



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

GUSTAVO PIMENTEL BITTENCOURT

**UM ESTUDO SOBRE A GERAÇÃO DE GRIDMAPS IRREGULARES PARA
LOCALIZAÇÃO DE ESTAÇÕES MÓVEIS EM AMBIENTES SEM FIO**

Recife
2020

GUSTAVO PIMENTEL BITTENCOURT

**UM ESTUDO SOBRE A GERAÇÃO DE GRIDMAPS IRREGULARES PARA
LOCALIZAÇÃO DE ESTAÇÕES MÓVEIS EM AMBIENTES SEM FIO**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Redes de Computadores e Sistemas Distribuídos

Orientador: Daniel Carvalho da Cunha

Recife

2020

Catálogo na fonte
Bibliotecária Nataly Soares Leite Moro, CRB15-861

B624e Bittencourt, Gustavo Pimentel
Um estudo sobre a geração de *gridmaps* irregulares para localização de estações móveis em ambientes sem fio / Gustavo Pimentel Bittencourt. – 2020. 75 f.: il., fig., tab.

Orientador: Daniel Carvalho da Cunha.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2020.
Inclui referências.

1. Redes de computadores e sistemas distribuídos. 2. *Location-Based Services* (LBS). 3. *Irregular gridmap*. 4. *Received Signal Strength* (RSS). I. Cunha, Daniel Carvalho da (orientador). II. Título

004.6 CDD (23. ed.) UFPE - CCEN 2021 – 170

GUSTAVO PIMENTEL BITTENCOURT

**“UM ESTUDO SOBRE A GERAÇÃO DE GRIDMAPS IRREGULARES PARA
LOCALIZAÇÃO DE ESTAÇÕES MÓVEIS EM AMBIENTES SEM FIO”**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 27 de novembro de 2020.

BANCA EXAMINADORA

Prof. Dr. Divanilson Rodrigo de Sousa Campelo
Centro de Informática / UFPE

Prof. Dr. Ugo Silva Dias
Departamento de Engenharia Elétrica / UnB

Prof. Dr. Daniel Carvalho da Cunha
Centro de Informática / UFPE
(Orientador)

À minha família, por me permitir conhecer o amor incondicional.

AGRADECIMENTOS

Agradeço à universidade a oportunidade de desenvolver minhas faculdades intelectuais. Agradeço aos professores todos os ensinamentos que me foram transmitidos, vocês foram essenciais nesta jornada. Sou grato, em particular, ao meu orientador, cujo apoio e incentivo foram indispensáveis para conclusão deste trabalho. Agradeço também a minha noiva, seu apoio, afeto e empatia foram fundamentais para que pudéssemos alcançar mais este objetivo. Agradeço, sobretudo, a minha família, sempre me apoiando em todas as minhas decisões, o amor de vocês me fortalece, enriquecendo o significado da minha existência. Muito obrigado.

"The structure of language determines not only thought, but reality itself."
(CHOMSKY; ANSHEN, 1986, p. 196)

RESUMO

A melhoria contínua e o aprimoramento dos sistemas para localização de estações móveis em ambientes sem fio constitui um dos principais desafios técnicos associados a ascensão do conceito de cidades inteligentes. Entre as técnicas de localização *outdoor* disponíveis, aquelas baseadas em *fingerprinting* de rádio frequência (*RF fingerprinting*) são capazes de fornecer uma boa acurácia para estimativa de localização, agregado a um baixo custo energético e de implementação. Em geral, seu funcionamento consiste em dividir uma área de cobertura em uma malha quadricular (*gridmap*) e definir um vetor com características de RF para cada célula dessa malha. O objetivo final é garantir que cada célula dessa malha possa ser unicamente identificada por meio de um vetor de características de RF que lhe foi atribuído. A principal característica dessa malha quadricular é que as células se encontram uniformemente espaçadas, além disso, a quantidade de células presentes em um *gridmap* determina a complexidade computacional associada ao modelo de localização. Recentemente, alguns estudos introduziram um conceito de *gridmap* que privilegia a distribuição de células em determinada região em detrimento de outras. Quando comparado aos *gridmaps* de malha quadricular, essa distribuição irregular das células na região de cobertura acarretou em uma diminuição da complexidade computacional necessária para assegurar que o erro associado a estimativa de localização não ultrapasse um determinado limite. Esse limite máximo para o erro é tipicamente imposto por órgãos reguladores, como ocorre com a *US Federal Communications Commission* (FCC), que impõe 50 metros como o erro máximo para chamadas de emergência. Diferentemente dos trabalhos anteriores, esse estudo propõe uma generalização sobre alguns dos conceitos que definem um *gridmap* irregular, introduzindo novas técnicas e algoritmos para automatizar seu processo de geração. Para validar a influência dos algoritmos de geração sobre a acurácia do sistema de localização, foram realizados experimentos utilizando o módulo LENA LTE da ferramenta *Network Simulator 3* (NS-3). Em seguida, foi efetuada uma avaliação com 4.617 medições RF, obtidas na cidade de Recife/PE para o padrão de telefonia móvel LTE com portadora na faixa de 1,8 GHz. Os resultados indicam que, quando confrontado com o *gridmap* de malha quadricular, o *gridmap* irregular proporcionou uma redução significativa na complexidade computacional necessária para satisfazer as exigências atuais da FCC para o erro de localização horizontal.

Palavras-chave: *Location-Based Services* (LBS); *irregular gridmap*; *RF fingerprinting*; *Received Signal Strength* (RSS); *Weighted K-Nearest Neighbor* (WK-NN).

ABSTRACT

The continuous improvement of mobile stations localization systems in wireless environments constitutes one of the major technical challenges to enable smart cities development. Among available outdoor localization techniques, those based on radio frequency fingerprinting (RF fingerprinting) are able to provide an effective location-estimation accuracy at low deployment and energetic costs. Generally speaking, the technique works by dividing the coverage area, creating a grid map, and defining an RF feature vector for each cell in that grid. The goal is to ensure that each cell in the grid can be uniquely identified by an RF vector. The main characteristic of a grid map is that cells are uniformly distributed. Besides, the number of cells determines the computational complexity associated with the localization model. Recently, some studies have introduced a grip map concept in which some cells are more concentrated in one region over the others. When compared to a regular grid map, this new irregular cell distribution over the coverage area has decreased the computational complexity when targeting a predetermined location-estimation accuracy requirement. The location-estimation accuracy requirements are typically enforced by government regulatory agencies, such as the US Federal Communications Commission (FCC), which imposes 50 meters as the maximum location-estimation error for emergency calls. Unlike previous works, this study proposes a generalization of some concepts that define an irregular grid map, introducing new techniques and algorithms to automate its generation. In order to evaluate the location-estimation accuracy provided by the irregular grid map generation algorithms proposed in this work, real and simulated experiments were performed. Simulations were carried out by using the LENA LTE module provided as an extension to the Network Simulator 3 (NS-3). Then, 4,617 real RF measurements were evaluated. Those measurements were collected in the city of Recife/PE for the Long-Term Evolution (LTE) standard with 1.8GHz carrier. When compared with regular grid maps, results for the irregular grid map have shown significant reduction in the computational complexity needed to satisfy current FCC location-estimation accuracy requirements.

Keywords: Location-Based Services (LBS); irregular gridmap; RF fingerprinting; Received Signal Strength (RSS); Weighted K -Nearest Neighbor (WK-NN).

LISTA DE FIGURAS

Figura 1 – Amostra hipotética do conjunto de medições <i>Radio Frequency</i> (RF) (<i>RF fingerprint</i>). Além das coordenadas espaciais, é possível observar n medições <i>Received Signal Strength</i> (RSS) referentes a um conjunto de n Estações Rádio Base (ERBs), sendo uma medição obtida para cada ERB.	25
Figura 2 – Segmentação da região de cobertura proporcionado através de uma malha quadricular. O conjunto de todas as coordenadas espaciais associadas ao centro das células é conhecido como <i>gridmap</i> , sendo esse conjunto, utilizado para representar computacionalmente essa região. .	26
Figura 3 – Arquitetura genérica para um sistema de localização baseado em <i>RF fingerprinting</i> . Na etapa de treinamento, as medições RF são generalizadas para toda a região de cobertura. Já na etapa de predição, é estabelecida uma correspondência entre o <i>RF fingerprint</i> observado para uma amostra e os <i>RF fingerprints</i> armazenados no <i>radiomap</i> , para estimar a posição da amostra.	27
Figura 4 – Topologia de rede estabelecida pelo padrão LTE/EPC. É possível perceber a distinção entre os elementos na camada <i>Long-Term Evolution</i> (LTE) (<i>LTE model</i> destacado em vermelho), daqueles presentes na camada <i>Evolved Packet Core</i> (EPC) (<i>EPC model</i> destacado em azul). . .	33
Figura 5 – <i>Radio Environment Map</i> (REM) obtido para o ambiente urbano simulado. Os triângulos identificam as torres com ERBs tri-setoriais, já os retângulos correspondem as edificações.	35
Figura 6 – Estações Móveis (EMs) distribuídas de modo a simular um <i>drive-testing</i> no perímetro urbano. Foram posicionadas 3.000 EMs, sendo 33% em ambientes <i>indoor</i>	36
Figura 7 – Exemplo de um <i>gridmap</i> regular com resolução r	40
Figura 8 – A taxa de crescimento do número de células é governada pela função quadrática $f(x) = x^2$. Em (a), (b) e (c), temos respectivamente $f(2) = 4$, $f(3) = 9$ e $f(4) = 16$	40
Figura 9 – Representação de um <i>gridmap</i> regular com 2.500 células. Os pontos brancos simbolizam o centro das células no <i>gridmap</i> , enquanto os triângulos representam as posições das ERBs.	41
Figura 10 – Diagrama em blocos para segmentação de <i>gridmaps</i> irregulares. O parâmetro c representa a quantidade total de células no <i>gridmap</i> , enquanto p corresponde a um percentual do total de células designadas às regiões prioritárias.	42

Figura 11 – Alternativas para distribuição uniforme de pontos sobre uma superfície: (a) Lattice regular. Três sequências de baixa discrepância em 2-dimensões, sendo (b) R , (c) Halton e (d) Sobol; (e) sequência amostrada de uma distribuição aleatória.	44
Figura 12 – <i>Gridmaps</i> obtidos com base em $c = 3.000$ células. (a) Geração por sequência quasi-aleatória R_2 . (b) Geração por sequência pseudo-aleatória.	45
Figura 13 – <i>Gridmaps</i> obtidos com base em $c = 2.500$ células utilizando: (a) malha quadricular (<i>gridmap</i> regular) e (b) sequência R_2 (<i>gridmap</i> irregular).	46
Figura 14 – Diagrama com o esquema final para criação de <i>gridmaps</i> irregulares. Os blocos A_1 e A_2 indicados no diagrama da Figura 14 foram substituídos respectivamente pelas implementações do <i>Farthest-First Traversal</i> (Algoritmo 1) e <i>Quasirandom Sequence R_2</i> (Algoritmo 2).	46
Figura 15 – <i>Gridmaps</i> irregulares obtidos com base em $c = 2.500$ células e $p\%$ de células dispostas nas regiões prioritárias, acompanhando a distribuição original das medições RF. (a) $p = 10\%$. (b) $p = 30\%$. (c) $p = 50\%$. (d) $p = 70\%$	47
Figura 16 – Sistema de localização utilizado nos experimentos. O elemento $\#PD(m, n^{th})$ atua como um filtro, estabelecendo uma correspondência entre o vetor <i>Propagation Delay</i> (PD) da amostra m e do n -ésimo elemento (n^{th}) do mapa de rádio (<i>radiomap</i>). A partir do <i>radiomap</i> filtrado, o elemento n^{th} selecionado é aquele que possuir a menor distância $d_{min}(m, n^{th})$ em relação a amostra m	51
Figura 17 – O gráfico mostra o <i>erro médio de localização para o gridmap μ</i> obtido com relação aos algoritmos S , R , Q e Λ . O algoritmo Λ foi aplicado utilizando diferentes valores percentuais p definidos para $p \in \mathbf{P}$	58
Figura 18 – Resultado para o diagrama de distância crítica <i>Critical Difference</i> (CD). O intervalo destacado mostra uma distância de $2d_c$ centrada no algoritmo S para <i>gridmaps</i> quadriculares, onde d_c representa a distância crítica da análise <i>post-hoc</i> Nemenyi calculada para o teste de Friedman.	61
Figura 19 – Distribuição para o erro de localização horizontal obtido quando um total de $c = 2500$ células foi utilizado. É possível perceber uma redução da média a medida que o valor do parâmetro p se aproxima de 1.0.	62
Figura 20 – Mapa de calor para o conjunto de medições RF reais obtidas através de <i>drive-testing</i> . A região de cobertura é definida pelo retângulo vermelho, compreendendo uma área de aproximadamente $3km^2$. As ERBs podem ser identificadas através dos polígonos na cor azul.	63
Figura 21 – (a) <i>Gridmap</i> de malha quadricular gerado pelo algoritmo S com $c = 2.500$ células. (b) <i>Gridmap</i> gerado pelo algoritmo Q com $c = 2.500$ células.	64

Figura 22 – Exemplo de <i>gridmaps</i> irregulares com $c = 2500$ células gerados para a região de cobertura. Em (a), (b), (c) e (d) podemos visualizar <i>gridmaps</i> onde uma percentual $p = 10\%$, $p = 20\%$, $p = 30\%$ e $p = 40\%$ das células foram designadas as regiões prioritárias.	65
Figura 23 – Diagramas CD para (a) $c = 2.500$, (b) $c = 10.000$ e (c) $c = 19.600$ células, obtidos pela análise <i>post-hoc</i> de Nemenyi para a média dos <i>ranks</i> calculados através do teste de Friedman.	67
Figura 24 – Erro médio de localização μ para os algoritmos S , $\Lambda_{p=10\%}$, $\Lambda_{p=20\%}$, $\Lambda_{p=30\%}$, $\Lambda_{p=40\%}$ e $\Lambda_{p=100\%}$. Os valores de μ foram calculados para 4.617 medições RF reais (<i>indoor</i> e <i>outdoor</i>), obtidas na cidade de Recife/PE.	68
Figura 25 – Distribuição do erro de localização horizontal para os algoritmos Q , R , S e Λ quando o parâmetro definido para o número de células foi $c = 2500$	69

LISTA DE TABELAS

Tabela 1	–	Parâmetros de simulação do módulo LENA LTE para as ERBs.	34
Tabela 2	–	Parâmetros do modelo de propagação utilizado na simulação.	37
Tabela 3	–	Amostra dos valores de RSS em dBm obtidos para as ERBs.	37
Tabela 4	–	Amostra dos parâmetros PD no conjunto de medições RF.	38

LISTA DE ABREVIATURAS E SIGLAS

ANN	<i>Artificial Neural Network</i>
BTS	<i>Base Transceiver Station</i>
CD	<i>Critical Difference</i>
EM	<i>Estação Móvel</i>
eNB	<i>Evolved Node-B</i>
EPC	<i>Evolved Packet Core</i>
ERB	<i>Estação Rádio Base</i>
ETU	<i>Extended Typical Urban Model</i>
FCC	<i>Federal Communications Commission</i>
GPS	<i>Global Positioning System</i>
GSM	<i>Global Mobile System Communication</i>
IBL	<i>Instance-Based Learning</i>
IDW	<i>Inverse Distance Weighting</i>
IMSI	<i>International Mobile Subscriber Identity</i>
K-NN	<i>K-Nearest Neighbors</i>
LBS	<i>Location-Based Service</i>
LTE	<i>Long-Term Evolution</i>
NG9-1-1	<i>Next Generation 9-1-1</i>
NS-3	<i>Network Simulator 3</i>
PD	<i>Propagation Delay</i>
REM	<i>Radio Environment Map</i>
RF	<i>Radio Frequency</i>
RSRP	<i>Reference Signal Received Power</i>
RSRQ	<i>Reference Signal Received Quality</i>
RSS	<i>Received Signal Strength</i>
RSSI	<i>Received Signal Strength</i>
SINR	<i>Signal-to-Interference-plus-Noise Ratio</i>
SVM	<i>Support Vector Machine</i>
UE	<i>User Equipment</i>
WK-NN	<i>Weighted K-Nearest Neighbors</i>

LISTA DE SÍMBOLOS

$\mathbf{C}$	Conjunto de valores para o número de células em um <i>gridmap</i>
$\mathbf{P}$	Conjunto de valores para a proporção de células em regiões prioritárias
S	Algoritmo para geração de <i>gridmaps</i> quadriculares
R	Algoritmo para geração de <i>gridmaps</i> pseudo-aleatórios
Q	Algoritmo para geração de <i>gridmaps</i> quasi-aleatórios
Λ	Algoritmo para geração de <i>gridmaps</i> adaptativos
Λ_p	Aplicação parcial do algoritmo Λ para $p \in \mathbf{P}$
$\mathbf{S}_c$	Conjunto de <i>gridmaps</i> gerados por S para valores $c \in \mathbf{C}$
$\mathbf{R}_c$	Conjunto de <i>gridmaps</i> gerados por R para valores $c \in \mathbf{C}$
$\mathbf{Q}_c$	Conjunto de <i>gridmaps</i> gerados por Q para valores $c \in \mathbf{C}$
Λ_{cp}	Conjunto de <i>gridmaps</i> gerados por Λ para valores $c \in \mathbf{C}$ e $p \in \mathbf{P}$
$\mathbf{G}$	Conjunto de <i>gridmaps</i> $\mathbf{G} = \mathbf{S}_c \cup \mathbf{R}_c \cup \mathbf{Q}_c \cup \Lambda_{cp}$
N_e	Número total de ERBs
N_g	Número de grupos de ERBs
$\mathbf{Z}$	Conjunto de medições RF
$\boldsymbol{\Omega}^{(n)}$	Vetor RSS para n -ésima amostra do conjunto $\mathbf{Z}$
$\boldsymbol{\Gamma}^{(n)}$	Vetor PD para n -ésima amostra do conjunto $\mathbf{Z}$
$\mathbf{E}^{(n)}$	Vetor posição para n -ésima amostra do conjunto $\mathbf{Z}$
$\boldsymbol{\Delta}^{(n)}$	Vetor diferença de PD entre grupos de ERBs
Υ	Conjunto <i>radiomap</i> cujo n -ésimo elemento é $(\mathbf{E}^{(n)}, \boldsymbol{\Omega}^{(n)}, \boldsymbol{\Gamma}^{(n)})$
$\mathbf{A}^m$	Medição RF para amostra m , corresponde ao par $(\boldsymbol{\Omega}^m, \boldsymbol{\Gamma}^m)$
$\boldsymbol{\Omega}^m$	Vetor RSS para uma amostra m
$\boldsymbol{\Gamma}^m$	Vetor PD para uma amostra m
$\boldsymbol{\Delta}^m$	Vetor diferença de PD para uma amostra m

Υ_r	Conjunto <i>radiomap</i> com espaço de busca reduzido
$d^{(n)}$	Distância Euclidiana entre os vetores $\mathbf{\Omega}^{(n)}$ e $\mathbf{\Omega}^m$
e_n	Erro de localização horizontal para n -ésima amostra
μ	Erro médio de localização para o <i>gridmap</i>
Δ_{x-y}^μ	Diferença de erro μ entre algoritmos ($x, y \in \{S, R, Q, \Lambda_p\}$, $x \neq y$)
μ_{max}	Maior melhora no erro médio μ observada entre algoritmos
μ_{min}	Menor melhora no erro médio μ observada entre algoritmos
$\bar{\mu}$	Média da melhora na acurácia μ observada entre algoritmos

SUMÁRIO

1	INTRODUÇÃO	17
1.1	CONTEXTUALIZAÇÃO	18
1.2	MOTIVAÇÃO	21
1.3	OBJETIVOS E RESULTADOS	22
1.4	ESTRUTURA DO DOCUMENTO	23
2	LOCALIZAÇÃO BASEADA EM RF FINGERPRINTING	24
2.1	CONJUNTO DE MEDIÇÕES RF	24
2.2	SEGMENTAÇÃO DA REGIÃO DE COBERTURA	25
2.3	ARQUITETURA DOS SISTEMAS DE LOCALIZAÇÃO	26
2.4	WEIGHTENED K-NEAREST NEIGHBORS (WK -NN)	28
2.5	TRABALHOS RELACIONADOS	29
3	NOVA PROPOSTA PARA GRIDMAPS IRREGULARES	32
3.1	CRIAÇÃO DO AMBIENTE SIMULADO	32
3.2	SIMULANDO O CONJUNTO DE MEDIÇÕES RF	37
3.3	ALGORITMOS PARA A GERAÇÃO DOS GRIDMAPS	39
3.4	NOTAÇÃO PARA ALGORITMOS E GRIDMAPS	48
3.5	DEFINIÇÃO PARA AS MEDIÇÕES RF	50
3.6	SISTEMA DE LOCALIZAÇÃO	50
4	AVALIAÇÃO E RESULTADOS	55
4.1	ERRO MÉDIO DE LOCALIZAÇÃO PARA O GRIDMAP	55
4.2	MÉTRICAS PARA AVALIAÇÃO DOS GRIDMAPS	56
4.3	RESULTADO DA AVALIAÇÃO DOS GRIDMAPS	58
4.4	CONJUNTO DE MEDIÇÕES RF REAIS	63
4.5	SELEÇÃO DOS GRIDMAPS	64
4.6	RESULTADO FINAL	66
5	CONCLUSÃO	71
	REFERÊNCIAS	72

1 INTRODUÇÃO

As primeiras discussões envolvendo o termo cidades inteligentes remontam a debates sobre políticas urbanas ocorridos nos Estados Unidos no começo da década de 90 (ROSATI; CONTI, 2016). Desde então, a idealização do conceito de cidades inteligentes vem sofrendo mudanças. Nas últimas décadas, o conceito deixou de ser puramente tecnológico, passando a considerar aspectos sociais e humanos, sendo força motriz para impulsionar um desenvolvimento sustentável que equilibre recursos naturais, crescimento econômico e qualidade de vida (PAHLAVAN; KRISHNAMURTHY; GENG, 2015). Apesar desse entendimento, não existe um consenso a respeito da definição precisa do termo cidades inteligentes. Na prática, ele pode ser encarado como um processo dinâmico que possibilita a cidade adaptar-se rapidamente às necessidades emergentes de sua população, garantindo uma maior resiliência e capacidade em responder rapidamente aos novos desafios (BIS, 2013).

Sob a perspectiva da Engenharia, as cidades inteligentes são concebidas por meio de sistemas computacionais distribuídos de forma pervasiva no ambiente físico, com capacidade para coletar, processar e correlacionar informações que impulsionem o desenvolvimento de serviços mais confiáveis e relevantes para os cidadãos (PAHLAVAN; KRISHNAMURTHY; GENG, 2015). É fundamental para as cidades inteligentes correlacionar informações de diferentes sistemas urbanos como os de mobilidade, habitacionais, administrativos, energéticos, educacionais, entre outros. A combinação desses sistemas é o que permite processar e analisar as informações para encontrar soluções ditas *inteligentes* para a cidade. Uma cidade digitalmente conectada poderia ter estações de medição instaladas em diversos locais para monitorar o nível de poluição do ar. Essa informação poderia ser utilizada, por exemplo, para fornecer diretrizes para restrição de tráfego quando determinados níveis de poluição, considerados prejudiciais à saúde, fossem atingidos.

Os cidadãos, por sua vez, fazem uso dos chamados aplicativos inteligentes, ou seja, aplicativos capazes de utilizar informação contextual como tempo, localização e sensores para as mais diversas finalidades. É por meio desses aplicativos que os cidadãos vivenciam as cidades inteligentes. Os aplicativos inteligentes podem permitir, entre outras coisas, monitorar o tráfego e avaliar alternativas de mobilidade dentro da cidade. Exemplos mais recentes incluem os aplicativos de *contact tracing*, com o desenvolvimento de protocolos que permitem rastrear eventuais contágios pelo SARS-CoV-2 (AHMED et al., 2020).

Diante dessa conjuntura, as tecnologias de localização desempenham um papel fundamental para aplicações que promovem o desenvolvimento das cidades inteligentes, tornando imprescindível o aprimoramento e a melhoria contínua dessas tecnologias. Essa necessidade surge já que diferentes serviços baseados em localização — *Location-Based Services* (LBSes) — possuem diferentes requisitos de acurácia. Por exemplo, a acurácia mínima necessária para localizar estações móveis em ambientes *outdoor* e *indoor* é

completamente diferente. Enquanto na primeira a margem de erro para uma estimativa de localização pode ser da ordem de dezenas ou até mesmo centenas de metros, na segunda, essa mesma exigência passa a ser da ordem de alguns poucos metros. Além disso, a demanda por tecnologias de localização cada vez mais precisas tem se tornado uma realidade, principalmente no contexto de cidades inteligentes. O desenvolvimento recente de tecnologias de transporte autônomo (*self-driving*) é um exemplo prático disto. Em seu cenário de aplicação, o erro para uma estimativa de localização pode variar na ordem de apenas alguns poucos centímetros (ROSS, 2016).

No intuito de atender a esse amplo espectro de aplicações, com exigências particulares com respeito ao erro associado à estimativa de localização, são utilizados diversos sensores. Os mais comuns são aqueles baseados em radiofrequência, como é o caso dos transceptores (*transceivers*) de GPS, WiFi, iBeacon (*Bluetooth*) e telefonia móvel (GSM, LTE, etc) (VO; DE, 2015). Algumas aplicações fazem uso, inclusive, de sensores mecânicos como é o caso do acelerômetro e do magnetômetro, com a finalidade de refinar a estimativa de localização para uma estação móvel (NGUYEN et al., 2019). A investigação sobre as características comportamentais desse conjunto de sensores, aliada ao emprego de métodos capazes de atender as diferentes exigências de acurácia, consiste no principal desafio que vem sendo colocado para as tecnologias de localização ao longo da última década.

1.1 CONTEXTUALIZAÇÃO

No contexto de localização *outdoor*, é indiscutível a popularidade do Sistema Global de Posicionamento — *Global Positioning System* (GPS). O GPS surgiu como um projeto estritamente militar na década de 70, mas sua comercialização, principalmente a partir do início da década de 90, impulsionou a disseminação dessa tecnologia (PAHLAVAN; KRISHNAMURTHY; GENG, 2015). Além da inegável popularidade e da maturidade do GPS, sua adoção em larga escala resultou em uma natural redução no custo de aquisição da tecnologia. Somado a isso, a capacidade de estimar a posição de objetos com acurácia da ordem de dezenas de metros supre a necessidade de boa parte das aplicações. Esses fatores contribuem para que o GPS ainda seja uma das principais alternativas para os serviços baseados em localização (DAO; RIZOS; WANG, 2002).

O GPS pertence a uma classe de técnicas de localização que fazem uso de argumentos geométricos para estimar a posição da Estação Móvel (EM). Essa tecnologia faz uso de diferenças no tempo de chegada dos sinais oriundos de diferentes satélites para estimar a posição relativa do objeto em relação a esses satélites, ou seja, o conjunto de satélites é utilizado como referencial inercial para inferir a posição do objeto. Esses parâmetros de tempo de chegada são utilizados para criar um conjunto de equações geométricas cuja solução pode ser obtida por meio de métodos matemáticos de aproximação como Newton-Raphson (VARAGNOLO et al., 2016) ou o de Nelder-Mead (GAO; HAN, 2012).

O mesmo ocorre com as técnicas baseadas em multilateração. Nesse caso, a potência do sinal recebido é utilizada para medir a quantidade de energia dissipada no trajeto entre o transmissor e o receptor. Como a quantidade de energia dissipada é função da distância percorrida pelo sinal eletromagnético, a diferença entre a potência do sinal transmitido e o recebido serve de insumo para modelos matemáticos de propagação eletromagnética capazes de determinar a distância percorrida pelo sinal e, conseqüentemente, a posição de um determinado objeto (CASSANO et al., 2009).

Por estarem baseadas em modelos matemáticos puramente determinísticos, essas técnicas são tipicamente ineficientes quando aplicadas *indoor* ou em ambientes urbanos densamente ocupados por edificações verticalizadas (*urban canyons*). Nesses ambientes, a presença natural de obstáculos causa a atenuação do sinal eletromagnético (*fading*) e introduz efeitos de sombreamento (*shadowing*) e propagação multipercurso (*multipath*), tudo isso associado à presença de ruído inerente ao meio de propagação, fato que acarreta na degradação da estimativa de localização fornecida por esses sistemas (SOUBIELLE et al., 2002). Esses efeitos concebem características não lineares ao problema, introduzindo uma maior imprevisibilidade à solução baseada em modelos puramente geométricos e acarretando em um maior erro na estimativa de localização.

No caso específico do GPS, vale destacar ainda o elevado consumo energético que pode ser impeditivo para algumas aplicações. No caso de aplicações baseadas em redes de sensores sem fio (*Wireless Sensors Network - WSN*), muitas vezes uma fonte perene para fornecer energia não é uma opção viável, ou mesmo a possibilidade de recarregar a bateria dos sensores (ABU-MAHFOUZ; HANCKE, 2017). Nesses casos, é necessário adotar estratégias para minimizar o consumo energético, prolongando a vida útil das baterias e impedindo-as de depletar rapidamente (ROUT; RATH; BHAGABATI, 2017).

Entre as técnicas de localização *outdoor*, aquelas baseadas em *Radio Frequency* (RF) *fingerprinting* também são capazes de fornecer uma acurácia na ordem de dezenas de metros, com um baixo custo energético e de implementação. A grande vantagem dessa técnica é sua capacidade de manter a acurácia mesmo em condições em que não existe linha de visada (*Non-Line Of Sight - NLOS*) entre o transmissor e o receptor (GUVENC; CHONG, 2009). Além disso, a implantação de técnicas de localização baseadas em RF *fingerprinting* pode tirar proveito da infraestrutura de comunicação já instalada, como acontece com as redes de telefonia móvel. Essa característica naturalmente barateia os custos de implantação, tornando essas tecnologias mais atrativas para serem incorporadas e ofertadas através de serviços baseados em localização.

O *modus operandi* das técnicas de localização baseadas em RF *fingerprinting* consiste em dividir uma determinada região de cobertura em áreas menores através de uma malha quadriculada (*gridmap*). A região de cobertura representa a área na qual se deseja localizar as estações móveis. Em seguida, é necessário obter um vetor de características de RF para cada célula do *gridmap* com relação a um conjunto de Estações Rádio Base (ERBs), cuja

potência de transmissão e localização são previamente conhecidas. Cada elemento desse vetor de características RF corresponde a um parâmetro extraído do sinal RF. Entre os principais parâmetros que é possível extrair de um sinal RF estão, por exemplo, o tempo de chegada (*Time of Arrival* - *ToA*), a diferença entre tempos de chegada (*Time Difference of Arrival* - *TDoA*), o ângulo de chegada (*Angle of Arrival* - *AoA*) e a intensidade de sinal recebido — *Received Signal Strength* (RSS). Esse processo resulta no que se costuma chamar mapa de rádio (*radiomap*) da região de cobertura.

O conceito fundamental por trás da utilização dos *radiomaps* apoia-se no fato de que os vetores de características RF são distintos para cada célula do *gridmap*. Esse aspecto, por sua vez, permite identificar unicamente uma célula do *gridmap* através de seu vetor de características RF, ou simplesmente, *RF fingerprint*. Como a posição geográfica de cada célula do *gridmap* é conhecida, já que existe uma correspondência única entre uma célula do *gridmap* e uma área real na região de cobertura, é possível determinar a localização de uma estação móvel comparando o *RF fingerprint* obtido pela EM em uma posição desconhecida, com aqueles armazenados no *radiomap*. O *RF fingerprint* presente no *radiomap* que melhor se assemelhar ao *RF fingerprint* fornecido pela EM, deve a princípio, fornecer uma boa estimativa da real posição dessa estação.

Tipicamente, as medições RF necessárias para viabilizar a construção dos mapas de rádio são obtidas por meio de duas técnicas: *drive-testing* ou *crowdsourcing*. Na primeira, um veículo motorizado é instrumentado com um equipamento específico com capacidade de mensurar os parâmetros RF (*scanner*). Esse veículo, por sua vez, se desloca na região de cobertura, captando e armazenando as medições RF obtidas em diferentes locais (KIM; MIN; YEO, 2014). Já a segunda técnica funciona a partir do engajamento de uma comunidade ou de um conjunto de usuários, que utiliza equipamentos pessoais, como *smartphones*, para obter as medições RF ao longo da região de cobertura por um determinado período de tempo (SATAKE, 2016).

Dada a natureza do processo de coleta, é impraticável, ou extremamente custoso, obter medições RF para cada célula do *gridmap*. Na prática, essas medições são frequentemente esparsas e encontram-se aglomeradas ou distribuídas ao redor de alguns poucos locais (ZEKAVAT; BUEHRER, 2011). Essa característica impede a criação de um mapeamento direto entre uma medição RF e uma célula no *gridmap*, já que ao final do processo de medição, muitas das células no *gridmap* permanecem sem um *RF fingerprint* associado. Sendo assim, é necessário empregar uma técnica para generalizar as medições RF para todas as células do *gridmap*. Entre as alternativas, pode-se utilizar um algoritmo de interpolação como *Inverse Distance Weighting* (IDW), um modelo de propagação eletromagnética, ou até mesmo algoritmos de regressão baseados em aprendizado de máquina (AHA; KIBLER; ALBERT, 1991). Dentre as opções, aquelas envolvendo aprendizado de máquina são particularmente indicadas pela sua capacidade de acomodar efeitos não lineares, como aqueles introduzidos pela presença de ruído ou demais fenômenos de propagação responsáveis por

causar a atenuação do sinal (LOVISOLO; CAMPOS, 2012).

Vários algoritmos de aprendizado de máquina podem ser empregados para criar o mapa de rádio, desde *Support Vector Machines* (SVMs) até *Artificial Neural Networks* (ANNs) (ZEKAVAT; BUEHRER, 2011). Um algoritmo comumente empregado para esse propósito é o *Weighted K-Nearest Neighbors* (WK-NN). Esse algoritmo é uma extensão do *K-Nearest Neighbors* (K-NN) (AHA; KIBLER; ALBERT, 1991). Devido a sua facilidade de implementação e a obtenção de bons resultados, o algoritmo WK-NN tem sido adotado como solução satisfatória para emprego em problemas de localização (BRIDA; MACHAJ; BENIKOVSKÝ, 2014; WANG et al., 2016).

1.2 MOTIVAÇÃO

A grande maioria dos trabalhos a respeito de técnicas de localização baseadas em *RF fingerprinting* concentram sua atenção na otimização de algoritmos de aprendizado de máquina no intuito de melhorar a acurácia da estimativa de localização fornecida (VO; DE, 2015; LIU et al., 2007; TIMOTEO et al., 2015; BRIDA; MACHAJ; BENIKOVSKÝ, 2014; WANG et al., 2016; YAN et al., 2017; PENG et al., 2019). Esses trabalhos tipicamente modelam a região de cobertura utilizando um *gridmap* de malha quadriculada, onde as células encontram-se igualmente espaçadas. Ainda são poucos os trabalhos que propõem investigar a influência da geometria do *gridmap* na acurácia dos sistemas de localização (KIM; MIN; YEO, 2014; MA; ZHANG; HAN, 2015; BITTENCOURT; URBANO; CUNHA, 2018).

Em 2018 nós propusemos um *gridmap* cuja geometria é modelada considerando a morfologia da malha viária dentro de um dado perímetro urbano (BITTENCOURT; URBANO; CUNHA, 2018). Na época, nós denominamos essa modelagem alternativa para geometria do *gridmap* de irregular (*irregular gridmap*). Essa nova abordagem trouxe benefícios, tanto em termos da acurácia quanto em termos de complexidade computacional. Esse benefício foi constatado levando em consideração medições RF reais realizadas para o padrão *Global Mobile System Communication* (GSM) na faixa de frequência de 1,8 GHz. Apesar dos resultados serem promissores, os autores alertaram que certas tendências associadas às medições RF utilizadas podem ter influenciado positivamente no resultado. O alerta diz respeito ao fato de as medições RF terem sido obtidas, exclusivamente, através de um procedimento de *drive-testing*. Essa característica pode ter interferido de forma a corroborar os resultados, já que todas as medições RF foram obtidas em ambiente *outdoor* e estavam associadas às vias da região de cobertura.

A principal hipótese levantada por nós para adoção do modelo de *gridmap* proposto é sustentada pelo fato de que, em ambientes urbanos densamente ocupados, verticalizados por edificações comerciais e residenciais, como acontece tipicamente em grandes centros metropolitanos, existe uma maior chance de as EMs estarem localizadas *outdoor*, em ruas, avenidas ou em regiões próximas, como calçadas e passarelas. Considerando esse cenário,

aplicações específicas, como aquelas focadas em mobilidade urbana, poderiam se beneficiar diretamente do modelo de *gridmap* irregular.

Outro resultado relevante em (BITTENCOURT; URBANO; CUNHA, 2018) mostra que, apesar dos modelos de *gridmap* irregulares oferecem um ganho de acurácia aos sistemas de localização, a maior vantagem para sua utilização é a redução na complexidade computacional. Essa redução fica nítida quando os dois modelos são comparados com relação a um determinado requisito de acurácia. A complexidade computacional para um sistema de localização é função do número de células presentes em um *gridmap*. Quanto menor esse número, menor a complexidade computacional associada ao sistema.

Conforme mencionado anteriormente, entre os requisitos impostos na construção de serviços de localização, um dos principais é a acurácia. O desempenho dos sistemas de localização é normalmente mensurado em termos da acurácia fornecida. Os experimentos conduzidos pelos autores em (BITTENCOURT; URBANO; CUNHA, 2018) mostraram que o modelo de *gridmap* irregular exigiu um número de células seis vezes menor que o *gridmap* regular para alcançar um requisito de acurácia na ordem de 50 metros.

1.3 OBJETIVOS E RESULTADOS

Diante do contexto apresentado, o presente trabalho procura generalizar o entendimento sobre o processo de concepção de *gridmaps* irregulares, oferecendo uma interpretação alternativa diferente daquela originalmente proposta por nós em (BITTENCOURT; URBANO; CUNHA, 2018). A hipótese original que motivou a concepção de *gridmaps* irregulares pode ser ampliada considerando que as medições de RF, implicitamente, carregam informações sobre como as EMs costumam se distribuir espacialmente. A principal vantagem por trás dessa nova interpretação é a capacidade de generalizar a região de cobertura, ou seja, a técnica funciona de forma idêntica independentemente do ambiente, seja ele urbano ou rural. A nova interpretação dispensa também as restrições de aplicabilidade do modelo original, cujo foco eram aplicações direcionadas a solução de problemas de mobilidade urbana. Os métodos apresentados neste trabalho para geração de *gridmaps* irregulares ampliam o espectro de problemas nos quais esse modelo possui empregabilidade, tornando-os adequados para as mais diversas aplicações.

No intuito de avaliar a eficácia dessa nova interpretação, foi proposta uma técnica de geração de *gridmaps* irregulares que leva em consideração a distribuição espacial original das medições de RF, sejam elas obtidas via *drive-testing* ou *crowdsourcing*. Como as evidências em (BITTENCOURT; URBANO; CUNHA, 2018) indicam que a principal vantagem para adoção do modelo de *gridmaps* irregular é a redução na complexidade computacional, o processo de geração de *gridmaps* foi parametrizado pelo número de células. A influência da distribuição espacial original das medições RF no *gridmap* resultante foi definida através de um parâmetro que representa um percentual do número total de células que devem ser dispostas nessas regiões. As demais células são distribuídas uniformemente pela região

de cobertura tal qual o efeito proporcionado por um *gridmap* de malha quadriculada. A possibilidade de controlar com precisão o parâmetro responsável por determinar o número de células, favorece uma comparação precisa entre os *gridmaps*, tanto sob a perspectiva da acurácia, quanto da complexidade computacional.

Para avaliar essa nova abordagem, foram realizados experimentos utilizando medições RF geradas pelo módulo LENA LTE do simulador *Network Simulator 3* (NS-3) (CCTC, 2011). Quando comparado ao *gridmap* tradicional, a técnica de criação de *gridmap* proposta foi capaz de gerar resultados estatisticamente relevantes em todas as simulações realizadas, com um nível de significância inferior a 0,05. Esse resultado foi conquistado quando um percentual de apenas 10% do total de células, foi distribuído sobre as regiões onde as medições RF foram originalmente obtidas. Já com relação ao erro médio da estimativa de localização, o modelo proposto foi capaz de garantir uma redução de até 27 metros no erro médio em relação ao modelo de *gridmap* regular.

Os resultados preliminares obtidos com as medições RF simuladas, foram confirmados utilizando um conjunto com 4.617 medições RF reais, obtidas na cidade de Recife/PE para o padrão de telefonia móvel *Long-Term Evolution* (LTE) com portadora na faixa de 1,8 GHz. Em diversas configurações possíveis de parâmetros, os *gridmaps* obtidos a partir dos algoritmos de segmentação propostos foram capazes de atender os requisitos de acurácia da *Federal Communications Commission* (FCC) para as chamadas realizadas aos serviços de emergência. Em contrapartida, o modelo de *gridmap* tradicional não foi capaz de atingir os requisitos em nenhuma das configurações testadas.

1.4 ESTRUTURA DO DOCUMENTO

Os Capítulos restantes desta dissertação encontram-se estruturados da seguinte forma: o Capítulo 2 introduz os sistemas de localização baseados em *RF fingerprinting*, detalhando sua arquitetura e descrevendo um regressor baseado no WK-NN que é comumente utilizado em tais sistemas. Ainda no Capítulo 2 serão apresentadas outras propostas que utilizaram modelos de segmentação irregular para *gridmaps*. O Capítulo 3 introduz a utilização do simulador NS-3 para criação das medições RF, além de descrever em detalhes os algoritmos de segmentação de *gridmap* propostos neste estudo. Em seguida, o capítulo introduz uma notação matemática para auxiliar a compreensão e a análise dos resultados. O Capítulo 4 apresenta detalhadamente o resultado dos experimentos, considerando as avaliações realizadas para as medições RF obtidas através do simulador NS-3. Por último, a nova técnica de segmentação de *gridmap* é avaliada considerando uma base de medições RF reais, trazendo como base comparativa os requisitos de acurácia da FCC. Ao final, o Capítulo 5 apresenta algumas considerações finais, fornecendo indicações para futuros trabalhos neste campo de atuação.

2 LOCALIZAÇÃO BASEADA EM RF FINGERPRINTING

Este Capítulo fornece uma introdução aos sistemas de localização baseados em *Radio Frequency* (RF) *Fingerprinting*, ou simplesmente, *RF fingerprinting*. A partir de uma arquitetura genérica para um sistema de localização, serão indicadas as macro etapas de processamento e os detalhes associados ao seu funcionamento. O objetivo é apresentar os principais elementos que compõem tais sistemas, desde o conjunto de medições RF, até o processo de segmentação da área de cobertura com a criação dos *gridmaps*. Por fim, serão discutidos alguns trabalhos que descrevem abordagens alternativas para segmentação da região de cobertura através de *gridmaps* irregulares.

2.1 CONJUNTO DE MEDIÇÕES RF

Um artefato fundamental dos sistemas de localização baseados em RF *fingerprinting* é o conjunto de medições RF. Cada elemento desse conjunto possui, além das coordenadas espaciais, parâmetros RF mensurados com relação a um conjunto de estações rádio base referenciais, cuja posição e potência de transmissão são conhecidas. Nesse trabalho, o termo Estação Rádio Base (ERB) será utilizado para se referir a qualquer elemento da infraestrutura de telefonia móvel responsável por prover a interface sem fio, permitindo que as Estações Móveis (EMs) acessem a rede de telefonia. Entre esses elementos, estão incluídos, por exemplo, os *Base Transceiver Stations* (BTSs) da infraestrutura de telefonia *Global Mobile System Communication* (GSM), além dos chamados *Evolved Node-Bs* (eNBs) do padrão de tecnologia *Long-Term Evolution* (LTE).

Entre os parâmetros RF utilizados para construção do conjunto de medições, os mais comuns são aqueles associados à potência do sinal. No padrão LTE, por exemplo, é comum obter medições de parâmetros como o *Reference Signal Received Power* (RSRP) e o *Reference Signal Received Quality* (RSRQ). O primeiro consiste no valor correspondente à potência, em unidades de *dBm*, do sinal recebido pela estação móvel em uma dada posição. Já o segundo valor, é tipicamente obtido em unidades de *dB* e representa uma medida adicional a respeito da qualidade do sinal.

Independentemente da tecnologia utilizada, seja LTE, GSM ou até mesmo WiFi, diversos protocolos de camada física utilizam o termo *Received Signal Strength* (RSS) para referir-se à potência do sinal em um dado ponto do espaço. Outro termo bastante comum nesses protocolos é o chamado *Received Signal Strength* (RSSI). Este último corresponde normalmente a uma versão quantizada do RSS, sendo muitas vezes disponibilizado pela camada física através dos protocolos de acesso ao meio. No intuito de simplificar o entendimento e definir uma nomenclatura única, este trabalho fará uso do acrônimo RSS para referir-se à potência do sinal recebido.

A Figura 1 traz o exemplo de uma amostra hipotética do conjunto de medições RF (*RF fingerprint*). Neste exemplo, o parâmetro RF observado corresponde a potência do sinal recebido (RSS) em relação a um conjunto de n ERBs. Além do RSS, é possível verificar a presença de duas coordenadas espaciais, sendo aqui representadas pela latitude e pela longitude. É importante notar que os parâmetros RF devem ser obtidos individualmente para cada uma das n ERBs que estão sendo consideradas.

Figura 1 – Amostra hipotética do conjunto de medições RF (*RF fingerprint*). Além das coordenadas espaciais, é possível observar n medições RSS referentes a um conjunto de n ERBs, sendo uma medição obtida para cada ERB.

GPS Latitude	GPS Longitude	ERB₁ RSS ₁	ERB₂ RSS ₂	ERB₃ RSS ₃	...	ERB_{N-1} RSS _{N-1}	ERB_N RSS _N
------------------------	-------------------------	--	--	--	-----	--	--

Fonte: Elaborada pelo autor (2020)

2.2 SEGMENTAÇÃO DA REGIÃO DE COBERTURA

Conforme definido anteriormente, a região de cobertura consiste na área na qual deseja-se localizar as EMs. O objetivo final de um sistema de localização é construir o *radiomap* da região de cobertura, em outras palavras, isso implica que deveria existir um *RF fingerprint* associado a cada uma das coordenadas espaciais dessa região.

A Seção 1.1 apresenta as principais dificuldades impostas aos sistemas de localização para alcançar esse objetivo de forma analítica, utilizando equações geométricas e modelos de propagação eletromagnética. Em suma, tais abordagens são extremamente sensíveis a flutuações do nível de sinal e acabam não funcionando bem sob a presença de ruído e demais efeitos de atenuação do sinal eletromagnético.

Uma abordagem alternativa utilizada pelos sistemas de localização baseados em *RF fingerprint* consiste em segmentar a região de cobertura a partir de uma malha quadricular e, em seguida, generalizar as medições RF para todas as células dessa malha. O processo de generalização é alcançado empregando, por exemplo, um algoritmo de interpolação como o *Inverse Distance Weighting* (IDW).

A Figura 2 mostra um exemplo de segmentação de uma região de cobertura utilizando uma malha quadricular. Computacionalmente, a região de cobertura é representada pela coordenada espacial associada ao centro de cada célula dessa malha. Sendo assim, o conjunto resultante de todas as coordenadas é utilizado para modelar computacionalmente a região de cobertura, sendo esse conjunto denominado *gridmap*.

Como o modelo de segmentação por malha quadricular produz uma distribuição uniforme das células, tal qual um lattice retangular, esse modelo de *gridmap* é referido nesta pela alcunha de *gridmap* regular. Existem outros modelos de segmentação da região de co-

bertura, inclusive aqueles propostos neste trabalho, desse modo, técnicas de segmentação alternativas serão introduzidas na Seção 2.5.

Figura 2 – Segmentação da região de cobertura proporcionado através de uma malha quadricular. O conjunto de todas as coordenadas espaciais associadas ao centro das células é conhecido como *gridmap*, sendo esse conjunto, utilizado para representar computacionalmente essa região.



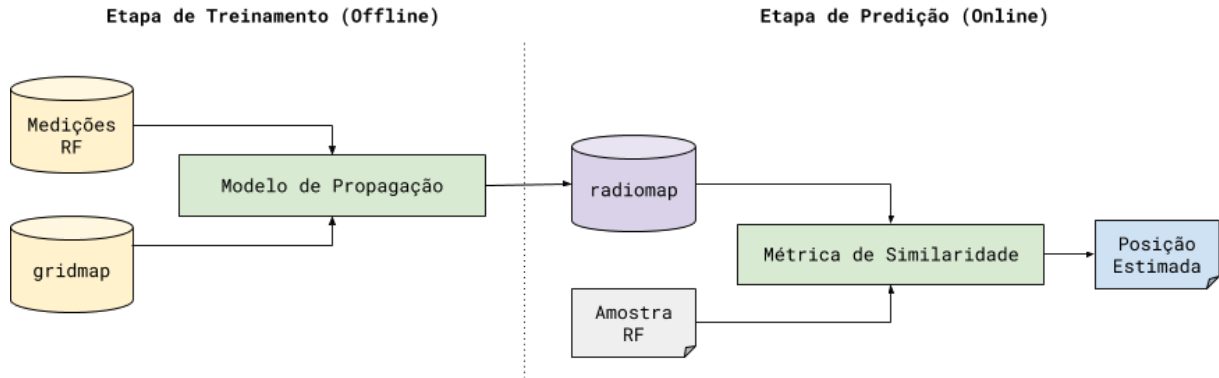
Fonte: Elaborada pelo autor (2020)

2.3 ARQUITETURA DOS SISTEMAS DE LOCALIZAÇÃO

A arquitetura geral de um sistema de localização baseado em *RF fingerprinting* pode ser dividida em duas etapas: *offline* e *online* (Figura 3). É papel da fase *offline* generalizar as medições RF originais para toda a área compreendida pela região de cobertura. Sendo assim, a fase *offline* recebe como entrada as medições RF e o *gridmap*, sendo esse último, um conjunto de coordenadas espaciais responsável por modelar a geometria da região de cobertura. O propósito geral da fase *offline* é construir um vetor de características RF (*RF fingerprint*) para identificar unicamente cada célula no *gridmap*. A posição de uma célula é tipicamente definida pela suas coordenadas geográficas (latitude e longitude). O *RF fingerprint* associado a célula é constituído por um ou mais parâmetros RF observados nessas coordenadas. Entre os parâmetros RF obtidos, o RSS é, sem dúvida, um dos mais comuns, sendo amplamente empregado em *Location-Based Services* (LBSes).

Em última instância, a operação de generalização efetuada na etapa *offline* pode ser definida como o processo de assinalar a cada célula do *gridmap* um *RF fingerprint* único. Esse processo é responsável por dar origem ao mapa de rádio (*radiomap*) da região de cobertura. A forma mais trivial de realizar essa generalização, seria obter empiricamente um vetor de RSS para cada coordenada geográfica do *gridmap*. Entretanto, o processo de obtenção de medições RSS ocorre através de técnicas como *drive-testing* (KIM; MIN;

Figura 3 – Arquitetura genérica para um sistema de localização baseado em *RF fingerprinting*. Na etapa de treinamento, as medições RF são generalizadas para toda a região de cobertura. Já na etapa de predição, é estabelecida uma correspondência entre o *RF fingerprint* observado para uma amostra e os *RF fingerprints* armazenados no *radiomap*, para estimar a posição da amostra.



Fonte: Elaborada pelo autor (2020)

YEO, 2014) ou *crowdsourcing* (SATAKE, 2016). Em ambos os casos, é praticamente inviável obter um vetor de medições RSS para todas as células do *gridmap*. Além de exaustivo, obter medições para todas as posições pode não ser factível por inúmeras razões, como a existência de propriedades privadas ou áreas inacessíveis. Como consequência, muitas células do *gridmap* permanecem sem um vetor RSS associado.

A solução tipicamente empregada pelos sistemas de localização baseados em *RF fingerprinting*, consiste em aplicar um algoritmo de regressão para derivar um vetor RSS para cada célula do *gridmap*. O problema deve ser modelado de modo que o algoritmo receba como entrada as coordenadas geográficas associadas a uma célula do *gridmap*, fornecendo como saída uma predição do vetor RSS naquela posição. Vários algoritmos de aprendizado de máquina podem ser empregados para esse propósito, desde máquinas de vetores de suporte, até redes neurais artificiais (ZEKAVAT; BUEHRER, 2011).

O mapa de rádio gerado na fase *offline* corresponde a uma das entradas utilizadas na etapa *online* do sistema de localização (Figura 3). O objetivo dessa etapa é fornecer uma estimativa de localização para a amostra RF. Diferentemente das medições RF, a amostra RF não possui coordenadas geográficas, apenas o vetor RSS observado com relação ao conjunto de ERBs consideradas pelo sistema. Para estimar a posição da amostra, a etapa *online* utiliza a distância Euclidiana como métrica de similaridade, determinando qual elemento do *radiomap* fornece a menor distância em relação à amostra. A distância Euclidiana é calculada tomando o vetor RSS da amostra e o vetor RSS de cada um dos elementos do mapa de rádio. A posição da amostra é então estimada como as coordenadas geográficas da célula do *radiomap*, cujo vetor RSS possui a menor distância Euclidiana com relação ao vetor RSS da amostra.

2.4 WEIGHTENED K-NEAREST NEIGHBORS (WK-NN)

Um algoritmo comumente utilizado na etapa de treinamento (Figura 3) consiste na variação do *K-Nearest Neighbors* (K-NN) conhecida como *Weighted K-Nearest Neighbors* (WK-NN). O algoritmo K-NN pertence a uma classe de algoritmos de aprendizado de máquina conhecida como *Instance-Based Learning* (IBL) ou aprendizado baseado em instâncias. Como o próprio nome sugere, o K-NN recebe como entrada uma amostra do conjunto de validação e escolhe K vizinhos mais próximos a essa amostra, ou seja, os K elementos do conjunto de treino que mais se aproximam da amostra de validação segundo alguma métrica (*nearest neighbors*). Entre as métricas utilizadas a distância Euclidiana é a mais comum, entretanto, outras métricas de distância podem ser utilizadas como Mahalanobis e Manhattan. Após selecionar os K vizinhos, o algoritmo extrai uma média aritmética desses elementos, fornecendo essa média como saída. Essa saída por sua vez, corresponde a predição oferecida pelo algoritmo (AHA; KIBLER; ALBERT, 1991). O funcionamento do algoritmo WK-NN é análogo ao K-NN, a diferença é que cada vizinho possui um grau de importância que é determinado usando como base o valor de W . O peso de cada vizinho é calculado tomando-se o inverso da distância entre ele e a amostra de validação, em seguida elevando-se o resultado à potência de W . No caso do WK-NN, a obtenção do valor de saída (predição) ocorre através de uma média ponderada que considera o peso calculado para cada vizinho.

A escolha do algoritmo WK-NN foi feita devido à sua facilidade de implementação, associada à obtenção de bons resultados. O algoritmo WK-NN tem sido adotado como solução satisfatória para emprego em problemas de localização (WANG et al., 2016; BRIDA; MACHAJ; BENIKOVSKÝ, 2014). No contexto desse trabalho, a saída fornecida pelo algoritmo WK-NN é definida por $RSS_{i,n}$ e corresponde à predição do algoritmo para a medição RSS referente a i -ésima ERB na posição determinada pela célula n . O valor para $RSS_{i,n}$ pode ser obtido através da seguinte expressão

$$RSS_{i,n} = \frac{\sum_{j=1}^k RSS_{i,j} \left(\frac{1}{d_{n,j}} \right)^w}{\sum_{j=1}^k \left(\frac{1}{d_{n,j}} \right)^w}, \quad (2.1)$$

em que k é o número de vizinhos, $RSS_{i,j}$ é o valor RSS para a i -ésima ERB do j -ésimo vizinho e $d_{n,j}$ corresponde a distância geodésica entre as coordenadas geográficas do j -ésimo vizinho e aquelas associadas a célula n . O valor w determina a influência da distância $d_{n,j}$ no termo $RSS_{i,j}$, indicando sua relevância para o resultado final. Perceba que caso o valor $w = 0$ seja escolhido, as distâncias não influenciarão no resultado obtido. Nesse caso específico, o algoritmo WK-NN se reduz ao K-NN.

Na Equação 2.1 o termo $d_{n,j}$ é calculado tomando a distância geodésica entre as coordenadas geográficas associadas as posições n e j . Para calcular $d_{n,j}$ foi utilizada a

relação do ângulo central para distância entre pontos na superfície de uma esfera

$$d_{n,j} = R \Theta_{n,j} , \quad (2.2)$$

em que R foi aproximado como o raio equatorial da Terra ($R = 6378,1 \text{ km}$) e $\Theta_{n,j}$ corresponde ao ângulo central, sendo seu *haversine* determinado pela fórmula

$$\text{hav}(\Theta_{n,j}) = \text{hav}(\varphi_j - \varphi_n) + \cos(\varphi_n)\cos(\varphi_j)\text{hav}(\lambda_j - \lambda_n) , \quad (2.3)$$

em que os valores φ_n e φ_j representam, respectivamente, as latitudes nos pontos n e j tomadas em radianos. Analogamente, λ_n e λ_j representam as longitudes dos pontos n e j também em radianos. Por fim, a função *haversine* para um ângulo θ é conhecida

$$\text{hav}(\theta) = \frac{1 - \cos(\theta)}{2} = \sin^2\left(\frac{\theta}{2}\right) . \quad (2.4)$$

Finalmente, a distância $d_{n,j}$ (Equação 2.2) em função da latitude e longitude pode ser obtida tomando o *haversine* inverso para $h = \text{hav}(\Theta_{n,j})$

$$d_{n,j} = R \Theta_{n,j} = R \text{arc hav}(h) = 2R \text{arc sen}(h) . \quad (2.5)$$

Por fim, vale destacar que as equações introduzidas neste capítulo descrevem o cerne do processo de regressão utilizado para generalizar as medições RF para todas as células do *gridmap*. Esse algoritmo é parte indispensável do sistema de localização que será apresentado na Seção 3.6 e discutido ao longo dos próximos capítulos.

2.5 TRABALHOS RELACIONADOS

A literatura a respeito da utilização de algoritmos de aprendizado de máquina para o processo de construção de mapas de rádio é bastante difundida (LIU et al., 2007; TIMOTEO et al., 2015; VO; DE, 2015; ZEKAVAT; BUEHRER, 2011). Em sua grande maioria, os trabalhos disponíveis fazem uso de *gridmaps* cujas células são uniformemente distribuídas sobre a região de cobertura, conforme uma malha quadriculada. Há ainda estudos cujo foco são as medidas de similaridade aplicadas na etapa de predição, como a distância de Mahalanobis e divergência de Kullback-Leibler (MONDAL; TURKKA; HENTTONEN, 2014). Entretanto, pouca investigação existe a respeito do impacto da escolha de *gridmaps* com diferentes geometrias na acurácia da estimativa de localização.

Em (MA; ZHANG; HAN, 2015), os autores propuseram um modelo de *gridmap* no qual a área das células é definida com base na sua distância para a ERB. O conceito explorado pelos autores reside na observação de que a qualidade do sinal nos arredores da ERB é maior do que em regiões mais distantes. Nas regiões vizinhas à ERB, o sinal é mais potente, o que reduz a influência do ruído e demais efeitos que causam atenuação do sinal. Por meio de uma geometria radial, o estudo propõe que a área designada para

as células no *gridmap* diminua à medida que elas se afastem da ERB. O objetivo dessa diminuição na área das células é aumentar a resolução nessas regiões. Ao passo que o sinal irradia a partir da ERB, além da perda de potência devido à própria propagação, o ruído e demais efeitos de atenuação tendem a impactar de forma mais severa a qualidade do sinal. Sendo assim, essas flutuações poderiam ser melhor capturadas utilizando uma área menor para as células nessas regiões. Os experimentos conduzidos no estudo avaliaram respectivamente dois e três níveis de segmentação para o *gridmap* proposto. Os resultados, obtidos através de simulação computacional, indicaram que o *gridmap* proposto forneceu uma melhora de até 16% na acurácia da estimativa de localização quando comparado com o *gridmap* de malha quadricular.

Em (KIM; MIN; YEO, 2014), os autores propuseram uma técnica para localização *out-door* baseada em *RSS fingerprinting* utilizando redes Wi-Fi, onde um teste de significância estatística foi desenvolvido para determinar se células adjacentes no *gridmap* podem, ou não, ser agrupadas. O método estatístico é aplicado entre os vetores RSS de duas células vizinhas para determinar sua similaridade, definindo assim, um critério de agrupamento. Esse algoritmo, quando aplicado ao conjunto de medições RSS obtidas para redes Wi-Fi, produziu uma segmentação irregular da região de cobertura. Embora os resultados pareçam promissores, os autores alertam que a melhora na acurácia da estimativa de localização se deve, em parte, a um aumento no número de células do *gridmap* promovido pelo processo aplicado. O método de segmentação do *gridmap* baseado em um teste de significância estatística introduz novas células, fato esse que introduziu um viés na comparação com o *gridmap* de malha quadricular. Apesar disso, o percentual de melhora na acurácia da estimativa de localização é positiva. Os resultados mostraram que a melhora na acurácia variou entre 25,7% até 30,1%. Já com relação ao *gridmap*, o modelo de segmentação proporcionou um aumento de 11,2% no número de células.

Outra abordagem envolvendo *gridmaps* foi apresentada por nós em (BITTENCOURT; URBANO; CUNHA, 2018), onde em vez de investigar a geometria das células, os autores exploraram a influência na forma como as células são distribuídas ao longo da região de cobertura. No referido estudo, optamos por distribuir as células do *gridmap* de forma a refletir a malha viária dentro da região de cobertura. A principal hipótese que corrobora a utilização desse modelo consiste no fato de que, em ambientes urbanos, os usuários de dispositivos móveis têm grande probabilidade de serem encontrados em ruas, avenidas, ou em calçadas e passarelas próximas a essas vias. Os resultados obtidos proporcionaram uma redução média de 69,7% no número de células, com acurácia similar ou até mesmo melhor que a apresentada pelos *gridmaps* de malha quadricular (melhoria de 4m em média). A redução significativa no número de células permitiu ainda aumentar a densidade das células nas regiões de interesse, melhorando ainda mais a acurácia do modelo proposto. Entretanto, os autores alertam que um determinado viés no conjunto de medições RF pode ter influenciado positivamente nos resultados. No estudo em questão, o conjunto de

medições RF foi criado através de um procedimento de *drive-testing*. Desta forma, todas as medições obtidas na região de cobertura estavam, inevitavelmente, posicionadas sobre a malha viária daquela região.

Uma tentativa de mitigar um eventual viés introduzido pelo conjunto de medições RF foi proposta e posteriormente avaliada por um trabalho independente. Em (ANDRADE, 2018), o autor utilizou um simulador de redes para criar conjuntos de medições RF com diferentes proporções, considerando medições presentes na malha viária e em regiões fora dela, incluindo o interior de edificações. O estudo corroborou com os resultados relativos a redução do número de células, formalizando as taxas de redução implicadas pela utilização do modelo proposto, em relação ao *gridmap* de malha quadricular. O estudo mostrou ainda que a densidade de ocupação urbana parece não influenciar no comportamento do *gridmap* com relação à acurácia. A redução do número de células também foi constatada para ambientes suburbanos e rurais. Por fim, o estudo afirma que para um percentual de medições em vias urbanas acima de 70%, o *gridmap* proposto é mais indicado. Para uma análise mais cuidadosa desse percentual, seria necessário considerar *gridmaps* com o mesmo número de células. Infelizmente, apenas um único par de *gridmaps* atendia a esse requisito, fato esse que limitou a análise.

3 NOVA PROPOSTA PARA GRIDMAPS IRREGULARES

Esse Capítulo tem como objetivo generalizar o modelo de segmentação para *gridmaps* irregulares introduzidos por (BITTENCOURT; URBANO; CUNHA, 2018). Para fins didáticos, primeiramente será apresentada uma ferramenta que permite simular a propagação RF — *Radio Frequency* (RF) — dentro de um perímetro urbano, dotado de diversos elementos tais como quarteirões, ruas, avenidas e edificações. Em seguida, serão apresentadas as novas propostas para os algoritmos de segmentação, tendo como base para experimentação a região de cobertura definida através do ambiente simulado.

3.1 CRIAÇÃO DO AMBIENTE SIMULADO

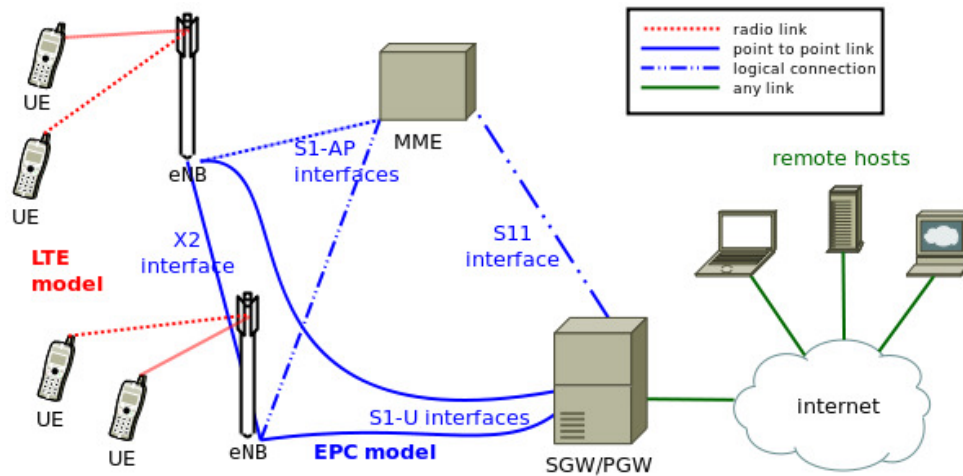
Para a realização dos experimentos, foi utilizado um simulador de eventos discretos para redes de computadores chamado *Network Simulator 3* (NS-3) (NSNAM, 2011). Entre as extensões disponíveis para esse simulador, o módulo LENA possui componentes que implementam os padrões *Long-Term Evolution* (LTE) e *Evolved Packet Core* (EPC) correspondentes a tecnologia de telefonia móvel de quarta geração (4G). O componente LENA LTE é responsável por simular a comunicação via rádio entre as Estações Rádio Base (ERBs) e as Estações Móveis (EMs). No contexto do módulo LENA, uma ERB é denotada por *Evolved Node-B* (eNB), enquanto uma EM é representada pelo termo *User Equipment* (UE). Apesar da nomenclatura do simulador, vamos utilizar os termos ERB e EM, respectivamente, para fazer referência a tais elementos.

Diferentemente do LENA LTE, o propósito do LENA EPC é simular a comunicação entre os elementos centrais da rede de telefonia celular. A Figura 4 ilustra a topologia de uma rede de comunicação 4G, na qual podemos identificar as fronteiras de atuação dos padrões LTE e EPC no sistema de comunicação. Para este trabalho, foi utilizado apenas o componente LTE do módulo LENA, já que, em última instância, ele é o responsável pela comunicação entre as ERBs e as EMs. O componente LTE permite simular os dados referentes à camada física da rede de acesso, incluindo observar a potência do sinal de RF (medida em *Watt*) ao longo de um determinado intervalo de tempo.

Uma propriedade importante do módulo LENA é a capacidade de adicionar edificações ao ambiente de simulação através de um recurso chamado *Building Models*. Também é possível configurar a perda de potência associada ao percurso para EMs localizadas em ambiente *indoor*, utilizando o *Building Propagation Model*. Entre os parâmetros disponíveis para simulação, estão o tipo de edificação (comercial, residencial ou empresarial), o material da edificação (madeira ou concreto), além de outras propriedades, tais como a presença de janelas, a quantidade de aposentos e o número de andares.

Para modelar um quarteirão, foi utilizada a classe *GridBuildingAllocator* que permite

Figura 4 – Topologia de rede estabelecida pelo padrão LTE/EPC. É possível perceber a distinção entre os elementos na camada LTE (*LTE model* destacado em vermelho), daqueles presentes na camada EPC (*EPC model* destacado em azul).



Fonte: CCTC (2011)

posicionar as edificações em uma área retangular bidimensional. Foram criados quarteirões com 6 edificações cada, seguindo uma distribuição 2×3 , ou seja, duas edificações no eixo x e 3 edificações no eixo y . Cada edificação, por sua vez, possui 3 apartamentos no eixo x e 5 apartamentos no eixo y , todos com uma dimensão de $10\text{ m} \times 10\text{ m}$. A distância entre duas edificações adjacentes é de 10 m . O número de andares em cada edificação é definido aleatoriamente dentro de uma faixa que vai de, no mínimo, 1 até, no máximo, 9 andares. As edificações foram classificadas como residenciais, possuindo paredes de concreto e janelas, segundo os parâmetros definidos através do *Building Models*.

Uma vez realizada a modelagem do quarteirão, é possível modelar o perímetro urbano. Para isso, foi utilizada a classe *GridBlockAllocator* para posicionar os quarteirões segundo uma distribuição 16×7 , ou seja, 16 quarteirões no eixo x e 7 no eixo y . A separação entre os quarteirões é feita por meio de avenidas cuja largura foi definida em 25 m . Por fim, o perímetro urbano resultante consiste em um retângulo com dimensões $1495\text{ m} \times 1340\text{ m}$, com uma área de aproximadamente 2 km^2 . Vale ressaltar que o perímetro urbano gerado corresponde à região de cobertura onde desejamos encontrar as EMs.

Para posicionar as ERBs, foi utilizada a classe *LteHexGridEnbTopologyHelper*, que distribui torres de telefonia com ERBs tri-setoriais segundo uma malha de geometria hexagonal. Para a região de cobertura, foram criadas 3 ERBs tri-setoriais, cada uma delas com 3 setores, com uma antena por setor. As torres de telefonia foram fixadas em posições conhecidas segundo os vértices de um triângulo equilátero de lado 750 m , de modo que as torres estão equidistantes 750 m entre si. Como as ERBs são tri-setoriais, cada antena cobre um setor de 120 graus em relação ao eixo central de cada uma das torres. Os parâmetros para a simulação de cada uma das antenas são apresentados na Tabela 1. Esses parâmetros foram obtidos com base nas suposições e considerações propostas no

documento 3GPP R4-092042 sobre simulações (3GPP, 2009).

Tabela 1 – Parâmetros de simulação do módulo LENA LTE para as ERBs.

Parâmetro	Valor
Modelo de Antena	Parabólica
Potência Emitida	46 dBm
Frequência da Portadora	2 GHz
Altura da Torre	30 m
Blocos de Recurso	25

Fonte: Elaborada pelo autor (2020)

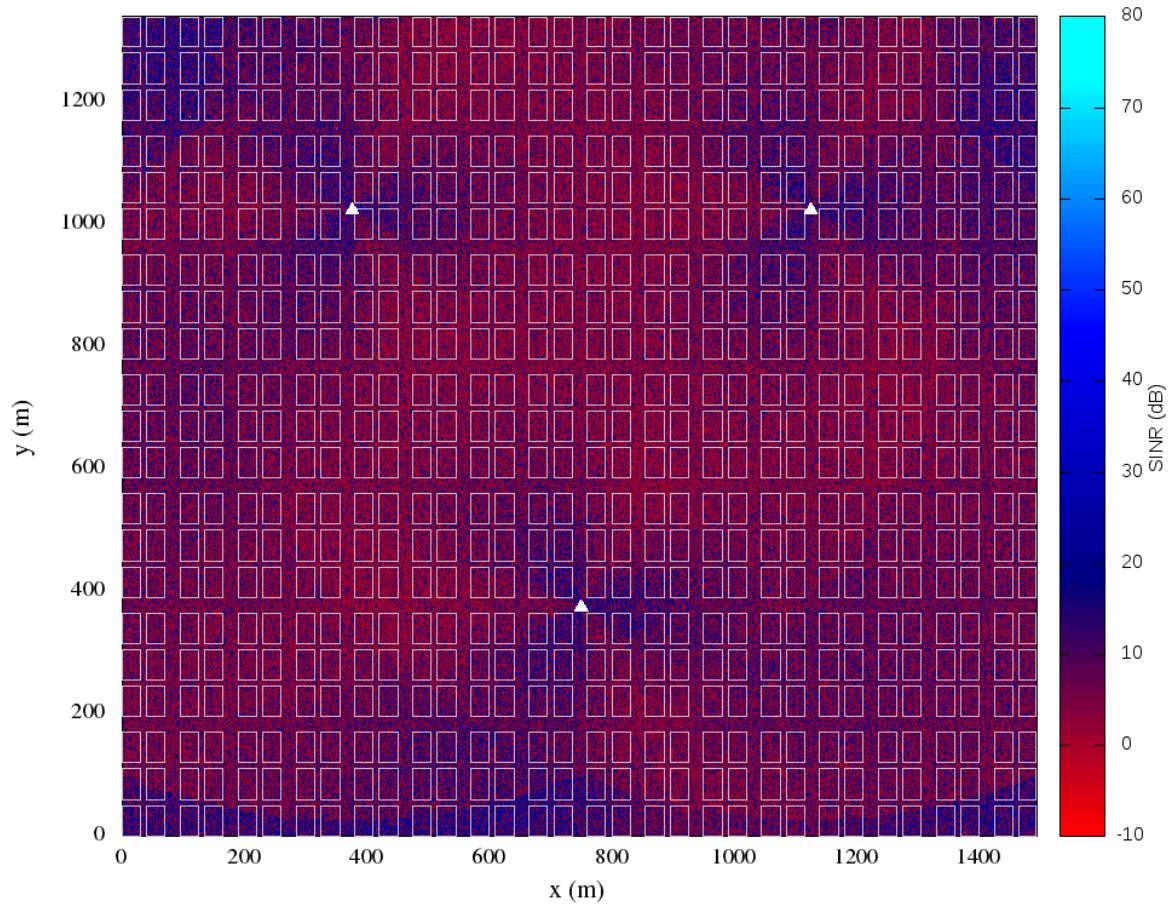
Outro recurso do módulo LENA permite gerar uma representação 2D da qualidade do sinal na região de cobertura. Essa representação, chamada de *Radio Environment Map* (REM), fornece uma visualização qualitativa de como a potência do sinal está distribuída ao longo da região de cobertura. O REM é construído com base na relação sinal-interferência-mais-ruído — *Signal-to-Interference-plus-Noise Ratio* (SINR) — do *downlink* em relação à ERB de melhor sinal em cada ponto do mapa. A Figura 5 mostra o REM para o perímetro urbano construído utilizando os recursos do módulo LENA.

Uma característica importante nas simulações com o LENA é o modelo de mobilidade utilizado. O módulo LENA LTE impõe a exigência de que todo objeto em uma simulação deve se deslocar conforme um modelo de mobilidade estabelecido. Para garantir que existam objetos estáticos, cuja posição não varia ao longo do tempo, a propriedade *ConstantPositionMobilityModel* deve ser atribuída a tais objetos. Sendo assim, os objetos que correspondem a edificações e a ERBs assumem esse valor para o seu modelo de mobilidade. Já para os objetos que representam as EMs, foi associada a propriedade *SteadyStateRandomWaypointMobilityModel*, garantindo que tais objetos sejam posicionados dentro de uma dada região específica com posição inicial e direção aleatórias.

Em última instância, o objetivo de uma simulação no módulo LENA LTE é obter um conjunto de medições RF que representa um procedimento de *drive-testing* ao longo da região de cobertura. Por esse motivo, as EMs foram posicionadas de forma aleatória dentro de uma região específica. As regiões escolhidas constituem um percurso dentro do perímetro urbano, de forma a espelhar o que ocorreria na prática. Como os percentuais de medições *indoor* e *outdoor* são importantes para o experimento, os terminais *outdoor* foram distribuídos aleatoriamente ao longo das avenidas que compõem o percurso, enquanto os terminais *indoor* foram posicionados de forma aleatória em cinco quarteirões vizinhos a essas avenidas. Os terminais espalhados nos quarteirões podem tanto estar dentro das edificações, quanto nos espaços entre edificações adjacentes.

Nas simulações realizadas, a velocidade de deslocamento das EMs foi definida constante em 3 km/h. Além disso, foram distribuídas 3.000 EMs ao longo do percurso definido.

Figura 5 – REM obtido para o ambiente urbano simulado. Os triângulos identificam as torres com ERBs tri-setoriais, já os retângulos correspondem as edificações.

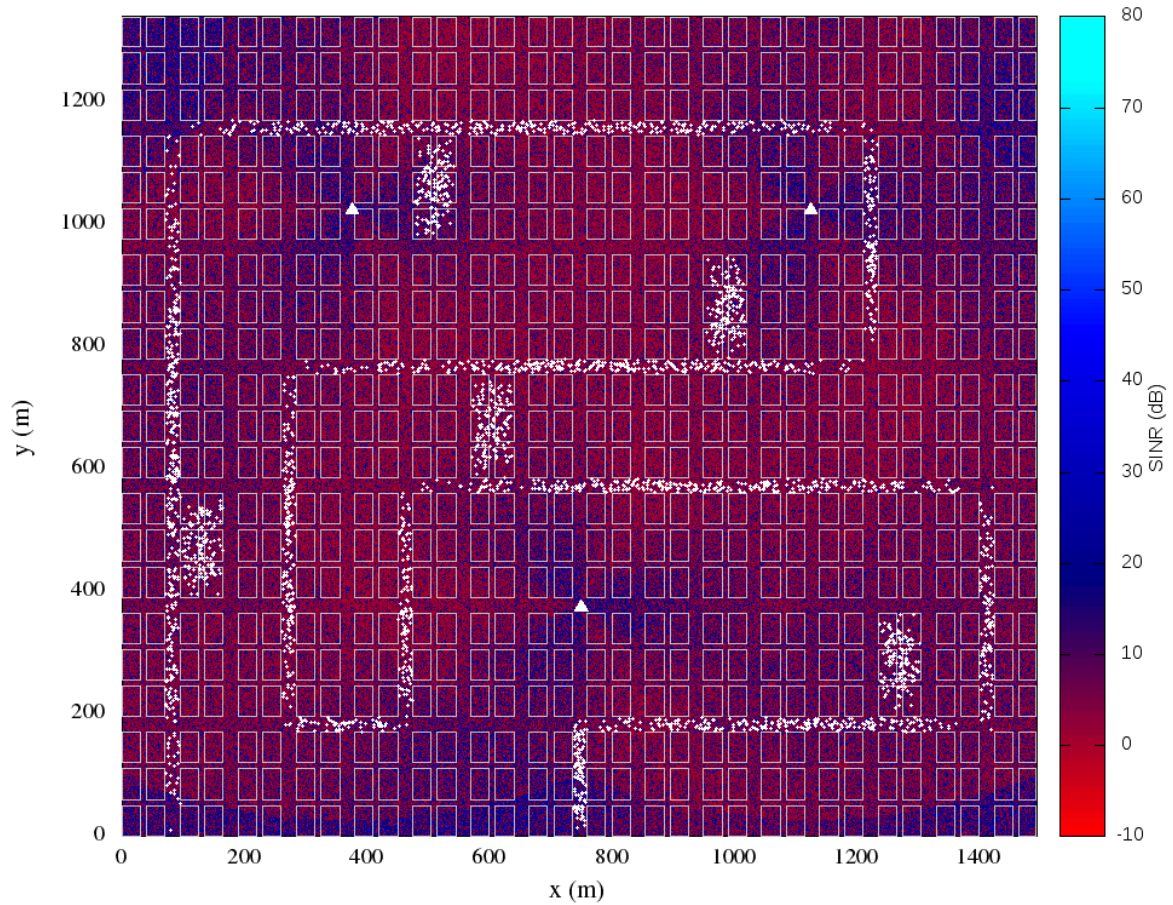


Fonte: Elaborada pelo autor (2020)

Esse valor foi escolhido visto que se aproxima do conjunto de medições RF utilizado no estudo (BITTENCOURT; URBANO; CUNHA, 2018). A Figura 6 ilustra a disposição das EMs ao longo do percurso estabelecido, onde as EMs estão representadas pelos pontos brancos. Vale indicar que, na figura em questão, 33% das EMs estão localizadas em ambientes *indoor*. Os percentuais de terminais *indoor* e *outdoor* são parametrizáveis, sendo obrigatoriamente definidos para cada simulação.

O modelo de propagação foi definido com base na classe *HybridBuildingsPropagationLossModel*. A documentação do módulo LENA LTE informa que essa classe combina diversos modelos de propagação para simular comunicações *outdoor*, *indoor*, de *outdoor* para *indoor* e de *indoor* para *outdoor* (CCTC, 2011). Os modelos de propagação utilizados pela classe são os modelos de Okumura Hata, COST231, ITU-R P.1411 (curta distância) e ITU-R P.1238 (ambiente *indoor*). Além disso, a classe *HybridBuildingsPropagationLossModel* tem o poder de combinar o modelo de propagação com as propriedades das edificações estabelecidas através dos *Building Models*.

Figura 6 – EMs distribuídas de modo a simular um *drive-testing* no perímetro urbano. Foram posicionadas 3.000 EMs, sendo 33% em ambientes *indoor*.



Fonte: Elaborada pelo autor (2020)

O modelo de propagação escolhido é capaz de levar em consideração perdas no percurso associadas à presença de paredes internas e externas presentes nas edificações, além dos efeitos de sombreamento estimados segundo uma distribuição log-normal com média nula e desvio padrão variável. O desvio padrão para as medições obtidas segundo o modelo de propagação do simulador é definido individualmente por ambiente, conforme indicado na Tabela 2. O elemento *Paredes Externas* refere-se a um modelo de propagação híbrido sendo utilizado para representar a comunicação entre um nó em ambiente *indoor* com outro em ambiente *outdoor*. O valor designado para representar a perda de potência quando o sinal atravessa paredes internas foi estabelecido em 10 dB , enquanto a altura média das edificações foi definida em 20 m . Esses valores foram escolhidos com base na especificação de referência utilizada para as simulações (3GPP, 2009; 3GPP, 2010a).

Quanto ao modelo de desvanecimento do sinal, foi selecionado o *Extended Typical Urban Model* (ETU), cujos parâmetros foram escolhidos conforme as especificações descritas no documento 3GPP TS 36.104 (3GPP, 2010b). A inclusão desse modelo gera flutuações

Tabela 2 – Parâmetros do modelo de propagação utilizado na simulação.

Elemento	σ (dB)
Ambiente <i>Outdoor</i>	4
Ambiente <i>Indoor</i>	3
<i>Paredes Externas</i>	10

Fonte: Elaborada pelo autor (2020)

nos níveis de sinal mensurados nas EMs que participam da simulação. Essas flutuações são esperadas e tem origem nos efeitos da propagação multipercurso, na qual diferentes componentes do sinal chegam ao destino com diferentes fases. Assim, o valor da potência recebida varia ao longo do tempo no decorrer de uma simulação.

3.2 SIMULANDO O CONJUNTO DE MEDIÇÕES RF

Uma vez definidos os parâmetros e os modelos do experimento, juntamente com a posição das ERBs, edificações e a distribuição das EMs, é possível iniciar o processo de simulação para a obtenção de um conjunto de medições RF. A dinâmica de uma simulação funciona da seguinte forma: a cada milissegundo, o módulo LENA disponibiliza dados de camada física, como o *Received Signal Strength* (RSS). Esses dados devem ser agrupados e, conforme a documentação, uma média aritmética deve ser transmitida para as camadas superiores da pilha de comunicação a cada 200 *ms*. Sendo assim, considerar o tempo total da simulação como 200 *ms* é suficiente para obter um valor aproximado do RSS para cada uma das EMs que participam da simulação.

Tabela 3 – Amostra dos valores de RSS em *dBm* obtidos para as ERBs.

x	y	RSS_1	RSS_2	RSS_3	RSS_4	RSS_5	RSS_6	RSS_7	RSS_8	RSS_9
75,9	1141,7	-144,5	-124,6	-143,6	-104,0	-86,4	-98,2	-146,4	-134,2	-138,9
81,8	404,9	-109,9	-95,7	-105,5	-116,4	-116,5	-84,1	-154,8	-149,3	-135,1
838,2	176,1	-87,6	-98,6	-85,9	-97,5	-119,0	-97,6	-110,0	-117,1	-97,9

Fonte: Elaborada pelo autor (2020)

Uma limitação do módulo LENA reside no fato de que, para uma determinada EM, o valor do RSS é fornecido apenas para a ERB à qual a EM está conectada. Como existem 9 ERBs, a construção do conjunto de medições exige que o nível de RSS para uma EM seja obtido para as 9 ERBs que fazem parte do experimento. Uma alternativa para essa restrição é executar 9 simulações, uma para cada ERB que compõe o experimento. Os valores de RSS para cada ERB devem então ser agregados para formar o conjunto de medições cujas entradas correspondem às ilustradas na Tabela 3, em que as colunas x e

y correspondem a posição da EM e as colunas RSS_1 a RSS_9 indicam os níveis de sinal, em dBm , com relação a cada uma das 9 ERBs.

Vale ressaltar que, para cada ERB, existem 200 medições em relação a uma determinada EM, sendo uma medição a cada 1 ms durante um período de 200 ms . O agrupamento das medições é feito com base no par (CellId, IMSI). O CellID é o parâmetro responsável por identificar unicamente a ERB, enquanto o *International Mobile Subscriber Identity* (IMSI) identifica unicamente a EM. Outra atividade efetuada na etapa de agregação é a conversão dos valores de RSS, originalmente medidos em *Watt*, para dBm . Essa conversão foi realizada por meio da seguinte expressão

$$P_{dBm} = 10 \log_{10} \left(\frac{P_W}{1\text{ mW}} \right) . \quad (3.1)$$

Além do RSS, outro parâmetro RF obtido foi o *Propagation Delay* (PD). Esse parâmetro corresponde a uma medida quantizada referente ao atraso de propagação do sinal eletromagnético. Esse parâmetro é calculado por intermédio da latência de propagação, considerando o tempo de viagem completa de ida e volta do sinal (*Round-Trip Time - RTT*), sendo utilizado para fornecer uma estimativa da distância entre a EM e a ERB. O parâmetro PD assume valores inteiros no intervalo de 0 a 56, onde cada incremento representa uma distância de 234 m . O valor máximo $PD = 56$ é utilizado para representar distâncias maiores que $13,1\text{ km}$. Esse parâmetro é obtido nas redes de telefonia baseadas no padrão LTE através de um canal de controle específico.

As restrições para obtenção do PD usando o simulador são análogas aquelas impostas para obtenção do RSS. Desse modo, foram executadas 9 simulações, cada uma delas considerando uma única ERB. Ao final, os valores obtidos para o PD foram agrupados para EM nas colunas PD_1 até PD_9 , conforme indicado na Tabela 4.

Tabela 4 – Amostra dos parâmetros PD no conjunto de medições RF.

x	y	PD_1	PD_2	PD_3	PD_4	PD_5	PD_6	PD_7	PD_8	PD_9
75.9	1141.7	6	6	6	4	4	4	3	3	3
81.8	404.9	4	4	4	1	1	1	6	6	6
838.2	176.1	6	6	6	7	7	7	1	1	1

Fonte: Elaborada pelo autor (2020)

Por fim, com base no procedimento de agregação apresentado e tomando como premissa uma distribuição de $1/3$ ($\sim 33,33\%$) de medições *indoor* e $2/3$ ($\sim 66,66\%$) de medições *outdoor*, foi gerado um conjunto de medições de RF com 3.000 registros.

3.3 ALGORITMOS PARA A GERAÇÃO DOS GRIDMAPS

Conforme discutido anteriormente, nas técnicas de localização baseadas em *RF fingerprinting*, a região de cobertura, ou seja, aquela na qual deseja-se localizar as EMs, é tipicamente modelada por meio de uma malha quadricular, também conhecida como *gridmap* (LOVISOLO; CAMPOS, 2012). Cada célula dessa malha é representada pelas coordenadas geográficas correspondentes a sua posição central. Para efeitos deste trabalho, qualquer modelagem da região de cobertura usando uma malha quadricular será chamada de *gridmap* regular. Diferentemente dos *gridmaps* regulares, os *gridmaps* irregulares consistem em modelagens alternativas dessa mesma região de cobertura. Essa modelagem alternativa, tenta tirar proveito de alguma característica a respeito da propagação eletromagnética dos sinais de RF na região (MA; ZHANG; HAN, 2015), ou exploram características qualitativas da região (KIM; MIN; YEO, 2014), promovendo um arranjo diferente das células.

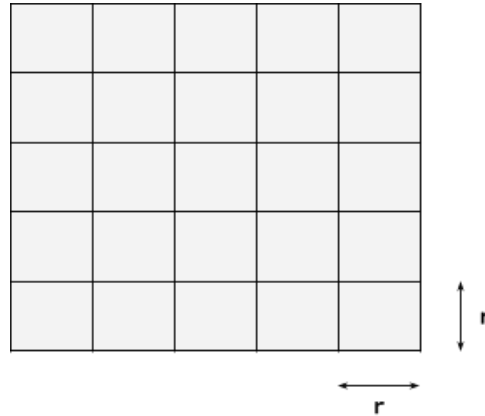
Outro aspecto relevante para se destacar é o impacto dos *gridmaps* nos sistemas de localização. Tipicamente, quanto maior for a quantidade de células utilizadas em um *gridmap*, menor tende a ser o erro associado à estimativa de localização, visto que é possível representar a região de cobertura com mais detalhes, ou seja, com uma maior resolução. Entretanto, a adição indiscriminada de células aumenta a complexidade computacional associada ao problema, pois o processamento computacional envolvido deve ser realizado individualmente para cada célula do *gridmap*, tanto durante a fase de treinamento (etapa *offline*), quanto posteriormente na fase de predição (etapa *online*).

A principal conclusão derivada dos experimentos envolvendo *gridmaps* irregulares aponta para o fator complexidade computacional, como a grande vantagem associada ao emprego desse tipo de modelagem (BITTENCOURT; URBANO; CUNHA, 2018). O estudo mostrou que, ao estabelecer 50 m como a faixa de erro médio associada a estimativa de localização, a utilização de um modelo de *gridmap* irregular proporcionou em uma redução de até seis vezes no número de células quando comparado ao *gridmap* regular.

Na literatura, os métodos para geração de *gridmaps* regulares são parametrizados através da *resolução*. Este parâmetro define o comprimento do lado associado a uma célula quadrada conforme indicado pela variável r na Figura 7. Em termos práticos, a *resolução* assume valores que variam em um intervalo de 5 a 30 m em média (LOVISOLO; CAMPOS, 2012). Entretanto, utilizar o parâmetro *resolução* no processo de geração de *gridmaps* irregulares pode não ser muito adequado (BITTENCOURT; URBANO; CUNHA, 2018). Como o processo de segmentação é diferente para cada modelo de *gridmap*, um determinado valor r para resolução pode gerar em um número diferente de células, dependendo se o modelo do *gridmap* é regular ou irregular. Essa característica dificulta a comparação entre modelos sob a perspectiva da complexidade computacional.

Nesse sentido, para oferecer um critério justo de comparação entre os diferentes modelos de *gridmap*, o primeiro passo é modificar seu processo de geração para utilizar, ao invés da resolução, o número de células como parâmetro. As evidências indicam que *gridmap*

Figura 7 – Exemplo de um *gridmap* regular com resolução r .

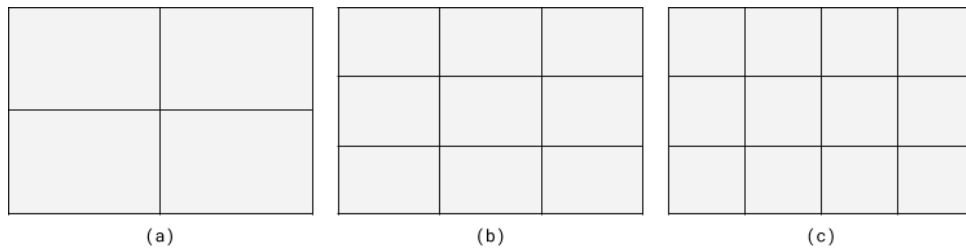


Fonte: Elaborada pelo autor (2020)

regular possui uma maior complexidade computacional, ou seja, um maior número de células (BITTENCOURT; URBANO; CUNHA, 2018), por esse motivo, o *gridmap* regular foi escolhido para balizar a taxa de crescimento do número de células.

A função que governa a taxa de crescimento do número de células em uma malha quadricular consiste na função quadrática $f(x) = x^2$, na qual $f(x)$ corresponde ao número de células. A Figura 8 evidencia esta característica, uma vez que é possível perceber que um incremento em uma unidade no número de células nos eixos verticais e horizontais acarreta em um aumento quadrático no número total de células.

Figura 8 – A taxa de crescimento do número de células é governada pela função quadrática $f(x) = x^2$. Em (a), (b) e (c), temos respectivamente $f(2) = 4$, $f(3) = 9$ e $f(4) = 16$.

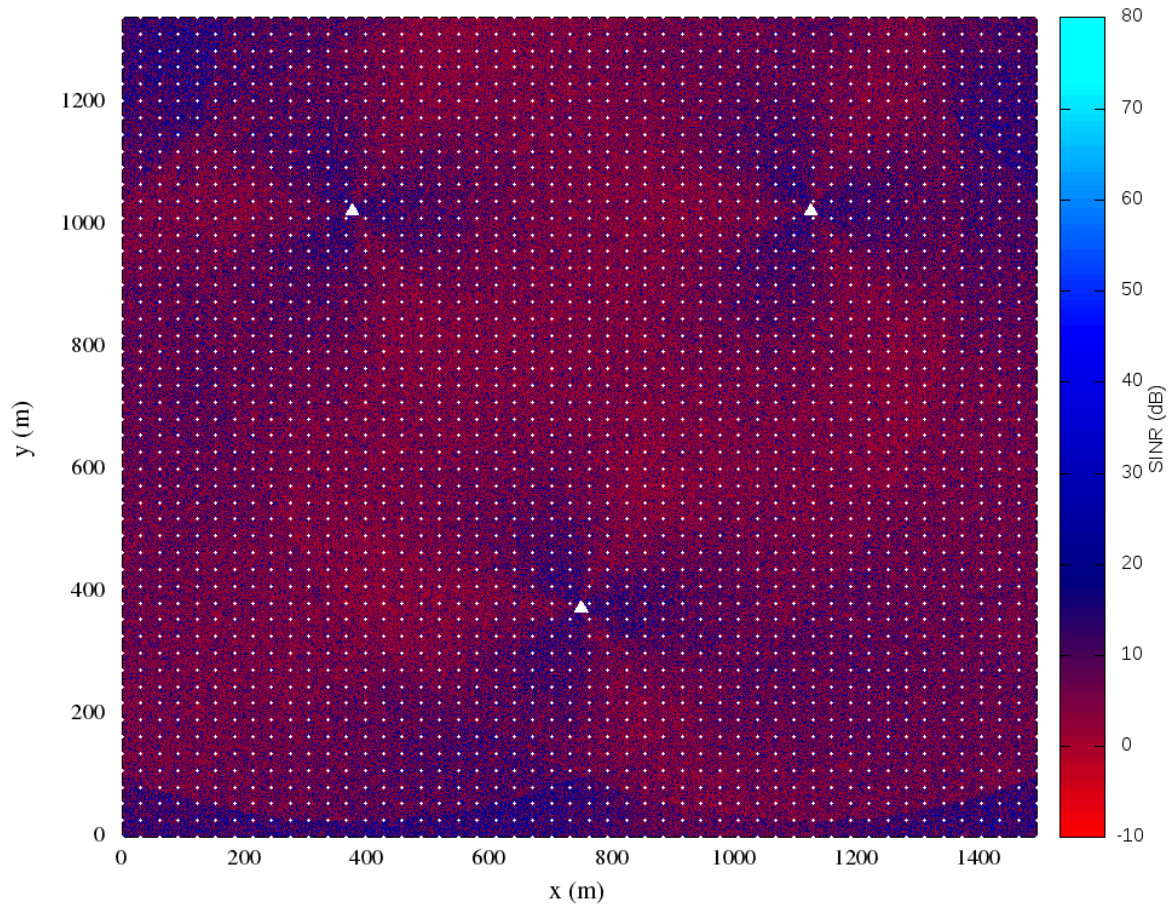


Fonte: Elaborada pelo autor (2020)

Em posse dessa compreensão sobre a taxa de crescimento das células em um *gridmap* regular, foi criado um algoritmo capaz de segmentar a região de cobertura através uma malha quadricular, onde o número de células foi definido como parâmetro para esse algoritmo. O método em questão, foi utilizado para segmentar o perímetro urbano obtido a partir do simulador NS-3 descrito na Seção 3.1. A Figura 9 ilustra o resultado obtido através do algoritmo para essa região quando o número de células selecionado foi 2.500. Os pontos brancos observados na imagem representam o centro das células, enquanto os triângulos indicam a posição das ERBs. Vale destacar que o *gridmap* regular definido por 2.500 células, possui uma *resolução* de aproximadamente 30 m.

Uma vez estabelecido o processo de segmentação que define um *gridmap* regular, onde o número de células é parametrizável, o próximo passo consiste em produzir um esquema de segmentação para o *gridmap* irregular. Nesse sentido, o único requisito definido até então é que a quantidade de células nesse modelo também seja parametrizável.

Figura 9 – Representação de um *gridmap* regular com 2.500 células. Os pontos brancos simbolizam o centro das células no *gridmap*, enquanto os triângulos representam as posições das ERBs.



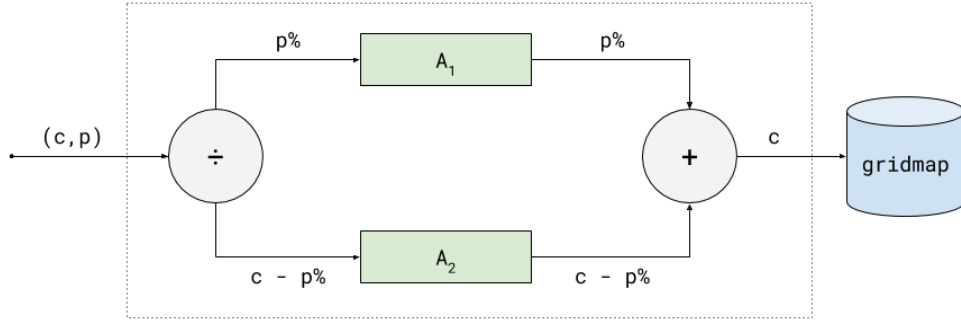
Fonte: Elaborada pelo autor (2020)

A generalização proposta para *gridmaps* irregulares neste trabalho considera a hipótese de que as medições RF utilizadas para treinar o modelo de localização, carregam consigo informação sobre a tendência de distribuição espacial das EMs. A princípio, essa informação não deveria ser descartada como acontece no caso da segmentação regular, mas sim transmitida para o modelo de localização através do *radiomap*.

No intuito de criar uma proposta de segmentação que transmita informação sobre a distribuição espacial das medições RF originais, foi idealizado um esquema de geração conforme ilustrado pelo diagrama da Figura 10. Nesse esquema, além do número total de células c , o processo de geração recebe como entrada um parâmetro p , que indica um percentual do total de células que devem ser posicionadas de modo a melhor representar

a distribuição espacial do conjunto de medições RF original. A região sobre a qual as medições RF se distribuem é chamada região prioritária.

Figura 10 – Diagrama em blocos para segmentação de *gridmaps* irregulares. O parâmetro c representa a quantidade total de células no *gridmap*, enquanto p corresponde a um percentual do total de células designadas às regiões prioritárias.



Fonte: Elaborada pelo autor (2020)

De maneira geral, o parâmetro p no diagrama da Figura 10, é responsável por particionar c em dois subconjuntos. O primeiro subconjunto, com $p\%$ elementos, corresponde a quantidade de células que devem ser distribuídas sobre as regiões prioritárias, sendo o bloco sinalizado por A_1 , o responsável realizar esse procedimento. Já o segundo subconjunto, com $c - p\%$ elementos, corresponde à quantidade de células que devem ser uniformemente distribuídas na região de cobertura, tal qual proporcionado por uma segmentação de malha quadricular, sendo esse último caso, incubido ao bloco A_2 .

O primeiro passo no intuito de satisfazer o esquema de segmentação definido pela Figura 10 é resolver o bloco A_1 . Para distribuir as i células nas regiões prioritárias, foi utilizada uma implementação do algoritmo *Farthest-First Traversal* (GONZALEZ, 1985). Esse algoritmo propõe uma solução gulosa (*greedy*) para um problema de otimização combinatorial conhecido na teoria dos grafos como *metric k-center* ou *metric facility location* (WILLIAMSON; SHMOYS, 2011). Informalmente, esse problema pode ser descrito sob a seguinte perspectiva: dado um conjunto S de locais em uma cidade, deseja-se instalar uma quantidade k de caixas eletrônicos de modo a minimizar a distância máxima necessária para chegar em um caixa eletrônico a partir de um local $j \in S$.

A noção intuitiva desse algoritmo é que os k elementos selecionados, devem, a princípio, ser aqueles que melhor aproximem a distribuição espacial original de S . Desse modo, o algoritmo inicia escolhendo arbitrariamente um ponto qualquer e segue selecionando pontos, de modo que o próximo ponto escolhido possui a maior distância para o conjunto de pontos previamente selecionados. Como critério de escolha, os novos pontos selecionados devem estar o mais distante possível daqueles pontos já escolhidos.

A formalização desse problema em teoria da computação descreve que para um grafo completo, não-direcionado $G = (V, E)$, onde V é o conjunto de vértices e E o conjunto de arestas, temos definida uma função *distance* como métrica espacial para o conjunto

V com $d_{ii} = 0$ e $d_{ij} = d_{ji}$ para qualquer $i, j \in V$, de modo que as distâncias satisfaçam a *desigualdade triangular*, ou seja, para $x, y, z \in V$ temos que $d_{xy} + d_{yz} \geq d_{xz}$. Para cada conjunto de vértices $S \subseteq V$, $d(i, S)$ corresponde a menor distância entre $i \in V$ e um vértice em S . O objetivo do problema é encontrar um conjunto S com cardinalidade $|S| = k$, de modo que $\max_{i \in V} d(i, S)$ seja minimizada.

O problema *metric k-center* é considerado NP-hard. Dessa forma, o processo para encontrar a solução ótima é computacionalmente custoso, possuindo um tempo de execução exponencial em função do tamanho da entrada. Existem alguns algoritmos de aproximação que fornecem soluções polinomiais no tempo de execução em função do tamanho da entrada e fornecem uma razão de aproximação de $\alpha = 2$. Seja $C^*(I)$, o custo da solução ótima para uma entrada I , enquanto $C(I)$ é o custo da solução aproximada. A fração $C^*(I)/C(I) \leq \alpha$ define a razão de aproximação entre o algoritmo ótimo e a solução otimizada, sendo o custo da solução otimizada avaliado no pior caso, ou seja, através do limitante superior definido segundo a notação *Big O*.

O Algoritmo 1 consiste em uma implementação do *Farthest-First Traversal* (GONZALEZ, 1985). Essa implementação fornece uma razão de aproximação $\alpha = 2$ para solucionar o problema *metric k-center* com uma complexidade no tempo limitada superiormente por $O(nk)$, em que n e k são, respectivamente, as cardinalidades dos conjuntos V e S . Na prática, o valor de k define a quantidade de iterações para a entrada n .

Algoritmo 1: Farthest-First Traversal

Entrada: $G = (V, E)$; k

Saída: S

- 1: $S \leftarrow \emptyset$
 - 2: $i \xleftarrow{R} V$ { *randomly select* $i \in V$ }
 - 3: $V \leftarrow V - \{i\}$
 - 4: $S \leftarrow S \cup \{i\}$
 - 5: **enquanto** $|S| \leq k$ **faça**
 - 6: $j \leftarrow \arg \max_{j \in V} d(j, S)$
 - 7: $V \leftarrow V - \{j\}$
 - 8: $S \leftarrow S \cup \{j\}$
 - 9: **fim do enquanto**
 - 10: **retorne** S
-

O Algoritmo 1 inicia selecionando arbitrariamente um vértice qualquer $i \in V$. Em seguida, o coloca no conjunto S de vértices escolhidos. Uma vez que um vértice é selecionado, ele é removido do conjunto V . Enquanto $|S| \leq k$, o algoritmo segue selecionando um novo vértice $j \in V$ de modo que a distância $d(j, S)$ seja maximizada. O algoritmo termina quando $|S| = k$, retornando o conjunto S que foi escolhido.

Existem outros algoritmos que podem ser utilizados para solucionar o problema *metric k-center* com razão de aproximação $\alpha = 2$ ou até inferior. Um exemplo é o algoritmo *parametric pruning* (VAZIRANI, 2003). Esse algoritmo tenta encontrar o conjunto inde-

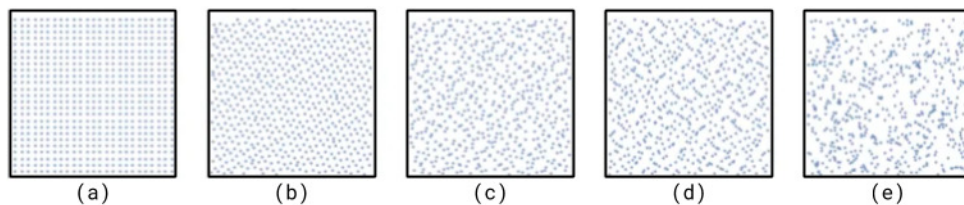
pendente máximo (*maximal*) a partir da construção G^2 do grafo $G(E, V)$. Esse método também oferece um fator de 2 para razão de aproximação. Atualmente, o melhor algoritmo de aproximação possui um fator de 1,488 (LI, 2011).

A implementação para o algoritmo *Farthest-First Traversal* fornecida, resolve o bloco superior (A_1) do esquema para segmentação de *gridmaps* irregulares mostrado no diagrama da Figura 14. Para concluir o procedimento de geração, falta ainda detalhar uma possível solução para o bloco inferior (A_2) mostrado no diagrama. Conforme detalhado anteriormente, o objetivo do bloco A_2 é distribuir as $c - p\%$ células restantes de maneira uniforme sobre a região de cobertura, de maneira similar aquela proporcionada pelo algoritmo de segmentação por malha quadricular.

A distribuição uniforme de um conjunto finito de pontos em uma dada área também recai na classe de problemas cuja solução depende de algoritmos de otimização. Em particular, essa classe de problemas é composta por problemas geométricos conhecidos como *packing problems* (SCHEITHAUER, 2017). Uma solução alternativa para esse tipo de problema é utilizar uma sequência quasi-aleatória de baixa discrepância como método para derivar deterministicamente um conjunto de pontos, de modo a reduzir a probabilidade de agrupamentos (discrepâncias), garantindo assim, que os pontos sejam distribuídos uniformemente no espaço (*blue-noise sampling*) (ROBERTS, 2018).

Uma extensão da sequência aditiva recursiva R para 2-dimensões foi escolhida para distribuir uniformemente o restante das células (ROBERTS, 2018). A principal motivação para escolha dessa sequência foram os resultados obtidos quando comparados a outras sequências de baixa discrepância como as sequências de Halton e Sobol (Figura 11). Outra vantagem da sequência R reside no fato de que, diferentemente das sequências de Sobol e Halton, sua extensão em d -dimensões não exige a seleção prévia de parâmetros. No caso das sequências de Halton e Sobol, algumas escolhas de parâmetros são problemáticas e acabam por não exibir uma baixa discrepância.

Figura 11 – Alternativas para distribuição uniforme de pontos sobre uma superfície: (a) Lattice regular. Três sequências de baixa discrepância em 2-dimensões, sendo (b) R , (c) Halton e (d) Sobol; (e) sequência amostrada de uma distribuição aleatória.



Fonte: Adaptado de ROBERTS (2018)

A extensão da sequência aditiva recursiva R em d -dimensões, sem introduzir a necessidade de seleção de parâmetros, é possível graças a uma generalização para a razão áurea introduzida por (KRÇADINAC, 2006). Seja ϕ_d a razão áurea generalizada para d -dimensões,

a sequência quasi-aleatória de baixa discrepância R_d é definida da seguinte forma

$$\begin{aligned} t_n &= s_0 + n\alpha \pmod{1}, \quad n = 1, 2, 3, \dots, \\ \alpha &= \left(\frac{1}{\phi_1}, \frac{1}{\phi_2^2}, \frac{1}{\phi_3^3}, \dots, \frac{1}{\phi_d^d} \right), \end{aligned} \quad (3.2)$$

em que t_n corresponde ao n -ésimo termo obtido através da sequência, s_0 é tido como o grau de liberdade e ϕ_d é a única raiz positiva do polinômio $x^{d+1} = x + 1$.

Algoritmo 2: n -termos da sequência quasi-aleatória R_2

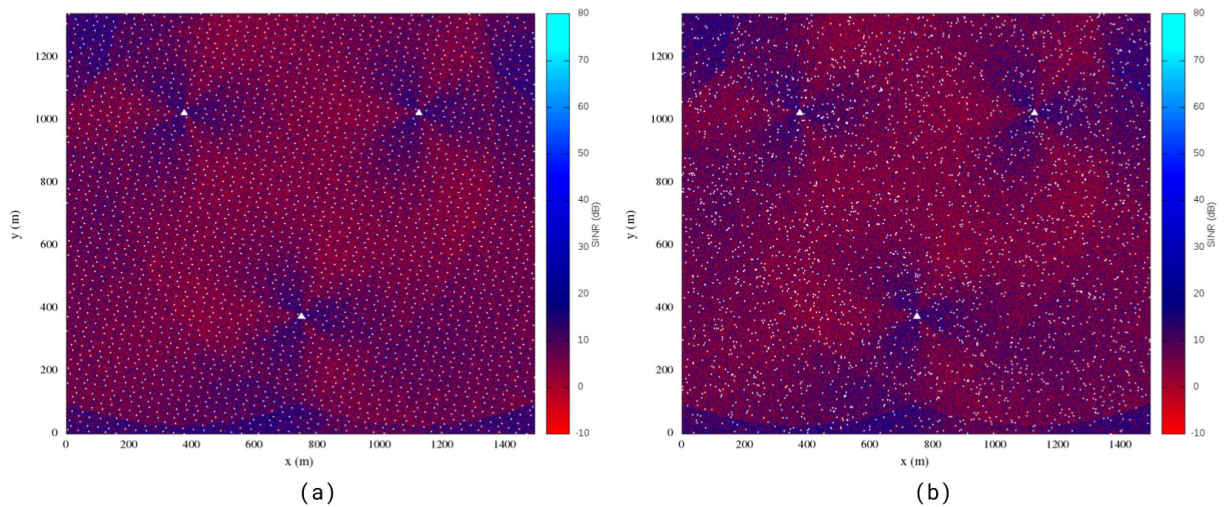
Entrada: n, w, h

Saída: S

- 1: $S \leftarrow \emptyset$
 - 2: $(s_{0_x}, s_{0_y}) \leftarrow (\frac{1}{2}, \frac{1}{2})$
 - 3: **enquanto** $i \leq n$ **faça**
 - 4: $(t_{i_x}, t_{i_y}) \leftarrow (s_{0_x} + i\alpha_1, s_{0_y} + i\alpha_2)$
 - 5: $(x_i, y_i) \leftarrow (wt_{i_x}, ht_{i_y})$
 - 6: $S \leftarrow S \cup \{ (x_i, y_i) \}$
 - 7: **fim do enquanto**
 - 8: **retorne** S
-

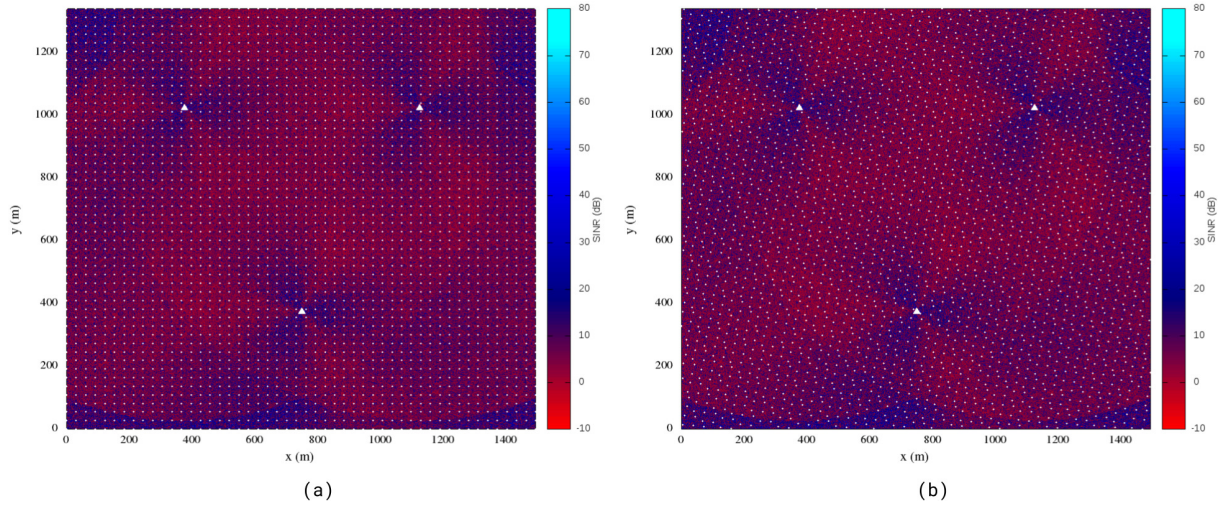
Vale salientar que o valor escolhido para s_0 não altera as características da sequência. É comum utilizar o valor $s_0 = 0$, mas definir esse valor pode oferecer um grau adicional de liberdade adicional, permitindo derivar valores deslocados para os elementos da sequência. No contexto desse trabalho, o valor de s_0 foi definido como $1/2$. Além disso, considerando $d = 2$, foi assumido que $\phi_1 = 1,61803398$ e $\phi_2 = 1,32471795$.

Figura 12 – *Gridmaps* obtidos com base em $c = 3.000$ células. (a) Geração por sequência quasi-aleatória R_2 . (b) Geração por sequência pseudo-aleatória.



Fonte: Elaborada pelo autor (2020)

Figura 13 – *Gridmaps* obtidos com base em $c = 2.500$ células utilizando: (a) malha quadricular (*gridmap* regular) e (b) sequência R_2 (*gridmap* irregular).



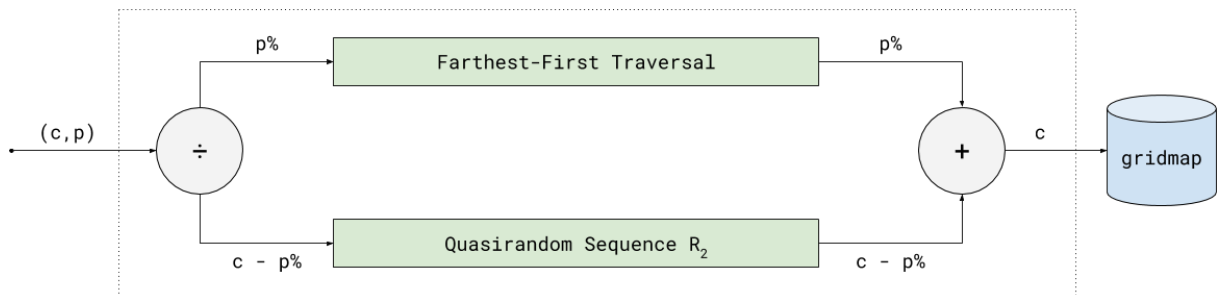
Fonte: Elaborada pelo autor (2020)

O procedimento utilizado para produzir n -termos dessa sequência, onde cada termo possui $d = 2$ dimensões, foi implementado no Algoritmo 2, com os parâmetros de entrada definidos por n , w e h . O parâmetro n está relacionado a quantidade de termos a serem gerados pela sequência. Os parâmetros w e h correspondem, respectivamente, aos limites de largura e altura da região de cobertura. O resultado é um vetor S com n elementos, onde cada elemento corresponde a um par de coordenadas $(x, y) \in \mathbb{R}^2$.

No intuito de comparar o resultado obtido através da sequência R_2 com aquele fornecido por uma sequência pseudo-aleatória, foi gerado um par de *gridmaps* com número de células $c = 3.000$. A Figura 12 ilustra a diferença entre a distribuição das células nos dois *gridmaps* obtidos. Fica claro, a partir dos *gridmaps* representados, que a sequência R_2 fornece como característica um menor agrupamento (baixa discrepância), diferentemente das características observadas no *gridmap* pseudo-aleatório.

A Figura 13 mostra dois *gridmaps* gerados a partir de $c = 2.500$ células. O primeiro foi

Figura 14 – Diagrama com o esquema final para criação de *gridmaps* irregulares. Os blocos A_1 e A_2 indicados no diagrama da Figura 14 foram substituídos pelas implementações do *Farthest-First Traversal* (Algoritmo 1) e *Quasirandom Sequence R_2* (Algoritmo 2).

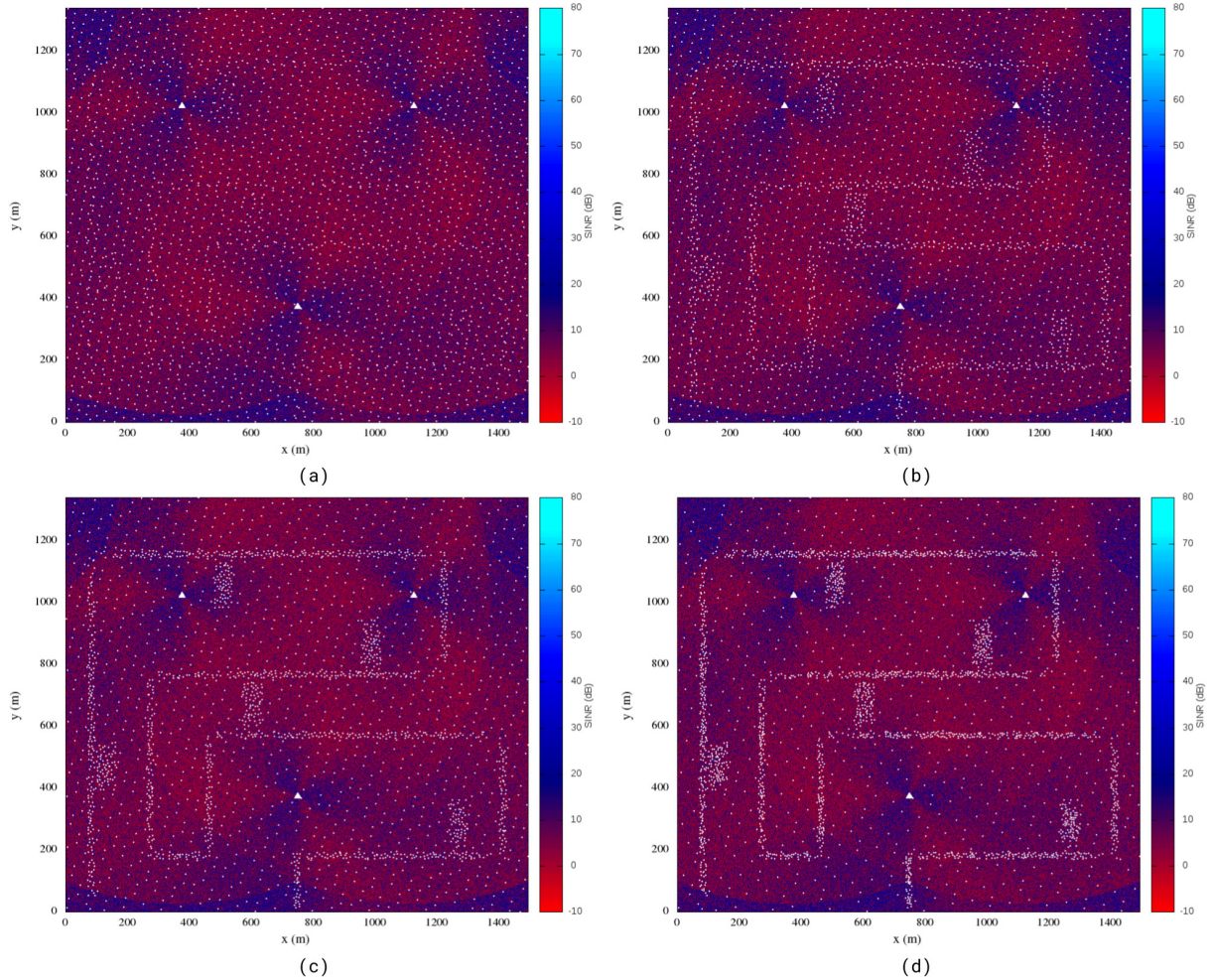


Fonte: Elaborada pelo autor (2020)

gerado utilizando o método tradicional para segmentação da área por meio de uma malha quadricular, ao passo que o segundo fez uso da sequência R_2 . A baixa discrepância da distribuição das células possibilita concluir que a sequência R_2 oferece uma distribuição uniforme comparável àquela do *gridmap* de malha quadricular, sendo uma escolha razoável para o esquema de geração de *gridmaps* irregulares.

O algoritmo obtido empregando a sequência R_2 constitui no elemento que faltava para finalizar o esquema de geração de *gridmaps* irregulares descrito pelo diagrama da Figura 10. O Algoritmo 2 corresponde a implementação proposta para o bloco A_2 apresentado no diagrama. Vale lembrar que para o bloco A_1 , foi utilizada a implementação para o *Farthest-First Traversal* detalhada no Algoritmo 1. A Figura 14 representa o diagrama final do esquema de geração de *gridmaps* irregulares percebe-se que as implementações dos blocos A_1 e A_2 foram substituídas respectivamente pelos Algoritmos 1 e 2.

Figura 15 – *Gridmaps* irregulares obtidos com base em $c = 2.500$ células e $p\%$ de células dispostas nas regiões prioritárias, acompanhando a distribuição original das medições RF. (a) $p = 10\%$. (b) $p = 30\%$. (c) $p = 50\%$. (d) $p = 70\%$.



Fonte: Elaborada pelo autor (2020)

A Figura 15 mostra 4 modelos de *gridmap* gerados por meio do sistema descrito

pelo diagrama em blocos da Figura 14. Os *gridmaps* em questão foram concebidos com $c = 2.500$ células cada. A diferença entre eles é a quantidade de células designadas às regiões prioritárias. Em (a), um total de 10% das células foram distribuídas nessas regiões. Por outro lado, em (b), (c) e (d) foram destacadas, respectivamente, 30%, 50% e 70% das células, aumentando a densidade de cobertura nas regiões prioritárias.

Definidos os procedimentos para geração de *gridmaps*, podemos unir esse recurso àqueles obtidos através do simulador NS-3, como a definição do perímetro urbano e o conjunto de medições RF, para então, propor um modelo definitivo para o sistema de localização que viabilize a comparação entre as diferentes propostas de *gridmaps*. Antes, entretanto, é essencial definir uma notação matemática para descrever precisamente esses objetos. Essa notação será importante no Capítulo 4, onde serão discutidos os resultados para acurácia do sistema, obtidos a partir da comparação entre modelos. Dessa forma, as próximas Seções deste Capítulo estão dedicadas a esse propósito.

3.4 NOTAÇÃO PARA ALGORITMOS E GRIDMAPS

Conforme discutido, os *gridmaps* são gerados basicamente segundo dois parâmetros: c e p . O primeiro corresponde ao número absoluto de células do *gridmap*, já o segundo corresponde a um percentual que representa a parcela das células do *gridmap* que serão destinadas a regiões prioritárias. Sejam $\mathbf{C}$ e $\mathbf{P}$ os conjuntos, respectivamente, dos quais os valores c e p são selecionados, tal que

$$\mathbf{C} = \{ 100 x^2 \} \quad x \in \mathbb{N}, \quad 5 \leq x \leq 14, \quad (3.3)$$

em que o conjunto $\mathbf{C}$, denotado pelo conjunto de Células, possui 10 diferentes valores para a quantidade de células no *gridmap*. Seguindo a equação $100 x^2$ e respeitando o intervalo definido para x teremos os valores 2.500, 3.600, até 19.600. Já para o caso de $\mathbf{P}$, temos o conjunto de Proporções, sendo definido por

$$\mathbf{P} = \{ 0,1 y \} \quad y \in \mathbb{N}, \quad 1 \leq y \leq 10, \quad (3.4)$$

compreendendo os elementos 0.1, 0.2 até 1.0, que correspondem respectivamente aos percentuais 10%, 20% até 100% intercalados por um fator de 10. Esses percentuais representam a quantidade de células que serão dispostas nas regiões prioritárias.

Para formular as estratégias de geração de *gridmaps*, foram definidos quatro algoritmos geradores, denotados por S , R , Q e Λ . Os três primeiros algoritmos recebem como entrada o número absoluto de células do *gridmap*, ou seja, $c \in \mathbf{C}$. Nesse sentido, sejam $S(c)$, $R(c)$ e $Q(c)$, aplicações dos algoritmos S , R e Q para um determinado valor de c , respectivamente. O algoritmo S corresponde àquele responsável pela geração de *gridmaps* usando uma malha quadricular. O algoritmo R concebe os *gridmaps* por meio de uma distribuição

pseudo-aleatória das células dentro da região de cobertura. Por fim, o algoritmo Q utiliza a sequência R_2 (quasi-aleatória) para criar os *gridmaps*.

A motivação por trás da simbologia adotada para S , R e Q remete ao procedimento para geração destes *gridmaps*. A notação S por exemplo, foi escolhida por remeter ao termo ***S*quare**, os *gridmaps* gerados através desse algoritmo serão detonados *gridmaps* quadriculares. Para o algoritmo R , a notação remete ao termo ***R*andom**, devido a uma tecnicidade desse método, os *gridmaps* gerados através do algoritmo R serão classificados como *gridmaps* pseudo-aleatórios. Já a notação para algoritmo Q por sua vez, foi derivada da palavra ***Q*uasirandom**, nesse sentido, os *gridmaps* gerados pelo algoritmo Q serão detonados por *gridmaps* quasi-aleatórios. Com exceção dos *gridmaps* quadriculares, classificados como *gridmaps* regulares, os *gridmaps* obtidos através dos demais algoritmos de segmentação encaixam-se na categoria de *gridmaps* irregulares.

Considere $\mathbf{S}_c$, $\mathbf{R}_c$ e $\mathbf{Q}_c$, três conjuntos distintos de *gridmaps*, gerados, respectivamente, por meio dos algoritmos S , R e Q . Sejam $\mathbf{S}_c$, $\mathbf{R}_c$ e $\mathbf{Q}_c$ definidos tal que

$$\begin{aligned}\mathbf{S}_c &= \{S(c)\} \quad c \in \mathbf{C} , \\ \mathbf{R}_c &= \{R(c)\} \quad c \in \mathbf{C} \text{ e} \\ \mathbf{Q}_c &= \{Q(c)\} \quad c \in \mathbf{C} ,\end{aligned}\tag{3.5}$$

em que a quantidade de elementos nesses conjuntos corresponde precisamente a cardinalidade do conjunto $\mathbf{C}$, ou seja, $|\mathbf{S}_c| = |\mathbf{R}_c| = |\mathbf{Q}_c| = |\mathbf{C}| = 10$.

Por fim, considere o algoritmo Λ , que define a estratégia de geração de *gridmaps* irregulares, ilustrada pelo diagrama em blocos apresentado na Figura 14. A principal característica identificada através dos resultados produzidos pelo algoritmo é sua capacidade de *adaptar* o *gridmap* ao conjunto de medições RF originais, por essa razão, os *gridmaps* gerados por esse algoritmo serão denotados por *gridmaps* adaptativos. Vale ressaltar que neste contexto, o termo adaptativo não faz referência a algoritmos adaptativos (ZAKNICH, 2005), logo, a expressão algoritmo para gerar *gridmaps* adaptativos, não implica, necessariamente, que o algoritmo de geração satisfaz a definição de (ZAKNICH, 2005).

Diferentemente dos algoritmos S , R e Q , a descrição da estratégia para implementação de Λ utiliza ambos os parâmetros c e p . Desse modo, uma aplicação específica do algoritmo Λ é representada por $\Lambda(c, p)$, em que $c \in \mathbf{C}$ e $p \in \mathbf{P}$. O conjunto $\mathbf{\Lambda}_{cp}$ que contém todos os *gridmaps* gerados por aplicações $\Lambda(c, p)$ é definido como

$$\mathbf{\Lambda}_{cp} = \{\Lambda(c, p)\} \quad c \in \mathbf{C} , \quad p \in \mathbf{P} ,\tag{3.6}$$

em que a cardinalidade de $\mathbf{\Lambda}_{cp}$ corresponde a $|\mathbf{\Lambda}_{cp}| = |\mathbf{C} \times \mathbf{P}| = 100$. Por fim, o conjunto união $\mathbf{G} = \mathbf{S}_c \cup \mathbf{R}_c \cup \mathbf{Q}_c \cup \mathbf{\Lambda}_{cp}$ corresponde ao conjunto de todos *gridmaps* avaliados nesse estudo. Vale destacar que, ao todo, serão avaliados 130 *gridmaps* distintos, conforme indicado pela cardinalidade do conjunto $\mathbf{G}$ ($|\mathbf{G}| = 130$).

3.5 DEFINIÇÃO PARA AS MEDIÇÕES RF

A Seção 3.2 descreve a obtenção das medições RF através de uma simulação utilizando o módulo LENA LTE do simulador NS-3. Em suma, além das coordenadas geográficas, cada registro no conjunto de medições RF possui associado um vetor de nível de sinal (RSS) e um vetor com os parâmetro de atraso na propagação (PD).

Ainda segundo a Seção 3.2, foi assumido na simulação que as torres de telefonia possuem antenas tri-setoriais. Sendo assim, apesar de sua posição espacial ser a mesma, as ERBs posicionadas nessas torres possuem azimutes diferentes. Por conveniência, um dado conjunto de ERBs cuja posição espacial é a mesma é chamado de grupo, sendo a quantidade de grupos de ERBs no experimento denotado por N_g . Analogamente, o número de ERBs em um grupo é denotado por N_e . O conjunto de medições RF é representado por $\mathbf{Z}$, tendo sua cardinalidade dada por $|\mathbf{Z}| = z$.

Seja $\mathbf{\Omega}^{(n)}$, o vetor que representa o conjunto de medições RSS para a n -ésima amostra coletada. O vetor $\mathbf{\Omega}^{(n)}$ é descrito em termos de N_g e N_e da seguinte forma

$$\mathbf{\Omega}^{(n)} = \{\omega_{ij}^{(n)}\} \quad i = 1, \dots, N_g \quad \text{e} \quad j = 1, \dots, N_e, \quad (3.7)$$

em que $\omega_{ij}^{(n)}$ corresponde ao valor RSS medido em relação à j -ésima ERB pertencente ao i -ésimo grupo. Analogamente, $\mathbf{\Gamma}^{(n)}$ consiste no vetor que representa a medição do parâmetro PD para a n -ésima amostra, sendo $\mathbf{\Gamma}^{(n)}$ dado por

$$\mathbf{\Gamma}^{(n)} = \{\gamma_{ij}^{(n)}\} \quad i = 1, \dots, N_g \quad \text{e} \quad j = 1, \dots, N_e, \quad (3.8)$$

em que $\gamma_{ij}^{(n)}$ corresponde ao valor PD obtido para a j -ésima ERB pertencente ao i -ésimo grupo. Por fim, combinando as coordenadas geográficas aos vetores $\mathbf{\Omega}^{(n)}$ e $\mathbf{\Gamma}^{(n)}$ temos que

$$\mathbf{Z} = \{(\mathbf{E}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Gamma}^{(n)}) \in \mathbb{R}^2 \times \mathbb{R}^N \times \mathbb{R}^N ; n = 1, \dots, z ; N = N_e N_g\}, \quad (3.9)$$

representa o vetor que contém o conjunto de medições RF, em que $\mathbf{E}^{(n)}$ corresponde às coordenadas geográficas da n -ésima amostra definida sobre $\mathbb{R}^2$ e $N = N_e N_g$ corresponde ao número total de ERBs definidas pelo experimento.

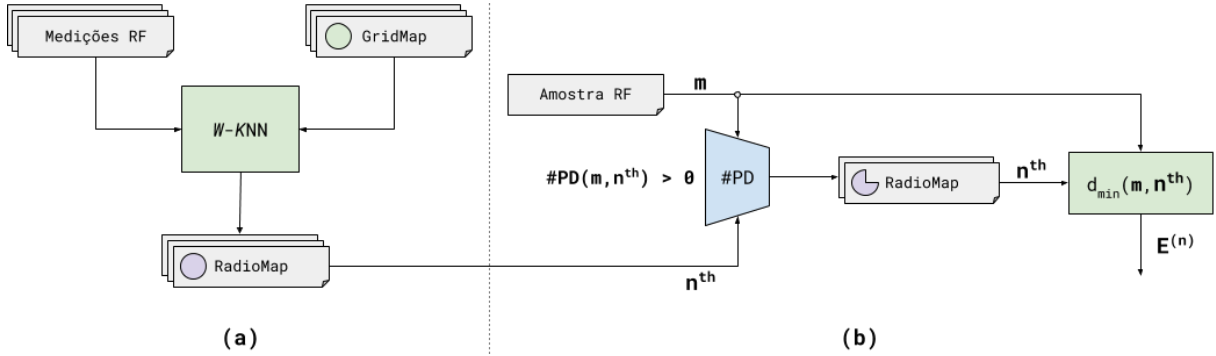
3.6 SISTEMA DE LOCALIZAÇÃO

O sistema de localização utilizado para avaliar os *gridmaps* gerados na Seção 3.3 segue as características gerais para sistemas de localização baseados em *RF fingerprinting*. Uma arquitetura genérica para sistemas de localização foi discutida na Seção 2.3.

A Figura 16 ilustra o diagrama em blocos para a instância do sistema de localização utilizado neste estudo. Esse foi o sistema empregado na avaliação dos *gridmaps* concebidos na Seção 3.3. Vale destacar que tal sistema possui as mesmas características e os elementos

arquiteturais apresentados na Seção 2.3. Conforme delineado na Seção 2.3, a etapa de treinamento, ou simplesmente etapa *offline*, é a responsável por conceber como saída o mapa de rádio (*radiomap*) da região de cobertura. Para alcançar esse objetivo, a etapa *offline* utiliza como entrada os pontos do *gridmap* juntamente com o conjunto de medições RF obtidas na região de cobertura (Figura 16(a)).

Figura 16 – Sistema de localização utilizado nos experimentos. O elemento $\#PD(m, n^{th})$ atua como um filtro, estabelecendo uma correspondência entre o vetor PD da amostra m e do n -ésimo elemento (n^{th}) do mapa de rádio (*radiomap*). A partir do *radiomap* filtrado, o elemento n^{th} selecionado é aquele que possui a menor distância $d_{min}(m, n^{th})$ em relação a amostra m .



Fonte: Elaborada pelo autor (2020)

O mapa de rádio gerado durante a etapa *offline* possui a mesma geometria do *gridmap*, ou seja, os elementos do *radiomap* estão dispostos espacialmente na mesma posição das células presentes no *gridmap*. Analogamente, assim como ocorre para as medições RF, um elemento do *radiomap* possui, além das coordenadas geográficas, um vetor RSS e um vetor PD. O *radiomap* é denotado por Υ , com cardinalidade $|\Upsilon| = v$, sendo definido por

$$\Upsilon = \{(\mathbf{E}^{(n)}, \mathbf{\Omega}^{(n)}, \mathbf{\Gamma}^{(n)}) \in \mathbb{R}^2 \times \mathbb{R}^N \times \mathbb{R}^N ; n = 1, \dots, v; N = N_e N_g\} . \quad (3.10)$$

Conforme apresentado na Seção 2.4, a instância do regressor baseado no algoritmo *Weighted K-Nearest Neighbors* (WK-NN) é responsável por derivar um vetor RSS para cada elemento do *radiomap*. O algoritmo WK-NN foi escolhido por apresentar bons resultados para sistemas de localização, o algoritmo fornece um bom equilíbrio entre o custo computacional e a acurácia obtida. Já o vetor PD pode ser computado deterministicamente, uma vez que a coordenada geográfica de um dado elemento no *radiomap* é conhecida. Nesse intuito, basta tomar a distância entre o elemento e a ERB. Em seguida, discretizar o valor para PD, conforme indicado na Seção 3.2.

A obtenção da estimativa de localização ocorre na fase *online*, também chamada de etapa de predição (Figura 16(b)). A estimativa de localização é sempre obtida em relação a uma amostra qualquer. Seja $\mathbf{A}^m$, uma amostra obtida por uma EM, para a qual se

deseja determinar a posição. Essa amostra também possui associado os vetores RSS (Ω^m) e PD (Γ^m) medidos em relação às ERBs do conjunto de medições RF

$$\mathbf{A}^m = \{(\Omega^m, \Gamma^m) \in \mathbb{R}^N \times \mathbb{R}^N ; N = N_e N_g\} , \quad (3.11)$$

em que os vetores Ω^m e Γ^m são análogos àqueles presentes no mapa de rádio Υ . Vale destacar ainda que diferentemente dos elementos em Υ , a amostra $\mathbf{A}^m$ não possui um vetor de coordenadas espaciais associado. As expressões que definem Ω^m e Γ^m são

$$\begin{aligned} \Omega^m &= \{\omega_{ij}^m\} \quad i = 1, \dots, N_g \quad , \quad j = 1, \dots, N_e \quad \text{e} \\ \Gamma^m &= \{\gamma_{ij}^m\} \quad i = 1, \dots, N_g \quad , \quad j = 1, \dots, N_e \quad . \end{aligned} \quad (3.12)$$

Um elemento importante da etapa de predição encontra-se ilustrado através do bloco #PD presente na Figura 16b. Conforme descrito na Seção 3.2, o parâmetro PD corresponde a uma medida quantizada referente ao atraso de propagação do sinal e pode ser utilizado para estimar a distância entre a EM e a ERB em intervalos de 234 m . Por exemplo, um valor $PD = 0$ indica que a EM está localizada em um raio de 0 a 234 m da ERB. Para um valor $PD = 1$, a EM está contida na circunscrita na região definida entre as circunferências de raio 234 m e 468 m , respectivamente.

Em suma, o objetivo desse bloco é utilizar o PD para fornecer uma estimativa preliminar da posição da EM. Essa estimativa permitiria reduzir o espaço de busca, atuando como um filtro para selecionar possíveis candidatos dentro dos elementos do *radiomap*. Tipicamente, reduzir o espaço de busca implica em uma diminuição do custo computacional. Caso essa seleção prévia não seja realizada, torna-se necessário comparar todos os registros do *radiomap* em relação à amostra no intuito de determinar aquela de maior semelhança. A utilização de filtros tende a melhorar o desempenho, minimizando o tempo de resposta do sistema de localização (LOVISOLO; CAMPOS, 2012).

Conforme a Seção 3.1, existem $N_g = 3$ torres de telefonia constituídas por $N_e = 3$ ERBs. Sendo assim, o valor do parâmetro PD obtido em relação a um grupo específico deve ser o mesmo, já que todas as ERBs em um grupo compartilham a mesma posição. Entretanto, cada ERB em um grupo possui uma antena tri-setorial com abertura de 120 graus apontando em uma direção específica. Devido a essa característica, nem sempre existirá uma linha de visada entre a ERB e a EM. Como efeito, o valor do parâmetro PD obtido para um determinado grupo pode variar. Por esse motivo, o bloco #PD realiza primeiramente um pré-processamento do vetor Γ^m de modo a escolher o menor valor do parâmetro PD para um determinado grupo de ERBs. A justificativa para essa escolha é que a ERB com o menor atraso deve possuir a melhor linha de visada em relação à EM. Esse pré-processamento é realizado tanto para o vetor PD da amostra Γ^m , quanto para

os vetores $\mathbf{\Gamma}^{(n)}$ dos elementos que compõem o *radiomap* (TIMOTEO, 2020).

$$\begin{aligned}\mathbf{\Gamma}^m &= \bigcup_{i=1}^{Ng} \min\{\gamma_{ij}^m\} \quad j = 1, \dots, N_e \\ \mathbf{\Gamma}^{(n)} &= \bigcup_{i=1}^{Ng} \min\{\gamma_{ij}^{(n)}\} \quad j = 1, \dots, N_e\end{aligned}\tag{3.13}$$

Em seguida, o bloco #PD computa os vetores diferença $\mathbf{\Delta}^{(n)}$ e $\mathbf{\Delta}^m$ em relação a cada grupo. O objetivo do vetor diferença é conferir maior robustez a seleção de elementos do *radiomap*, minimizando a influência caso as amostras tenham sido obtidas em ambiente *indoor*. Os efeitos de atenuação da potência do sinal nesses ambientes são mais acentuados, acarretando em um maior atraso na propagação, o que leva a valores mais elevados do parâmetro PD. A utilização do vetor diferença de PD para todas as amostras reduz a ocorrência desse tipo de problema (TIMOTEO, 2020).

$$\begin{aligned}\mathbf{\Delta}^m &= \{\gamma_u^m - \gamma_v^m\} \quad u = v = 1, \dots, N_g ; u \neq v ; u < v \\ \mathbf{\Delta}^{(n)} &= \{\gamma_u^{(n)} - \gamma_v^{(n)}\} \quad u = v = 1, \dots, N_g ; u \neq v ; u < v\end{aligned}\tag{3.14}$$

Com base nos valores do vetor diferença, o Algoritmo 3 é utilizado para eleger os elementos do mapa de rádio que serão selecionados para comparação. O algoritmo recebe como entrada o vetor $\mathbf{\Delta}^m$, em seguida procura por correspondências dos elementos em todos os vetores $\mathbf{\Delta}^{(n)}$ do *radiomap*, caso haja uma correspondência, o elemento do *radiomap* é selecionado. No algoritmo os elementos do *radiomap* são representados pelo conjunto $\mathbf{\Upsilon}$, sendo a n -ésima amostra desse conjunto denotada por $\mathbf{\Upsilon}^{(n)}$. Vale ressaltar ainda que, caso o filtro não seja capaz de selecionar nenhum elemento, todos os elementos do *radiomap* serão fornecidos como resultado do algoritmo.

Algoritmo 3: Seleção de correspondências do parâmetro PD

Entrada: $\mathbf{\Upsilon}$, $\mathbf{\Delta}^m$

Saída: $\mathbf{\Upsilon}_r$ {espaço de busca reduzido}

- 1: **enquanto** $n \leq |\mathbf{\Upsilon}|$ **faça**
 - 2: **se** $\mathbf{\Delta}^m \cap \mathbf{\Delta}^{(n)} \neq \emptyset$ **então**
 - 3: $\mathbf{\Upsilon}_r \leftarrow \mathbf{\Upsilon}_r \cup \{\mathbf{\Upsilon}^{(n)}\}$
 - 4: **fim do se**
 - 5: **fim do enquanto**
 - 6: **se** $\mathbf{\Upsilon}_r = \emptyset$ **então**
 - 7: $\mathbf{\Upsilon}_r \leftarrow \mathbf{\Upsilon}$
 - 8: **fim do se**
 - 9: **retorne** $\mathbf{\Upsilon}_r$
-

Após a seleção inicial das células do *radiomap* através do Algoritmo 3, é realizada uma comparação entre os elementos do conjunto *radiomap* reduzido $\mathbf{\Upsilon}_r$ e a amostra $\mathbf{A}^m$

utilizando a distância Euclidiana como métrica de similaridade. A distância Euclidiana é calculada entre o vetor $\mathbf{\Omega}^{(n)}$ do *radiomap* e o vetor $\mathbf{\Omega}^m$ da amostra.

$$d^{(n)} = \sqrt{\sum_{i=1}^{N_g} \sum_{j=1}^{N_e} \left(\omega_{ij}^{(n)} - \omega_{ij}^m \right)^2} \quad (3.15)$$

Finalmente, será escolhido o elemento dentro do espaço de busca reduzido $\mathbf{\Upsilon}_r$ que possuir a menor distância Euclidiana $d^{(n)}$. O vetor posição $\mathbf{E}^{(n)}$ deste elemento escolhido corresponderá a estimativa de localização fornecida pelo sistema.

4 AVALIAÇÃO E RESULTADOS

Este Capítulo traz a avaliação dos *gridmaps* classificados anteriormente com relação a algumas métricas que serão definidas. O *erro médio de localização para o gridmap* será utilizado como métrica de acurácia para balizar a comparação entre modelos. Por fim, os resultados obtidos a partir da simulação, serão verificados para um conjunto de medições de *Radio Frequency* (RF) reais obtidas na cidade de Recife/PE.

4.1 ERRO MÉDIO DE LOCALIZAÇÃO PARA O GRIDMAP

O impacto da geometria dos *gridmaps* no sistema de localização foi avaliado através de um método estatístico de validação cruzada conhecido como *k-fold*. Esse método consiste em uma técnica de reamostragem que particiona aleatoriamente o conjunto original das amostras em k subconjuntos mutuamente exclusivos, onde cada subconjunto possui aproximadamente a mesma quantidade de elementos (KUHN; JOHNSON, 2016). A cada iteração do método *k-fold*, um subconjunto é reservado para etapa de validação enquanto os demais $(k - 1)$ subconjuntos são destinados a etapa de treinamento. Entre as iterações, o método rotaciona qual subconjunto será utilizado na etapa de validação do modelo, evitando repetições. O método encerra após k iterações, fornecendo como saída a precisão do modelo com base nos erros obtidos em cada iteração.

Os experimentos realizados consideraram um valor de $k = 10$ para o método *k-fold*. A métrica utilizada no processo de validação cruzada foi o *erro médio de localização para o gridmap* (μ). O valor de μ corresponde a um dos valores utilizados para avaliar o desempenho dos *gridmaps*. O erro médio μ é definido tal que

$$\mu = \frac{1}{k} \sum_{i=1}^k e_i , \quad (4.1)$$

em que e_i é o erro médio calculado para o i -ésimo subconjunto de validação. Por sua vez, a expressão para o erro médio e_i é definida como

$$e_i = \frac{1}{T} \sum_{n=1}^T e_{i,n} , \quad (4.2)$$

em que $e_{i,n}$ corresponde ao erro na estimativa de localização associado a n -ésima amostra pertencente ao i -ésimo subconjunto de validação e T , representa o número de amostras no i -ésimo subconjunto de validação. Por fim, o erro $e_{i,n}$ é dado por

$$e_{i,n} = \left| \mathbf{E}^{(n)} - \hat{\mathbf{E}}^{(n)} \right| , \quad (4.3)$$

em que $\hat{\mathbf{E}}^{(n)}$ corresponde à posição estimada para a amostra e $\mathbf{E}^{(n)}$ corresponde à posição real associada a amostra de validação. A diferença entre $\mathbf{E}^{(n)}$ e $\hat{\mathbf{E}}^{(n)}$ é obtida por meio da distância geodésica, definida na Equação 2.2. A distância geodésica é calculada entre a posição real $\mathbf{E}^{(n)}$ e a posição estimada $\hat{\mathbf{E}}^{(n)}$.

Ao longo das próximas Seções, o *erro médio de localização para o gridmap* (μ) será amplamente utilizado para descrever os resultados experimentais. No intuito de simplificar a compreensão dos resultados, μ será muitas vezes descrito utilizando expressões como *erro médio de localização*, ou simplesmente *erro médio*.

4.2 MÉTRICAS PARA AVALIAÇÃO DOS GRIDMAPS

Conforme definido na Seção 3.4, $\mathbf{S}_c$, $\mathbf{R}_c$, $\mathbf{Q}_c$ e Λ_{cp} correspondem, respectivamente, aos conjuntos de *gridmaps* obtidos a partir da aplicação dos algoritmos geradores S , R , Q e Λ . Vale lembrar que o algoritmo S (*Square*) é responsável segmentar a região de cobertura através de uma malha quadricular, sendo assim os *gridmaps* gerados por tal algoritmo serão chamados *gridmaps* quadriculares. Analogamente, os algoritmos R (*Random*) e (*Quasirandom*) condizem, respectivamente, com os *gridmaps* pseudo-aleatórios e quasi-aleatórios. Por fim, o algoritmo Λ descreve o método de segmentação apresentado na Figura 14, os *gridmaps* obtidos por Λ são denominados *gridmaps* adaptativos.

Os algoritmos S , R , Q e Λ , recebem como entrada dois argumentos c e p , de modo que $c \in \mathbf{C}$ e $p \in \mathbf{P}$. Um dos principais objetivos desta seção é comparar os resultados de μ obtidos com a utilização dos *gridmaps* dos conjuntos $\mathbf{S}_c$, $\mathbf{R}_c$, $\mathbf{Q}_c$ com relação aqueles obtidos para os *gridmaps* do conjunto Λ_{cp} . Entretanto, a cardinalidade $|\Lambda_{cp}| = |\mathbf{C} \times \mathbf{P}|$, difere daquela obtida para os demais conjuntos $|\mathbf{S}_c| = |\mathbf{R}_c| = |\mathbf{Q}_c| = |\mathbf{C}|$. Para permitir tais comparações, foram criadas p partições do conjunto Λ_{cp} da seguinte forma

$$\Lambda_p = \{\Lambda_p(c)\} \quad c \in \mathbf{C}, p \in \mathbf{P}, \quad (4.4)$$

em que Λ_p corresponde a uma aplicação parcial (*currying*) da função de geração Λ com p predeterminado, sendo $c \in \mathbf{C}$ o único argumento variável. Dessa forma, a cardinalidade de uma partição p assume o valor $|\Lambda_p| = |\mathbf{C}|$, o que possibilita aplicar o conceito do erro médio μ para comparação entre diferentes estratégias de geração de *gridmaps*.

A partir de μ é possível derivar outras métricas úteis para comparação entre *gridmaps*. Suponha, por exemplo, que para um determinado número de células $c = 2.500$, que satisfaz a relação de pertinência $c \in \mathbf{C}$, deseja-se saber qual foi a melhora no erro médio μ , quando o algoritmo Λ_p foi introduzido no lugar de S . Para avaliar essas situações, foi introduzida a notação Δ_{x-y}^μ , que representa o conjunto diferença de erro médio μ entre *gridmaps*, computado entre duas estratégias x e y de geração de *gridmaps*, tal que $x, y \in \{S, R, Q, \Lambda_p\}$ e $x \neq y$. O conjunto Δ_{x-y}^μ permite investigar a melhora no erro médio μ entre dois algoritmos de geração, para todos os valores $c \in \mathbf{C}$ utilizados no experimento.

Considerando o exemplo de comparação entre os algoritmos S e Λ_p , teremos a seguinte notação para o conjunto diferença de erro médio μ entre *gridmaps*

$$\Delta_{S-\Lambda_p}^\mu = \{ |\mu_{S(c)} - \mu_{\Lambda_p(c)}| \mid c \in \mathbf{C} \}. \quad (4.5)$$

O conjunto Δ_{x-y}^μ permite derivar três métricas para comparar diferentes estratégias de geração. As duas primeiras correspondem as métricas μ_{max} e μ_{min} , que são utilizadas para avaliar, respectivamente, o maior e o menor melhora em relação ao erro médio de localização μ , obtidos para duas estratégias diferentes de geração de *gridmaps*, para todos os valores de células $c \in \mathbf{C}$ utilizados nos experimentos. As expressões para μ_{max} e μ_{min} , obtidas a partir do conjunto Δ_{x-y}^μ , estão indicadas a seguir.

$$\begin{aligned} \mu_{max} &= \max\{\Delta_{x-y}^\mu\} & x, y \in \{S, R, Q, \Lambda_p\} ; x \neq y \\ \mu_{min} &= \min\{\Delta_{x-y}^\mu\} & x, y \in \{S, R, Q, \Lambda_p\} ; x \neq y \end{aligned} \quad (4.6)$$

Assim como no exemplo anterior, sejam S e Λ_p dois algoritmos de geração. Em outras palavras, a métrica μ_{max} permite observar, para qual valor c , a opção pelo algoritmo Λ_p proporcionou a maior melhora em relação ao erro médio μ . No caso da métrica μ_{min} , o processo é análogo, nesse caso, o objetivo é revelar para qual valor c houve a menor melhora em relação ao erro médio μ . Considerando ainda o exemplo S e Λ_p , uma interpretação válida para o valor obtido através de μ_{max} e μ_{min} é dizer que, a opção pelo algoritmo Λ_p , conferiu uma melhora no erro médio μ de no máximo $\mu_{max} = X m$ e no mínimo $\mu_{min} = Y m$. As expressões de μ_{max} e μ_{min} para os algoritmos S e Λ_p é dada em função de Δ_{x-y}^μ conforme indicado a seguir.

$$\begin{aligned} \mu_{max} &= \max\{\Delta_{S-\Lambda_p}^\mu\} \\ \mu_{min} &= \min\{\Delta_{S-\Lambda_p}^\mu\} \end{aligned} \quad (4.7)$$

Outra métrica derivada a partir do erro médio μ é o $\bar{\mu}$. Essa métrica representa a média aritmética dos elementos do conjunto Δ_{x-y}^μ . O valor $\bar{\mu}$ pode ser interpretado como a melhora em relação ao médio μ , considerando duas diferentes estratégias de geração de *gridmaps*. A expressão para obter o valor $\bar{\mu}$ foi definida da seguinte forma.

$$\bar{\mu} = \frac{1}{|\Delta_{x-y}^\mu|} \sum_{i \in \Delta_{x-y}^\mu} i \quad (4.8)$$

Uma possível interpretação para métrica $\bar{\mu}$ é obtida da seguinte forma. Considerando o exemplo dos algoritmos de geração S e Λ_p , pode-se afirmar que a opção por Λ_p ao invés S fornece uma melhora no erro médio μ da ordem de $\bar{\mu} = X m$.

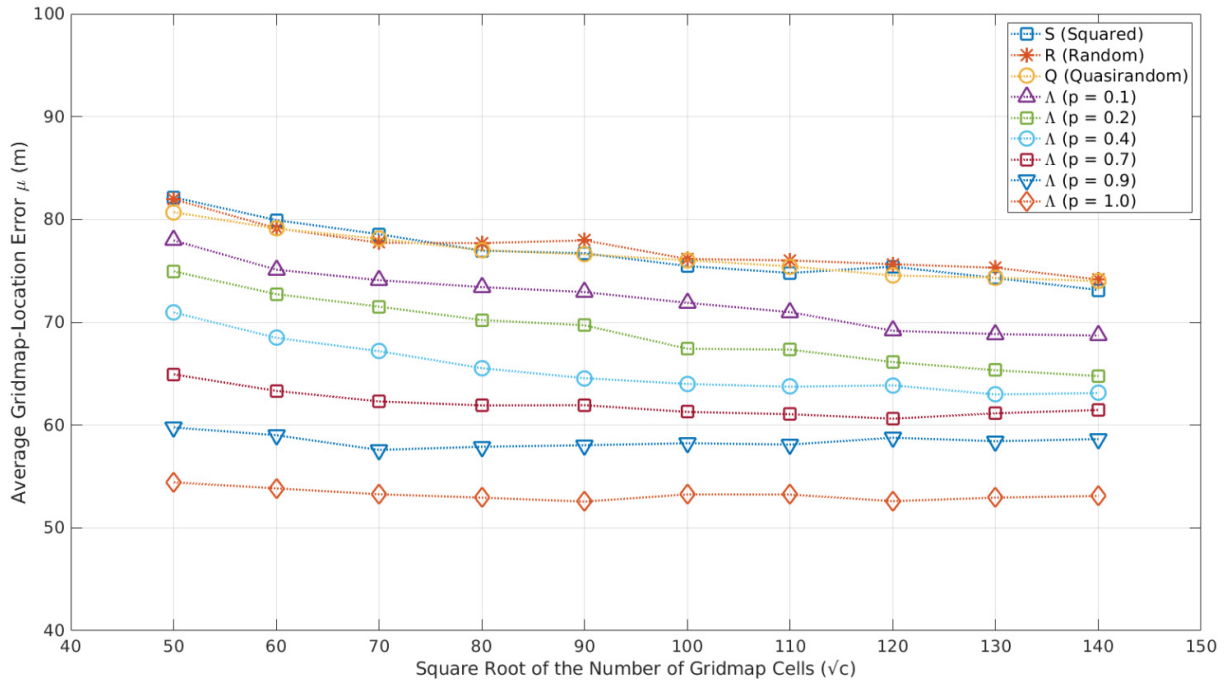
No restante deste Capítulo, os valores de μ , μ_{max} , μ_{min} e $\bar{\mu}$ serão utilizados como métricas para comparar as diferentes estratégias de geração de *gridmaps* S , R , Q e Λ_p .

4.3 RESULTADO DA AVALIAÇÃO DOS GRIDMAPS

A instância do sistema de localização apresentado na Seção 3.6 foi utilizada para avaliar o conjunto de *gridmaps* $\mathbf{G}$ definido na Seção 3.4. Cada *gridmap* $g \in \mathbf{G}$ foi avaliado através do método de validação cruzada *k-fold* com $k = 10$. O conjunto de medições RF utilizado corresponde aquele obtido através do simulador *Network Simulator 3* (NS-3) como descrito na Seção 3.2. A representação dos elementos desse conjunto foi apresentada na Seção 3.5. Além de μ , os parâmetros μ_{max} , μ_{min} e $\bar{\mu}$ foram utilizados como métricas para comparar as diferentes estratégias de geração de *gridmaps*.

A seleção dos parâmetros w e k para o regressor baseado no algoritmo *Weighted K-Nearest Neighbors* (WK-NN) foi realizada por um procedimento conhecido como *tuning*. Os parâmetros k e w correspondem a números naturais que foram buscados respectivamente nos intervalos $[0 - 12]$ e $[0 - 3]$. Para isso, a técnica de validação cruzada com $k = 10$ foi utilizada considerando um *gridmap* de malha quadricular com 2.500 células. Os valores de $k = 9$ e $w = 1$ foram selecionados, pois proporcionaram a maior redução no erro médio de localização μ , detalhado na Seção 4.1.

Figura 17 – O gráfico mostra o erro médio de localização para o *gridmap* μ obtido com relação aos algoritmos S , R , Q e Λ . O algoritmo Λ foi aplicado utilizando diferentes valores percentuais p definidos para $p \in \mathbf{P}$.



Fonte: Elaborada pelo autor (2020)

A Figura 17 mostra o comportamento do erro médio de localização para o *gridmap* μ com relação ao número de células. É possível perceber que praticamente não existe diferença no valor de μ observado para os *gridmaps* gerados pelos algoritmos S , R e Q . Em teoria, esperava-se que o algoritmo S resultasse em um menor valor de μ , já que

ele é o que distribui de maneira mais uniforme as células na região de cobertura. Em contrapartida, o algoritmo R deveria apresentar um valor de μ superior aquele obtido pelo algoritmo S , já que o método R tende a gerar agrupamentos de células em regiões arbitrárias. Entretanto, a observação realizada aponta que os algoritmos S , R e Q , parecem não influenciar significativamente no valor de μ observado.

Diferentemente do observado para os algoritmos S , R e Q , o algoritmo de geração para *gridmaps* adaptativos Λ influenciou nos valores de μ obtidos. Ainda com base na Figura 17, é possível perceber que um incremento de apenas 10% na densidade de células sobre as regiões prioritárias ($p = 0.1$), promoveu uma melhora no valor de μ da ordem de $\bar{\mu} = 4.43m$ quando comparado ao algoritmo S . Sendo o melhor resultado $\mu_{max} = 6.24m$, observado quando o número de células utilizado foi $c = 14400$.

No caso onde 100% das células foram concentradas na região de interesse ($p = 1.0$), a melhora no erro médio μ foi de $\bar{\mu} = 23.55m$, quando comparado ao algoritmo S . Os valores de μ_{max} e μ_{min} foram respectivamente $27.76m$ e $20.05m$. A melhoria máxima para o erro médio μ (μ_{max}) foi observado para o *gridmap* com $c = 2500$ células, enquanto o *gridmap* com $c = 19600$ células proporcionou a menor melhoria no erro médio de localização μ . Essa observação mostra que, à medida que o número de células cresce, os valores de μ tendem a convergir. Nesses casos, a partir de um determinado limiar, introduzir mais células não necessariamente implica em uma redução no valor de μ .

A principal hipótese para essa observação deriva do fato que o *radiomap* fornece uma generalização do conjunto de medições RF. Desse modo, existe um equilíbrio entre a quantidade de células no *gridmap* e a quantidade de amostras RF no conjunto de medições. À medida que o número de células para uma dada área de cobertura aumenta, as células tendem a ficar concentradas, ou seja, mais próximas uma das outras. Durante a construção do *radiomap*, no caso da utilização do *K-Nearest Neighbors* (K-NN) como regressor, o valor do vetor *RF fingerprint* para uma dada célula precisa ser generalizado a partir de k amostras do conjunto de medições RF. As k amostras escolhidas são aquelas que possuem a menor distância em relação à posição da célula. Para regiões onde existe uma alta concentração no número de células, os k vizinhos escolhidos por células adjacentes tendem a ser sempre os mesmos. Assim, o valor do vetor *RF fingerprint* atribuído a tais células possui pouca, ou quase nenhuma, diferença. Essa baixa variância entre os vetores *RF fingerprint* de células adjacentes, tende a limitar o processo de generalização promovido pelo algoritmo WK-NN. Essa saturação limita a capacidade de classificação dos elementos do *radiomap*, fazendo com que o valor de μ permaneça constante.

Sendo assim, deve existir uma relação ótima entre a área total da região de cobertura, o número de células e a quantidade de medições RF. O incremento indiscriminado no número de células apenas satura o problema, sem necessariamente implicar em mudanças para o valor de μ . Apesar desta constatação, encontrar a relação e um eventual equilíbrio entre esses três elementos não fez parte do escopo deste estudo, de modo que esse problema

pode ser endereçado futuramente em outros trabalhos.

A saturação no número de células fica evidente, por exemplo, para o caso específico do algoritmo $\Lambda_{p=1.0}$. Na Figura 17, é possível perceber que o valor de μ permaneceu praticamente constante para todos os *gridmaps*, independentemente do número de células. Vale destacar ainda que o modelo de *gridmap* gerado pelo algoritmo $\Lambda_{p=1.0}$ corresponde aquele detalhado em (BITTENCOURT; URBANO; CUNHA, 2018). A diferença é que, no artigo em questão, todas as amostras do conjunto de medições RF haviam sido obtidas em ambiente *outdoor*. Já para os experimentos conduzidos neste estudo, o conjunto de medições foi formado com cerca de 1/3 das medições obtidas em ambiente *indoor*.

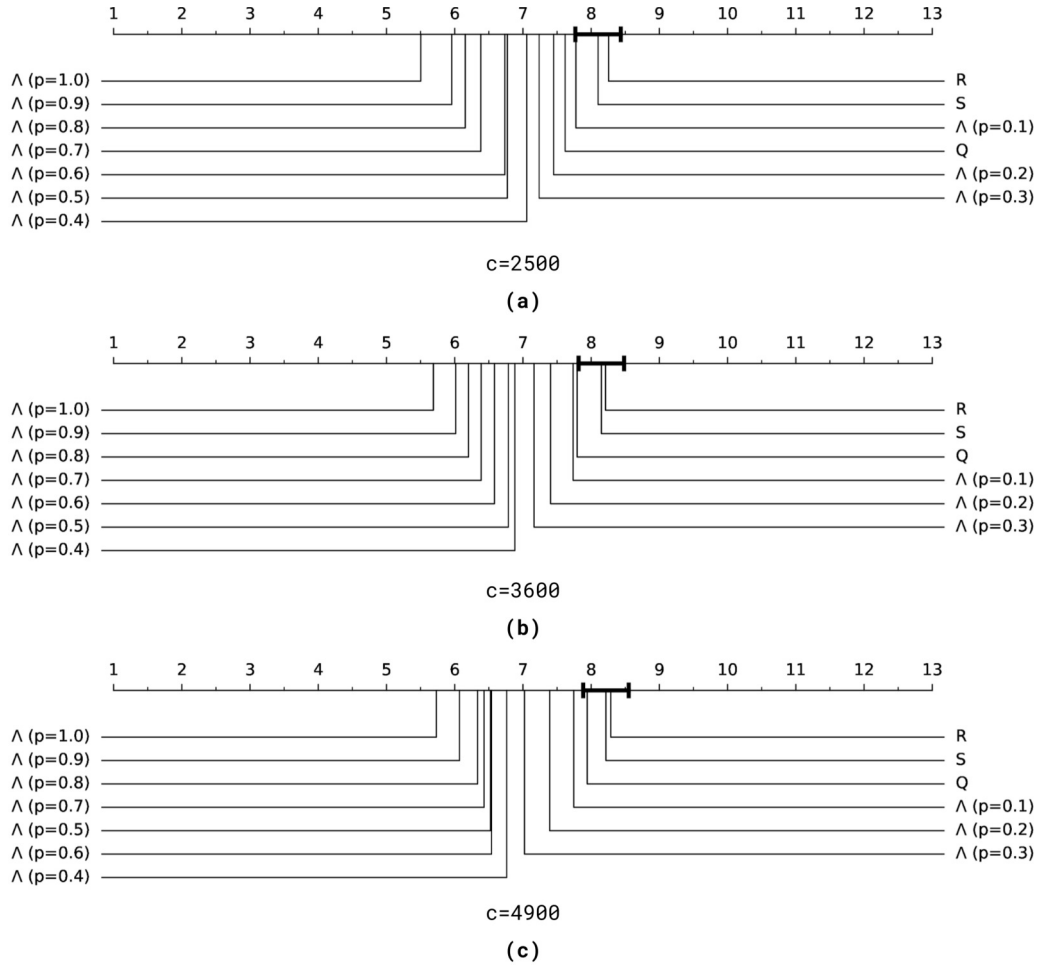
No intuito de verificar se os diferentes valores de acurácia observados são estatisticamente relevantes, foi aplicado o teste pareado não paramétrico de Friedman (DANIEL, 1990). Esse método estende o teste dos postos sinalizados de Wilcoxon (*Wilcoxon Signed Rank*) para múltiplos grupos (WILCOXON, 1945). A hipótese nula H_0 é que não existe diferença entre os grupos, enquanto a hipótese alternativa H_1 assume que existe ao menos uma diferença entre os grupos testados.

O teste de Friedman foi aplicado da seguinte considerando 13 grupos, onde cada grupo representa um algoritmo para geração de *gridmaps* — 3 grupos formados pelos algoritmos S , R e Q , mais 10 grupos derivados das aplicações parciais de $p \in \mathbf{P}$ em Λ . A quantidade de testes realizados seguindo a definição para os elementos do conjunto $\mathbf{C}$, conforme descrito na Seção 3.4. Desse modo, a quantidade de testes foi definido pela cardinalidade do conjunto de células $|C| = 10$, ou seja, foram realizadas ao todo 10 instâncias do teste de Friedman entre os grupos, uma para cada valor $c \in \mathbf{C}$. O nível de significância α foi definido como 0,05 para cada instância do teste e os *p-value* obtidos em todas elas foram iguais a zero (*p-value* = 0). Ou seja, os resultados dos testes de Friedman aplicados indicam que a hipótese nula H_0 pode ser rejeitada, concluindo que ao menos um dos algoritmos de geração de *gridmaps* difere dos demais.

Para permitir uma comparação dos grupos em pares, foi utilizada a análise Nemenyi *post-hoc* para o teste de Friedman (NEMENYI, 1963). O teste Nemenyi *post-hoc* foi utilizado para obter o diagrama de diferença crítica, *Critical Difference* (CD), para cada *gridmap* $c \in \mathbf{C}$ em relação aos algoritmos de geração (Figura 18). O diagrama CD fornece uma visualização da significância estatística observada comparando a diferença entre a média dos *ranks* obtidos através do teste de Friedman (DEMSAR, 2006).

O intervalo destacado na Figura 18 corresponde a uma distância equivalente a $2d_c$ centrada no algoritmo S para criação de *gridmaps* quadriculares. O valor de d_c corresponde a distância crítica calculada através da análise *post-hoc* de Nemenyi para o teste de Friedman. A imagem contém uma amostra dos diagramas CD obtidos na análise, os diagramas representam o comportamento dos grupos respectivamente para $c = 2.500$, $c = 3.600$ e $c = 4.900$ células. Essa visualização foi escolhida pelo fato do algoritmo S ser usualmente empregado nos sistemas de localização, logo, a comparação dos outros

Figura 18 – Resultado para o diagrama de distância crítica CD. O intervalo destacado mostra uma distância de $2d_c$ centrada no algoritmo S para *gridmaps* quadriculares, onde d_c representa a distância crítica da análise *post-hoc* Nemenyi calculada para o teste de Friedman.



Fonte: Elaborada pelo autor (2020)

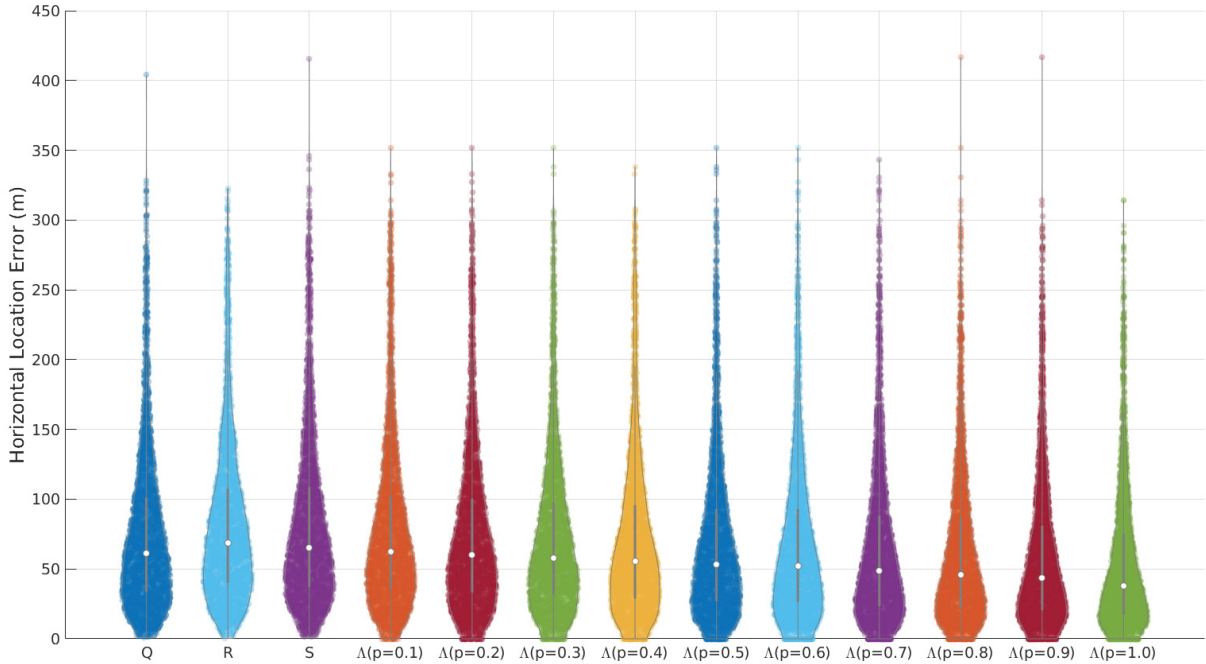
algoritmos em relação a S é desejada.

Em geral, as análises comparativas mostraram que não é possível observar uma relevância estatística entre os algoritmos S , R e Q . A única exceção a essa observação ocorreu para o algoritmo Q para os casos com $c = 2.500$ e $c = 3.600$ células. Observando a Figura 18(a) percebe-se que Q está fora da região crítica definida por S . Já na Figura 18(b), verifica-se que o algoritmo Q aparece no limiar da distância crítica para S , posicionado praticamente na fronteira, mas ligeiramente acima do valor d_c definido pelo experimento. Para os demais casos, o algoritmo Q não se mostrou estatisticamente relevante com relação a S e R , isso pode ser observado já a partir de $c = 4.900$ células como indicado na Figura 18(c). Assim, podemos inferir que para valores de c acima de 4.900 células, o algoritmo Q não demonstra ser estatisticamente relevante. Comparando este resultado com o gráfico para o erro médio μ mostrado na Figura 17, é possível perceber que quando $c = 2.000$ o algoritmo Q teve um desempenho ligeiramente melhor que S e R . Especificamente para o algoritmo Q , o caso $c = 2.500$ células corresponde ao ponto onde a maior diminuição

no erro médio μ em relação a S foi observada. Comparando Q e S o valor $\mu_{max} = 1.46m$ corresponde exatamente ao ponto $c = 2.500$. Apesar disso, o erro médio μ observado para Q nos experimentos é praticamente o mesmo daquele observado para os algoritmos S e R , essa característica corrobora para o resultado do teste estatístico, indicando que não existe uma relevância estatística entre S , R e Q .

Outra conclusão importante derivada da observação dos diagramas CD parece indicar que, para o algoritmo Λ , a escolha de um parâmetro $p = 0, 1$ (10%) já é capaz de produzir resultados estatisticamente relevantes para o erro médio μ observado. Isso foi observado em todos os diagramas CD com exceção do diagrama para $c = 2.500$ células (Figura 18(a)). Nesse caso específico, o algoritmo $\Lambda_{p=0,1}$ ficou na fronteira da distância crítica d_c em relação a S . O valor de d_c calculado pelo teste estatístico foi 0,33 e a distância obtida entre os algoritmos S e $\Lambda_{p=0,1}$ foi 0,32. Em contrapartida, as Figuras 18(b) e 18(c) o algoritmo $\Lambda_{p=0,1}$ já se mostra estatisticamente relevante, seguindo essa tendência para os demais diagramas CD obtidos no experimento. Ainda no caso do algoritmo $\Lambda_{p=0,1}$, o gráfico do erro médio μ mostrado na Figura 17 revelou uma melhora em média $\bar{\mu} = 4,43m$ quando o algoritmo $\Lambda_{p=0,1}$ foi escolhido para o sistema de localização no lugar do algoritmo S . A maior e a menor melhora no erro médio μ foram $\mu_{max} = 6,23m$ e $\mu_{min} = 3,58m$, obtido respectivamente para os *gridmaps* com $c = 14400$ e $c = 10000$ células.

Figura 19 – Distribuição para o erro de localização horizontal obtido quando um total de $c = 2500$ células foi utilizado. É possível perceber uma redução da média a medida que o valor do parâmetro p se aproxima de 1.0.



Fonte: Elaborada pelo autor (2020)

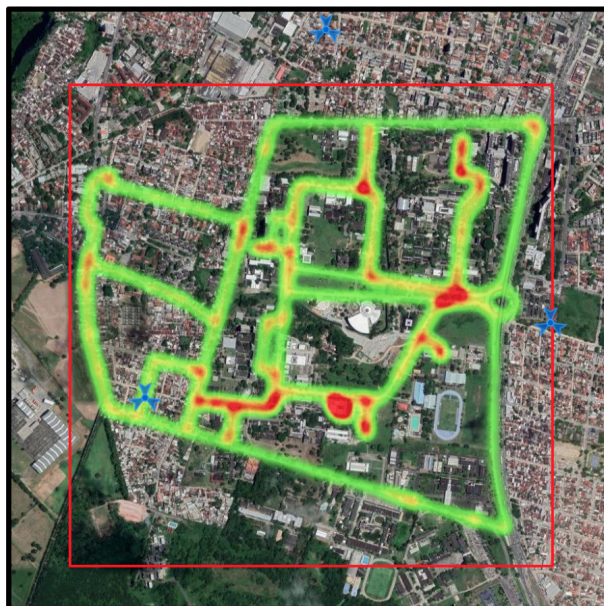
Por fim, a distribuição para o erro de localização definido na Equação 4.3 foi exibido em um gráfico de violino e encontra-se ilustrado na Figura 19. O erro de localização

apresentado em 4.3 também é conhecido como *erro de localização horizontal* e consiste no erro de localização obtido individualmente para cada amostra do conjunto de validação. A imagem em questão mostra a distribuição do erro de localização horizontal para todos os algoritmos quando um total de $c = 2500$ células foi utilizado. É possível perceber que o aumento do parâmetro p acarreta em um achatamento da distribuição em torno de valores mais baixos para erro de localização horizontal.

4.4 CONJUNTO DE MEDIÇÕES RF REAIS

No intuito validar os resultados observados na Seção 4.3, foi utilizado um conjunto de medições RF obtidas em um perímetro urbano na cidade de Recife/PE. Esta região corresponde ao *campus* urbano da Universidade Federal de Pernambuco (UFPE), localizado no bairro da Várzea na zona oeste do Recife. A região de cobertura possui uma área de aproximadamente 3 km^2 , sendo o conjunto de medições RF obtido para o padrão de telefonia *Long-Term Evolution* (LTE), com portadora na faixa de 1,8 GHz. Assim como na simulação, os parâmetros RF coletados corresponderam ao *Received Signal Strength* (RSS) e ao *Propagation Delay* (PD), tais parâmetros foram obtidos tanto em ambiente *indoor*, quanto *outdoor*. A descrição para este conjunto de medições segue as mesmas definições apresentadas na Seção 3.5. Para manter a fidelidade com a simulação, foi selecionada uma fração de $1/3$ ($\sim 33\%$) de medições em ambientes *indoor*, os demais $2/3$ ($\sim 66\%$) correspondem a medições *outdoor* obtidas majoritariamente através da técnica de *drive-testing*, sendo coletadas, ao todo, 4.617 medições.

Figura 20 – Mapa de calor para o conjunto de medições RF reais obtidas através de *drive-testing*. A região de cobertura é definida pelo retângulo vermelho, compreendendo uma área de aproximadamente 3 km^2 . As ERBs podem ser identificadas através dos polígonos na cor azul.



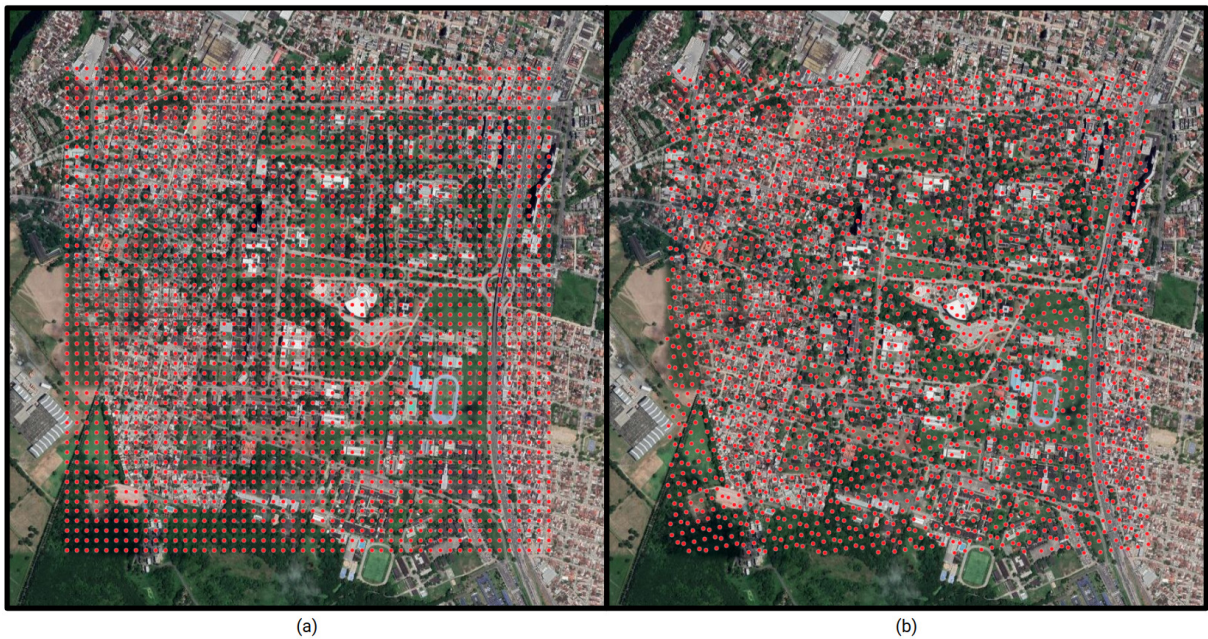
Fonte: Elaborada pelo autor (2020)

A Figura 20 ilustra o mapa de calor do conjunto com 4.617 medições RF que foi utilizado. Assim como na simulação do NS-3, as medições foram obtidas com relação a nove ERBs, posicionadas em três torres distintas. Cada ERB possui uma antena tri-setorial com abertura de 120 graus. As ERBs foram reunidas em 3 grupos distintos, cada um com 3 ERBs, seguindo a mesma especificação detalhada em Seção 3.5. Na figura em questão, esses elementos foram representados por meio de polígonos azuis, onde o círculo representa a torre e os polígonos que partem dele indicam a abertura das antenas. Por fim, as linhas traçadas delimitam a região de cobertura.

4.5 SELEÇÃO DOS GRIDMAPS

Para efeitos de comparação, foram gerados *gridmaps* pelos algoritmos S , R , Q e Λ , definidos na Seção 3.3. Para os algoritmos S , R e Q , os *gridmaps* foram concebidos considerando apenas o número de células c , como parâmetro de entrada ($c \in \mathbf{C}$). No caso do algoritmo Λ , além de c , foi também utilizado como entrada o percentual p de células destacadas para as regiões prioritárias ($p \in \mathbf{P}$). O sistema de localização utilizado é idêntico àquele descrito na Seção 3.6, no qual as medições RF do NS-3 foram avaliadas.

Figura 21 – (a) *Gridmap* de malha quadricular gerado pelo algoritmo S com $c = 2.500$ células. (b) *Gridmap* gerado pelo algoritmo Q com $c = 2.500$ células.



Fonte: Elaborada pelo autor (2020)

As Figuras 21(a) e 21(b) mostram, respectivamente, os *gridmaps* gerados pelos algoritmos S e Q com $c = 2.500$ células. Apesar da utilização de distâncias geodésicas, é possível perceber que a distribuição das células apresenta as mesmas características dos *gridmaps* apresentados na Seção 3.3. Em outras palavras, o algoritmo Q mantém uma

baixa discrepância, garantindo uma distribuição uniforme das células quando comparado ao *gridmap* de malha quadricular gerado pelo algoritmo S .

Outra visualização interessante pode ser feita na Figura 22, que contém os *gridmaps* adaptativos gerados através do algoritmo Λ . Na imagem em questão, os *gridmaps* em (a), (b), (c) e (d) foram gerados, respectivamente, utilizando uma proporção p de 10%, 20%, 30% e 40% para as células posicionadas em regiões prioritárias. O parâmetro para o número total de células utilizado foi $c = 2500$ para os quatro *gridmaps*.

Figura 22 – Exemplo de *gridmaps* irregulares com $c = 2500$ células gerados para a região de cobertura. Em (a), (b), (c) e (d) podemos visualizar *gridmaps* onde uma percentual $p = 10\%$, $p = 20\%$, $p = 30\%$ e $p = 40\%$ das células foram designadas as regiões prioritárias.



Fonte: Elaborada pelo autor (2020)

Vale destacar que, assim como efetuado no experimento com medições RF do NS-3, foram gerados, ao todo, 130 *gridmaps*. A construção dos *gridmaps* utilizou todas as combinações possíveis dos parâmetros $p \in \mathbf{P}$ e $c \in \mathbf{C}$, aplicadas a todos os algoritmos R , Q , S e Λ , conforme discutido na Seção 3.4.

Por fim, Os *gridmaps* foram avaliados em um sistema de localização cujos parâmetros e definições obedecem àqueles delineados na Seção 3.6. Os parâmetros observados nos experimentos correspondem aqueles discutidos nas Seções 4.1 e 4.2. Assim como discutido na Seção 4.3, a seleção dos parâmetros w e k para o WK-NN foi realizada via *tuning*. Os parâmetros $w \in \mathbb{N}$ e $k \in \mathbb{N}$ foram buscados utilizando o método de validação cruzada k -fold com $k = 10$, respectivamente, nos intervalos $[0 - 12]$ e $[0 - 3]$. Os valores de $k = 3$ e $w = 0$ foram selecionados, pois proporcionaram a maior redução no *erro médio de localização para o gridmap* (μ), detalhado na Seção 4.1.

4.6 RESULTADO FINAL

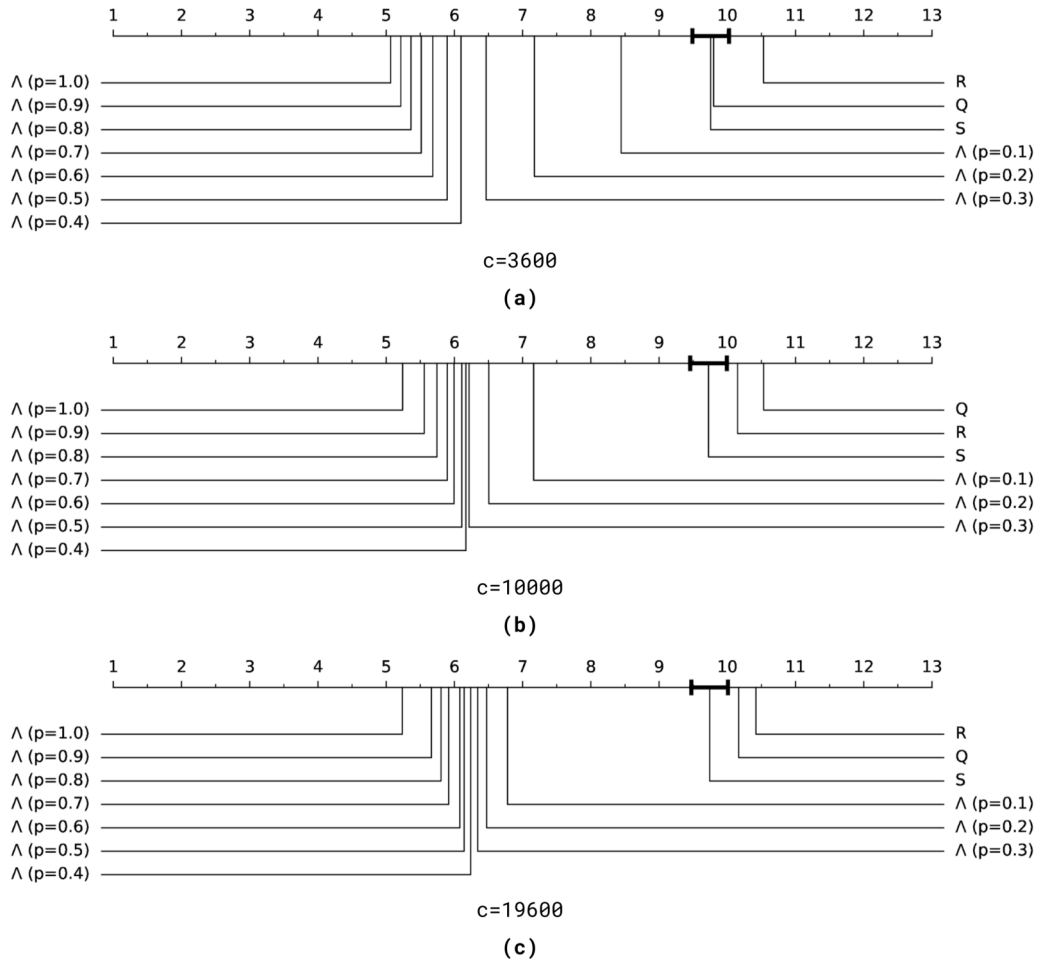
O experimento utilizando os *gridmaps* foi conduzido para o conjunto de medições RF reais de forma análoga aquele realizado na Seção 4.3. Os *gridmaps* foram criados utilizando os algoritmos S , R , Q e Λ , para toda combinação possível de parâmetros no qual $c \in \mathbf{C}$ e $p \in \mathbf{P}$, seguindo a descrição apresentada na Seção 3.4.

Para validar se a relevância estatística obtida para os *gridmaps* adaptativos condiz com aquela observada no conjunto de medições RF do simulador NS-3, foi aplicado o teste de hipótese de Friedman para múltiplos grupos, onde cada grupo representa um determinado algoritmo de geração, em um procedimento análogo aquele discutido na Seção 4.3. A hipótese nula (H_0) é que não existe diferença entre os grupos, já a hipótese alternativa (H_1), indica que ao menos um dos grupos difere significativamente dos demais. O nível de significância do experimento foi definido com $\alpha = 0.05$ e o p -value obtido em todas as instâncias do teste foi igual a zero (p -value = 0). Diante desse cenário, a hipótese H_0 pode ser rejeitada e a análise segue aplicando o teste *post-hoc* de Nemenyi para obter o diagrama de diferença crítica (CD) entre os grupos testados.

A Figura 23 contém uma amostra dos resultados da análise *post-hoc* realizada entre grupos de algoritmos. O intervalo destacado nos diagramas CD corresponde a uma distância equivalente a duas vezes a distância crítica ($2d_c$) centrada no algoritmo de geração de *gridmaps* quadriculares S . Assim como obtido para as medições RF simuladas, é possível verificar que a partir da escolha de um parâmetro $p = 0.1$ (10%), o algoritmo Λ já torna-se significativamente relevante em relação ao algoritmo S , esse fato foi observado em todos os experimentos conduzidos e corrobora o resultado encontrado na Seção 4.3.

Um critério comumente utilizado para avaliar um sistema de localização é verificar sua aderência em relação a algum parâmetro de projeto, tipicamente a acurácia. Os requisitos do *Next Generation 9-1-1* (NG9-1-1), definidos pela *Federal Communications Commission* (FCC), correspondem a exigências dos órgãos reguladores do governo dos EUA para as

Figura 23 – Diagramas CD para (a) $c = 2.500$, (b) $c = 10.000$ e (c) $c = 19.600$ células, obtidos pela análise *post-hoc* de Nemenyi para a média dos *ranks* calculados através do teste de Friedman.



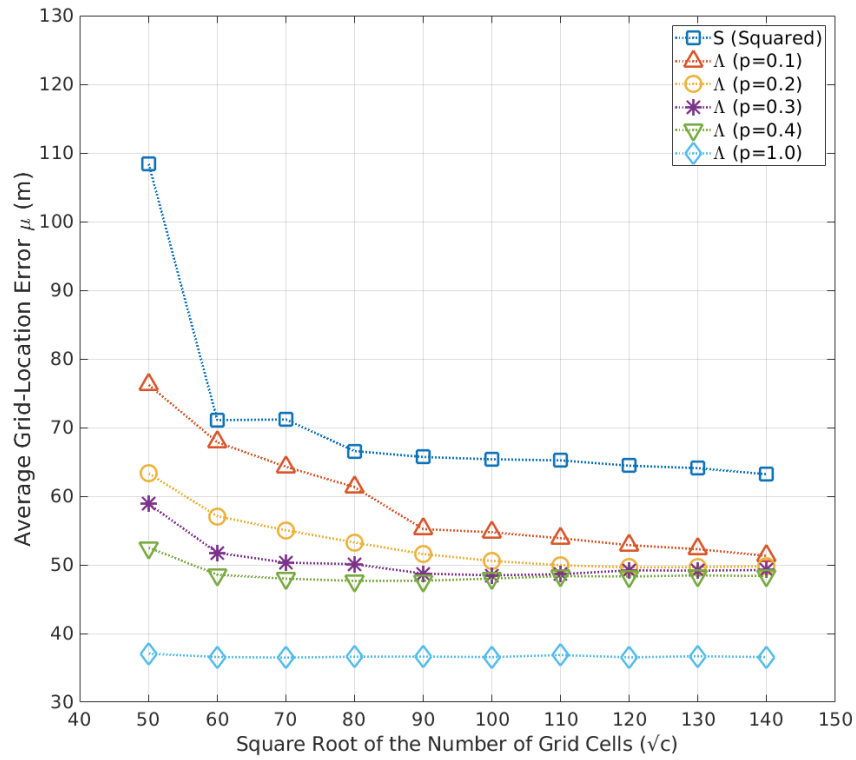
Fonte: Elaborada pelo autor (2020)

chamadas de emergência através das operadoras que prestam serviço de telefonia móvel. As diretrizes mais recentes para o erro de localização horizontal foram introduzidas no documento *Fourth Report and Order*, em matéria do *Wireless E911 Location Accuracy Requirements*, no qual são definidos os *benchmarks* de acurácia para o erro de localização horizontal a serem alcançados pelos prestadores de serviço de comunicação móvel ao longo dos próximos anos (FCC, 2015). Esse relatório determina que em até 8 anos, as operadoras do sistema de telefonia móvel devem ser capazes de fornecer um erro de localização horizontal (*indoor* e *outdoor*) de no máximo 50 m em até 80% das chamadas aos serviços de emergência naquele país. Ainda segundo o documento, os requisitos válidos em 2020, que correspondem ao 5º *benchmark*, exige que para 70% das chamadas realizadas a tais serviços, o erro de localização horizontal obtido seja de no máximo 50 m.

Com base no valor do erro obtido para as amostras de validação utilizadas nos experimentos, foi estabelecido um procedimento de inferência estatística através de intervalo de confiança para determinar quais *gridmaps* são capazes de satisfazer as exigências para o erro de localização horizontal impostas pela FCC. Conforme delineado anteriormente, a

Equação 4.3 define o erro de localização horizontal obtido para cada amostra do conjunto de validação. Sendo assim, os valores de erro $e_{i,n}$ obtidos para as amostras de validação foram transformados em uma distribuição binomial, onde o critério de sucesso foi definido através da obtenção de um valor para o erro igual, ou inferior a 50 m ($e_{i,n} \leq 50\text{ m}$). Por fim, o parâmetro populacional foi definido como sendo a proporção de amostras cujo critério de sucesso foi satisfeito. Nas análises realizadas, o nível de confiança utilizado correspondeu a $\gamma = 0,95$, ou seja, 95%.

Figura 24 – Erro médio de localização μ para os algoritmos S , $\Lambda_{p=10\%}$, $\Lambda_{p=20\%}$, $\Lambda_{p=30\%}$, $\Lambda_{p=40\%}$ e $\Lambda_{p=100\%}$. Os valores de μ foram calculados para 4.617 medições RF reais (*indoor* e *outdoor*), obtidas na cidade de Recife/PE.



Fonte: Elaborada pelo autor (2020)

O 5° *benchmark* definido pela FCC determina que o erro de localização horizontal seja de no máximo 50 m para 70% das chamadas aos serviços de emergência (FCC, 2015). Para avaliar esse cenário segundo a perspectiva do intervalo de confiança, o parâmetro populacional foi definido como 70%. Os resultados indicam que o algoritmo de geração mais eficiente capaz de satisfazer esse requisito foi $\Lambda_{p=0,4}(c = 2.500)$, no qual 40% das células foram destinadas às regiões prioritárias. Por mais eficiente, subentende-se o algoritmo com menor complexidade computacional. O algoritmo $\Lambda_{p=0,4}$ precisou de 2.500 células para cumprir o requisito, que corresponde ao menor valor c para o número de células utilizado nos experimentos ($c \in \mathbf{C}$). O intervalo de confiança obtido para $\Lambda_{p=0,4}(c = 2.500)$ foi $0,6971 \leq p_e \leq 0,7106$, onde p_e corresponde ao valor do parâmetro populacional ($p_e = 70\%$). Vale destacar que um baixo número de células implica em uma menor com-

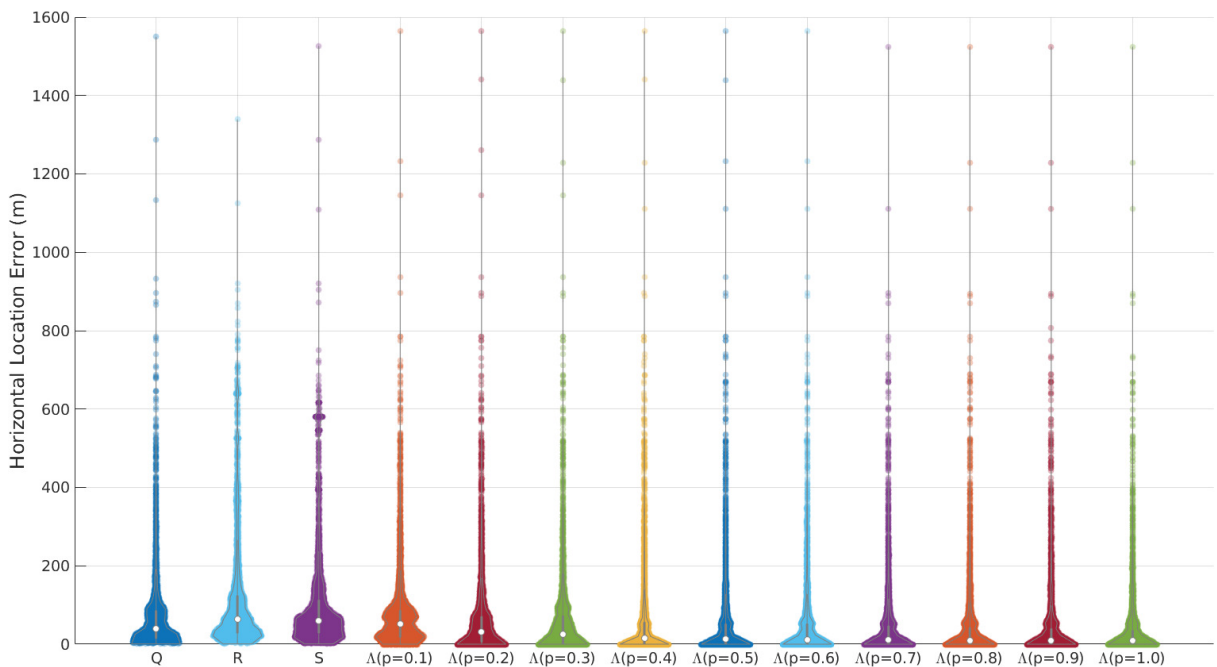
plexidade computacional para o sistema de localização.

Outro *gridmap* capaz de satisfazer a exigência para $p_e = 70\%$ foi $\lambda_{p=0,3}(c = 4.900)$, nesse caso, uma menor quantidade de células (30%) foi destinada a região prioritária. O intervalo de confiança para o parâmetro p_e foi $0,6967 \leq p_e \leq 0,7101$. A partir de $c = 8.100$ é possível atingir o objetivo utilizando $p = 0,2$ (20% das células), com intervalo de confiança $0,6865 \leq p \leq 0,7000$. Por fim, posicionando apenas 10% das células nas regiões prioritárias, o *gridmap* com $c = 19.600$ células foi capaz de satisfazer as exigências no intervalo de confiança $0,6912 \leq p_e \leq 0,7048$. Sendo assim, dependendo dos requisitos de projeto para o sistema de localização, uma seleção adequada de p e c pode ser utilizada de modo a satisfazer eventuais exigências para o erro de localização horizontal.

Ainda considerando o intervalo de confiança, no caso em que o parâmetro populacional utilizado foi $p_e = 80\%$, apenas o algoritmo $\Lambda_{p=1,0}$, com 100% das células posicionadas nas regiões prioritárias, foi capaz de atender as exigências para p_e . O critério p_e foi satisfeito para qualquer combinação de $c \in \mathbf{C}$. Essa exigência corresponde ao quinto *benchmark* previsto pelo documento *Fourth Report and Order*, na temática do *Wireless E911 Location Accuracy Requirements*. Segundo o relatório da FCC, o 5º *benchmark* está previsto para entrar em vigor em 2021 (FCC, 2015). Vale destacar que esse modelo de *gridmap* corresponde àquele utilizado em (BITTENCOURT; URBANO; CUNHA, 2018).

A Figura 24 mostra o gráfico para o *erro médio de localização do gridmap* (μ) definido na Seção 4.1, para os algoritmos S , $\Lambda_{p=10\%}$, $\Lambda_{p=20\%}$, $\Lambda_{p=30\%}$, $\Lambda_{p=40\%}$ e $\Lambda_{p=100\%}$. Esses algoritmos correspondem aqueles onde os critérios de estimativa por intervalo de confiança

Figura 25 – Distribuição do erro de localização horizontal para os algoritmos Q , R , S e Λ quando o parâmetro definido para o número de células foi $c = 2500$.



Fonte: Elaborada pelo autor (2020)

foram satisfeitos, com exceção do algoritmo S para o *gridmap* de malha quadricular. Vale destacar que o algoritmo S , em nenhuma das suas possíveis configurações dos parâmetros p e c , foi capaz de atender às exigências da FCC para o erro de localização horizontal quando o conjunto de medições RF real foi utilizado.

No intuito de fornecer uma estimativa para distribuição do erro de localização horizontal obtido através das amostras de validação, foi utilizado um gráfico de violino conforme indicado na Figura 25. A imagem foi obtida para a faixa de experimentos conduzidos com o parâmetro $c = 2500$ escolhido para a quantidade de células do *gridmap*. Assim como observado anteriormente, a distribuição do erro de localização horizontal tende a se concentrar em valores mais próximos a zero, à medida que a proporção de células designadas as regiões prioritárias aumenta.

Por fim, é possível constatar que, além da escolha do *gridmap*, o conjunto de medições RF possui influência direta na acurácia para um sistema de localização. Essa constatação reforça a importância de investigações futuras que possibilitem uma melhor compreensão da influência deste elemento em *Location-Based Services* (LBSes).

5 CONCLUSÃO

As técnicas apresentadas para os algoritmos de geração de *gridmaps* permitem generalizar o modelo de *gridmap* exposto em (BITTENCOURT; URBANO; CUNHA, 2018). Originalmente, o modelo de *gridmap* apresentado só possui aplicabilidade em aplicações que envolvem o deslocamento urbano, já que o modelo de segmentação utilizado segue a malha viária dentro do perímetro urbano. O presente trabalho possibilitou expandir a ideia original dos autores, ao apresentar algoritmos que, além de levarem em consideração a distribuição espacial do conjunto de medições de *Radio Frequency* (RF), também buscam generalizar as medições RF para toda a região de cobertura, a exemplo do que acontece com um sistema de localização baseado puramente em um *gridmap* de malha quadrangular.

Os benefícios da utilização de uma técnica de segmentação alternativa para a região de cobertura também foram comprovados. Utilizando o algoritmo Λ para geração de *gridmaps* adaptativos (*gridmap* irregular) foi possível atender aos requisitos de acurácia para localização horizontal da *Federal Communications Commission* (FCC) em ambientes sem fio com 2.500 células (FCC, 2015). Vale destacar que 2.500 era a menor configuração possível para o parâmetro c associado ao número de células. Em contrapartida, o algoritmo S , responsável por segmentar a região de cobertura em uma malha quadrangular, não foi capaz de alcançar os requisitos de acurácia para o erro de localização horizontal em nenhuma das seleções de parâmetro utilizadas no experimento. Essa característica corrobora para que a adoção de modelos de segmentação alternativos para região de cobertura, que explorem diferentes características dos elementos que compõem um sistema de localização, possam garantir uma redução significativa não apenas na acurácia, mas principalmente na complexidade computacional associada ao modelo de localização.

As melhorias propostas por intermédio da exploração de características geométricas dos *gridmaps* parecem promissoras, indicando que a influência deste elemento dos sistemas de localização impacta diretamente na acurácia e na complexidade computacional. Além da geometria do *gridmap*, este estudo apontou que deve haver um equilíbrio ótimo entre o *gridmap*, seu número de células e os elementos do conjunto de medições RF. É fato que um conjunto de medições RF, principalmente aquele obtido através de técnicas de *crowdsourcing*, carrega consigo informações relevantes sobre a mobilidade real das Estações Móveis (EMs) dentro de uma dada região, revelando padrões e até mesmo tendências de deslocamento. Os sistemas de localização baseados em *gridmap* ignoram essa informação, fornecendo a mesma relevância para qualquer local dentro da região de cobertura. Nesse sentido, um possível objeto de estudo seria desenvolver um ferramental matemático capaz de expressar tais propriedades, permitindo escolhas mais adequadas para os parâmetros associados aos elementos do sistema de localização, otimizando a acurácia ou a complexidade e minimizando os efeitos que induzem à saturação do modelo.

REFERÊNCIAS

- 3GPP. *Simulation Assumptions and Parameters for FDD HeNB RF Requirements. Alcatel-Lucent, picoChip Designs, and Vodafone*. [S.l.], 2009.
- 3GPP. *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects*. [S.l.], 2010a.
- 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*. [S.l.], 2010b.
- ABU-MAHFOUZ, A.; HANCKE, G. ALWadHA Localisation Algorithm: Yet More Energy Efficient. *IEEE Access*, v. 5, p. 6661–6667, 03 2017.
- AHA, W.; KIBLER, D.; ALBERT, M. Instance-Based Learning Algorithms. *Machine Learning*, v. 6, p. 37–66, 01 1991.
- AHMED, N.; MICHELIN, R.; XUE, W.; RUJ, S.; MALANEY, R.; KANHERE, S.; SENEVIRATNE, A.; HU, W.; JANICKE, H.; JHA, S. A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, PP, p. 1–1, 07 2020.
- ANDRADE, F. S. *Avaliação de um Modelo de Grid Irregular para Localização Outdoor Utilizando Simulação*. Dissertação (Graduação) — Universidade Federal de Pernambuco (UFPE), 2018.
- BIS, G. B. D. for B. I. . S. *Smart Cities: Background Paper*. 2013. Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246019/bis-13-1209-smart-cities-background-paper-digital.pdf>.
- BITTENCOURT, G.; URBANO, A.; CUNHA, D. A proposal of an RF Fingerprint-based Outdoor Localization Technique using Irregular Grid Maps. In: *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.l.: s.n.], 2018. p. 1–6.
- BRIDA, P.; MACHAJ, J.; BENIKOVSKÝ, J. A Modular Localization System as a Positioning Service for Road Transport. *Sensors (Basel, Switzerland)*, v. 14, p. 20274–96, 11 2014.
- CASSANO, E.; FLORIO, F.; RANGO, F. D.; MARANO, S. A Performance Comparison Between ROC-RSSI and Trilateration Localization Techniques for WPAN Sensor Networks in a Real Outdoor Testbed. In: *2009 Wireless Telecommunications Symposium*. [S.l.: s.n.], 2009. p. 1 – 8.
- CCTC, C. T. d. T. d. C. 2011. Disponível em: <<http://networks.cttc.es/mobile-networks/software-tools/lena/>>.
- CHOMSKY, N.; ANSHEN, R. *Biography of an Idea*. Mt. Kissco, NY: Moyer Bell Limited, 1986. 196 p.
- DANIEL, W. W. *Applied Nonparametric Statistics*. 2. ed. [S.l.]: PWS-KENT, 1990. 262–274 p.

- DAO, D.; RIZOS, C.; WANG, J. Location-Based Services: Technical and Business Issues. *GPS Solutions*, v. 6, p. 169–178, 12 2002.
- DEMSAR, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 01 2006.
- FCC, F. C. C. *Wireless E911 Location Accuracy Requirements*. [S.l.: s.n.], 2015. 74-104 p.
- GAO, F.; HAN, L. Implementing the Nelder-Mead Simplex Algorithm with Adaptive Parameters. *Computational Optimization and Applications*, v. 51, p. 259–277, 05 2012.
- GONZALEZ, T. F. Clustering to Minimize the Maximum Intercluster Distance. *Theoretical Computer Science*, Elsevier BV, v. 38, p. 293–306, 1985.
- GUVENC, I.; CHONG, C.-C. A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques. *Communications Surveys Tutorials, IEEE*, v. 11, p. 107 – 124, 01 2009.
- KIM, J.-H.; MIN, K. S.; YEO, W.-Y. A Design of Irregular Grid Map for Large-Scale Wi-Fi LAN Fingerprint Positioning Systems. *The Scientific World Journal*, Hindawi Limited, v. 2014, p. 1–13, 2014.
- KRÇADINAC, V. A New Generalization of the Golden Ratio. *The Fibonacci Quarterly*, v. 44, p. 335–340, 11 2006.
- KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. [S.l.]: Springer, 2016.
- LI, S. A 1.488 Approximation Algorithm for the Uncapacitated Facility Location Problem. In: *Automata, Languages and Programming*. [S.l.]: Springer Berlin Heidelberg, 2011. p. 77–88.
- LIU, H.; DARABI, H.; BANERJEE, P.; LIU, J. Survey of Wireless Indoor Positioning Techniques and Systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, v. 37, p. 1067 – 1080, 12 2007.
- LOVISOLO, L.; CAMPOS, R. RF Fingerprinting Location Techniques. In: *Handbook of Position Location: Theory, Practice and Advances*. [S.l.: s.n.], 2012. p. 487 – 520. ISBN 9780470943427.
- MA, H.; ZHANG, L.; HAN, Y. An Efficient RF Fingerprint Positioning Algorithm Based on Uneven Grid Layout. In: *6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)*. [S.l.: s.n.], 2015.
- MONDAL, R.; TURKKA, J.; HENTTONEN, T. Performance Evaluation of MDT Assisted LTE RF Fingerprint Framework. In: *2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*. [S.l.: s.n.], 2014. p. 33–37. ISBN 978-1-4799-2231-4.
- NEMENYI, P. *Distribution-free Multiple Comparisons*. Tese (Doutorado) — Princeton University, 1963.
- NGUYEN, K.; AKRAM, R. N.; MARKANTONAKIS, K.; LUO, Z.; WATKINS, C. Location Tracking Using Smartphone Accelerometer and Magnetometer Traces. In: . [S.l.: s.n.], 2019. p. 1–9. ISBN 978-1-4503-7164-3.

NSNAM. 2011. Disponível em: <<https://www.nsnam.org/>>.

PAHLAVAN, K.; KRISHNAMURTHY, P.; GENG, Y. Localization Challenges for the Emergence of the Smart World. *IEEE Access*, v. 3, p. 3058–3067, 12 2015.

PENG, B.; SECO-GRANADOS, G.; STEINMETZ, E.; FROHLE, M.; WYMEERSCH, H. Decentralized Scheduling for Cooperative Localization With Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, v. 68, p. 4295–4305, 04 2019.

ROBERTS, M. *The Unreasonable Effectiveness of Quasirandom Sequences*. 2018. Disponível em: <<http://extremelearning.com.au/unreasonable-effectiveness-of-quasirandom-sequences/>>.

ROSATI, U.; CONTI, S. What is a Smart City Project? An Urban Model or A Corporate Business Plan? *Procedia - Social and Behavioral Sciences*, v. 223, p. 968–973, 06 2016.

ROSS, P. Centimeter-Level GPS Positioning for Cars. *IEEE Spectrum*, 2016. Disponível em: <<https://spectrum.ieee.org/cars-that-think/transportation/sensors/centimeterlevel-gps-positioning-for-cars>>.

ROUT, S.; RATH, A.; BHAGABATI, C. Energy Efficient Dynamic Node Localization Technique in Wireless Sensor Networks. *Indian Journal of Science and Technology*, v. 10, p. 1–8, 04 2017.

SATAKE, T. Precise and Fast Interactive Area QoE Management Framework Toward 5G Era. In: *2016 17th International Telecommunications Network Strategy and Planning Symposium (Networks)*. [S.l.]: IEEE, 2016.

SCHEITHAUER, G. *Introduction to Cutting and Packing Optimization*. [S.l.]: Springer, 2017. 227-244 p.

SOUBIELLE, J.; FIJALKOW, I.; DUVAUT, P.; BIBAUT, A. GPS Positioning in a Multipath Environment. *Signal Processing, IEEE Transactions on*, v. 50, p. 141 – 150, 02 2002.

TIMOTEO, R. *Combinação Dinâmica de Regressores Aplicada à Localização de Usuários Móveis em Redes Sem Fio*. Tese (Doutorado) — Universidade Federal de Pernambuco (UFPE), 2020.

TIMOTEO, R.; SILVA, L.; CUNHA, D.; CAVALCANTI, G. An Approach using Support Vector Regression for Mobile Location in Cellular Networks. *Computer Networks*, v. 95, p. 51–61, 12 2015.

VARAGNOLO, D.; ZANELLA, F.; CENEDESE, A.; PILLONETTO, G.; SCHENATO, L. Newton-Raphson Consensus for Distributed Convex Optimization. v. 61, p. 994–1009, 04 2016.

VAZIRANI, V. V. *Approximation Algorithms*. [S.l.]: Springer, 2003. 47-53 p.

VO, Q.; DE, P. A Survey of Fingerprint-Based Outdoor Localization. *IEEE Communications Surveys & Tutorials*, v. 18, p. 1–1, 06 2015.

WANG, Q.; FENG, Y.; ZHANG, X.; SUN, Y.; LU, X. IWKNN: An Effective Bluetooth Positioning Method Based on Isomap and WKNN. *Mobile Information Systems*, v. 2016, p. 1–11, 01 2016.

WILCOXON, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, Elsevier BV, v. 1, n. 6, p. 80–83, 1945.

WILLIAMSON, D. P.; SHMOYS, D. B. *The Design of Approximation Algorithms*. [S.l.]: Cambridge University Press, 2011. 37–39 p.

YAN, J.; ZHAO, L.; TANG, J.; CHEN, Y.; CHEN, R.; CHEN, L. Hybrid Kernel Based Machine Learning Using Received Signal Strength Measurements for Indoor Localization. *IEEE Transactions on Vehicular Technology*, v. 67, p. 2824–2829, 11 2017.

ZAKNICH, A. *Principles of Adaptive Filters and Self-learning Systems*. [S.l.]: Springer, 2005. 1-3 p.

ZEKAVAT, S.; BUEHRER, R. *Handbook of Position Location: Theory, Practice, and Advances*. [S.l.: s.n.], 2011.