|r) = \frac{p(r|l)p(l)}{p(r)} \tag{5}$$

where $p(l)$ is the *prior* function and $p(r)$ is a normalizing factor given by the total probability theorem:

$$p(r) = \int p(r|l')p(l')dl' \tag{6}$$

Equation (6) refers to the continuous case, in which l' represents each possible coordinate uniformly distributed over the area. Although it is computationally unfeasible to calculate this integral analytically, an approximation to the discrete form can be done [19]. That is, the area can be divided into many discrete coordinates as possible, treated here as the reference points (RPs). Likewise, the likelihood function is computed for each RP given. Thus, Equation (5) can be rewritten as:

$$p(l_i|r) = \frac{p(r|l_i)p(l_i)}{\sum\limits_{j=1}^{m} p(r|l_j)p(l_j)} \tag{7}$$

where m is the number of RPs and $p(l_i|r)$ is the posterior function that relates the measure of RSS r with location l_i, in which $i \in \{1, 2, ..., m\}$.

The equations developed so far take into account one RSS sample from one AP only. However, $n > 1$ APs are considered in practical situations to improve accuracy, as it generates fewer ambiguities among the candidate RPs for the estimated location. In this case, the n-dimensional RSS vector $\mathbf{r}$ is adopted instead of the one-dimensional r. Another strategy to improve accuracy is to collect a sufficient number of RSS samples and take their mean for the estimation. According to the *strong law of large numbers*, the sample mean $\bar{\mathbf{r}}$ tends to its true value $\mu_{\bar{\mathbf{r}}}$, as well as the *Tchebycheff's condition* states that the sample variance $\bar{\sigma}_n^2$ tends to zero as $n \to \infty$ [20]. Thereby, as variance diminishes, accuracy is improved due to fewer ambiguities in the estimation calculus. The likelihood function already presented in Equation (4) can be rewritten as:

$$p(\mathbf{r}|l) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp\left\{-\frac{(\mathbf{r} - \mu_{\mathbf{r}})^T\Sigma^{-1}(\mathbf{r} - \mu_{\mathbf{r}})}{2}\right\} \tag{8}$$

where Σ is the covariance matrix and $\mu_{\mathbf{r}}$ the vector with expected values for the RSS at location l. Moreover, considering the RSS data provided by different APs as statistically independent, Equation (7) becomes:

$$p(l_i|\mathbf{r}) = \frac{\prod\limits_{k=1}^{n} p(r_k|l_i)p(l_i)}{\sum\limits_{j=1}^{m}\prod\limits_{k=1}^{n} p(r_k|l_j)p(l_j)} \tag{9}$$

One way to estimate the location is to determine the *maximum a posteriori estimate* $\hat{\ell}_{MAP}$, which simply gives the RP coordinate l_i that maximizes $p(l_i|\mathbf{r})$ in Equation (9):

$$\hat{\ell}_{MAP} = \underset{l_i}{\arg\max}\, p(l_i|\mathbf{r}) \tag{10}$$

This estimation is usually easy to determine [21] as well as it needs less computational effort. Besides, the mathematical principles are quite similar to that of used by the Nearest-Neighbors (NN) method [19], as seen in the work of [5].

3.2. Construction of the Simulation Environment

To illustrate how our simulation model works, a simple scenario is built according to Figure 2. The power signal P_t is transmitted by the AP and values of RSS are registered into the variable r described in Equation (2) concerning test points (1) and (2). The random component of the model σ is generated artificially using an adequate function from Octave [22]. Also, the test points are generated by the use of an inbuilt function that replicates uniform distributions. In the example, the test points coordinates in meters are $TP(1) = (2.74, 9.77)$ and $TP(2) = (6.22, 4.00)$. The vectors r generated are $r_1 = -100$ (dBm) and $r_2 = -94$ (dBm), which corresponds to the TPs (1) and (2), respectively. Next, these values are used to compute the posterior functions described in Equation 9 for each RP. The results for $TP(1)$ were $p(l_1|r_1 = -100) = 0.69$ and $p(l_2|r_1 = -100) = 0.31$. For $TP(2)$, it followed that $p(l_1|r_2 = -94) = 0.41$ and $p(l_2|r_2 = -94) = 0.59$. Finally, from Equation (10), the estimated locations for the tests (1) and (2) were $\hat{\ell}_{MAP}(1) = (2.5, 5)$ and $\hat{\ell}_{MAP}(2) = (7.5, 5)$ respectively. As expected, the estimations correspond to the RP locations given in Figure 2. Indeed, it can be visually verified that $TP(1)$ is closer to $RP(1)$ whereas $TP(2)$ is closer to $RP(2)$. Thus, for this specific example, the tests were classified correctly in terms of the RPs neighborhood, although the localization error can be still considered large.
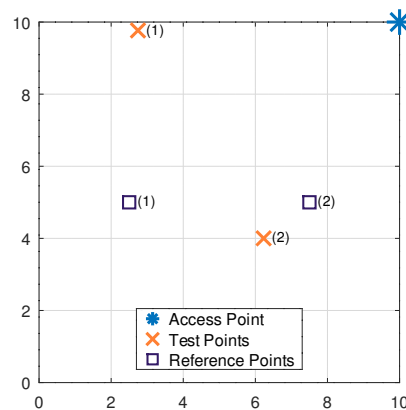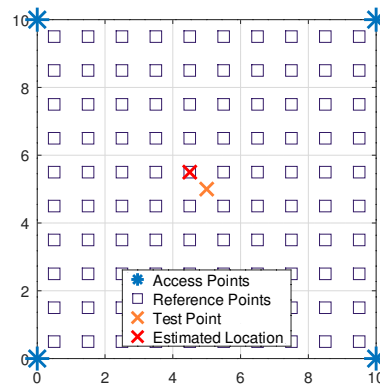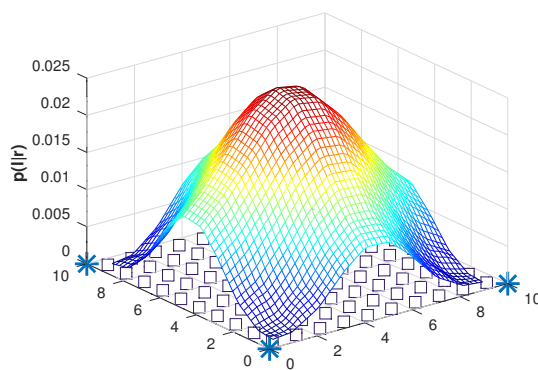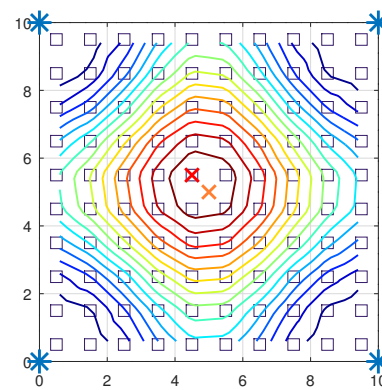

Figure 2. Example of scenario: 1 AP, 2RPs and 2 Test Points. Also, the environment parameters are set to $P_t = -12$ dBm, $PL(d_0) = 68.5$ dB, $\sigma = 4.4$ dB and $\alpha = 1.8$.

(a) Environment plan.

(b) Surface plot.

(c) Contour plot.

Figure 3. Simulated scenario with 4 APs, 100 RPs and 1 Test Point.

Another way to observe how locations are estimated is to see the values of $p(l|\mathbf{r})$ distributed over the environment. A new proposed scenario is described in Figure 3a , in which only one test point is analyzed. Figures 3b and 3c depict the probability that a RSS vector $\mathbf{r}$ is associated with each RP by means of surface and contour plots respectively. As one can verify, the test point located at coordinates $(5,5)$ is estimated as the RP with coordinates $(4.5, 5.5)$, whose probability $p(l|\mathbf{r})$ associated is the maximum found among the other RPs.

4. Analysis of the Impact of Design Factors on the System Accuracy

In this section, we analyze the accuracy using our proposed model by varying the main design factors of the positioning system. The metric chosen for evaluating the performance is the 95% interval of cumulative error distribution, calculated by taking the distance between the ground-truth position of each test point and the corresponding estimated location.

The use of the error at 95% is particularly interesting because of its representative character along the entire area of study. The average error, in contrast, can instill a false notion that the error calculated is sufficiently small. An example of a concrete case occurs when most of the error is concentrated in a central part of the area, and the error at the borders is considerably large. This way, the average error is expected to represent the central area but not the entire room.

The extraction of environment parameters is the first step to build a consistent model. For WLAN-based systems, the technology often deployed is either WiFi or Bluetooth. Log-distance path loss models eventually fit better with Bluetooth technology [23,24], although it can be also applied to WiFi or any other wireless technologies. In this work, we chose the parameters based on some experiments of our research group, whose scenario is composed of Bluetooth Low Energy (BLE)

devices, channel diversity, and high RSS variability. There are, of course, plenty of scenarios presented in many works where one can extract useful data from the environment. The work of Nikoukar et al. [25], for instance, compares different scenarios for Bluetooth-based systems and provides typical parameters that can be both useful and insightful for model-based studies. In our case, we obtain the environment parameters using the maximum likelihood estimates (MLEs) [17].

The environment parameters used in the next simulations are listed in Table 1.

Table 1. Environment parameters.

Parameter	Value
α	1.8
σ	4.4 dB
$PL(d_0)$	60 dB
Area	100 m^2

4.1. Impact of Location Tests

In a dynamic indoor scenario, one question that arises naturally is the number of location tests (or test points) over whom the localization error should be calculated. It happens that the number of location tests might not be sufficient to express the true localization error. A few tests distributed over an indoor area may not represent the entire environment, even if they have a uniform distribution. In other words, if one performs different rounds of tests using only a few test points, the results obtained can be surprisingly different among each other. To overcome this issue, a reasonable number of location tests must be set before any further detailed analysis with the scenario is done.

Figures 4a and 4b describe two scenarios with 100 and 1,000 test points, respectively. It can be observed the more scattered pattern with fewer tests, which may not represent the entire area on average.
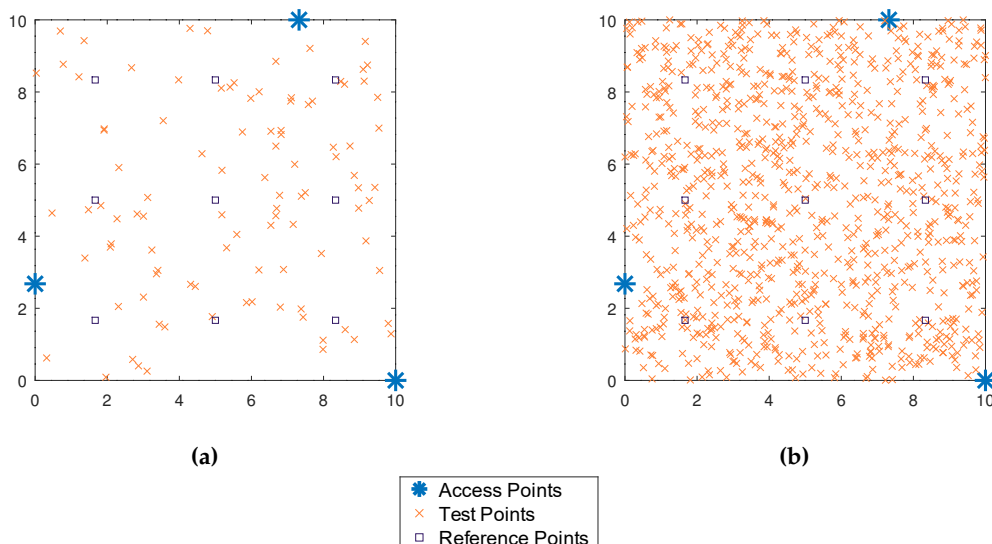

Figure 4. Simulated scenarios with (**a**) 100 and (**b**) 1,000 test points.

To demonstrate the differences in accuracy, simulations ranging from 100 to 2,000 test points are performed. For each set of them, one hundred simulation rounds are done to extract the mean error and its mean standard deviation. In this example, 3 APs, 9 RPs, and 5 samples per test are considered. By simply inspecting the graphics in Figures 5a and 5b, a decreasing rate tendency of the error standard deviation with the number of test points is verified. This suggests an "ideal" value for the number of tests that make error predictions more reliable. Thus, considering a threshold of 0.2 m for the error

standard deviation and the computational effort of performing too many tests, the number of 1,000 test points is established as a reasonable choice for the next simulations proposed.
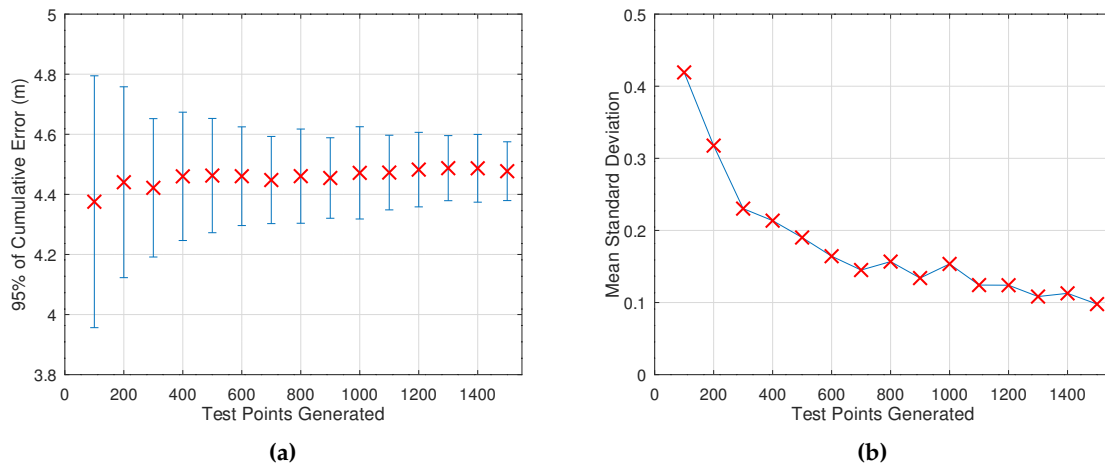

Figure 5. Variation of 95% cumulative error with number of tests in terms of (**a**) error bar and (**b**) mean standard deviation.

4.2. Impact of Reference Points

Intuitively, the more reference points per area, the more accurate a probabilistic-based localization system is. Indeed, a greater density of RPs tends to make the test points estimates closer to their true locations. Figure 6a illustrates what occurs in this case.
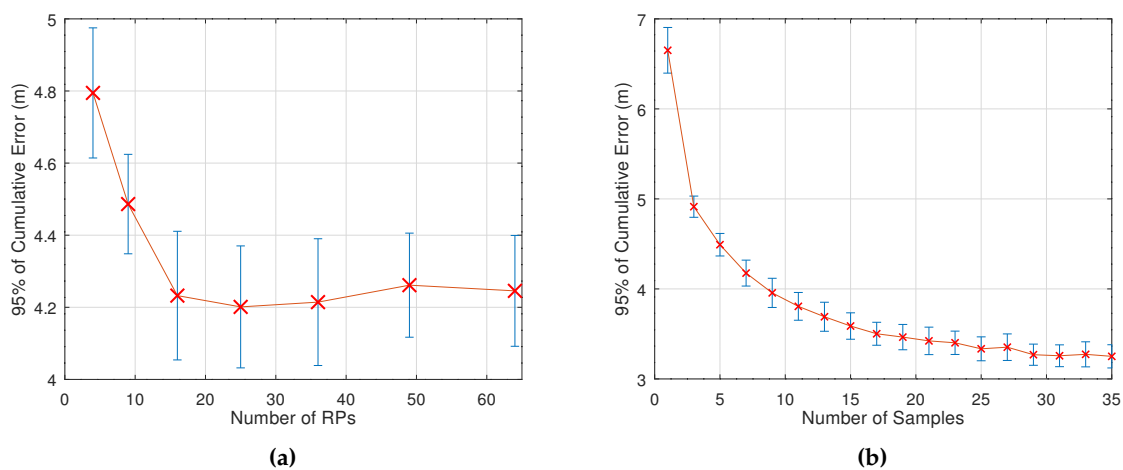

Figure 6. Variation of 95% cumulative error with (**a**) number of RPs and (**b**) number of samples. The "X" symbols indicate the average errors and the bars the mean standard deviation.

As it can be observed, the error does not vary much from 16 RPs onwards. This is somewhat expected due to the intrinsic variability of the log-distance path loss model. In other words, as the distance between closer RPs decreases to a certain point, the probabilistic-based positioning algorithm begins to have difficulty in estimating more precisely the RP closer to the true location of the test.

In terms of computational efficiency, the simulation is faster as the number of RPs decreases, once fewer comparisons are executed to cover all the tests requested. By the way, a bridge could be built between the RPs presented here and the training points of fingerprinting-based systems using the k-Nearest Neighbors (k-NN) localization method. As the essence of these methods is similar,

decreasing the number of RPs is equivalent to decreasing the number of training points, which reduces the training time in the offline phase, as well as the computational effort of performing real-time localization.

4.3. Impact of Samples Collected

Because of the high RSS variability, a technique usually employed for improving localization systems accuracy in practice consists of collecting more than one RSS sample during a determined time window [26,27]. This way, by taking the mean of these samples, for instance, the uncertainty in estimations is decreased, as mentioned before. Figure 6b shows this situation, in which the error is decreased as the number of samples increases. It can be visually verified that the improvement in localization error is minimum from twenty samples per test on. In this case, the error decreasing rate also diminishes with the number of samples. These results indicate that one must choose a value for the "samples" variable that is optimal from the point of view of both accuracy and real-time positioning applications.

4.4. Impact of Access Points Placement

The APs placement is another factor that plays a relevant role when accuracy is a measure of performance. In what follows, three different AP configurations, represented by Figures 7a, 7b and 7c, are proposed and their accuracy analyzed.
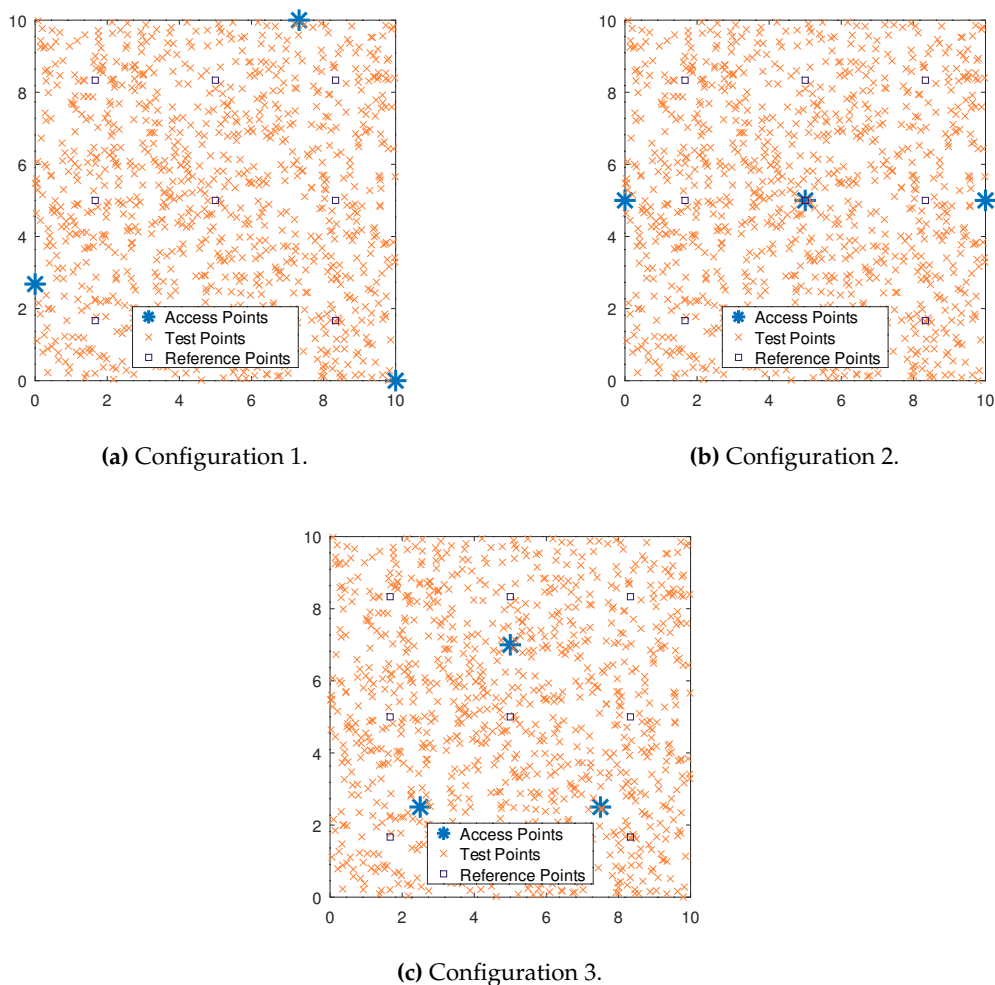

(a) Configuration 1.

(b) Configuration 2.

(c) Configuration 3.

Figure 7. Simulated scenarios with 1,000 test points.

The simulation results are depicted in Figure 8. It can be verified the difference in terms of localization error caused by the different AP configurations. The way the APs are arranged in the environment can generate more or fewer ambiguities when estimating the location of a test point. In this case, specifically, the triangular format of Configuration 1 is a much more accurate choice than the placement proposed by Configuration 2. On the other hand, Configuration 3 has an intermediate accuracy compared to the others. The results show the impact a simple arrangement of APs has on the accuracy of IPS. In principle, there is not an evident pattern that can produce smaller errors, which reinforces the need for simulating each given scenario extensively.
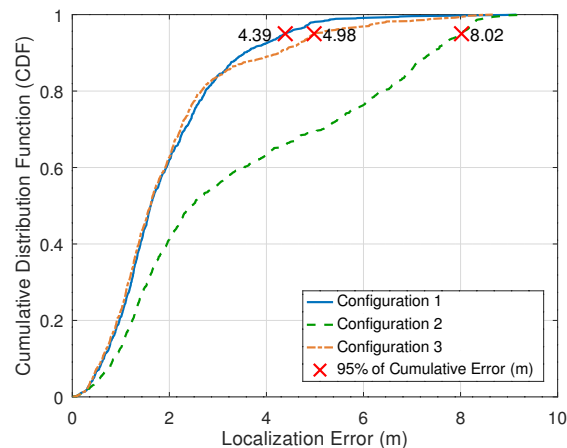

Figure 8. Variation of localization error under the perspective of three different configurations.

4.5. Method for the Best Design Choice

With the knowledge of how the factors relate to the positioning error, a natural path is to find the best factor combination which achieves a determined accuracy given as input. In other words, a method that seeks to achieve the localization error requested with a minimum number of RPs and samples collected per test.

Our proposed method, in this case, consists of some basic steps and is represented by the flowchart depicted in Figure 9. Given the environment lognormal path loss model parameters and its area, the number of APs, and the positioning error required, the proposed simulation sequence can be developed as follows:

1. **Search for a reliable number of tests to get robust results**. Reliability means the positioning error has either little or no variance around its mean when the simulation is run at different times for the same number of tests. The tolerance – mean standard deviation – can be set according to a certain application requirement.
2. **Establishment of limits for the number of samples and RPs**. As we know there is little gain in accuracy with considerably large values for these factors, superior limits can be set to restrict the possible combinations.
3. **Sweep the combination possibilities, keeping number of RPs as small as possible**. As we shall verify later, this factor has the largest impact on the run time simulation. Thus, it should be the last variable to change (if necessary).
4. **Search for the best AP configuration**. With combinations well mounted, the arrangement of APs becomes the variable of interest. For each combination of RPs and samples per test, the positioning error is calculated for every possible AP placement. If the result that gives the smallest error is either smaller than or equal to the required, then the best AP configuration was achieved. In other words, no more simulations are needed, once the influencing parameters involved are already optimized.
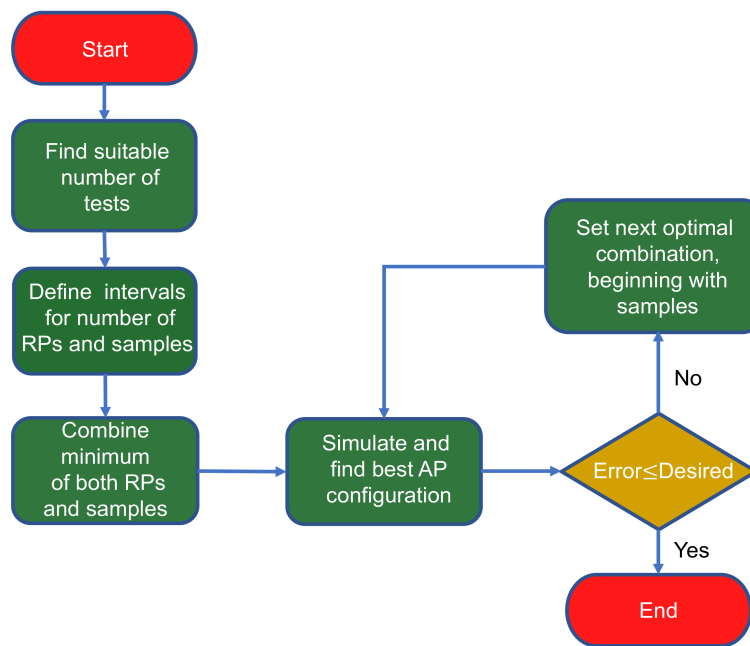
Figure 9. Steps to find the best configuration for a given environment.

5. Performance Evaluation

To evaluate the performance of our probabilistic-based approach, as well as our method for finding the best design choice, a case study is proposed. To simplify, the same environment parameters in Table 1 and a number of 3 APs are used here. Also, the desired localization error is set to 3.0 m, considering the 95% point of the cumulative distribution function for the error. The number of tests is set to 1,000 due to the environment features already analyzed before.

The number of RPs is restricted to $\#RP \in \{4, 9, 16, 25\}$, and the number of samples to $\#SMP \in \{1, 5, 10\}$. As can be seen, the total number of possibilities is 12. For each possible combination, a simulation is executed to find the AP placement which gives the minimum error. The optimization algorithm employed at this point is the brute force just to demonstrate the improvement in accuracy by the method described here. By taking the possibilities of AP placement over the environment as n_c uniformly distributed coordinates, it is possible to allocate and test n APs given as input. The total possible configurations to analyze is the choice of n out of n_c possibilities. In this case study, n_c is 16 and n is 3. Thus, the total number of combinations τ is given by:

$$\tau = \binom{16}{3} = 560 \tag{11}$$

According to Equation (11), the simulation environment is run 560 times and outputs the configuration with minimum localization error.

In Table 2, the results for each parameter combination are shown. As one can verify, the combination which better fits the requirement of 3.0 m for the localization error is the one with 16 RPs and 10 samples per test, illustrated in Figure 10a. A deeper analysis about this configuration is seen in Figures 10b and 10c, in which different AP placements can produce even great differences in errors among each other. The "x" axis represents a map of indexes to the possible coordinates for the given scenario, generated by the function *nchoosek* in Octave. At this point, it can be verified that the results from the 95% of cumulative errors were quite more dispersed than the results from the usual average. This shows that the first metric mentioned, besides promoting a better picture of the error distribution, also can provide a clearer contrast among the configurations tested.

Table 2. 95% of cumulative error results obtained for different configurations of 3 APs.
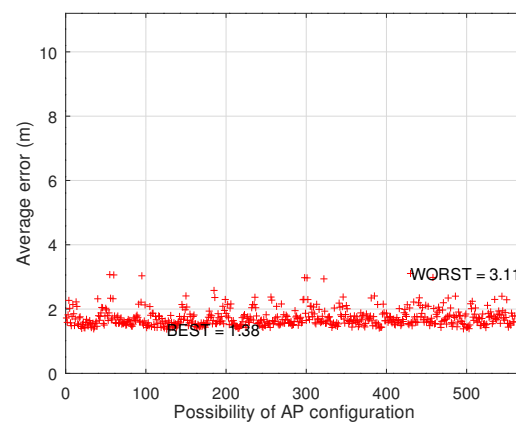
Sequence	#RP	#SMP	Localization Error (m)	
			Best	**Worst**
1	4	1	5.96	8.72
2	4	5	4.01	8.75
3	4	10	3.60	8.75
4	9	1	5.86	9.38
5	9	5	3.65	9.87
6	9	10	3.06	9.65
7	16	1	6.00	9.57
8	16	5	3.50	10.13
9	16	10	2.86	10.20
10	25	1	6.00	9.70
11	25	5	3.44	10.01
12	25	10	2.71	10.38
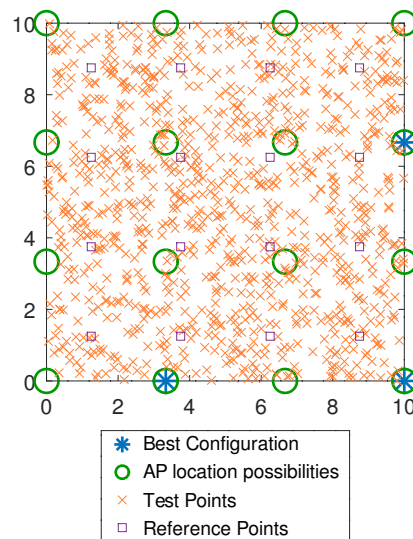
(a)
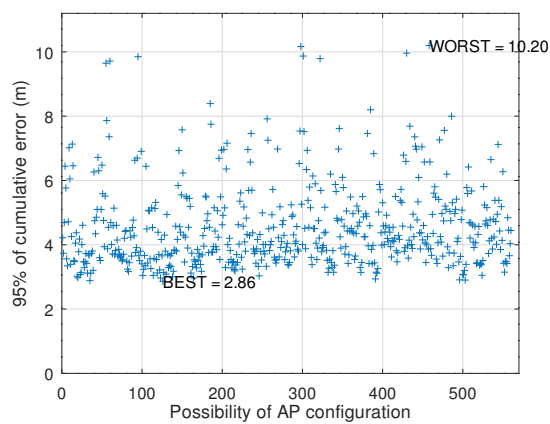
(b)

(c)

Figure 10. Simulation results in a scenario with 3 APs, 16 RPs and 10 samples per test: (**a**) Configuration with minimum localization error (95%); (**b**) Results of 560 possible AP placements (95%); (**c**) Results of 560 possible AP placements (average).

In addition, it was observed that the configurations whose simulations processed only 1 sample per test gave very different AP placement patterns for the best accuracy found. Similar results were obtained when 4 RPs were used in simulations. On the other hand, from 9 RPs and 5 samples per test onwards, all configurations which resulted in a minimum localization error have the same pattern of what is depicted in Figure 10a, being different from each other by a simple rotation of the room.

In order to check the variability in localization error, the configuration proposed in Figure 10a was run 100 times and the results were, on average, 2.94 m for the 95% of cumulative error with a mean standard deviation of 0.14 m. That is, a localization error of (2.94 ± 0.14) m, which constitutes a befitting value around the required of 3.0 m.

Although the method can achieve the main goal of finding the combination with minimum localization error, there is a clear need to analyze the simulation time performance. At first, one can verify the influence of each factor on the simulation run time in Figure 11. The scenario is the same as the one provided in Table 2. For this analysis, the simulation environment was developed using Octave-5.2.0 on a Sony Vaio laptop (Windows 10, 64-bit operating system, 2.70 GHz Intel i7-7500U Processor and 8 GB RAM). Also, the time metric considered was the CPU time.
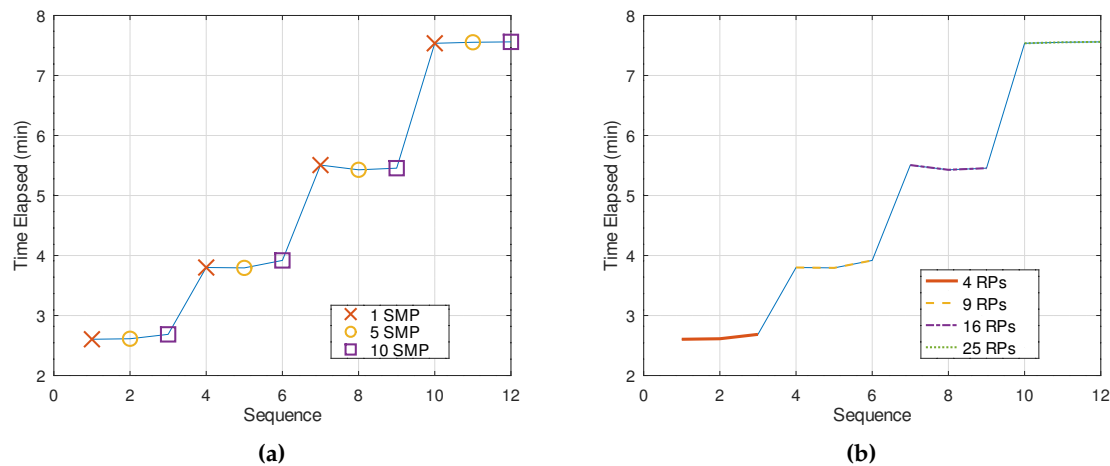

Figure 11. Time elapsed for each sequence: (**a**) Influence of samples per test; (**b**) Influence of RPs.

The influence of samples per test on the simulation time is negligible compared to the number of RPs. As expected, as this last factor increases, the simulation run time increases linearly, since the number of probability calculations is proportional to the number of RPs. On the other hand, the contribution from the number of samples is just a single calculation mean of RSS values for each test. This way, the time complexity for each sequence would be $\mathcal{O}(m)$ concerning m RPs and $\mathcal{O}(1)$ due to a given number of samples per test. This evidences the priority in changing the number of samples per test firstly, as it little influences the simulation run time.

Another way to verify the influence of the number of RPs is depicted in Figure 12. The linear behavior between time and number of RPs can be verified in a clearer way. In addition, one can think of the total time elapsed on the execution of the method proposed, that is, the cumulative time elapsed. In this case, it has a time complexity of $\mathcal{O}(m^2)$. Indeed, for a function with a linear trend, its cumulative sum reflects the sum of an arithmetic progression, which is quadratic.
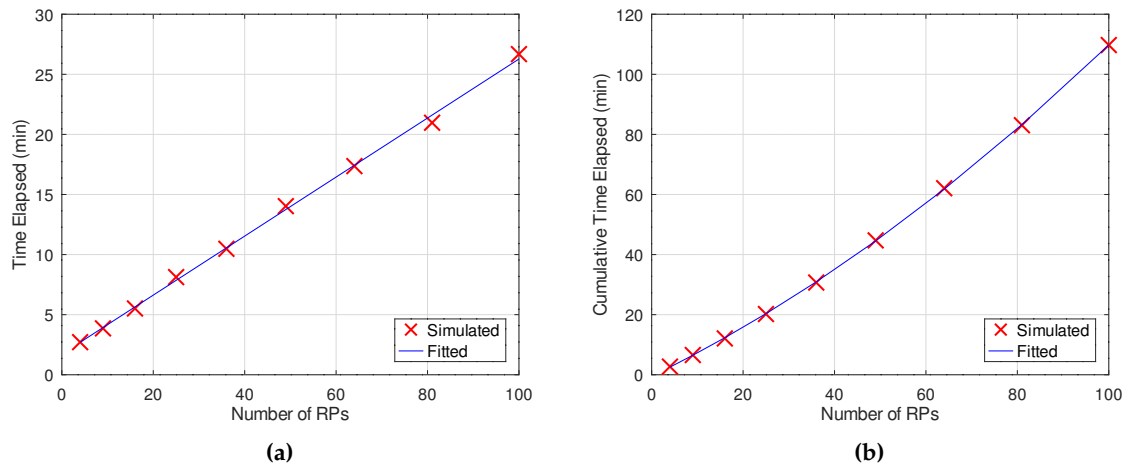
Figure 12. Influence of the number of RPs on time elapsed: (**a**) linear trend; (**b**) quadratic trend; A fixed number of 10 samples per test is considered.

It is important to highlight that all of these simulations consider a fixed number of AP location possibilities. However, this also affects the simulation run time, as the more possible locations available to be allocated, the more configurations one has to simulate. For the same scenario depicted in Figure 10a, simulations ranging from 4 to 64 AP allocation spaces were run, and some important results can be seen in Table 3. Although the localization error decreases as n_c increases, the total possible configurations that must be run increases with $\binom{n_c}{3}$. In this case, there has to be a trade-off between the goal of reducing localization error and the computational effort to be deployed.

Table 3. 95% of cumulative error results obtained for 3 APs, 16 RPs and 10 samples per test. Different values for the AP allocation spaces are simulated.

AP Allocation Spaces (n_c)*	$\tau = \binom{n_c}{3}$	Localization Error (m)	
		Best	Worst
4	4	3.06	3.23
9	84	3.04	9.93
16	560	2.92	10.14
25	2,300	2.79	10.00
36	7,140	2.75	10.16
49	18,424	2.73	10.20
64	41,664	2.70	10.29

*AP Allocation Spaces (n_c) represents the number of possible places (or coordinates) that the APs can assume over the environment. In this case, we consider them to be uniformly distributed from the area limits.

These results also show that when there are more options for the APs to be allocated in the environment, the chances are that the localization error can diminish substantially. On the other hand, the error is supposed to reach a threshold value around which there is no significant improvement. Figure 13 depicts better this situation. Visually, a reasonable choice would be that in which the error reaches the value of 2.75 m with 36 possibilities of allocation spaces. In this case, the simulation run time needed is around one hour. More allocation spaces possibilities would then overload the simulation unnecessarily.
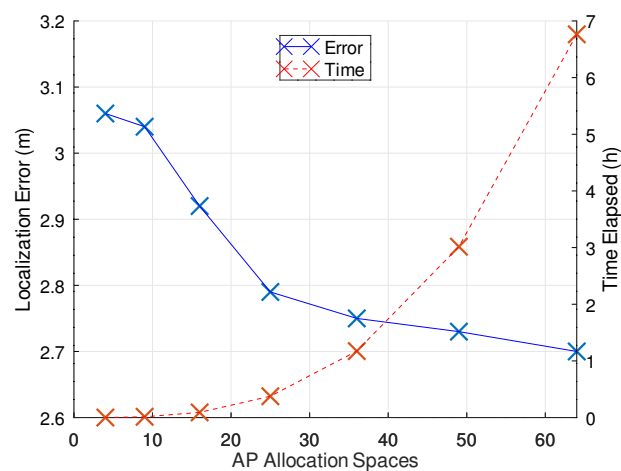
Figure 13. Influence of the number of AP allocation spaces on 95% of cumulative error (solid line) and CPU time elapsed (dashed line). Scenario considered: 3 APs, 16 RPs and 10 samples per test.

6. Conclusions

In this paper, we propose a method that is rooted in a probabilistic modeling approach to optimize the use of the most relevant factors that have an impact on the accuracy of indoor positioning systems. From the reasonable assumption that a log-distance path loss model can represent the RSS variation over an entire indoor area, we analyze the impact of the number of reference points, the number of samples collected per test, and the access points arrangement on the system accuracy. Knowing, thus, how factors relate to each other, a simple algorithm to go through the combinations of these factors until the best configuration is found is proposed. Moreover, a real concern about the number of tests necessary to represent the dynamics of an IPS and provide a reliable localization error is grounded and discussed. Not less important, the 95th percentile of the cumulative distribution function for the error is introduced as a more unique and representative metric than the common average error.

The results obtained show that the combination set of design factors can be fairly reduced so the required accuracy can be still achieved without any penalties. This way, when restrictions of computational cost and real-time applications are an issue, the method is able to provide a more efficient usage of the system's inputs.

For future work, we plan to develop an efficient algorithm to help optimize the method proposed in order to reduce its current time complexity and then compare the results with the common existing optimization approaches. In addition, we expect to validate the simulated results in a real testbed experiment, also adding the effects of obstructions to the model.

References

1. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2007**, *37*, 1067–1080.
2. Soares Lima, M.W.; Fernandes de Oliveira, H.A.B.; dos Santos, E.M.; de Moura, E.S.; Costa, R.K.; Levorato, M. Efficient and Robust WiFi Indoor Positioning Using Hierarchical Navigable Small World Graphs. 2018 IEEE 17th International Symposium on Network Computing and Applications (NCA), 2018, pp. 1–5.
3. Farid, Z.; Nordin, R.; Ismail, M. Recent Advances in Wireless Indoor Localization Techniques and System. *Journal Comp. Netw. and Communic.* **2013**, *2013*, 185138:1–185138:12.
4. Gu, Y.; Lo, A.; Niemegeers, I. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys Tutorials* **2009**, *11*, 13–32.
5. Bahl, P.; Padmanabhan, V.N. RADAR: an in-building RF-based user location and tracking system. Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual

Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064), 2000, Vol. 2, pp. 775–784 vol.2.

6. Youssef, M.; Agrawala, A. The Horus WLAN Location Determination System. Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services; Association for Computing Machinery: New York, NY, USA, 2005; MobiSys '05, p. 205–218. doi:10.1145/1067170.1067193.

7. He, S.; Chan, S..G. Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons. *IEEE Communications Surveys Tutorials* **2016**, *18*, 466–490.

8. Kaemarungsi, K.; Krishnamurthy, P. Modeling of indoor positioning systems based on location fingerprinting. IEEE INFOCOM 2004, 2004, Vol. 2, pp. 1012–1022 vol.2.

9. Shinsuke Hara.; Takahiro Fukumura. Determination of the placement of anchor nodes satisfying a required localization accuracy. 2008 IEEE International Symposium on Wireless Communication Systems, 2008, pp. 128–132.

10. Zhao, Y.; Zhou, H.; Li, M. Indoor Access Points Location Optimization Using Differential Evolution. 2008 International Conference on Computer Science and Software Engineering, 2008, Vol. 1, pp. 382–385.

11. Ying He.; Weixiao Meng.; Lin Ma.; Zhian Deng. Rapid deployment of APs in WLAN indoor positioning system. 2011 6th International ICST Conference on Communications and Networking in China (CHINACOM), 2011, pp. 268–273.

12. Farkas, K.; Huszák, A.; Gódor, G. Optimization of Wi-Fi Access Point Placement for Indoor Localization. *Journal IIT (Informatics & IT Today)* **2013**, *1*, 28–33.

13. Aomumpai, S.; Kondee, K.; Prommak, C.; Kaemarungsi, K. Optimal placement of reference nodes for wireless indoor positioning systems. 2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2014, pp. 1–6.

14. Rajagopal, N.; Chayapathy, S.; Sinopoli, B.; Rowe, A. Beacon placement for range-based indoor localization. 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2016, pp. 1–8.

15. Sharma, R.; Badarla, V. Geometrical Optimization of A Novel Beacon Placement Strategy for 3D Indoor Localization. 2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2018, pp. 1–6.

16. Jia, M.; Khattak, S.B.A.; Guo, Q.; Gu, X.; Lin, Y. Access Point Optimization for Reliable Indoor Localization Systems. *IEEE Transactions on Reliability* **2019**, pp. 1–13.

17. Roos, T.; Myllymäki, P.; Tirri, H. A Statistical Modeling Approach to Location Estimation. *IEEE Trans. Mob. Comput.* **2002**, *1*, 59–69.

18. Rappaport, T. *Wireless Communications: Principles and Practice*, 2nd ed.; Vol. 2, Prentice-Hall: Upper Saddle River, N.J, 2002; p. 707.

19. Honkavirta, V.; Perälä, T.; Ali-Löytty, S.; Piché, R. A comparative survey of WLAN location fingerprinting methods. *2009 6th Workshop on Positioning, Navigation and Communication* **2009**, pp. 243–251.

20. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 3rd ed.; McGraw-Hill, 1991; p. 666.

21. de Coulon, F. *Signal Theory and Processing*, 1st. ed.; Artech House: Dedham, MA, 1986.

22. Eaton, J.W.; Bateman, D.; Hauberg, S.; Wehbring, R. *GNU Octave version 5.2.0 manual: a high-level interactive language for numerical computations*, 2020.

23. Zhao, X.; Xiao, Z.; Markham, A.; Trigoni, N.; Ren, Y. Does BTLE measure up against wifi? A comparison of indoor location performance. *20th European Wireless Conference, EW 2014* **2014**, pp. 263–268.

24. Wu, T.; Xia, H.; Liu, S.; Qiao, Y. Probability-Based Indoor Positioning Algorithm Using iBeacons. *Sensors* **2019**, *19*, 5226.

25. Nikoukar, A.; Abboud, M.; Samadi, B.; Güneş, M.; Dezfouli, B. Empirical analysis and modeling of Bluetooth low-energy (BLE) advertisement channels. 2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), 2018, pp. 1–6.

26. Faragher, R.; Harle, R. An Analysis of the Accuracy of Bluetooth Low Energy for Indoor Positioning Applications. Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014), 2014, pp. 201–210.

27. Faragher, R.; Harle, R. Location Fingerprinting With Bluetooth Low Energy Beacons. *IEEE Journal on Selected Areas in Communications* **2015**, *33*, 2418–2428.