FEDERAL UNIVERSITY OF AMAZONAS - UFAM

INSTITUTE OF COMPUTING - ICOMP

POST-GRADUATE PROGRAM IN INFORMATICS - PPGI

Site Survey-free Indoor Positioning System based on RSSI Diversity and Sequential Least Squares Programming

Bráulio Henrique Orion Uchôa Veloso Pinto

Manaus - AM

August 2025

Bráulio Henrique Orion Uchôa Veloso Pinto

Site Survey-free Indoor Positioning System based on RSSI Diversity and Sequential Least Squares Programming

Thesis submitted to the Post-graduate Program in Informatics of the Institute of Computing of the Federal University of Amazonas for the degree of Doctor of Science. Concentration area: Computer Science.

Advisor

Horácio A. B. Fernandes de Oliveira, D.Sc.

FEDERAL UNIVERSITY OF AMAZONAS - UFAM

INSTITUTE OF COMPUTING - ICOMP

Manaus - AM

August 2025

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

P659s Pinto, Bráulio Henrique Orion Uchôa Veloso
Site Survey-free Indoor Positioning System based on RSSI Diversity and Sequential Least Squares Programming / Bráulio Henrique Orion Uchôa Veloso Pinto. - 2025.
81 f. : il., p&b. ; 31 cm.

Orientador(a): Horácio Antonio Braga Fernandes de Oliveira.
Tese (doutorado) - Universidade Federal do Amazonas, Programa de Pós-Graduação em Informática, Manaus, 2025.

1. Bluetooth low-energy. 2. Indoor positioning. 3. Path loss model. 4. RSSI - Received Signal Strength Indicator. 5. Optimization. I. Oliveira, Horácio Antonio Braga Fernandes de. II. Universidade Federal do Amazonas. Programa de Pós-Graduação em Informática. III. Título

---

Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"SITE SURVEY-FREE INDOOR POSITIONING SYSTEM BASED ON RSSI DIVERSITY AND SEQUENTIAL LEAST SQUARES PROGRAMMING"

BRÁULIO HENRIQUE ORION UCHÔA VELOSO PINTO

Tese de Doutorado defendida e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Horácio Antônio Braga Fernandes de Oliveira - **Presidente**

gov.br Documento assinado digitalmente
YURI FREITAS ASSAYAG
Data: 29/08/2025 11:20:44-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Yuri Freitas Assayag - **Membro Externo**

gov.br Documento assinado digitalmente
FELIPE LEITE LOBO
Data: 28/08/2025 10:04:36-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Felipe Leite Lobo - **Membro Externo**

Profa. Dra. Eulanda Miranda dos Santos - **Membro Interno**

Prof. Dr. Eduardo Freire Nakamura - **Membro Interno**

Manaus, 20 de agosto de 2025.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Horácio Antônio Braga Fernandes de Oliveira**, **Professor do Magistério Superior**, em 22/08/2025, às 07:18, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

sei! assinatura eletrônica

Documento assinado eletronicamente por **Eduardo Freire Nakamura**, **Professor do Magistério Superior**, em 22/08/2025, às 15:28, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

Documento assinado eletronicamente por **Eulanda Miranda dos Santos**, **Professor do Magistério Superior**, em 22/08/2025, às 16:56, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015 .

A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2751837** e o código CRC **F3648638**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Referência: Processo nº 23105.034614/2025-89 SEI nº 2751837

Acknowledgements

First of all, I would like to thank God for the daily graces I have been given, and my family for all the support and patience delivered during this journey.

I want to thank the Federal University of Amazonas, represented by the Institute of Computing, for all the support and infrastructure provided for my research. A special thanks to my advisor prof. Dr. Horácio Fernandes for all his precious time spent with me in our meetings, his guidance, and his ability to extract the best of me.

I would also like to thank the support given by the Foundation for Research Support of the State of Amazonas (FAPEAM) and the Coordination for the Improvement of Higher Education Personnel (CAPES), that allowed this work to be carried out.

And last but not least, I want to thank all my research colleagues for their partnership and valuable tips, and to all my friends who somewhat helped me achieve my goals.

Have the courage to follow your heart and intuition. They somehow already know what you truly want to become.

Steve Jobs

Site Survey-free Indoor Positioning System based on RSSI Diversity and Sequential Least Squares Programming

Author: Bráulio Henrique Orion Uchôa Veloso Pinto
Advisor: Horácio A. B. Fernandes de Oliveira, D.Sc.

Resumo

Sistemas de localização em ambientes fechados (IPSs) são essenciais para viabilizar serviços baseados em localização, como varejo, saúde e logística. No entanto, os IPSs existentes frequentemente exigem etapas de *fingerprinting* offline que são trabalhosas e demoradas, limitando a escalabilidade e a adaptabilidade em cenários internos dinâmicos. Esta tese apresenta o OPTIMAPS, uma solução inovadora independente de coleta de *fingerprints* que utiliza a tecnologia *Bluetooth Low Energy* (BLE). O OPTIMAPS emprega um modelo de perda de percurso logarítmica, cujos parâmetros são automaticamente determinados a partir da geometria do cenário e da análise de diversidade de sinal, eliminando assim a necessidade de coleta prévia de RSSI em campo. Durante a operação *online*, o sistema combina uma estimativa baseada no vizinho mais próximo

(NN) — escolhida por sua eficiência computacional e robustez como estimador inicial — com Programação Sequencial de Mínimos Quadrados (SLSQP) para otimização com restrições não-lineares, refinando as estimativas de posição dentro de um círculo de restrição empiricamente ajustado à granularidade do cenário experimental. Uma das principais inovações do OPTIMAPS é a adoção da métrica de Chebyshev para quantificar a dissimilaridade entre vetores de RSSI, que, juntamente com a análise da média das distâncias par-a-par, fornece melhor discriminação de sinais e aumenta a acurácia da localização. O sistema foi rigorosamente validado em um cenário real de larga escala — um cenário de 720 m² com 15 pontos de acesso BLE e 148 locais de teste — utilizando um conjunto de dados publicamente disponível. O OPTIMAPS alcançou um erro médio de localização (APE) de 2.65 metros, competitivo com o estado-da-arte, e superou métodos similares sem fingerprinting considerando-se a métrica de erro normalizado por densidade de pontos de acesso. Ademais, o OPTIMAPS demonstrou escalabilidade computacional linear em relação à área do cenário de testes, uma vantagem crítica sobre abordagens meta-heurísticas, como algoritmos genéticos, que apresentam crescimento quadrático ou maior em relação à demanda por recursos computacionais. A análise de desempenho espacial revelou que o refinamento via SLSQP proporciona os maiores ganhos em ambientes complexos e ricos em multipercurso, com reduções de erro superiores a 30% nas salas mais desafiadoras e impacto irrisório em espaços abertos. A quantificação do "efeito de tamanho" e testes estatísticos adicionais reforçaram a robustez e a aplicabilidade prática do OPTIMAPS. De modo geral, o OPTIMAPS se mostra uma solução escalável, precisa e eficiente em consumo de energia, que elimina a necessidade de coleta de *fingerprints* e se posiciona como uma proposta altamente

promissora para implantação em cenários de ambientes fechados reais, diversos e dinâmicos.
Palavras chave: Bluetooth low-energy, Localização em ambientes fechados, Modelo de perda de percurso, RSSI, Otimização.

Site Survey-free Indoor Positioning System based on RSSI Diversity and Sequential Least Squares Programming

Author: Bráulio Henrique Orion Uchôa Veloso Pinto
Advisor: Horácio A. B. Fernandes de Oliveira, D.Sc.

Abstract

Indoor positioning systems (IPSs) are essential for enabling location-based services in complex environments such as retail, healthcare, and logistics. However, existing IPSs often require labor-intensive and time-consuming offline fingerprinting phases that limit scalability and adaptability in dynamic indoor settings. This thesis presents OPTIMAPS (Optimized Positioning Technique Integrating Model-based and Pairwise Selection), a novel site survey-free solution leveraging Bluetooth Low Energy (BLE) technology. OPTIMAPS utilizes a log-distance path loss model, with parameters automatically identified from scenario geometry and signal diversity analysis, thus eliminating the need for pre-deployment RSSI collection. During online operation, the system combines a nearest-neighbor (NN) estimation — chosen for its computational efficiency and robustness as

an initializer — with Sequential Least Squares Programming (SLSQP) for constrained nonlinear optimization, refining position estimates within a restriction circle empirically matched to testbed granularity. A major innovation in OPTIMAPS is the adoption of the Chebyshev metric for quantifying RSSI vector dissimilarity, which, together with mean pairwise distance analysis, leads to superior signal discrimination and enhances positioning accuracy. The system was rigorously validated in a large-scale real-world scenario — a 720 m² testbed with 15 BLE access points and 148 testing locations — using a publicly available dataset. OPTIMAPS achieved an average positioning error (APE) of 2.65 meters, competitive with state-of-the-art techniques, and outperformed comparable survey-free methods on anchor density-normalized error. Furthermore, OPTIMAPS demonstrated linear computational scaling with environmental size, a critical advantage over metaheuristic optimization approaches such as genetic algorithms, which exhibit quadratic or worse growth in resource demand. Spatially resolved performance analysis revealed that the SLSQP refinement yields the greatest improvements in complex, multipath-rich environments, with error reductions surpassing 30% in the most challenging rooms, and negligible overhead in open spaces. Effect size quantification and statistical testing further established OPTIMAPS' robustness and practical deployability. Overall, OPTIMAPS proves to be a scalable, accurate, and energy-efficient solution that eliminates the need for site surveys, positioning it as a highly promising option for real-world deployment across diverse and dynamic indoor scenarios.
Keywords: Bluetooth low-energy, Indoor positioning, Path loss model, RSSI, Optimization.

List of Figures

Figure 1 – Comparative illustration of experimental design paradigms in indoor positioning research. 31
Figure 2 – Statistical decision regions for the paired t-test under a two-tailed alternative hypothesis. 33
Figure 3 – Statistical decision framework for method comparison in indoor positioning research. 37
Figure 4 – Positioning system pipeline divided into an offline phase and an online phase. 40
Figure 5 – Experimental testbed floor plan: 11 rooms, 3 halls, 15 access points, and 148 testing locations (adapted from Assayag et al. (2024a)). 49
Figure 6 – Hardware utilized for collecting RSSI data: receiver devices (a) and transmitter devices (b) (Assayag et al., 2024a). 49
Figure 7 – Positioning Error Comparison: Nearest-Neighbor vs SLSQP Optimization by log-distance path loss parameters (ρ_0: 40-80 dB, steps of 5 dB); α_L: 1-7. 53
Figure 8 – Mean of the pairwise distances by log-distance path loss parameters (ρ_0: 40–80 dB, steps of 5 dB; α_L: 1–7). 55

Figure 9 – Average positioning error in perspective: Nearest-Neighbor(NN), SLSQP(+), Mean with Euclidean and SLSQP(+) (Mean-Euclidean), and Mean with Chebyshev and SLSQP(+) (Mean-Chebyshev). 57
Figure 10 – Cumulative error distribution by algorithm and calculated parameters for selected ρ_0 values. 59
Figure 11 – Statistical performance comparison between Mean-Chebyshev (with SLSQP(+)) and pure SLSQP(+) for each ρ_0. 62

List of Tables

Table 1 – Performance comparison among BLE-based indoor positioning systems. 11
Table 2 – Performance comparison among optimization algorithms with parameters $(\rho_0, \alpha_L) = (55, 4.25)$. 51
Table 3 – Performance comparison among estimation techniques with parameters $(\rho_0, \alpha_L) = (55, 4.25)$. 52
Table 4 – Average positioning error (APE) by technique: for each given ρ_0, there is a most effective α_L associated. 56
Table 5 – Performance comparison between NN (the reference model) and OPTIMAPS. 58
Table 6 – Wilcoxon-signed rank testing and effect size results for median and mean comparisons between Mean-Chebyshev and SLSQP(+) methods ($\alpha = 0.05$). 61
Table 7 – Algorithm performance summary across 148 unique labels using parameters $\rho_0 = 55.0$ dB, $\alpha_L = 4.25$. 63
Table 8 – Spatial area classification system and label distribution. Areas are categorized as enclosed rooms or open halls based on architectural characteristics. 64

Table 9 – Room-level algorithm performance comparison ordered by NN baseline error (ascending). Negative improvement percentages indicate areas where NN outperforms SLSQP(+). 65
Table 10 – Correlation analysis between NN baseline error and SLSQP(+) improvement percentage. 66

List of abbreviations and acronymns

AN Anchor Node

AP Access Point

APE Average Positioning Error

BFGS Broyden–Fletcher–Goldfarb–Shanno

BLE Bluetooth Low-Energy

CNN Convolutional Neural Network

DE Differential Evolution

GNSS Global Navigation Satellite Systems

GPS Global Positioning System

IPS Indoor Positioning System

KNN K-Nearest Neighbor

LS Least Squares

NB Naïve Bayes

NLOS Non-Line-Of-Sight

NN Nearest Neighbor

PSO Particle Swarm Optmization

RF Radio Frequency

RP Reference Point

RSSI Received Signal Strength Indicator

SA Simulated Annealing

SLSQP Sequential Least Squares Programming

TRF Trust-Region Reflective

WKNN Weighted K-Nearest Neighbor

List of symbols

α significance level

α_L path loss exponent

ρ_0 path loss at reference distance

d_0 reference distance at 1 meter

P_t Beacon transmitted power

Contents

1 INTRODUCTION 1

1.1 Context 1

1.2 Problem 3

1.3 Objectives 4

1.4 Structure of the Thesis 5

2 RELATED WORK 6

3 THEORETICAL BACKGROUND 13

3.1 Indoor Positioning Technologies 13

3.1.1 Wi-Fi-based Positioning 14

3.1.2 Bluetooth Low Energy (BLE)-based Positioning 15

3.2 Radio Propagation Modeling 16

3.2.1 The Log-distance Path Loss Model 16

3.3 RSSI Diversity and Characterization 18

3.3.1 Pairwise Distance Metrics in Signal Space 20

3.3.1.1 Euclidean Distance 21

3.3.1.2 Chebyshev Distance 21

3.4 Positioning Estimation Methods 22

3.4.1 Nearest-Neighbor (NN) 22

3.4.2 K-Nearest Neighbors (KNN) 23

3.4.3 Least Squares (LS) 24

3.5 Optimization Techniques 25

3.5.1 Broyden–Fletcher–Goldfarb–Shanno (BFGS) 25
3.5.2 Trust-Region Reflective (TRF) Algorithm 26
3.5.3 Particle Swarm Optimization (PSO) 27
3.5.4 Simulated Annealing (SA) 27
3.5.5 Differential Evolution (DE) 28
3.5.6 Sequential Least Squares Programming (SLSQP) 29
3.6 Statistical Analysis of Location Errors in IPSs 30
3.6.1 Experimental Design Considerations 31
3.6.2 Statistical Inference for Paired Comparisons 32
3.6.3 Effect Size Analysis 35
3.6.3.1 Cohen's d for Paired Comparisons 35
3.6.3.2 Contextual Effect Analysis 35
3.6.4 Statistical Decision Framework 36
3.6.5 Synthesis and Recommendations 37

4 SYSTEM OVERVIEW 39
4.1 Proposed Method 39
4.1.1 Offline phase 40
4.1.2 Online phase 43
4.2 Experimental framework 45
4.2.1 Scenario 46
4.2.2 Testbed and Data Acquisition 46
4.2.3 Data Collection Protocol 46
4.2.4 Dataset Content and Structure 47
4.2.5 Experimental Design and System Architecture 47
4.2.6 Research Value and Utility 48

4.2.7 Experiment Visualization and Summary 48

5 RESULTS AND DISCUSSIONS 50
5.1 Choice of optimization algorithm 50
5.2 Optimization with the SLSQP algorithm 52
5.3 Choice of model parameters and performance analysis 54
5.4 Error analysis by room type 62
5.4.1 Data preprocessing 63
5.4.2 Algorithm performance comparison 63
5.4.3 Spatial mapping and room-based aggregation 64
5.4.4 Room-level performance analysis 65
5.4.5 Correlation analysis: baseline performance versus optimization benefit 66

6 CONCLUSIONS 68
6.1 Limitations And Future Work 71
6.2 Published Papers 73

Bibliography 76

1 Introduction

1.1 Context

Indoor positioning systems (IPSs) remain a prominent research area driven by the demand for location-based services as retail, healthcare, and logistics. In contrast to outdoor systems that use satellite signals like the Global Positioning System (GPS), indoor environments pose challenges due to signal obstructions and multipath propagation. As a result, researchers have investigated technologies to improve indoor positioning accuracy, with Bluetooth Low Energy (BLE) standing out for its low power consumption and widespread use in mobile devices (Abed et al., 2022; Xiao et al., 2024; Zhuang et al., 2022).

Traditional IPS methodologies often involve offline training phases, such as Wi-Fi and BLE fingerprinting, where extensive data collection is required to create a radio map. However, these approaches can be time-consuming and impractical in dynamic environments where conditions frequently change (Jang and Kim, 2019). Hybrid positioning systems that integrate both Wi-Fi and BLE technologies have been proposed to leverage the strengths of each method. By combining the high accuracy of Wi-Fi with the low power consumption of BLE, these systems can provide robust indoor positioning solutions. For example, a hybrid approach may use Wi-Fi for precise localization in areas with

dense AP coverage while employing BLE for broader coverage in less dense areas (Amr et al., 2021). This flexibility allows for improved performance across various indoor environments without the need for extensive pre-training. Recent innovations have emphasized fully online systems based on path loss models, removing the necessity for prior training (Ali et al., 2019; Assayag et al., 2023).

Furthermore, advanced algorithms like genetic algorithms and optimization techniques are being explored to enhance indoor positioning systems (Wang et al., 2022; Xie et al., 2023; Zhao et al., 2024). These methods optimize positioning parameters, enhancing accuracy and efficiency while reducing extensive offline training needs. Integrating these algorithms with machine learning models has demonstrated promising results in boosting indoor positioning system performance.

Based on the aforementioned topics, this thesis presents OPTIMAPS (Optimized Positioning Technique Integrating Model-based and Pairwise Selection), a novel indoor positioning solution that utilizes a path loss model and Sequential Least Squares Programming (SLSQP) to enhance real-time accuracy with BLE technology. The newly site survey-free approach overcomes the limitations of conventional methods, offering a more adaptable indoor positioning solution. Also, the deployed system is validated in a large-scale real-world scenario, demonstrating its practical feasibility. Specifically, a log-distance path loss model is employed to estimate distances from Received Signal Strength Indicator (RSSI) values from BLE access points (APs), with parameters adjusted for signal diversity. Additionally, SLSQP continuously refines position estimates, adapting to real-time changes in the indoor environment without extensive pre-collected data. This approach enhances positioning accuracy and demonstrates efficient

online processing in complex indoor settings.

1.2 Problem

Although fingerprint-based systems are usually accurate, the offline or training phase is very labor-intensive and time-consuming, since it demands a significant amount of time to gather reliable and enough RSSI samples for every selected point in large, indoor scenarios (Liu et al., 2020).

To partially overcome the problem of intensive site-survey effort, some solutions rely on the log-distance path loss models, which describe, on average, the signal propagation throughout an indoor environment (Rappaport, 2002). Geometrical approaches, such as trilateration and multilateration techniques, are very efficient in this regard but have lower accuracy (Subedi and Pyun, 2020; Zafari et al., 2019). Probability-based approaches, in turn, are usually more accurate than geometrical ones, but they are not as precise as fingerprint methods for larger and more complex environments (Man et al., 2020).

As one can see, the efficiency in the site-survey effort is essential for the feasibility of IPSs, notably regarding time spent on training and energy consumption. Naturally, many studies are concerned with continuously improving classic solutions from an innovative perspective while keeping the positioning error at competitive levels.

1.3 Objectives

The primary objective is to develop a novel positioning system that eliminates the need for a time-consuming training phase while maintaining competitive accuracy and online processing time compared to existing solutions in similar scenarios. To achieve this, several specific objectives are outlined:

1. Formulate a novel approach for determining log-distance path loss model parameters exclusively through analysis of scenario geometry and RSSI dissimilarity, eliminating any reliance on on-site signal fingerprinting;

2. Design and implement an adaptive, two-stage positioning pipeline that leverages both pairwise signal diversity metrics and constrained nonlinear optimization (SLSQP), enabling real-time positional refinement in dynamic, multipath-rich indoor environments;

3. Perform a detailed error and sensitivity analysis across spatial sub-regions, environmental conditions, and algorithmic parameters, identifying the scenarios where the survey-free approach excels or presents limitations, and proposing guidelines for practical deployment.

4. Demonstrate the scalability of the proposed framework by validating it in a large-scale, labeled testbed with diverse device placements, and establishing open-source resources for future research reproducibility.

1.4 Structure of the Thesis

The remainder of this paper is organized as follows. Chapter 2 presents the related work. Chapter 3 expresses the background theory needed to understand the fundamental parts of our system. Chapter 4 outlines the proposed method, including the characterization of the experimental testbed. Chapter 5 presents the results and discussions. Chapter 6 draws the conclusion and presents the expectations for the next steps of the research.

2 Related Work

Recent advancements in BLE-based IPSs are attributed to lower power consumption and global infrastructure proliferation. Solutions range from fingerprinting methods to less time-consuming training techniques, depending on the application and permissible positioning error. This section reviews the literature on these methodologies, outlining the infrastructure, techniques employed, and achieved accuracy.

Subhan et al. (2019) present an experimental study on RSSI for distance and position estimation in indoor environments. Conducted in a 100 m^2 indoor area with four fixed BLE modules, the study collected extensive RSSI data at 100 grid points. By applying the logarithmic path loss model and adjusting radio propagation constants, the authors aimed to reduce distance estimation errors. They compared position estimation techniques, including fingerprinting-based K-Nearest Neighbor (KNN), multilateration, and MinMax, finding KNN achieved the highest accuracy with an average error of 1.02 meters, followed by MinMax at 2.12 meters and multilateration at 3.18 meters. The study highlights the importance of optimizing RSSI measurements and environmental constants for accurate BLE-based indoor positioning.

Ho and Chan (2020) present a decentralized indoor positioning system using BLE, eliminating the need for pre-training. It employs RSSI-based distance

estimation but improves it with a continuous on-the-fly training process that updates signal propagation parameters during beacon communication. A key innovation is the dynamic adjustment of distance estimation models without predefined calibration. Experiments in real indoor environments show an average positioning error (APE) of 1.5 meters, with 80% of results under 2 meters and 90% under 3 meters.

Nikodem and Szeliński (2021) introduce a BLE-based indoor positioning method that leverages channel diversity by utilizing extended advertisements to improve positioning accuracy. Their system integrates channel diversity and selection with the log-distance path loss model and Weighted Multilateration algorithm for optimal performance. The authors demonstrate that employing 40 channels substantially improves positioning accuracy compared to only using 3 primary channels. In a 100 m^2 test area, the IPS achieves an average error of 1.77 meters, though it necessitates a calibration phase.

Daniş et al. (2023) propose a probabilistic framework for indoor tracking using BLE beacons, employing Hidden Markov Models and the forward algorithm for position estimation based on RSSI data from stationary sensors. The study involves constructing probabilistic radio-frequency maps through fingerprinting and histogram interpolation, and modeling state transitions with Gaussian blur masks. An offline phase collects RSSI measurements to calibrate the maximal filter, which preprocesses data to mitigate multipath effects. Experiments in a 364 m^2 office environment with twelve BLE sensors showed that the maximal filter significantly improved accuracy, achieving mean position estimation errors as low as 2.14 meters. The findings demonstrate the method's effectiveness in real-world scenarios, making it a valuable contribution to the

field of indoor positioning.

Szyc et al. (2023) present three RSSI-based positioning algorithms (Ring, Sectional, and Mass) designed for large industrial areas with limited infrastructure. These algorithms estimate device positions without relying on fingerprinting or propagation models. The experimental testbed is a 1,600 m^2 cowshed divided into sections, with 10 BLE anchors and approximately 100 cows each wearing BLE devices. The Ring B algorithm achieves the best positioning accuracy with an average error of 7.89 meters, while the Sectional algorithm excels in section localization with an average Chebyshev error of 0.93 sections. The system operates in real-time using aggregated RSSI measurements, without significant offline training, and demonstrates improved positioning accuracy compared to traditional methods.

Beigi and Shah-Mansouri (2024) propose a comprehensive indoor positioning algorithm that integrates Wi-Fi and BLE devices to achieve high accuracy. The system utilizes RSSI-based localization, enhanced by a Kalman filter for denoising, and categorizes the filtered RSSI values into distinct classes using the KNN algorithm. These processed data are then fed into a recurrent neural network to estimate positions with high precision. The experimental setup involves a 28 m^2 room with two BLE beacons and one Wi-Fi access point, collecting RSSI data using ESP32 devices. The proposed algorithm achieves an APE of 61.29 cm, demonstrating a 56% improvement over existing methods. The system effectively captures the complex interactions between RSSI and positions, providing a robust solution for accurate indoor positioning.

Pinto and Oliveira (2024) introduce SeALS (Selection Strategy of Access Points with Least Squares Estimation), an innovative RSSI-based indoor posi-

tioning system utilizing BLE access points. SeALS reduces training and enhances accuracy by selecting the strongest RSSI components and combining them with the Least Squares (LS) estimation method. Validated in a large-scale, real-world scenario with a 45 m × 16 m area, 15 access points, and 148 testing locations, SeALS demonstrated an APE of 2.88 meters. The system showed up to 13% improvement in accuracy over pure OLS and up to 30% over the KNN technique. The system's efficiency is particularly notable in environments with high AP density, making it a viable solution for scenarios requiring minimal training effort and high accuracy.

Wu et al. (2024) evaluate various BLE-based indoor positioning techniques, including KNN, Weighted K-Nearest Neighbor (WKNN), Naïve Bayes (NB), RSSI-based Neural Network, and a novel Convolutional Neural Network (CNN) method. The study involves a training phase where RSSI data from eight BLE beacons installed in a 12m x 6m office space are collected at 28 training points and 12 testing points. The CNN approach, which transforms RSS data into an image-like format for feature extraction, achieved the best performance with an average positioning accuracy of 1.22 meters. The experimental setup and results demonstrate the potential of deep learning, particularly CNNs, to enhance the accuracy of indoor positioning systems using BLE technology.

Assayag et al. (2024b) propose an innovative indoor positioning system that employs Particle Swarm Optimization (PSO) and signal propagation models to improve positioning accuracy of mobile devices using BLE. Unlike traditional fingerprint-based IPSs, the MIPS-PSO system requires no prior training or parameter knowledge, eliminating extensive offline data collection. The system achieves an APE of 2.57 meters, a 40% improvement over trilateration-based IPSs.

The experimental testbed, on the second floor of a school covering an area of 720 m^2, includes 15 fixed anchor nodes (ANs) and approximately 150 reference points spaced 2 meters apart. This setup allowed the collection of 15,000 signal samples using 11 different mobile devices to validate the system's performance. Another contribution from the authors (Assayag et al., 2024c) details a new IPS leveraging BLE technology and RSSI in a log-distance signal propagation model optimized with a genetic algorithm to improve mobile device positioning accuracy. The GASP-IPS system requires minimal pre-deployment efforts and no prior signal data collection, depending only on the floor plan layout and fixed AN coordinates. This system achieves an APE of 2.52 meters, exceeding the performance of traditional approaches using fixed propagation model values. The experimental framework remains consistent with their previous paper.

In summary, the reviewed literature on BLE indoor positioning methods reveals various RSSI data collection approaches, some bypassing the training phase. These studies employ machine learning, deep learning, multilateration, genetic optimization, and adaptive modeling techniques, achieving positioning errors typically between 1 to 3 meters, depending on validation scenario features. This indicates a trend towards more efficient and accurate BLE-based indoor positioning solutions.

Table 1 depicts the main features regarding the aforementioned positioning systems, from which The proposed OPTIMAPS is compared with. For a fair analysis, the accuracy measure proposed by Morgan (2024) is utilized, which takes into account infrastructure parameters such as testbed size and the number of APs, in addition to the pure position error:

$$\%E_{AD} = E_{ABS}^2 \times \frac{N_A}{A} \times 100 \tag{2.1}$$

where $\%E_{AD}$ is the Anchor Density (AD) percentage error, E_{ABS} is the absolute error, N_A is the number of anchor nodes (access points), and A is the testbed area. A lower value for $\%E_{AD}$ indicates higher system accuracy.

The table presents a comparison of the methods utilized, validation scenario size, number of APs, site-survey effort, APE, and AD percentage error. The system designated as "hybrid" employs a combination of range-free and model-based methodologies. Training effort is classified as "low" or "high" based on the average number of offline training points gathered relative to the indoor area. In this context, high training effort refers to a scenario in which fingerprints are collected uniformly across the entire area, whereas low training effort is characterized by selecting specific points within the scenario. An effortless training scenario is then defined as one in which no fingerprint collection occurs.

Table 1 – Performance comparison among BLE-based indoor positioning systems.

System	Method	Testbed Size	Number of APs	Training Effort[(1)]	Average Error	$\%E_{AD}$[(2)]
Subhan et al. (2019)	Fingerprinting	100 m^2	4	High	1.02 m	4.16
Nikodem and Szeliński (2021)	Model-based	100 m^2	4	Low	1.77 m	12.53
Danış et al. (2023)	Fingerprinting	364 m^2	12	High	2.14 m	15.10
Szyc et al. (2023)	Hybrid	1,600 m^2	10	Low	7.89 m	38.91
Beigi and Shah-Mansouri (2024)	Fingerprinting	28 m^2	3	High	0.61 m	3.99
Pinto and Oliveira (2024)	Model-based	720 m^2	15	Low	2.88 m	17.28
Wu et al. (2024)	Fingerprinting	72 m^2	8	High	1.22 m	16.54
Assayag et al. (2024b)	Model-based	720 m^2	15	None	2.57 m	13.76
Assayag et al. (2024c)	Model-based	720 m^2	15	None	2.52 m	13.23
OPTIMAPS	Model-based	720 m^2	15	None	2.65 m	14.63

(1) Data collection effort in the offline phase
(2) Anchor density percentage error

One can verify that OPTIMAPS outperforms nearly all systems with low training profiles concerning the AD percentage error metric. Only two

fingerprinting-based methods exhibit higher accuracy but require significant training efforts during the offline phase. Furthermore, OPTIMAPS achieved comparable APE results across all studies presented that utilized the same RSSI dataset and indoor scenario. This underscores the feasibility and competitiveness of the proposed solution within BLE technology, particularly in relation to the constraints of site-survey efforts.

Among the approaches that require no specific training effort and utilize the same experimental testbed, it is noteworthy that, while some methods surpass OPTIMAPS in terms of average positioning error, they do so at a significantly higher computational cost during the estimation phase. Specifically, both methods proposed by Assayag et al. employ genetic algorithms, whose typical computational complexity is $\mathcal{O}(k \times p \times n_p)$, with k denoting the number of iterations, p the number of particles, and n_p the number of optimized parameters. As the scenario size increases, both k and p tend to scale proportionally, potentially leading to an overall quadratic or even worse runtime. In contrast, OPTIMAPS utilizes a nearest-neighbor (NN) step with complexity $\mathcal{O}(n)$, where n is the number of reference points (which scales with the area), followed by SLSQP, which runs at $\mathcal{O}(k \times n_p^3)$ with k iterations and a small n_p. This means that, as the environment grows, OPTIMAPS maintains an essentially linear complexity with respect to area, representing a considerable advantage in terms of computational efficiency, energy consumption, and practical deployment on resource-constrained hardware.

3 Theoretical Background

This chapter presents the theoretical background upon which the proposed solution is based. Firstly, commonly deployed indoor positioning technologies are described, focusing on those based on Wi-Fi and BLE. Additionally, the main aspects of the log-distance path loss model are highlighted, which typically describes how the RSSI behaves in indoor environments. Subsequently, several metrics are introduced to characterize RSSI diversity in a given scenario. Positioning estimation methods are also discussed, emphasizing Nearest-Neighbor (NN), KNN, and LS approaches. Finally, some standard optimization techniques are presented, with particular attention given to the SLSQP algorithm.

3.1 Indoor Positioning Technologies

The inability of Global Navigation Satellite Systems (GNSS), such as GPS, to operate effectively indoors has catalyzed extensive research into Indoor Positioning Systems (IPS). An IPS is any system that attempts to determine the location of an object or person within a building or other enclosed space (Zafari et al., 2019). While numerous technologies exist (e.g., Ultra-Wideband, magnetic fields, computer vision), those based on Radio Frequency (RF) signals from commodity

hardware are particularly attractive due to their low cost and ease of deployment.

The most prevalent paradigm for RF-based IPS is fingerprinting, which typically involves two distinct phases:

1. **Offline (Training) Phase:** This is a labor-intensive calibration phase where a detailed "radio map" of the target environment is constructed. A surveyor systematically moves through the area, stopping at a grid of known locations, termed Reference Points (RPs). At each RP, the device collects RSSI measurements from all detectable transmitters (e.g., Wi-Fi APs, BLE beacons). This data, often comprising the mean and other statistics of the RSSI values from each transmitter, is stored in a database. The resulting entry for each RP is a vector of signal strength indicators, which serves as a unique "fingerprint" for that physical location.

2. **Online (Positioning) Phase:** In this operational phase, a user with a mobile device wishes to determine their location. The device scans for nearby RF signals and measures their current RSSI values, forming a real-time fingerprint. This measured fingerprint is then passed to a positioning algorithm, which compares it against the pre-computed fingerprints in the radio map database. By finding the best match or combination of matches, the algorithm estimates the user's current coordinates.

3.1.1 Wi-Fi-based Positioning

The widespread deployment of IEEE 802.11 (Wi-Fi) networks in commercial and residential buildings makes Wi-Fi a de facto infrastructure for indoor positioning. The RADAR system was a pioneering work that demonstrated the feasibility

of using Wi-Fi fingerprinting for accurate room-level localization (Bahl and Padmanabhan, 2000).

The primary advantage of Wi-Fi is leveraging existing infrastructure, thus minimizing deployment costs. However, Wi-Fi-based positioning faces significant challenges. RSSI values are notoriously unstable due to complex indoor propagation phenomena such as multipath fading (where the signal reaches the receiver via multiple paths, causing constructive and destructive interference), shadowing from static and dynamic obstacles (e.g., walls, furniture, people), and Non-Line-of-Sight (NLOS) conditions. Furthermore, device heterogeneity poses a problem, as different Wi-Fi chipsets and antenna orientations can result in disparate RSSI readings at the same location.

3.1.2 Bluetooth Low Energy (BLE)-based Positioning

Introduced as part of the Bluetooth 4.0 specification, BLE was engineered for short-range communication with minimal power consumption. This has made it a compelling technology for granular indoor positioning. The typical architecture involves deploying small, battery-powered transmitters known as beacons, which periodically broadcast advertising packets on specific channels. These packets contain a unique identifier and can be detected by any BLE-enabled device, such as a smartphone.

Compared to Wi-Fi, BLE offers several advantages. The low power profile allows beacons to operate for years on a single coin-cell battery, simplifying deployment and maintenance. Their low cost enables dense deployments, which can lead to higher spatial resolution and accuracy (Faragher and Harle, 2012).

However, BLE is not without its drawbacks. Its signals, typically in the 2.4 GHz band, are also susceptible to multipath, shadowing, and interference. The lower transmission power and shorter range mean that human body absorption can significantly attenuate the signal, introducing considerable variance.

3.2 Radio Propagation Modeling

Understanding the relationship between distance and signal strength is fundamental to both model-based positioning (like multilateration) and interpreting the data used in fingerprinting. While indoor environments are too complex for perfect deterministic models, empirical models provide a valuable approximation.

3.2.1 The Log-distance Path Loss Model

The log-distance path loss model is a widely used mathematical model that estimates the attenuation or loss of signal strength as it travels from a transmitter to a receiver over a distance. It takes into account factors such as the distance between the transmitter and receiver, the frequency of the wireless signal, and the environment in which the signal propagates. The model is based on the empirical observation that signal strength decreases logarithmically with distance (Rappaport, 2002).

In the context of indoor localization, the log-distance path loss model is applied to estimate the distance between a transmitter (e.g., an access point or a beacon) and a receiver (e.g., a mobile device). By measuring the received

signal strength and using the log-distance path loss model, it becomes possible to estimate the distance between the transmitter and receiver. This distance estimation is then used as input for localization algorithms such as the ones which triangulate or trilaterate the position of the receiver within the indoor environment.

The log-distance path loss model is typically represented by the following equation:

$$r = P_t - \rho_0 - 10\alpha_L \log\left(\frac{d}{d_0}\right) + X_\sigma \tag{3.1}$$

In this equation, r is the received signal strength indicator (RSSI), measured in dBm, d is the distance between the transmitter and receiver, d_0 is a reference distance, P_t is the AP transmitted power, ρ_0 is the path loss at the reference distance, α_L is the path loss exponent, and X_σ represents additional factors such as shadowing or fading.

One can summarize the main parameters of the model as follows:

- *Path loss exponent* (α_L): The path loss exponent α_L is a parameter that characterizes the rate at which the signal strength decreases with distance. It is typically determined through empirical measurements or simulation studies specific to the indoor environment under consideration. The value of α_L depends on various factors, including the frequency of the wireless signal, the nature of the indoor environment (e.g., presence of walls, furniture), and the propagation characteristics of the wireless technology used (e.g., Wi-Fi, Bluetooth).

- *Reference distance* (d_0) and *path loss at reference distance* (ρ_0): The reference distance d_0 is a predefined distance at which the path loss is known or

assumed. The path loss at the reference distance ρ_0 is the path loss value corresponding to the reference distance. These values are typically obtained through calibration or measurement campaigns in the specific indoor environment of interest.

- *Additional factors* (X_σ): The additional factors X_σ in the log-distance path loss model account for variations or deviations from the idealized path loss behavior. These factors may include shadowing, fading, or other environmental conditions that affect the signal propagation in indoor environments. The specific form and inclusion of these factors depend on the model being used and the characteristics of the environment.

It is important to note that the log-distance path loss model provides a simplified representation of signal propagation in indoor environments and may not capture all the intricacies of real-world conditions (Vallet García, 2020). Hence, researchers often combine the path loss model with other techniques, such as fingerprinting, angle-of-arrival, or time-of-flight measurements, to improve the accuracy of indoor positioning systems.

3.3 RSSI Diversity and Characterization

The temporal and spatial variability of RSSI is a double-edged sword. While it creates unique fingerprints, its instability is a primary source of error. Therefore, characterizing this diversity using statistical metrics is crucial for building robust positioning systems (Kaemarungsi and Krishnamurthy, 2004). For a set of RSSI samples collected at a single point over time, the following metrics are essential:

- **Mean (μ):** The arithmetic average of the RSSI samples. It is the most common feature used in a fingerprint, representing the central tendency of the signal strength at a location.

- **Standard Deviation (σ) and Variance (σ^2):** These metrics quantify the dispersion or volatility of the signal. A high variance indicates an unstable signal, possibly due to dynamic multipath or interference. This information is valuable; for instance, in a positioning algorithm, measurements from transmitters with high historical variance can be down-weighted, as they are less reliable.

- **Higher-Order Moments:** Skewness measures the asymmetry of the RSSI distribution, while kurtosis measures its "tailedness." A non-zero skewness might indicate persistent NLOS conditions, while high kurtosis could suggest infrequent but strong interference spikes. These can be used to build more sophisticated signal models beyond a simple Gaussian assumption.

While the statistical metrics described above are essential for characterizing the RSSI distribution at a single reference point and building a robust radio map, their primary utility in a fingerprinting system is to form the basis for comparison during the online positioning phase. When a mobile device captures a real-time RSSI vector, the fundamental task of the positioning algorithm is to determine which of the pre-recorded fingerprints in the radio map it most closely resembles.

This resemblance is not evaluated subjectively but is quantified mathematically using pairwise distance metrics. These metrics operate in the N-dimensional signal space, where N is the number of observed transmitters, and

compute a scalar value that represents the dissimilarity between the live measurement and each stored fingerprint. The selection of an appropriate distance metric is a critical design choice, as it dictates how "closeness" is defined and directly impacts the system's accuracy and robustness to signal noise. The following subsections detail the most common metrics employed for this purpose.

3.3.1 Pairwise Distance Metrics in Signal Space

In the context of RSSI fingerprinting, a core task of the online positioning phase is to quantify the dissimilarity between a newly measured RSSI vector and the set of pre-recorded fingerprint vectors stored in the radio map. This quantification is achieved using a distance metric, which calculates a scalar value representing the "distance" between two vectors in the N-dimensional signal space, where N is the number of transmitters (e.g., Access Points). The choice of metric is critical as it directly influences which reference points are considered "close" and, consequently, the final position estimate. These metrics are a specific application of the more general Minkowski distance, or L_p norm, which is defined for two vectors $\mathbf{x}$ and $\mathbf{y}$ as $L_p = (\sum_{i=1}^{N} |x_i - y_i|^p)^{1/p}$ (Duda et al., 2012). In the following subtopics, the two most common instances used in localization will be discussed: Euclidean ($p = 2$) and Chebyshev ($p \to \infty$).

Let the real-time measured RSSI vector be $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ and a stored fingerprint vector from the j-th reference point be $\mathbf{s}_j = (s_{j1}, s_{j2}, \ldots, s_{jN})$.

3.3.1.1 Euclidean Distance

The Euclidean distance is the most widely used metric in fingerprinting systems, largely due to its intuitive geometric interpretation as the straight-line or "as-the-crow-flies" distance between two points in signal space (Bahl and Padmanabhan, 2000). It is the L_2 norm of the difference between the two vectors. The Euclidean distance, D_E, between the live vector $\mathbf{r}$ and a stored vector $\mathbf{s}_j$ is calculated as:

$$D_E(\mathbf{r}, \mathbf{s}_j) = \sqrt{\sum_{i=1}^{N}(r_i - s_{ji})^2} = ||\mathbf{r} - \mathbf{s}_j||_2 \tag{3.2}$$

The metric squares the difference for each transmitter's RSSI, sums these squares, and takes the square root. This process ensures that all transmitters contribute to the final distance value. A key characteristic of the Euclidean distance is its sensitivity to the magnitude of errors. A large deviation in the RSSI from a single transmitter will result in a large squared term, significantly increasing the overall distance. While this provides a holistic measure of dissimilarity, it also makes the metric susceptible to outlier measurements, where a single spurious reading can dominate the distance calculation and potentially lead to an incorrect match.

3.3.1.2 Chebyshev Distance

The Chebyshev distance, also known as the maximum value distance or the L_∞ norm, offers an alternative perspective on dissimilarity. Instead of aggregating the differences across all dimensions, it defines the distance between two vectors as the single greatest difference along any coordinate dimension (Youssef and Agrawala, 2003). The Chebyshev distance, D_C, is calculated as:

$$D_C(\mathbf{r}, \mathbf{s}_j) = \max_i |r_i - s_{ji}| = ||\mathbf{r} - \mathbf{s}_j||_\infty \tag{3.3}$$

This metric effectively identifies the "worst-case" deviation between the live RSSI vector and a stored fingerprint. Its primary advantage is its robustness to a situation where multiple transmitters show small deviations but one transmitter shows a large, anomalous deviation. The Euclidean distance would be heavily skewed by this single large error, whereas the Chebyshev distance would simply report that maximum error. Conversely, its main drawback is that it completely ignores the information from all other transmitters. If two fingerprints have the same maximum deviation from the live vector but one is much closer on all other dimensions, the Chebyshev distance will consider them equally dissimilar. The choice between Euclidean and Chebyshev, therefore, depends on the expected noise characteristics of the environment and the desired behavior of the matching algorithm.

3.4 Positioning Estimation Methods

After collecting a real-time RSSI vector, an algorithm must process it to produce a coordinate estimate. This section details three fundamental deterministic algorithms.

3.4.1 Nearest-Neighbor (NN)

The Nearest-Neighbor algorithm is the most straightforward fingerprinting method. It compares the live RSSI vector with every fingerprint in the radio

map and identifies the single RP that is "closest" in signal space. The physical coordinates of this best-matching RP are then returned as the estimated position. The notion of "closeness" is defined by a distance metric, with the Euclidean distance being the most common:

$$D_j = \left(\sum_{i=1}^{N} (r_i - s_{ij})^2 \right)^{1/2} \tag{3.4}$$

where D_j is the signal distance to the j-th RP, N is the number of transmitters in the fingerprint, r_i is the live RSSI from transmitter i, and s_{ij} is the stored mean RSSI from transmitter i at RP j. While simple and computationally fast in the online phase, NN is highly susceptible to noise and outliers; a single erroneous RSSI reading can cause it to select the wrong RP.

3.4.2 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors algorithm is a robust extension of NN that mitigates its sensitivity to outliers (Bahl and Padmanabhan, 2000). Instead of relying on a single best match, KNN identifies the K reference points with the smallest signal space distances to the live measurement. The final position estimate is then computed as the coordinate corresponding to the mode of these K neighbors.

A more refined version is the Weighted K-Nearest Neighbors (WKNN) algorithm. In WKNN, the contribution of each of the K neighbors is weighted, typically by the inverse of its signal distance. This gives more influence to closer neighbors in signal space, further improving accuracy. The estimated position $(\hat{x}, \hat{y})$ is calculated as:

$$(\hat{x}, \hat{y}) = \sum_{j=1}^{K} w_j (x_j, y_j), \quad \text{where} \quad w_j = \frac{1/D_j}{\sum_{k=1}^{K} 1/D_k} \tag{3.5}$$

Here, (x_j, y_j) are the coordinates of the j-th nearest neighbor and D_j is its signal distance. The choice of K is a critical tuning parameter, representing a trade-off between noise suppression and localization resolution.

3.4.3 Least Squares (LS)

Unlike fingerprinting, the Least Squares method is a model-based approach that falls under the category of multilateration. It first attempts to convert RSSI measurements into distance estimates using an inverted form of the path loss model (Equation 3.1). With distance estimates d_i from M transmitters at known locations (x_i, y_i), the user's unknown location (x, y) must satisfy a system of geometric equations:

$$(x - x_i)^2 + (y - y_i)^2 = d_i^2, \quad \text{for } i = 1, \ldots, M \tag{3.6}$$

Due to the high error in RSSI-to-distance conversion, this system is typically inconsistent. The LS method seeks to find the location (x, y) that best fits the equations by minimizing the sum of the squared errors (residuals). The Non-Linear Least Squares (NLLS) formulation is:

$$\underset{x,y}{\text{minimize}} \sum_{i=1}^{M} \left(\sqrt{(x - x_i)^2 + (y - y_i)^2} - d_i \right)^2 \tag{3.7}$$

This is a non-convex optimization problem that requires an iterative solver. A common simplification involves linearizing the system, which allows

for a closed-form solution but can amplify errors. The NLLS approach, while more complex, is generally more accurate.

3.5 Optimization Techniques

Positioning, particularly using model-based approaches like LS, is fundamentally an optimization problem: we seek to find the parameters (coordinates) that best minimize a cost function (e.g., the sum of squared errors) subject to certain constraints (e.g., the user must be within the building's boundaries).

3.5.1 Broyden–Fletcher–Goldfarb–Shanno (BFGS)

The BFGS algorithm is one of the most popular quasi-Newton methods for unconstrained nonlinear optimization (Nocedal and Wright, 2006). Newton's method finds the minimum of a function by using its second derivatives (the Hessian matrix), but computing the Hessian can be expensive. Quasi-Newton methods, like BFGS, circumvent this by iteratively building an approximation of the inverse Hessian matrix using only first-derivative (gradient) information.

At each iteration, BFGS performs the following steps:

1. Compute a search direction $\mathbf{p}_k = -B_k^{-1}\nabla f(\mathbf{x}_k)$, where B_k is the current approximation of the Hessian.

2. Perform a line search to find an optimal step size α_k and update the position: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$.

3. Update the inverse Hessian approximation B_{k+1}^{-1} using the change in position and the change in the gradient.

BFGS is known for its superlinear convergence rate and is highly effective for smooth, unconstrained problems. While the basic algorithm is for unconstrained problems, variants like L-BFGS-B can handle simple box constraints (bounds), making it applicable to positioning problems where the search space is a known rectangle.

3.5.2 Trust-Region Reflective (TRF) Algorithm

Trust-region methods are another class of powerful algorithms for nonlinear optimization. Instead of choosing a search direction and then a step size (like line search methods), a trust-region method defines a "region" around the current iterate where a model of the objective function (typically quadratic) is considered trustworthy. It then takes a step to the minimizer of the model within this trust region (Conn et al., 2000).

The Trust-Region Reflective (TRF) algorithm is a specific implementation designed for large-scale, bound-constrained problems. At each step, it solves a trust-region subproblem. If the proposed step hits a boundary, it is "reflected" back into the feasible domain. The size of the trust region is adjusted based on the agreement between the model and the actual objective function. If the model was a good predictor, the region is expanded; if not, it is shrunk. This makes the method robust, especially for ill-conditioned problems. It is a common choice for solving non-linear least-squares problems subject to bounds.

3.5.3 Particle Swarm Optimization (PSO)

PSO is a population-based stochastic optimization technique inspired by the social behavior of bird flocking or fish schooling (Kennedy and Eberhart, 1995). The algorithm maintains a population, or "swarm," of candidate solutions, called "particles." Each particle "flies" through the multi-dimensional search space, adjusting its position based on its own best-known position and the entire swarm's best-known position.

The velocity and position of each particle are updated in each iteration. The velocity update for a particle is influenced by three components:

- Its current inertia (tendency to maintain its current direction).
- The "cognitive" component: the particle is drawn towards its own personal best position found so far (p_{best}).
- The "social" component: the particle is drawn towards the global best position found by any particle in the swarm (g_{best}).

PSO is a metaheuristic, meaning it makes few assumptions about the problem being optimized and can search large spaces. This makes it suitable for complex, non-convex, and non-differentiable cost functions like those found in positioning, where traditional gradient-based methods might get stuck in local minima.

3.5.4 Simulated Annealing (SA)

Simulated Annealing (SA) is a probabilistic metaheuristic for global optimization, inspired by the process of annealing in metallurgy, where a material is heated

and then slowly cooled to increase the size of its crystals and reduce its defects (Kirkpatrick et al., 1983).

The algorithm starts with a random solution and an initial high "temperature." In each iteration, it considers a move to a neighboring solution. If the new solution is better, the move is always accepted. If the new solution is worse, it may still be accepted with a certain probability, which is a function of the current temperature and the magnitude of the worsening. This probability of accepting worse solutions is the key feature of SA: it allows the algorithm to escape local optima, especially at high temperatures. As the algorithm progresses, the temperature is gradually decreased according to a "cooling schedule." At lower temperatures, the probability of accepting worse moves decreases, causing the algorithm to converge towards a good solution. Its main strength is its proven convergence to the global optimum (given a sufficiently slow cooling schedule), but it can be computationally slow.

3.5.5 Differential Evolution (DE)

Differential Evolution (DE) is another powerful population-based metaheuristic, similar in spirit to genetic algorithms, designed for global optimization over continuous spaces (Storn and Price, 1997). DE works by maintaining a population of candidate solutions and creating new candidates by combining existing ones.

For each member of the population (the "target vector") in each generation, a "mutant vector" is created by taking three other random members from the population, say $\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3}$. The mutant is generated as $\mathbf{v} = \mathbf{x}_{r1} + F \cdot (\mathbf{x}_{r2} - \mathbf{x}_{r3})$, where F is a scaling factor. Then, a "trial vector" is created by "crossover," mixing

the parameters of the mutant vector with the target vector. If this trial vector yields a better value of the objective function than the target vector, it replaces the target in the next generation. This simple yet powerful mechanism of using population differences to perturb solutions makes DE very effective at navigating complex search spaces and finding global optima.

3.5.6 Sequential Least Squares Programming (SLSQP)

Sequential Least Squares Programming (also known as Sequential Quadratic Programming) is a state-of-the-art iterative method for solving constrained Non-Linear Programming (NLP) problems (Nocedal and Wright, 2006). It is designed to handle problems of the general form:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \tag{3.8}$$

$$\text{subject to} \quad g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, J \quad \text{(inequality constraints)} \tag{3.9}$$

$$h_k(\mathbf{x}) = 0, \quad k = 1, \ldots, K \quad \text{(equality constraints)} \tag{3.10}$$

The core idea of the algorithm, originally proposed by Kraft (Kraft, 1988), is to solve the complex non-linear problem by iteratively solving a sequence of simpler subproblems. At each iteration k, the algorithm approximates the original problem with a Quadratic Program (QP). It does this by forming a quadratic approximation of the problem's Lagrangian function and a linear approximation of the constraints around the current iterate $\mathbf{x}_k$. This QP subproblem can be solved efficiently to find a search direction $\mathbf{p}_k$. A line search is then performed along this direction to find a step size α_k that makes sufficient progress in reducing the objective function or a merit function, leading to the next iterate: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$.

For the NLLS positioning problem (Equation 3.7), SLSQP is particularly well-suited. The objective function is the sum of squared residuals, and we can add inequality constraints to ensure the estimated position lies within the known physical boundaries of the building (e.g., $x_{min} \leq x \leq x_{max}$). Its robustness and ability to handle both equality and inequality constraints make it a powerful tool for refining location estimates.

3.6 Statistical Analysis of Location Errors in IPSs

The rigorous evaluation of IPS performance necessitates sophisticated statistical methodologies to quantify measurement uncertainty and establish meaningful comparisons between competing approaches. Location error analysis presents unique challenges due to the inherently spatial nature of positioning data, potential correlations between measurements, and the need to distinguish between systematic and random error components. This section presents a comprehensive statistical framework that addresses these challenges while providing robust methods for performance assessment and comparative analysis.

The fundamental question underlying any comparative study in indoor positioning concerns whether observed performance differences between methods represent genuine systematic variations or merely reflect random measurement variability. Traditional statistical approaches, when applied without consideration of the spatial and paired nature of positioning data, may lead to incorrect conclusions regarding system performance. This analysis framework addresses these concerns through appropriate test selection, assumption validation, and effect size quantification.

3.6.1 Experimental Design Considerations

The validity and power of statistical comparisons in indoor positioning research depend critically on experimental design decisions, particularly regarding the spatial configuration of test locations and the temporal sequence of measurements. Two primary design paradigms emerge from the literature: independent sampling designs and paired comparison designs.

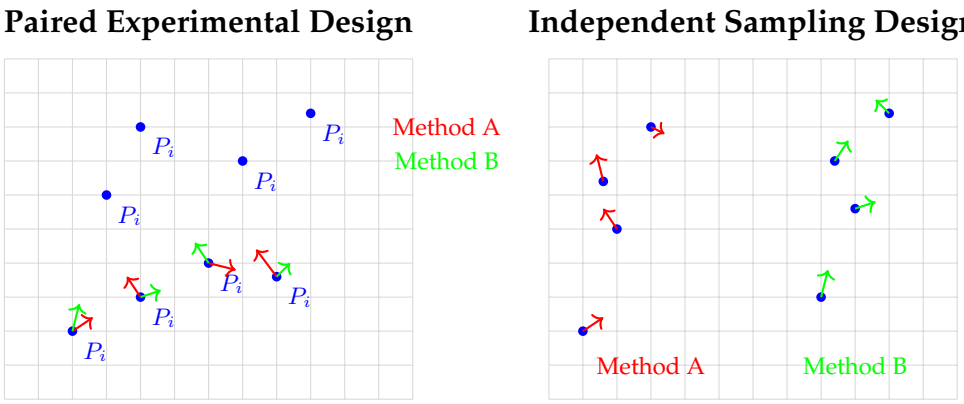

Figure 1 – Comparative illustration of experimental design paradigms in indoor positioning research.

Paired comparison designs, wherein both positioning methods are evaluated at identical spatial locations, offer several theoretical and practical advantages. The elimination of location-specific confounding factors substantially increases statistical power by reducing between-subject variability. This design paradigm enables the isolation of method-specific effects from spatial heterogeneity, leading to more precise estimates of performance differences.

In contrast, independent sampling designs, where different spatial locations are used for each method evaluation, may provide broader generalizability but at the cost of reduced statistical power and potential introduction of location-

related bias. The choice between these paradigms involves fundamental tradeoffs between internal validity and external generalizability.

The temporal aspects of data collection present additional considerations. Sequential measurements at the same location may introduce temporal correlation due to environmental changes, infrastructure modifications, or equipment drift. Randomized measurement sequences can mitigate some of these effects, while repeated measurements can provide estimates of temporal variability.

3.6.2 Statistical Inference for Paired Comparisons

When positioning methods are evaluated using paired experimental designs, the statistical analysis focuses on the distribution of pairwise differences rather than the individual error distributions. This approach offers substantial advantages in terms of statistical power and interpretability.

For each location i, the performance difference between methods is quantified as:

$$D_i = d_{1i} - d_{2i} \tag{3.11}$$

where d_{1i} and d_{2i} represent the positioning errors for methods 1 and 2 at location i, respectively. The central question in paired analysis concerns whether the expected value of these differences differs significantly from zero.

Under the assumption that the differences D_i follow a normal distribution, the paired t-test provides an optimal testing framework (Student, 1908). The test statistic is constructed as:

$$t = \frac{\bar{D} - \mu_0}{s_D/\sqrt{n}} \tag{3.12}$$

where $\bar{D}$ represents the sample mean of differences, s_D denotes the sample standard deviation of differences, n is the number of paired observations, and $\mu_0 = 0$ under the null hypothesis of no difference between methods.

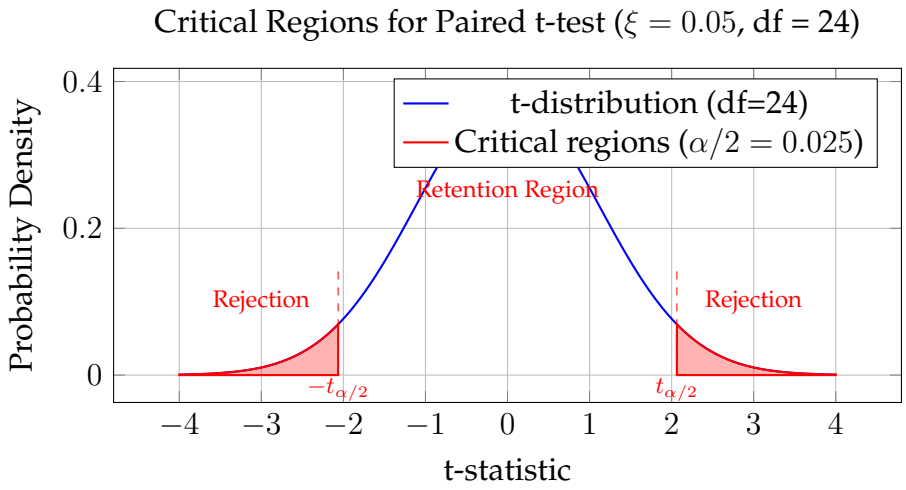

Figure 2 – Statistical decision regions for the paired t-test under a two-tailed alternative hypothesis.

This test statistic follows a t-distribution with $(n-1)$ degrees of freedom under the null hypothesis. The confidence interval for the mean difference provides both point and interval estimates of the performance differential:

$$\bar{D} \pm t_{\alpha/2,n-1} \cdot \frac{s_D}{\sqrt{n}} \tag{3.13}$$

where α is the significance level.

The validity of the paired t-test depends on several critical assumptions: normality of the difference distribution, independence of paired observations,

and absence of extreme outliers. Violation of these assumptions may necessitate alternative approaches.

When the normality assumption is violated, non-parametric alternatives provide robust inference procedures. The Wilcoxon signed-rank test evaluates whether the median of the differences differs significantly from zero without requiring distributional assumptions beyond symmetry around the median (Wilcoxon, 1945).

The test procedure involves ranking the absolute values of non-zero differences and computing the sum of ranks corresponding to positive differences:

$$W^+ = \sum_{D_i>0} R_i \tag{3.14}$$

where R_i represents the rank of $|D_i|$ among all absolute differences. For large sample sizes, the test statistic can be standardized using a normal approximation:

$$Z = \frac{W^+ - \mu_{W^+}}{\sigma_{W^+}} \tag{3.15}$$

where $\mu_{W^+} = \frac{n(n+1)}{4}$ and $\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$.

The sign test provides the most robust non-parametric alternative, requiring only that the differences be independent and that the probability of positive and negative differences be equal under the null hypothesis. The test statistic simply counts the number of positive differences:

$$S = |\{i : D_i > 0\}| \tag{3.16}$$

Under the null hypothesis, this statistic follows a binomial distribution with parameters n and $p = 0.5$.

3.6.3 Effect Size Analysis

Effect size quantification provides a measure of practical significance beyond statistical significance, particularly crucial when comparing algorithm performance with large sample sizes where even trivial differences may achieve statistical significance (Cohen, 1988; Lakens, 2013).

3.6.3.1 Cohen's d for Paired Comparisons

For paired algorithm comparisons, we employ Cohen's d for dependent samples:

$$d = \frac{\bar{d}}{s_d} \tag{3.17}$$

where $\bar{d}$ represents the mean of paired differences and s_d is the standard deviation of these differences. Cohen's conventional interpretation guidelines classify effect sizes as small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$).

3.6.3.2 Contextual Effect Analysis

Effect sizes are calculated both globally and within spatial subregions to identify areas where algorithmic improvements are most pronounced. This approach reveals whether performance gains are consistent across the deployment en-

vironment or concentrated in specific scenarios. The combination of statistical significance testing and effect size quantification ensures that algorithm selection is based on both detectability and practical relevance of performance differences.

3.6.4 Statistical Decision Framework

The selection of appropriate statistical methods depends on several key factors that must be carefully evaluated for each comparative study. A systematic decision framework emerges from the theoretical considerations presented above.

The primary decision point concerns the experimental design: whether positioning methods are evaluated at identical spatial locations (enabling paired analysis) or at different locations (requiring independent sample methods). This fundamental design choice has profound implications for statistical power, required sample sizes, and the validity of conclusions.

Secondary decisions involve the verification of distributional assumptions. Normality testing of error differences (for paired designs) or individual error distributions (for independent designs) determines whether parametric or non-parametric methods are appropriate. The Shapiro-Wilk test provides good performance for smaller samples ($n < 50$), while the Kolmogorov-Smirnov test is more appropriate for larger samples (Massey, 1951; Shapiro and Wilk, 1965).

Homoscedasticity (equal variances) represents an additional assumption for independent sample comparisons. Levene's test and Bartlett's test provide formal procedures for variance equality assessment, though visual inspection through box plots or residual plots often provides valuable diagnostic information.

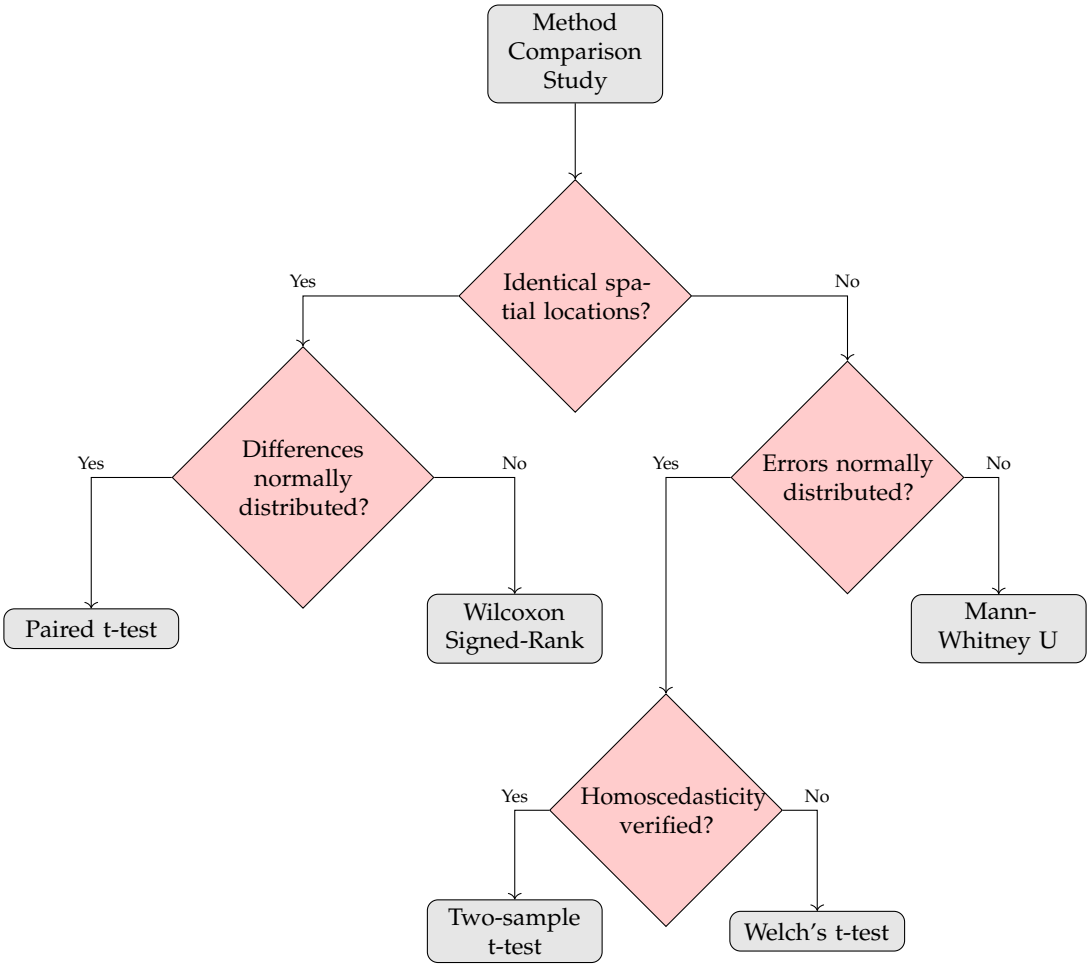

Figure 3 – Statistical decision framework for method comparison in indoor positioning research.

3.6.5 Synthesis and Recommendations

The statistical analysis of location errors in indoor positioning systems requires careful consideration of multiple interacting factors. The paired experimental design emerges as the preferred approach when feasible, offering substantial advantages in terms of statistical power, control of confounding variables, and precision of effect size estimates.

The theoretical framework presented here demonstrates that appropriate

statistical method selection depends critically on experimental design characteristics, distributional properties of positioning errors, and the specific research questions under investigation. Parametric methods (paired t-test, two-sample t-test) provide optimal power when their assumptions are satisfied, while non-parametric alternatives (Wilcoxon signed-rank test, Mann-Whitney U test) offer robust alternatives when distributional assumptions are violated.

Effect size quantification and confidence interval estimation provide essential information beyond traditional hypothesis testing, enabling assessment of practical significance and precision of estimates.

The spatial and temporal complexities inherent in positioning data necessitate careful attention to assumptions of independence and the potential presence of outliers.

The integration of appropriate experimental design, statistical method selection, and effect size quantification provides a foundation for advancing the scientific understanding of indoor positioning technology performance.

4 System Overview

This chapter presents the proposed solution, from fundamentals to application. It outlines the proposed system's steps, and depicts the experimental testbed based on a real-world indoor layout.

4.1 Proposed Method

The so-called OPTIMAPS consists of offline and online phases. In the offline phase, parameters of the log-distance path loss model are established to reflect the average propagation characteristics of the environment. For each test point requiring location estimation, expected RSSI vectors are calculated using the established model. Notably, this offline phase demands only the AP ground-truth positions along the floor plan, with no RSSI data collection for training. In the online phase, a nearest-neighbor estimation algorithm is optimized through SLSQP to generate the final position estimate.

Figure 4 illustrates the data flow diagram. Each step is described in the subsequent topics.

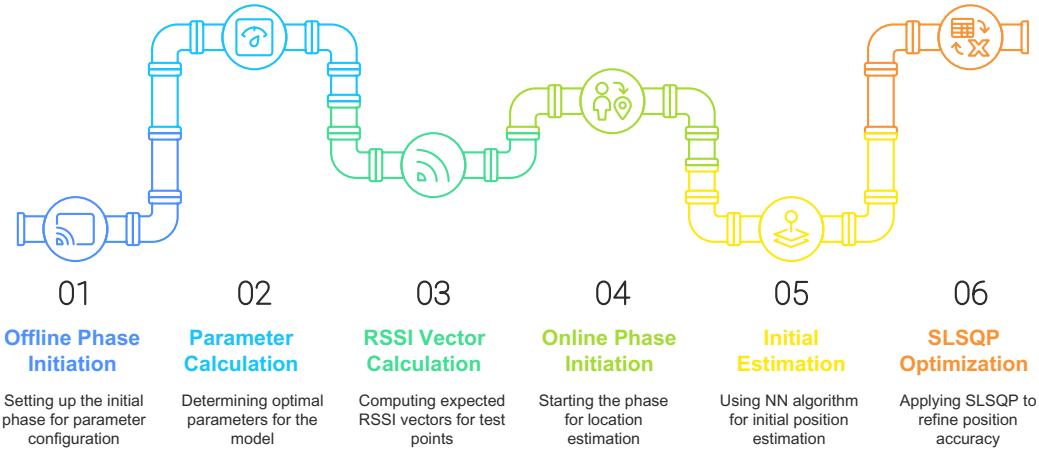

Figure 4 – Positioning system pipeline divided into an offline phase and an online phase.

4.1.1 Offline phase

IPSs typically utilize this phase for fingerprint collection; however, the presented system focuses on configuring log-distance path loss model parameters during the offline phase. While fingerprint-based schemes require significant time and effort for site surveys, the aim is to eliminate this inconvenience entirely.

Having presented the model fundamentals in the previous section, the goal is to identify suitable pairs for ρ_0 and α_L to characterize the signal, given that the transmitter power P_t of each BLE beacon is fixed by the hardware at 0 dBm. Typically, these parameters are estimated using regression techniques on offline collected RSSI samples (Zafari et al., 2019). Still, the proposed solution is to determine these parameters based solely on the geometric characteristics of the scenario and the hypothesis of RSSI dissimilarity.

The geometric features are defined by the physical dimensions and coordinates of each access point. Using the log-distance path loss model, the expected

RSSI at each test point can be calculated, with the expected value of the model's random variable being zero – $X_\sigma \sim \mathcal{N}(\mu,\, \sigma^2)$. Each component of the resulting vector corresponds to the signal received by a specific access point.

The RSSI dissimilarity hypothesis asserts that increased distances between expected RSSI result in more accurate location estimation. In a nearest-neighbor approach, for instance, the nearest RSSI vector is more distinct under varied signal conditions than when signals are similar. An appropriate dissimilarity measure enables the selection of a suitable combination (ρ_0, α_L), ultimately yielding a more accurate estimation. A common method to calculate this measure is centered on the pairwise distances between RSSI vectors (Zerzucha and Walczak, 2012). Therefore, an indicator of the distribution of these distances should identify the most effective pair $(\hat{\rho_0}, \hat{\alpha_L})$.

Candidate pairs (ρ_0, α_L) are established with 40 dB $\leq \rho_0 \leq 80$ dB, and $1 \leq \alpha_L \leq 7$, which consist of intervals sufficiently large to encompass typical reported values in the literature concerning the previously mentioned hardware constraints (Assayag et al., 2024b; Seidel and Rappaport, 1992). Subsequently, the pairwise distances for each set of parameters are calculated. In this context, the `pdist` function from the SciPy library (Virtanen et al., 2020) is employed. The relevant metric parameters considered include the Euclidean and Chebyshev distances, as outlined in the previous section.

Essentially, the Euclidean metric quantifies the straight-line distance between two points in Euclidean space. For two n-dimensional points $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$, the Euclidean distance is defined as:

$$d_{Euc}(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2} \tag{4.1}$$

This metric meets the axioms of a metric space, making it the standard for evaluating true physical distances without anisotropic distortions or obstructions. Many IPS algorithms, such as fingerprinting and ranging-based localization, use Euclidean distance to assess the "closeness" between measured signal characteristics (RSSI values in the system) from an unknown position and a reference database (Torres-Sospedra et al., 2015).

The Chebyshev distance between two vectors X and Y is defined as the maximum absolute difference along any coordinate dimension. Mathematically:

$$d_{Cheb}(X, Y) = max\{|x_1 - y_1|, |x_2 - y_2|, ..., |x_n - y_n|\} \tag{4.2}$$

Unlike the Euclidean metric, the Chebyshev distance prioritizes the largest coordinate difference. It is especially beneficial when movement or error propagation is constrained along orthogonal axes. In IPS applications, particularly in grid-based localization or maze-like propagation models, the Chebyshev metric provides a simplified error measure. For instance, if obstacles constrain movement to horizontal or vertical paths, using the maximum deviation as a metric can effectively represent worst-case errors (Pu and You, 2018).

After selecting a metric, a diversity measure is required to quantify the distance between RSSI vectors. In this context, the mean of pairwise distances is selected as a metric for the average distances. A higher mean indicates greater dissimilarity among the expected RSSI vectors.

The offline phase concludes with the determination of maximum signal diversity or dissimilarity, along with the associated (ρ_0, α_L) recovered. For the mean, the maximum indicates the most dissimilar average RSSI vectors.

4.1.2 Online phase

By using the theoretically most effective pair $(\hat{\rho_0}, \hat{\alpha_L})$ and the corresponding expected RSSI vectors for each test point, the initial online location is estimated based on the received RSSI vector. A nearest-neighbor (NN) algorithm for this estimation is employed, although other methods can be also applicable.

The use of the NN as the initial estimation in the online phase is primarily motivated by its conceptual simplicity and computational efficiency. The NN method rapidly identifies the candidate position whose model-based RSSI vector is most similar to the current observation, exhibiting linear complexity relative to the number of reference points. This characteristic makes it highly suitable for real-time applications and scalable indoor environments.

Furthermore, when integrated with a scenario-specific path loss model and signal diversity maximization, NN provides a reliable and robust starting point for further optimization. It is less prone to initialization errors or convergence instability compared to more complex methods, ensuring that the subsequent SLSQP refinement begins from a physically meaningful and plausible location estimate. This synergy between a straightforward, well-founded initial guess and an adaptive optimization phase enhances both the speed and accuracy of the OPTIMAPS system.

The objective of this optimization step is to improve accuracy by utilizing the initial position estimation from the nearest-neighbor algorithm as input, in conjunction with a constraint circle to limit the search for the optimized solution. In particular, the SLSQP search is restricted to a circular area centered at $(\hat{x_0}, \hat{y_0})$, which denotes the initial estimation derived from the offline phase, with a radius R of 2 meters.

A circular (Euclidean) constraint more accurately reflects physical distances in real indoor environments, where signal propagation and movement are not biased to grid axes. The log-distance path loss model and most spatial RSSI diversity metrics (e.g., Euclidean, Chebyshev) are inherently isotropic. By confining the search area to a circular region, we ensure that all directions surrounding the initial estimate are considered equally, thereby reducing the potential for directional bias that may arise from a square or Manhattan (axis-aligned) restriction.

The value of 2 meters was selected based on the empirical spacing between test points in the experimental setup, which is approximately 2 meters. This choice ensures that the likely true position, even in the worst-case scenario (initial NN error), falls within the specified range. This approach reduces the risk of overlooking the correct location while simultaneously maintaining a compact optimization area, thereby enhancing the speed of convergence.

Optimization using SLSQP requires an objective function for minimization. For a robust estimation, the online RSSI must closely match the expected RSSI from the offline phase, which leads to:

$$f_{obj}(\rho_0, \alpha_L, x_k, y_k) = \sqrt{\sum_{j=1}^{n} \left(r_j + \rho_0 + 10\alpha_L \log_{10} \left(\sqrt{(x_k - x_j)^2 + (y_k - y_j)^2 + H^2} \right) \right)^2} \tag{4.3}$$

where r_j is the online RSSI from the $j-th$ AP, (x_j, y_j) denotes the coordinates of the $j-th$ AP, n indicates the total number of APs, (x_k, y_k) specifies the candidate locations within the restriction circle $(x_k - \hat{x_0})^2 + (y_k - \hat{y_0})^2 \leq R^2$ and the scenario dimensions, with k signifying the number of iterations executed by the optimiza-

tion algorithm, $(\hat{x_0}, \hat{y_0})$ representing the initial online estimation, and H denoting the vertical distance between the receiver and transmitter devices. Regarding the algorithm implementation, the `minimize` function from the SciPy library is used, with the corresponding metric parameter set to `"slsqp"`.

It is worth mentioning that the SLSQP algorithm aims to identify the optimal 4-tuple $(\rho_0, \alpha_L, x_k, y_k)$ that minimizes the objective function. During the optimization process, the parameters ρ_0 and α_L are continuously estimated for each online RSSI, which may not correspond with the pair $(\hat{\rho_0}, \hat{\alpha_L})$ estimated offline. The final location estimation is then expressed as:

$$(\hat{x}, \hat{y}) = \underset{(x_k, y_k)}{\operatorname{argmin}} \{f_{obj}(\rho_0, \alpha_L, x_k, y_k)\} \tag{4.4}$$

4.2 Experimental framework

To evaluate the system, a large-scale real-world testbed was employed, with RSSI data collected from real BLE devices. The corresponding dataset can be accessed in (Assayag et al., 2024a). This article describes the conception, implementation, and distribution of a comprehensive indoor positioning dataset created with BLE technology, based on RSSI measurements. The dataset was gathered to fill critical gaps in existing positioning resources, particularly the lack of real-world, large-scale RSSI data with explicit spatial labeling and anchor metadata, supporting both model-based and fingerprinting localization research.

4.2.1 Scenario

The experiments were conducted in a school facility in Manaus, Brazil (GPS: -3.088334, -59.964559), spanning about 720 m^2. The environment features 11 classrooms and 3 halls, with varied room dimensions and substantial wall partitions, replicating actual signal attenuation and multipath conditions typical of indoor scenarios.

4.2.2 Testbed and Data Acquisition

- **Test Points and Layout:** 148 unique positions, roughly 2 m apart, distributed throughout all rooms and halls. Placement started at 0.5 m from corner walls and extended linearly.
- **Mobile BLE Devices:** 10 manufactured beacons, identical in hardware/components, positioned variably (on arm, pocket, backpack) to represent practical use cases.
- **Access Points:** 15 fixed BLE nodes, mostly ceiling-mounted at 3.0 m, with some hall anchors at 2.5 m, installed to reflect realistic wiring and accessibility constraints. Access points locations are explicitly mapped and provided as metadata.

4.2.3 Data Collection Protocol

At each test point, every beacon transmitted BLE advertising packets at a frequency of 100 ms. The 15 access points, acting as BLE scanners, listened and measured RSSI on all three BLE advertising channels, forwarding the results to a

central gateway using a 900 MHz wireless connection. Each second, an average RSSI measurement was recorded per anchor, per device, per location—producing robust, noise-suppressed data. The team exclusively handled the environment during collection to avoid confounding influences.

4.2.4 Dataset Content and Structure

Each device has a corresponding CSV file, with rows describing:

- RSSI values for each of the 15 access points (WAP-#ID, dBm);
- Test point label ("LABEL", 150 unique IDs);
- X, Y coordinates (meters, origin defined on the floor map);
- Device ID;
- Room or hall ID.

4.2.5 Experimental Design and System Architecture

The physical infrastructure was purpose-built for energy efficiency (BLE beacons avoid WiFi), and designed for realistic operational challenges, such as power access and unobtrusive anchor installation. The research team opted to avoid WiFi dependencies, instead transmitting anchor data by sub-GHz wireless links to a central server (Intel i7, 16GB RAM). All 10 beacons were built to identical standards, ensuring data comparability.

Bluetooth advertising events on channels 37, 38, and 39 facilitated comprehensive signal scanning, while the anchors' storage bus enabled high-frequency

packet capture. To eliminate transient noise, each RSSI vector reflects averaging over ten BLE packets (1 s duration at 100 ms intervals).

4.2.6 Research Value and Utility

The dataset is a versatile resource for the scientific community:

- Enables fingerprinting and propagation modeling with high-precision ground truth;
- Supports comparative studies of positioning algorithms, error modeling, and spatial analysis;
- Facilitates machine learning applications in regression and classification for device localization;
- Allows studies involving device heterogeneity, environmental variability, and real-world deployment constraints.

4.2.7 Experiment Visualization and Summary

Figure 5 displays the testbed floor plan: a 45 m $\times$ 16 m area with 11 rooms and 3 halls. The infrastructure is based on BLE technology, consisting of scanners (access points) and beacons (wearables) that transmit at 0 dBm, as depicted in Figure 6. There are 148 points, spaced 2 meters apart and uniformly distributed, to collect RSSI samples from 15 APs. Each point represents a specific location on the floor plan. Also, approximately 100 samples using 10 different receiver devices are collected in each point (sample counts range from 7 to 20 per device

per testing location).

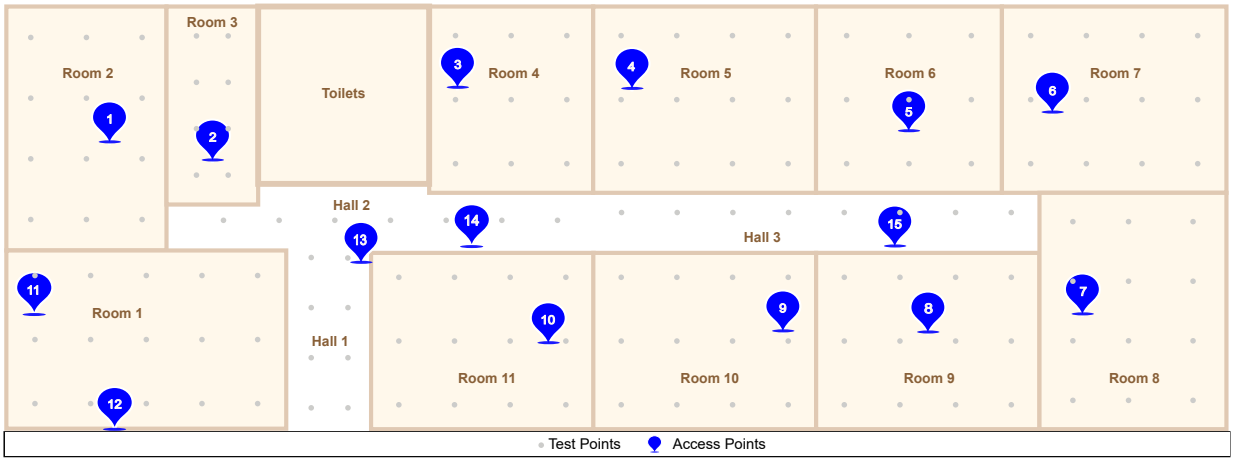

Figure 5 – Experimental testbed floor plan: 11 rooms, 3 halls, 15 access points, and 148 testing locations (adapted from Assayag et al. (2024a)).

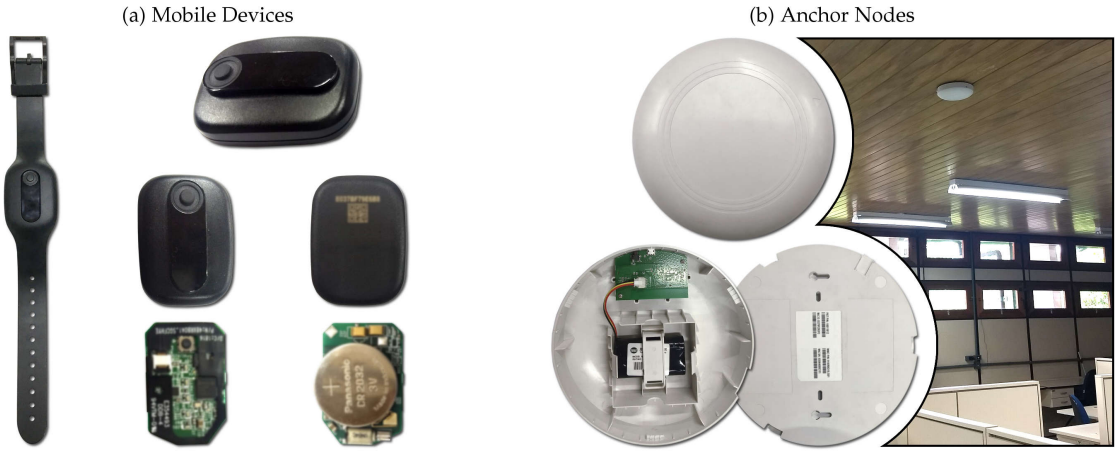

Figure 6 – Hardware utilized for collecting RSSI data: receiver devices (a) and transmitter devices (b) (Assayag et al., 2024a).

Extract, Load, and Transform (ELT) processes, system modeling, and tests were conducted using Python 3.11.11 on an HP Pro Mini Desktop (Windows 11 Pro, 64-bit; 2.30 GHz Intel i3-12300T; 8 GB RAM). The code developed for this thesis is stored in a repository on GitHub (Pinto, 2025). It contains all Jupyter notebooks required to reproduce the results presented in this thesis.

5 Results and Discussions

This chapter presents the main results of the proposed method, utilizing the APE for comparison with other techniques. The analysis examines how distance error varies with the $\alpha - \rho_0$ parameter combinations, demonstrating the superior performance of the SLSQP algorithm in enhancing the reference model. Additionally, the APE performance is assessed across the metric combinations employed for the pairwise distances in each pair $\alpha_L - \rho_0$. The cumulative error distribution for the most effective performances is then highlighted as an alternative means of visualizing positioning accuracy. Finally, the proposed system is compared to similar methods from recent literature to evaluate its applicability.

5.1 Choice of optimization algorithm

Among various optimization methods, the SLSQP algorithm is notable for its efficiency in execution time, as the average error remains comparable across the different approaches. For the purpose of comparison, Table 2 illustrates the overall APE and execution time for each of the listed optimization algorithms employed to enhance positioning accuracy. The APE is calculated across all test points and all BLE beacons, while execution time is expressed as the average

processing time per test point and per device.

Table 2 – Performance comparison among optimization algorithms with parameters $(\rho_0, \alpha_L) = (55, 4.25)$.

Optimization Algorithm	**APE (m)**	**Time Complexity**	**Execution Time (ms)**
PSO	2.68	$O(p \times k \times n_p)$	298.61
SA	2.72	$O(k \times n_p)$	617.09
DE	2.65	$O(p \times k \times n_p)$	1,587.97
TRF	2.72	$O(k \times n_p^2)$	255.47
L-BFGS-B	2.65	$O(k \times n_p^2)$	360.59
SLSQP	2.65	$O(k \times n_p^3)$	30.47

p: Population size (number of particles/individuals, for population-based algorithms);
k: Number of iterations (generations, cooling steps, or optimizer steps);
n_p: Number of parameters or variables being optimized (e.g., spatial coordinates, model parameters).

Under the same conditions, specifically under identical constraints and with a maximum number of iterations set to 100 (where applicable), the SLSQP algorithm produces the smallest APE and the shortest execution time. This indicates a performance that is more than eight times faster than that of the second-best algorithm, TRF. Although the processing time differences for positioning might not pose a problem, the SLSQP algorithm is the superior choice in terms of power consumption.

A closer examination of time complexity reveals that population-based metaheuristics, such as PSO and DE, require increasingly vast computational resources as the size of the environment, the number of anchors, or the dimensionality of the search space grows. This often renders these methods impractical for real-time, large-scale indoor positioning. In contrast, gradient-based algorithms such as TRF, BFGS, and particularly SLSQP, maintain high efficiency provided the number of optimization parameters remains small — typically four in this application. Notably, SLSQP excels in handling constrained nonlinear problems, offering both rapid convergence and robust constraint management,

and is especially effective when started from a strong initial estimate within a limited search region.

5.2 Optimization with the SLSQP algorithm

One important consideration when utilizing the SLSQP algorithm is the set of parameters passed to the corresponding Python function as arguments. The initial guess for the position estimate is derived from the estimation provided by the NN algorithm during the offline phase. There is also the option of incorporating the non-linear constraint, represented by the restriction circle mentioned previously. To illustrate the possible implementation scenarios, Table 3 depicts the APE results considering the log-distance parameters $(\rho_0, \alpha_L) = (55, 4.25)$. The SLSQP(-) denotes the algorithm without the restriction circle, whereas SLSQP(+) incorporates it.

Table 3 – Performance comparison among estimation techniques with parameters $(\rho_0, \alpha_L) = (55, 4.25)$.

Technique	**APE (m)**	**Execution time (ms)**
NN	2.89	1.15
SLSQP(-)	2.82	34.59
SLSQP(+)	2.65	30.47

One can verify that SLSQP(+) yields the best performance, indicating that the refinement of the position estimate with SLSQP enhances system accuracy. Furthermore, the restriction circle plays a critical role in reducing the APE. In terms of execution time, a single position estimate using SLSQP(+) takes a total of 30.47 ms.

For varying parameter sets α_L and ρ_0 in the path loss model, the reference model, such as the nearest-neighbor algorithm, will yield different APEs, as the expected RSSI vector at each test point varies. Visualizing APE behavior against the combination $\alpha_L - \rho_0$ is crucial for examining the relationship between parameters and average error, and for evaluating how the SLSQP(+) algorithm improves system accuracy. This is illustrated by Figure 7, where positioning errors are calculated for each model parameter combination α_L from 1 to 7 (0.25 steps) and ρ_0 from 40 to 80 dB (2.5 dB steps). To improve the readability of the graphics, ρ_0 is presented at each 5 dB step.

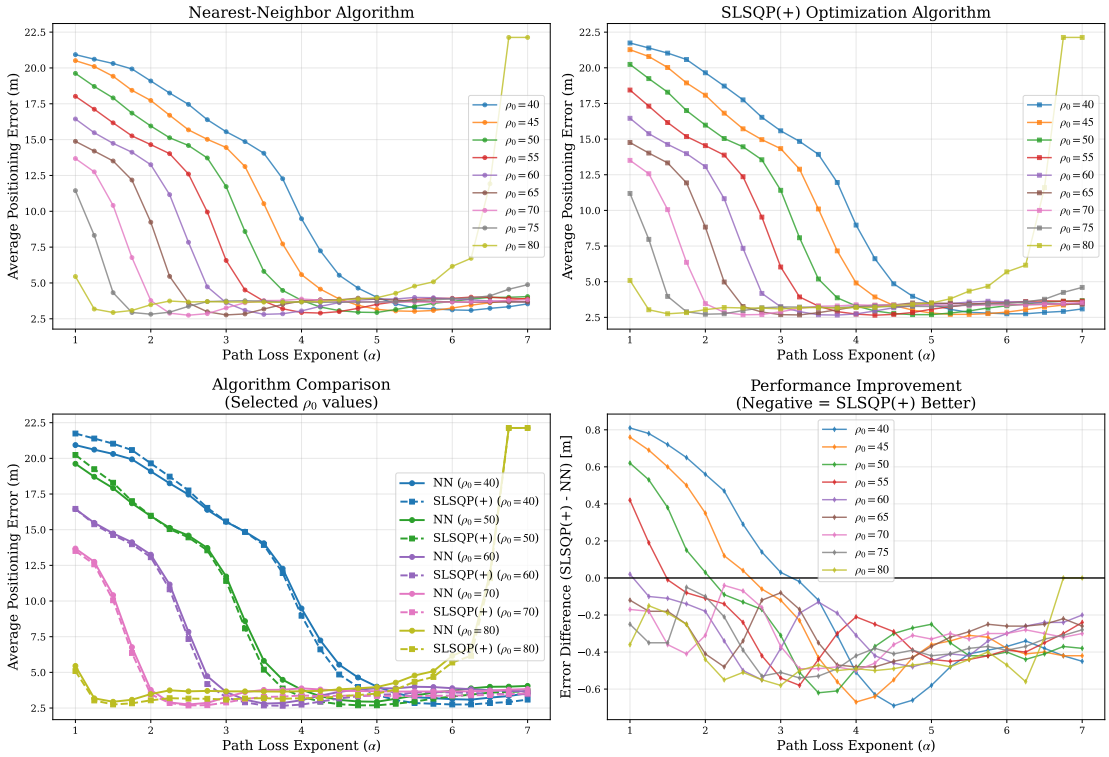

Figure 7 – Positioning Error Comparison: Nearest-Neighbor vs SLSQP Optimization by log-distance path loss parameters (ρ_0: 40-80 dB, steps of 5 dB); α_L: 1-7.

For extreme parameter combinations, such as $\alpha_L = 1.0$ and $40 \leq \rho_0 \leq 60$

(dB), the accuracy of both NN and SLSQP(+) exceeds 15 meters. As ρ_0 increases, positioning error decreases, allowing for a broader range of α_L. This supports the concept that a greater ρ_0 necessitates a smaller α_L to effectively represent potential signal loss.

Furthermore, with an APE below 3 meters being considered desirable, the SLSQP(+) consistently outperforms the reference model – NN across all scenarios. It is noteworthy that by providing the SLSQP algorithm with an initial location estimation from the NN, the optimization technique typically achieves enhanced position estimation, particularly with suitable combinations of ρ_0 and α_L.

5.3 Choice of model parameters and performance analysis

After analyzing APE behavior with the combination $\alpha_L - \rho_0$, the primary objective is to identify the parameter combinations that yield a more accurate system. Three key metrics are employed to evaluate RSSI diversity within the physical scenario. The first two metrics calculate pairwise distances for the expected RSSI vectors at candidate test points: Euclidean distance and Chebyshev distance. The remaining metric aggregates the pairwise distance data into a single indicator: mean distance. Figure 8 illustrates the behavior of the mean of the pairwise distances with the combinations (α_L, ρ_0). In addition, Table 4 and Figure 9 demonstrate the performance of the three metrics when integrated with SLSQP(+) compared to the pure NN and SLSQP(+) techniques.

For $40\ dB \leq \rho_0 \leq 65\ dB$, the Mean-Chebyshev approach using SLSQP(+)

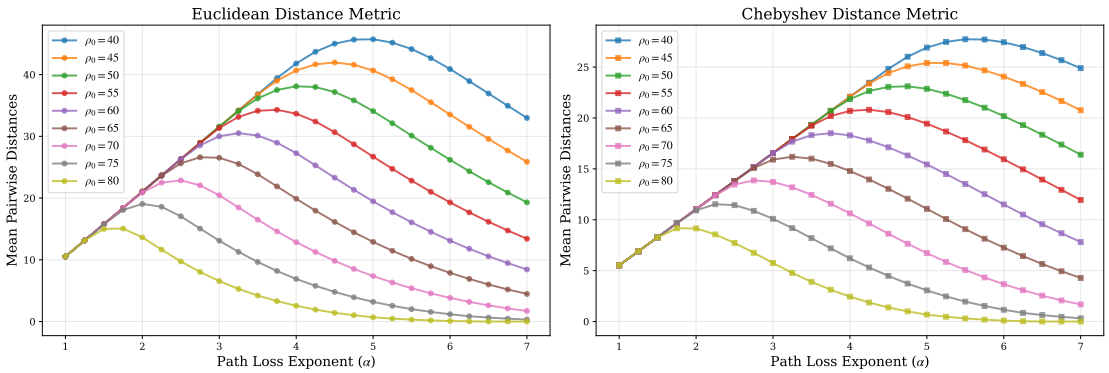

Figure 8 – Mean of the pairwise distances by log-distance path loss parameters (ρ_0: 40–80 dB, steps of 5 dB; α_L: 1–7).

yields the best optimization results, achieving a minimum APE of 2.65 meters. Conversely, for $\rho_0 > 65\ dB$, the Mean-Euclidean approach attains the best values for APE, minimizing at 2.67 meters for $\rho_0 = 67.5$ dB and $\alpha_L = 2.75$. However, for $\rho_0 \leq 65\ dB$, the Euclidean choice exhibits low performance concerning the APE, with values falling below 3 meters.

Overall, the Mean-Chebyshev choice emerges as a more robust alternative by leveraging superior performance in terms of positioning accuracy on average. The advantage of the Chebyshev metric over the Euclidean metric may stem from its focus on the maximum difference in the components of the RSSI vector, which renders it more reliable (Pu and You, 2018). In contrast, the Euclidean metric is more susceptible to capturing noise features. Consequently, the Chebyshev metric enhances dissimilarity among candidate points, thereby improving the measurement of signal diversity in indoor environments.

If one examines the results more closely, the difference between the proposed OPTIMAPS (defined as the combination of Mean-Chebyshev metrics and the SLSQP(+) algorithm) and the NN (used as the reference model) may not appear significant at first glance. However, a more accurate comparison must

Table 4 – Average positioning error (APE) by technique: for each given ρ_0, there is a most effective α_L associated.

Technique	ρ_0 (dB)	α_L	APE (m)	Technique	ρ_0 (dB)	α_L	APE (m)
	40.0	6.25	3.09		40.0	6.00	2.75
	42.5	5.75	3.03		42.5	5.75	2.70
	45.0	5.50	3.02		45.0	5.25	2.71
	47.5	5.00	2.99		47.5	5.25	2.69
	50.0	5.00	2.94		50.0	4.75	2.69
	52.5	4.50	2.92		52.5	4.50	2.67
	55.0	4.25	2.90		55.0	4.25	2.65
	57.5	3.75	2.86		57.5	4.00	2.65
Nearest-Neighbor	60.0	3.50	2.81	SLSQP(+)	60.0	3.75	2.66
	62.5	3.25	2.77		62.5	3.25	2.66
	65.0	3.00	2.77		65.0	3.25	2.67
	67.5	2.75	2.73		67.5	2.75	2.67
	70.0	2.50	2.75		70.0	2.50	2.68
	72.5	2.25	2.77		72.5	2.25	2.68
	75.0	2.00	2.82		75.0	2.00	2.72
	77.5	1.75	2.86		77.5	1.75	2.73
	80.0	1.50	2.94		80.0	1.50	2.75
	40.0	5.00	3.42		40.0	5.50	2.85
	42.5	4.75	3.38		42.5	5.25	2.84
	45.0	4.50	3.33		45.0	5.25	2.71
	47.5	4.25	3.32		47.5	5.00	2.71
	50.0	4.00	3.30		50.0	4.75	2.69
	52.5	4.00	2.92		52.5	4.50	2.67
	55.0	3.75	2.90		55.0	4.25	2.65
	57.5	3.50	2.89		57.5	4.00	2.65
Mean-Euclidean	60.0	3.25	2.90	Mean-Chebyshev	60.0	3.75	2.66
SLSQP(+)	62.5	3.00	2.86	SLSQP(+)	62.5	3.50	2.68
	65.0	2.75	2.87		65.0	3.25	2.67
	67.5	2.75	2.67		67.5	3.00	2.68
	70.0	2.50	2.68		70.0	2.75	2.70
	72.5	2.25	2.68		72.5	2.50	2.72
	75.0	2.00	2.72		75.0	2.25	2.75
	77.5	1.75	2.73		77.5	2.00	2.79
	80.0	1.75	2.83		80.0	1.75	2.83

consider the same log-distance parameters for both the NN and OPTIMAPS approaches. In this context, since OPTIMAPS employs the NN algorithm to establish the initial guess for location estimation, the parameter combinations serving as the reference are determined during the pairwise distance step, which is utilized for both the NN and SLSQP(+) estimation steps. Table 5 presents a

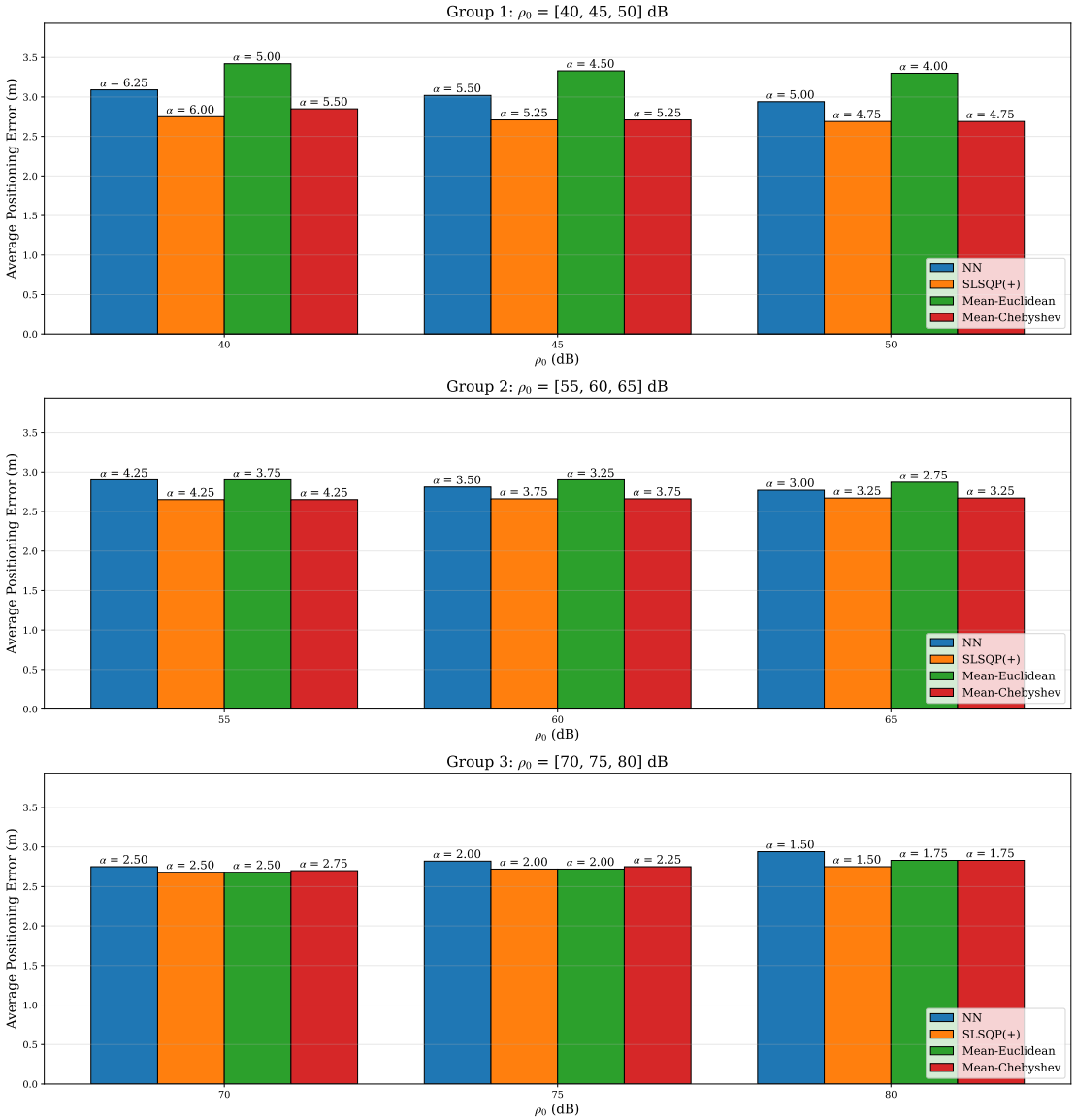

Figure 9 – Average positioning error in perspective: Nearest-Neighbor(NN), SLSQP(+), Mean with Euclidean and SLSQP(+) (Mean-Euclidean), and Mean with Chebyshev and SLSQP(+) (Mean-Chebyshev).

performance comparison in this regard. It is evident that OPTIMAPS enhances the NN approach in all scenarios, with percentage improvements ranging from 5.56% to 12.58%.

An alternative visualization of the results is provided by the cumulative

Table 5 – Performance comparison between NN (the reference model) and OPTIMAPS.

ρ_0 (dB)	α_L	APE (m)		Relative Improvement (%)
		NN	OPTIMAPS	
40.0	5.50	3.26	2.85	12.58
42.5	5.25	3.22	2.84	11.80
45.0	5.25	3.05	2.71	11.14
47.5	5.00	2.99	2.71	9.36
50.0	4.75	2.96	2.69	9.12
52.5	4.50	2.92	2.67	8.56
55.0	4.25	2.90	2.65	11.07
57.5	4.00	2.87	2.65	7.67
60.0	3.75	2.85	2.66	6.67
62.5	3.50	2.86	2.68	9.46
65.0	3.25	2.84	2.67	5.99
67.5	3.00	2.85	2.68	5.96
70.0	2.75	2.86	2.70	5.59
72.5	2.50	2.88	2.72	5.56
75.0	2.25	2.96	2.75	7.09
77.5	2.00	3.02	2.79	7.61
80.0	1.75	3.08	2.83	8.11

error distribution relative to the best performances. Figure 10 displays eight scenarios for the parameter ρ_0 that yield the lowest APEs. Notably, the ground-truth value of ρ_0 lies between 55 and 60 dB (Assayag et al., 2024b,2), which corresponds to the best results achieved. The findings support the notion that the combination of Mean and Chebyshev metrics outperforms other candidate combinations throughout most of the error distribution, establishing it as the most effective choice for determining the $\alpha_L - \rho_0$ combination that results in a lower APE.

Although the APE results obtained through the Mean-Chebyshev method combined with SLSQP(+) optimization are closer to the most effective values,

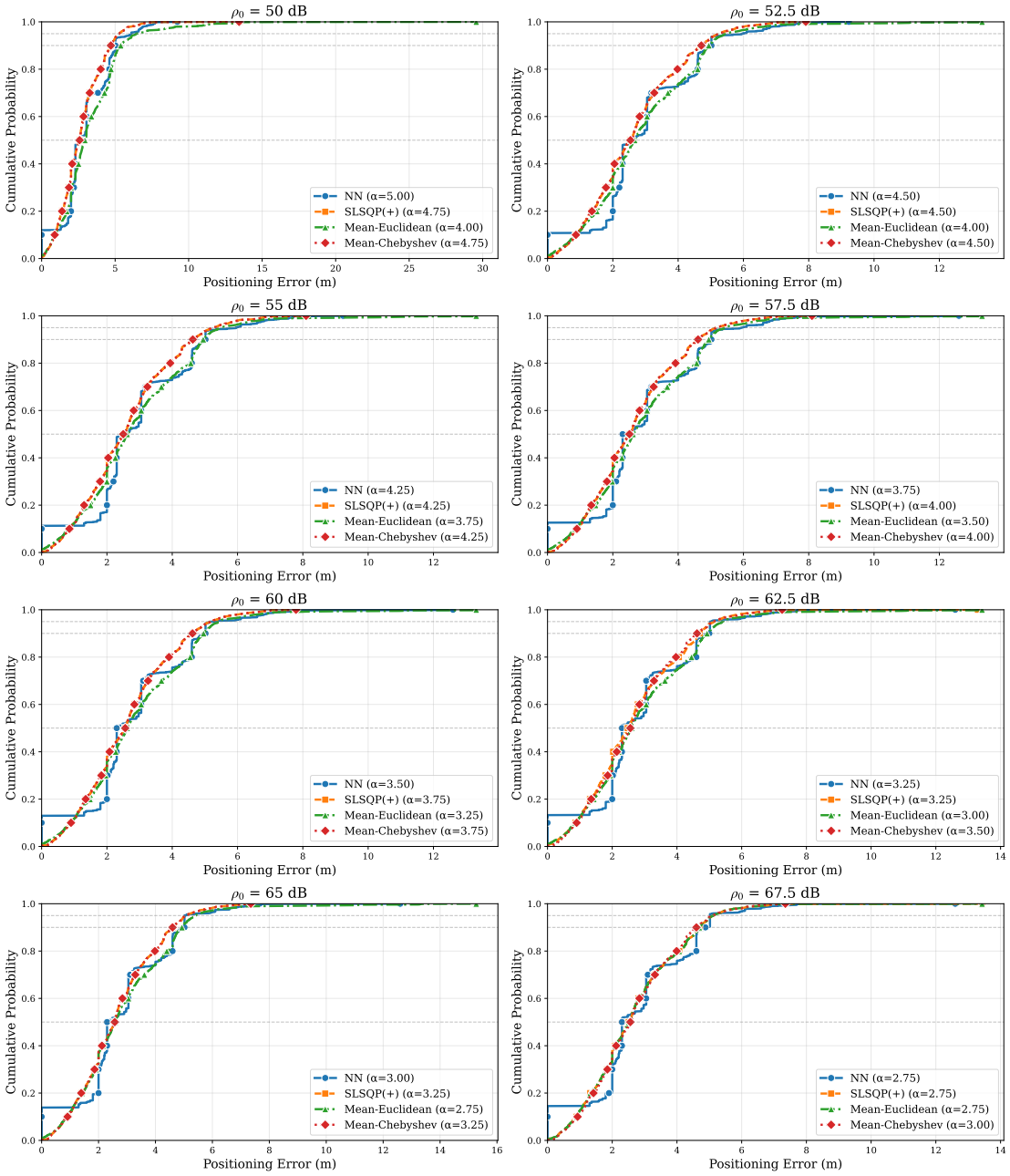

Figure 10 – Cumulative error distribution by algorithm and calculated parameters for selected ρ_0 values.

this method does not consistently yield the minimum possible APE. Conversely, to conduct a more comprehensive and robust analysis, a statistical evaluation

of the error distribution for both Mean-Chebyshev and SLSQP(+) methods is essential.

Given the error vectors displayed by the two aforementioned methods, it is possible to compare the differences between their means. Since the estimations correspond to the same testing points for both methods, a paired test is appropriate. For normally distributed paired differences, paired t-tests compare mean positioning errors. For non-normally distributed differences, Wilcoxon signed-rank tests compare median differences. Both approaches assess algorithm superiority but target different location parameters. In this specific case, to assess the normality assumption of the error differences, the Shapiro-Wilk and Kolmogorov-Smirnov tests were employed. The results indicated that the normality assumption was violated for all possible ρ_0 values, necessitating the use of the non-parametric Wilcoxon signed-rank test for statistical analysis. Table 6 presents the obtained results.

The statistical tests indicate that both estimation methods perform similarly regarding their medians; however, significant differences are noted for extreme values of ρ_0. Additionally, when differences are observed, the practical divergences between means are considered small based on the quantification of effect size. Figure 11 illustrates the presented results.

The four-panel analysis indicates that both algorithms demonstrate nearly identical mean positioning errors, characterized by a U-shaped performance curve, with optimal accuracy observed at ρ_0 = 52.5-55 dB. Although large sample sizes (n $\approx$ 1420) permit the detection of statistically significant differences at various noise levels, the practical differences are minimal to be deemed significant.

In this context, the effect size analysis provides compelling evidence for

Table 6 – Wilcoxon-signed rank testing and effect size results for median and mean comparisons between Mean-Chebyshev and SLSQP(+) methods ($\alpha = 0.05$).

ρ_0 (dB)	Mean Difference (m)	p-value	Significant	Effect Size (Cohen d)	Interpretation
40.0	0.11	6.67×10^{-4}	Yes	0.090	Small Effect
42.5	0.14	9.39×10^{-4}	Yes	0.087	Small Effect
45.0	0.00	–	No	–	–
47.5	0.01	8.13×10^{-1}	No	0.006	Small Effect
50.0	0.00	–	No	–	–
52.5	0.00	–	No	–	–
55.0	0.00	–	No	–	–
57.5	0.00	–	No	–	–
60.0	0.00	–	No	–	–
62.5	0.02	6.64×10^{-1}	No	0.012	Small Effect
65.0	0.00	–	No	–	–
67.5	0.01	8.71×10^{-1}	No	0.004	Small Effect
70.0	0.02	1.26×10^{-1}	No	0.041	Small Effect
72.5	0.04	6.44×10^{-3}	Yes	0.072	Small Effect
75.0	0.03	3.80×10^{-3}	Yes	0.077	Small Effect
77.5	0.07	4.51×10^{-5}	Yes	0.108	Small Effect
80.0	0.08	1.80×10^{-7}	Yes	0.138	Small Effect

algorithmic equivalence, with Cohen's d values consistently remaining below the 0.2 threshold, indicative of small effects across all noise conditions. The p-value visualization illustrates the disconnect between statistical significance and practical relevance, where even highly significant results ($p < 10^{-6}$) correspond to negligible effect sizes. The paired differences analysis reveals bidirectional performance variations without a systematic advantage for either algorithm, while the constrained zero effect sizes for specific ρ_0 values (45, 50, 52.5, 55, 57.5, 60, 65 dB) reinforce the conclusion of equivalence. When converted to practical positioning units, the mean absolute differences amount to mere centimeters, representing a negligible impact for real-world applications.

Therefore, OPTIMAPS, enhanced by the Mean-Chebyshev approach, is capable of delivering highly effective performance in nearly all scenarios, exhibiting minimal variations in extreme values of ρ_0. This characteristic makes

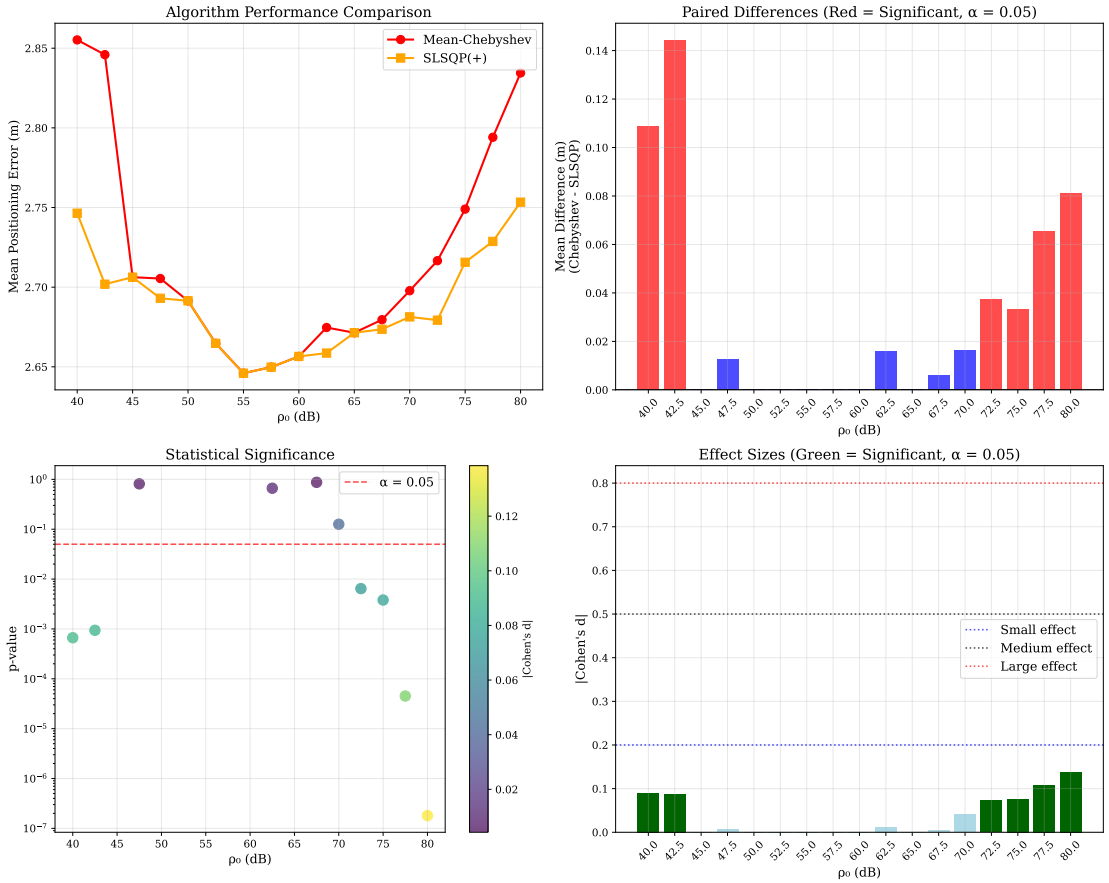

Figure 11 – Statistical performance comparison between Mean-Chebyshev (with SLSQP(+)) and pure SLSQP(+) for each ρ_0.

it a suitable and robust choice for indoor solutions that do not require a prior training dataset, relying solely on the knowledge of the environment's geometry and the locations of the deployed access points.

5.4 Error analysis by room type

This section provides a thorough per-label error analysis, spatial aggregation by room type, and examines correlation patterns between baseline algorithm

performance and optimization advantages.

5.4.1 Data preprocessing

The analysis utilized the complete test measurement dataset comprising 148 unique location labels distributed across 10 devices.

All analysis focused on the parameter configuration $\rho_0 = 55.0\,\text{dB}$ and $\alpha_L = 4.25$, enabling direct algorithmic comparison under identical conditions.

5.4.2 Algorithm performance comparison

Table 7 presents the comprehensive performance metrics for both algorithms analyzed.

Algorithm	**Mean Error (m)**	**Standard Deviation (m)**
Nearest Neighbor (NN)	2.90	1.67
SLSQP(+)	2.65	1.18
Improvement	**0.25**	**—**
Relative Improvement	**8.62%**	**—**

Table 7 – Algorithm performance summary across 148 unique labels using parameters $\rho_0 = 55.0\,\text{dB}$, $\alpha_L = 4.25$.

The SLSQP(+) algorithm demonstrated superior average performance with a mean positioning error reduction of 8.62% compared to the Nearest Neighbor baseline. However, this aggregate improvement masked significant spatial heterogeneity, with individual label improvements ranging from substantial gains exceeding 30% to locations where NN maintained superior performance.

Analysis of the error distributions revealed that SLSQP(+) achieved lower error variance (1.18 m vs 1.67 m), indicating more consistent performance across diverse spatial contexts. The comparison showed SLSQP(+) outperforming NN in approximately 76% of locations, with the remaining 24% favoring the simpler NN approach.

5.4.3 Spatial mapping and room-based aggregation

The 148 unique labels were systematically mapped to spatial areas using a deterministic prefix-based classification system, as detailed in Table 8.

Label Prefix	**Area Classification**	**Area Type**	**Test Points**
16	Room 1	Enclosed	15
41	Room 2	Enclosed	12
57	Room 3	Enclosed	8
40	Room 4	Enclosed	9
53	Room 5	Enclosed	12
39	Room 6	Enclosed	9
47	Room 7	Enclosed	12
51	Room 8	Enclosed	12
45	Room 9	Enclosed	12
43	Room 10	Enclosed	12
46	Room 11	Enclosed	12
63	Hall 1	Open	8
61	Hall 2	Open	7
62	Hall 3	Open	8
Total			**148**

Table 8 – Spatial area classification system and label distribution. Areas are categorized as enclosed rooms or open halls based on architectural characteristics.

This classification enabled environment-specific performance analysis,

revealing fundamental differences in algorithm effectiveness between enclosed and open spaces.

5.4.4 Room-level performance analysis

Aggregation by spatial area revealed distinct algorithmic performance patterns across environment types. Table 9 presents the comprehensive room-level analysis.

Area	Type	Labels	NN Error (m)	SLSQP(+) Error (m)	Improvement (%)
Hall 1	Open	8	1.54	1.75	−13.64
Hall 3	Open	8	1.63	1.94	−19.02
Hall 2	Open	7	2.03	1.97	2.96
Room 5	Enclosed	12	2.22	2.68	−20.72
Room 4	Enclosed	15	2.38	2.45	−2.94
Room 6	Enclosed	9	2.68	2.78	−3.73
Room 1	Enclosed	9	2.93	2.83	3.41
Room 2	Enclosed	12	3.05	3.08	−0.98
Room 7	Enclosed	12	3.11	3.28	−5.47
Room 9	Enclosed	12	3.20	2.23	30.31
Room 10	Enclosed	12	3.44	2.10	38.95
Room 3	Enclosed	8	3.46	3.02	12.72
Room 11	Enclosed	12	3.62	2.81	22.38
Room 8	Enclosed	12	4.05	3.39	16.30

Table 9 – Room-level algorithm performance comparison ordered by NN baseline error (ascending). Negative improvement percentages indicate areas where NN outperforms SLSQP(+).

The analysis revealed a clear algorithmic dichotomy based on spatial characteristics:

1. Open areas consistently demonstrated lower baseline positioning errors (mean NN error: 1.73 m) but showed mixed or negative responses to SLSQP(+) optimization. The average improvement for halls was -8.85%, indicating that the optimization approach introduced unnecessary complexity in these simpler environments;

2. Enclosed rooms exhibited higher baseline positioning challenges (mean NN error: 3.10 m) but demonstrated substantial benefits from SLSQP(+) optimization, with an average improvement of 10.22%. The most dramatic improvements occurred in the most challenging environments:

   - Room 10: 38.95% improvement (3.44 m $\rightarrow$ 2.10 m)
   - Room 9: 30.31% improvement (3.20 m $\rightarrow$ 2.23 m)
   - Room 11: 22.38% improvement (3.62 m $\rightarrow$ 2.81 m)
   - Room 8: 16.30% improvement (4.05 m $\rightarrow$ 3.39 m)
   - Room 3: 12.72% improvement (3.46 m $\rightarrow$ 3.02 m)

5.4.5 Correlation analysis: baseline performance versus optimization benefit

A critical finding emerged from the systematic analysis of the relationship between NN baseline performance and SLSQP(+) optimization benefit. The statistical analysis revealed a strong positive correlation between NN average error and SLSQP(+) improvement percentage.

Statistical Measure	**Value**
Pearson Correlation Coefficient	0.754
Two-sided p-value	0.0018
Statistical Significance	Yes ($p < 0.01$)
Correlation Strength	Strong Positive
Sample Size	14 areas
Confidence Level	99%

Table 10 – Correlation analysis between NN baseline error and SLSQP(+) improvement percentage.

This analysis confirmed that areas with the most significant positioning challenges (indicated by poor NN performance) experienced substantial benefits from optimization. In contrast, regions that exhibited excellent NN performance showed degradation when subjected to the more complex SLSQP(+) approach. In other words, the SLSQP algorithm exhibits improved performance in environments characterized by high RSSI variability, as the NN, which relies exclusively on a fixed-parameter log-distance model, is unable to deliver accurate location estimations.

6 Conclusions

This thesis introduced **OPTIMAPS**, an innovative, site-survey-free indoor positioning system (IPS) that leverages a log-distance path loss model combined with Sequential Least Squares Programming (SLSQP) optimization to achieve robust, real-time positioning using Bluetooth Low Energy (BLE) technology. The proposed system directly addresses critical barriers faced by traditional IPS solutions — most notably, the heavy reliance on labor-intensive and time-consuming offline fingerprinting phases. By eliminating the requirement for pre-deployment RSSI sampling and relying solely on known anchor node positions and scenario geometry, OPTIMAPS establishes a new, practical paradigm for scalable and accurate indoor localization.

The system was extensively validated in a demanding large-scale real-world environment: a 720 m^2 school facility composed of 11 classrooms, 3 halls, 15 BLE access points, and 148 well-labeled testing locations. The use of multiple BLE receiver devices and thousands of RSSI samples provided a rigorous and reproducible evaluation. OPTIMAPS produced a mean positioning error (APE) of 2.65 meters — comparable to or exceeding the accuracy of other state-of-the-art solutions in the literature, especially among those requiring no site-survey. Crucially, assessment through anchor density-normalized error further demonstrated that OPTIMAPS outperforms nearly all low-training systems in

terms of precision and efficiency.

A major contribution of this thesis is the design of a two-stage pipeline that separates offline scenario-dependent modeling from adaptive online estimation. In the offline phase, rather than collecting RSSI fingerprints, suitable parameters for the log-distance path loss model are discovered via geometric analysis and maximization of signal vector diversity. The Chebyshev metric was empirically and theoretically shown to provide superior quantification of RSSI dissimilarity, producing the lowest average errors; the combination of the Mean-Chebyshev metric with SLSQP yielded the minimum APE values throughout the tested parameter space. This approach both minimizes initial deployment effort and provides a principled framework for parameter selection even in highly dynamic or previously unmapped environments.

For online estimation, OPTIMAPS employs a nearest-neighbor (NN) strategy for the initial position guess, which is then refined using SLSQP optimization within a constraint circle. This pipeline integrates high computational efficiency (the NN step is $\mathcal{O}(n)$; SLSQP is $\mathcal{O}(k)$, where n — the number of reference points — and k — the number of iterations — increase proportionally to the testbed area, yet remain fast for typical IPS dimensions) with strong local adaptation to real-time signal anomalies. Experimental results, which include comprehensive statistical analyses (paired testing, effect size, cumulative error distributions) confirm that SLSQP(+) consistently outperforms both NN baselines and alternative metaheuristic algorithms, reducing APE by up to 12.6% compared to NN and achieving inference in just 30 ms per estimate.

A granular, spatially-structured analysis revealed essential nuances: while open areas (halls) already exhibit low baseline errors with little benefit from

optimization, enclosed and multipath-rich rooms experience dramatically improved positioning accuracy—with error reductions exceeding 30% in the most challenging locations, and a strong positive correlation between optimization gain and baseline difficulty. As a result, OPTIMAPS not only matches or exceeds the performance of a simpler reference algorithm as NN in easy scenarios, but also delivers its greatest improvements precisely where traditional methods are least reliable.

In direct comparison with recent literature — including systems tested on the identical public BLE benchmark dataset — OPTIMAPS achieves competitive or superior accuracy relative to survey-free, model-based solutions based on population optimization, while requiring substantially less computational and hardware resources. Unlike fingerprinting or deep learning alternatives, OPTIMAPS enables deployment in fully dynamic environments, with zero site-survey overhead and superior scalability.

In summary, the results of this thesis demonstrate that practical, accurate, and computationally efficient indoor positioning systems can be achieved without the need for site surveys or labor-intensive training. By integrating geometric modeling, signal diversity optimization, and advanced constrained estimation, OPTIMAPS bridges the performance gap with traditional methods, thereby facilitating practical, scalable, and robust indoor positioning system solutions for a wide array of real-world applications.

6.1 Limitations And Future Work

Despite the substantial advancements presented in this thesis, OPTIMAPS leaves several fundamental and challenging research opportunities available for future exploration. Addressing these opportunities could significantly enhance the practicality, robustness, and scientific impact of survey-free indoor positioning.

1. **Adaptation to dynamically changing access point infrastructures**

   This work assumes that the positions of all BLE APs are fixed and known a priori. However, real-world scenarios — ranging from large-scale commercial deployments to ad-hoc or "bring-your-own-device" networks — are characterized by the dynamic (re-)positioning of APs, temporary outages, and heterogeneous anchor geometries. As a challenging direction, future work should develop algorithms for:

   - *Online AP self-localization*: enabling the system to estimate or track anchor positions in real time, using only RSSI measurements and device movement data, without external calibration.
   - *Robust optimization under baseline uncertainty*: formulating the entire location pipeline to tolerate, adapt, and estimate under time-varying anchor location priors (e.g., probabilistic anchors).

2. **Integration of rich, multi-modal signal and physical environment modeling**

   The log-distance path loss model, while computationally efficient, does not account for the highly non-linear and heterogeneous nature of indoor radio propagation affected by walls, furniture, human mobility, and multipath effects. Advancing beyond this limitation requires:

- *Physics-informed machine learning*: Leveraging ray tracing, CAD-based layout data, or hybrid deep learning models that take as input the architectural blueprint and known material properties to predict RSSI or other signal features at arbitrary positions.
- *Transfer learning across buildings*: Enabling rapid adaptation of the model to entirely new settings using minimal data, possibly by meta-learning or domain adaptation.
- *Online adaptation*: Real-time updating of the signal propagation model based on ongoing measurements, enabling robust operation through environmental changes (e.g., moving people, furniture).

3. **Heterogeneous and uncalibrated device ecosystem**

   The current study uses BLE devices with relatively homogeneous hardware characteristics. In real deployments, variability in transmit power, antenna directionality, and hardware calibration across device manufacturers and types can severely degrade accuracy. Addressing these requires:

   - *Cross-device calibration-free methods*: Designing algorithms that can infer and adapt to unknown device-specific biases in RSSI readings — possibly through unsupervised calibration or federated learning.
   - *Device identification and anomaly detection*: Automatically recognizing and compensating for rogue or malfunctioning devices whose measurements do not fit global trends.

4. **Toward room-level semantics and contextual awareness**

   Most current approaches, including OPTIMAPS, focus purely on (x,y) position. Important open questions include:

- *Integration with semantic building maps*: How can signals be mapped not just to coordinates but to meaningful zones or activities?
- *Fusion with additional modalities*: Combining BLE-based estimation with vision, inertial sensors, or crowd-sourced floor plans for human-centric applications such as personalized navigation or asset tracking.

5. **Open benchmarking, privacy, and real-world long-term deployment**

   True evaluation of site-survey-free systems demands longitudinal, open, and privacy-preserving deployments. Future work should:

   - *Benchmark across diverse real-world environments* (malls, hospitals, airports), with long-term temporal and seasonal effects.
   - *Design privacy-preserving data collection and inference protocols* ensuring user location and device data cannot be exploited or tracked by adversaries.

In summary, this thesis represents a significant advancement toward survey-free, practical indoor positioning. However, addressing the outstanding challenges will necessitate progress in unsupervised spatial learning, robust multi-modal modeling, distributed computation, and privacy-preserving protocols — areas that present opportunities for impactful future research.

6.2 Published Papers

As a result of the research conducted over the past few years, three papers have been published:

1. Pinto, B.H.O.U.V.; de Oliveira, H.A.B.F.; Souto, E.J.P. Factor Optimization for the Design of Indoor Positioning Systems Using a Probability-Based Algorithm. *J. Sens. Actuator Netw*. 2021, 10, 16. (Pinto et al., 2021b)

2. Pinto, B.; Barreto, R.; Souto, E.; de Oliveira, H. Robust RSSI-based Indoor Positioning System using K-means Clustering and Bayesian Estimation. *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24462-24470, 1 Nov.1, 2021. (Pinto et al., 2021a)

3. Pinto, B., Oliveira, H. Online RSSI selection strategy for indoor positioning in low-effort training scenarios. Computing 106, 2059–2077 (2024). (Pinto and Oliveira, 2024)

The first paper, published during the Master's program, focuses on identifying the factors that significantly influence the accuracy of Bayesian estimation and proposes a simulation approach for designing an IPS utilizing probability-based algorithms. This ultimately serves as the foundation for transforming the offline training phase into an "artificial" scenario for this thesis.

The second paper, initially submitted during the Master's program and subsequently refined at the Doctorate level, proposes a new IPS that uses k-means clustering to classify scenarios into various log-distance path loss models. It also leverages Bayesian inference for position estimation. In this context, the first step towards minimizing the training effort is undertaken, with careful consideration of the method's execution time and scalability.

The third paper, submitted and published during the Doctorate, introduces a novel IPS that selects the strongest RSSI from each AP and estimates the position using the ordinary least squares method. The effort required during the

training phase is significantly reduced while maintaining reasonable levels of accuracy. In this work, the strongest RSSI utilized to minimize the average error serves as the foundation for considering the Chebyshev metric in the presented thesis, which ultimately emerges as the most effective choice for OPTIMAPS among the options listed.

Collectively, the aforementioned publications contributed key concepts and methodologies that underpin this thesis, enabling the development of an indoor positioning system that eliminates the need for fingerprint collection during the offline phase while achieving high efficiency in both accuracy and execution time.

Bibliography

Abed, F. A., Hamza, Z. A., and Mosleh, M. F. (2022). Indoor positioning system based on wi-fi and bluetooth low energy. In *2022 8th International Engineering Conference on Sustainable Technology and Development (IEC)*, pages 136–141.

Ali, M. U., Hur, S., and Park, Y. (2019). Wi-fi-based effortless indoor positioning system using iot sensors. *Sensors*, 19(7).

Amr, M. N., Elattar, H. M., Azeem, M. H. A. E., and Badawy, H. E. (2021). An enhanced indoor positioning technique based on a novel received signal strength indicator distance prediction and correction model. *Sensors*, 21:719.

Assayag, Y., Oliveira, H., Lima, M., Junior, J., Preste, M., Guimarães, L., and Souto, E. (2024a). Indoor environment dataset based on rssi collected with bluetooth devices. *Data in Brief*, 55:110692.

Assayag, Y., Oliveira, H., Souto, E., Barreto, R., and Pazzi, R. (2023). Adaptive path loss model for ble indoor positioning system. *IEEE Internet of Things Journal*, 10(14):12898–12907.

Assayag, Y., Oliveira, H., Souto, E., Barreto, R., and Pazzi, R. (2024b). A model-based ble indoor positioning system using particle swarm optimization. *IEEE Sensors Journal*, 24(5):6898–6908.

Assayag, Y., Souto, E., Barreto, R., Carvalho, M., Pazzi, R., and Fernandes, H. (2024c). Efficient exploration of indoor localization using genetic algorithm

and signal propagation model. *Computing*, 107(1).

Bahl, P. and Padmanabhan, V. N. (2000). Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 2, pages 775–784 vol.2.

Beigi, K. and Shah-Mansouri, H. (2024). An intelligent indoor positioning algorithm based on wi-fi and bluetooth low energy. In *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, page 1–6. IEEE.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition.

Conn, A. R., Gould, N. I., and Toint, P. L. (2000). *Trust region methods*. Siam.

Daniş, F. S., Ersoy, C., and Cemgil, A. T. (2023). Probabilistic indoor tracking of bluetooth low-energy beacons. *Performance Evaluation*, 162:102374.

Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Faragher, R. and Harle, R. (2012). An analysis of the accuracy of bluetooth low energy for indoor positioning applications. *Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+)*, pages 201–210.

Ho, Y. H. and Chan, H. C. (2020). Decentralized adaptive indoor positioning protocol using bluetooth low energy. *Computer Communications*, 159:231–244.

Jang, B. and Kim, H. (2019). Indoor positioning technologies without offline fingerprinting map: A survey. *IEEE Communications Surveys & Tutorials*, 21(1):508–525.

Kaemarungsi, K. and Krishnamurthy, P. (2004). Modeling of indoor positioning

systems based on location fingerprinting. *IEEE INFOCOM 2004*, 2:1012–1022.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

Kraft, D. (1988). A software package for sequential quadratic programming. *DFVLR-FB*, 88(28):1–69.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863.

Liu, F., Liu, J., Yin, Y., Wang, W., Hu, D., Chen, P., and Niu, Q. (2020). Survey on wifi-based indoor positioning techniques. *IET Communications*, 14(9):1372–1383.

Man, D., Bing, L., and Lv, J. (2020). Indoor localization algorithm based on attribute-independent weighted naive bayesian. *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies*.

Massey, Jr., F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78.

Morgan, A. A. (2024). On the accuracy of ble indoor localization systems: An assessment survey. *Computers and Electrical Engineering*, 118:109455.

Nikodem, M. and Szeliński, P. (2021). Channel diversity for indoor localization using bluetooth low energy and extended advertisements. *IEEE Access*, 9:169261–169269.

Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Pinto, B. (2025). Site survey-free rssi-based indoor positioning system. https://github.com/brauliopinto/optimaps-indoor-positioning.git. Accessed: 31-07-2025.

Pinto, B., Barreto, R., Souto, E., and Oliveira, H. (2021a). Robust rssi-based indoor positioning system using k-means clustering and bayesian estimation. *IEEE Sensors Journal*, 21(21):24462–24470.

Pinto, B. and Oliveira, H. (2024). Online rssi selection strategy for indoor positioning in low-effort training scenarios. *Computing*, 106(6):2059–2077.

Pinto, B. H. O. U. V., de Oliveira, H. A. B. F., and Souto, E. J. P. (2021b). Factor optimization for the design of indoor positioning systems using a probability-based algorithm. *Journal of Sensor and Actuator Networks*, 10(1).

Pu, Y.-C. and You, P.-C. (2018). Indoor positioning system based on ble location fingerprinting with classification approach. *Applied Mathematical Modelling*, 62:654–663.

Rappaport, T. S. (2002). *Wireless Communications: Principles and Practice*, volume 2. Prentice-Hall, Upper Saddle River, N.J, 2nd edition.

Seidel, S. and Rappaport, T. (1992). 914 mhz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40(2):207–217.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11:341–359.

Student (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

Subedi, S. and Pyun, J.-Y. (2020). A survey of smartphone-based indoor positioning system using rf-based wireless technologies. *Sensors*, 20(24).

Subhan, F., Khan, A., Saleem, S., Ahmed, S., Imran, M., Asghar, Z., and Bangash, J. I. (2019). Experimental analysis of received signals strength in bluetooth low energy (ble) and its effect on distance and position estimation. *Transactions on Emerging Telecommunications Technologies*, 33(2).

Szyc, K., Nikodem, M., and Zdunek, M. (2023). Bluetooth low energy indoor localization for large industrial areas and limited infrastructure. *Ad Hoc Networks*, 139:103024.

Torres-Sospedra, J., Montoliu, R., Trilles, S., Belmonte, O., and Huerta, J. (2015). Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, 42(23):9263–9278.

Vallet García, J. M. (2020). Characterization of the log-normal model for received signal strength measurements in real wireless sensor networks. *Journal of Sensor and Actuator Networks*, 9(1).

Virtanen, P., Gommers, R., and Oliphant *et al.*, T. E. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272.

Wang, W., Qing-shan, Z., Wang, Z., Zhao, X., and Yanfang, Y. (2022). Research on indoor positioning algorithm based on saga-bp neural network. *IEEE Sensors Journal*, 22:3736–3744.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Wu, C., Wong, I.-C., Wang, Y., Ke, W., and Yang, X. (2024). Experimental study of bluetooth indoor positioning using rss and deep learning algorithms. *Mathematics*, 12(9):1386.

Xiao, K., Hao, F., Zhang, W., Li, N., and Wang, Y. (2024). Research and implementation of indoor positioning algorithm based on bluetooth 5.1 aoa and aod.

Sensors, 24(14).

Xie, S., Yu, X., Guo, Z., Zhu, M., and Han, Y. (2023). Multi-output regression indoor localization algorithm based on hybrid grey wolf particle swarm optimization. *Applied Sciences*, 13(22).

Youssef, M. and Agrawala, A. (2003). Wlan location determination via clustering and probability distributions. *IEEE PerCom*, pages 143–150.

Zafari, F., Gkelias, A., and Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3):2568–2599.

Zerzucha, P. and Walczak, B. (2012). Concept of (dis)similarity in data analysis. *TrAC Trends in Analytical Chemistry*, 38:116–128.

Zhao, Y., Wang, T., Miao, Q., Jin, Y., and Wang, R. (2024). Research on indoor and outdoor positioning switching algorithm based on improved pso-bp. *Measurement Science and Technology*, 35:086313.

Zhuang, Y., Zhang, C., Huai, J., Li, Y., Chen, L., and Chen, R. (2022). Bluetooth localization technology: Principles, applications, and future trends. *IEEE Internet of Things Journal*, 9(23):23506–23524.