
**ARCADE: Uma Metodologia Orientada a
Dados para Modelagem e Aprimoramento da
Cobertura de Redes Celulares**

Daniel Ricardo Cunha Oliveira



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia
2026

Daniel Ricardo Cunha Oliveira

**ARCADE: Uma Metodologia Orientada a
Dados para Modelagem e Aprimoramento da
Cobertura de Redes Celulares**

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Flávio de Oliveira Silva, Ph.D.

Uberlândia

2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

O48
2026

Oliveira, Daniel Ricardo Cunha, 1974-
ARCADE: Uma metodologia orientada a dados para modelagem
e aprimoramento da cobertura de redes celulares [recurso
eletrônico] / Daniel Ricardo Cunha Oliveira. - 2026.

Orientador: Flávio de Oliveira Silva.

Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-
graduação em Ciência da Computação.

Modo de acesso: Internet.

DOI <http://doi.org/10.14393/ufu.te.2026.384>

Inclui bibliografia.

1. Computação. I. Silva, Flávio de Oliveira, 1970-, (Orient.). II.
Universidade Federal de Uberlândia. Pós-graduação em Ciência da
Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Tese, 12/2026, PPGCO				
Data:	06 de Maio de 2026	Hora de início:	09:30	Hora de encerramento:	13:51
Matrícula do Discente:	12113CCP001				
Nome do Discente:	Daniel Ricardo Cunha Oliveira				
Título do Trabalho:	ARCADE: Uma Metodologia Orientada a Dados para Modelagem e Aprimoramento da Cobertura de Redes Celulares				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Sistemas de Computação				
Projeto de Pesquisa de vinculação:	-----				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Rafael Pasquini - FACOM/UFU, Silvio Ereno Quincozes - Unipampa, Daniel Nunes Corujo - Universidade de Aveiro, Cristiano Bonato Both - CCET/Unisinus e Flávio de Oliveira Silva - Universidade do Minho, orientador do(a) candidato(a).

Os examinadores participaram desde as seguintes localidades: Flávio de Oliveira Silva - Braga/Portugal, Silvio Ereno Quincozes - Alegrete/RS, Daniel Nunes Corujo - Aveiro/Portugal, Cristiano Bonato Both - São Leopoldo/RS. O aluno e os outros membros participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Flávio de Oliveira Silva, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(á) Discente a palavra para a exposição do seu trabalho. A duração da apresentação do(a) Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Cristiano Bonato Both, Usuário Externo**, em 08/05/2026, às 16:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Silvio Ereno Quincozes, Usuário Externo**, em 08/05/2026, às 18:34, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rafael Pasquini, Professor(a) do Magistério Superior**, em 08/05/2026, às 19:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Daniel Nunes Corujo, Usuário Externo**, em 09/05/2026, às 14:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Flávio de Oliveira Silva, Usuário Externo**, em 11/05/2026, às 08:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7263524** e o código CRC **ED41DFA3**.

Este trabalho é dedicado aos que, através da ciência, e usando o conhecimento como ferramenta de transformação, tornam a sociedade mais igualitária e justa.

Agradecimentos

Há muitos a agradecer e pouco espaço para fazer jus a todos que contribuíram para que eu pudesse alcançar essa etapa, mas minimamente gostaria de mencionar:

Minha esposa e melhor amiga, prof. Alessandra Sousa Araújo, M.Sc., que também é uma acadêmica dedicada e contribui para a disseminação do conhecimento científico em nossa sociedade.

Minha mãe Maria Tereza Salete Cunha Oliveira, minha sogra Eleuza Sousa Araújo e meus irmãos Tatiana e Fábio.

Meu orientador e amigo Prof. Flávio de Oliveira Silva, que sempre me incentivou nas minhas conquistas acadêmicas e nunca me deixou desistir.

Prof. Rui Aguiar, do Instituto de Telecomunicações da Universidade de Aveiro (Portugal), que calorosamente me recebeu na instituição e contribuiu muito para o andamento desta tese, juntamente com os demais colegas do IT que, além de me orientarem no trabalho, me acolheram como bons amigos durante minha estadia em Aveiro.

Todos meus colegas da Faculdade de Computação da Universidade Federal de Uberlândia, além da equipe da secretaria do Programa de Pós-Graduação em Ciência da Computação, que também sempre me apoiaram no meu trabalho.

“Perseverance is not a long race; it is many short races one after the other.”
(Walter Elliot, político escocês)

Resumo

A evolução das redes móveis e o aumento de sua complexidade têm tornado o ajuste fino do ambiente de Radiofrequência (RF) cada vez mais complexo. Com isso, há um esforço dos órgãos padronizadores e da indústria para a automação de processos trabalhosos e complexos, em especial o de aprimoramento da cobertura da *Radio Access Network* (RAN) de um sistema celular. Algumas propostas, como a funcionalidade *Minimization of Drive Tests* (MDT), incluída nas últimas *releases* do *Third Generation Partnership Project* (3GPP), visam à obtenção de dados crus da rede de modo a viabilizar o desenvolvimento de ferramentas que automatizem um trabalho que até então é realizado por engenheiros especializados. Porém, como observado na prática, os dados de cobertura de RF são muitas vezes esparsos e não fornecem por si só informações abrangentes sobre a cobertura das células de modo a viabilizar uma análise criteriosa e metódica. Os principais problemas que levam à perda de eficiência de cobertura são os chamados *overshooting*, quando a célula cobre mais do que deveria, interferindo e reduzindo a capacidade das células vizinhas, e *undershooting*, quando a cobertura é muito menor do que deveria, causando falhas de serviço e subutilização dos recursos de redes. A proposta deste trabalho é apresentar a metodologia *Adaptative Radio Coverage Analysis through Data-driven Evolution* (ARCADE) que, baseando-se em dados de RF obtidos por quaisquer métodos (como MDT, *drive tests* ou *crowdsourcing*), permite a determinação de uma configuração melhorada da cobertura de rede por meio de um processo evolutivo, de modo que possa ser futuramente automatizável dentro do próprio sistema. A proposta também tem como premissa que não há dados de projeto disponíveis (como altura de torres, direcionamento ou modelos de antenas, ou base de dados de terrenos), permitindo uma implementação auto-contida no próprio sistema celular. Isso elimina a dependência da confiabilidade e acuracidade desses dados e a necessidade de aquisição e integração dessas bases, simplificando a implementação da solução, além dos desvios dos vários modelos matemáticos teóricos de predição de propagação normalmente usados hoje. Este trabalho pode ser dividido em dois problemas distintos a serem investigados sucessivamente. O primeiro deles vem da necessidade de se ter um modelamento adequado do ambiente de RF

com base em dados muitas vezes esparsos, dependendo do volume de usuários aderentes à funcionalidade ou ao aplicativo. Também temos que considerar que nos dados haverá pontos considerados como *outliers*, provenientes de erros de leitura ou processamento, ou mesmo de medidas muito esporádicas fora do contexto desejado (como por exemplo, de dispositivos de usuários localizados em terraços altos) e que podem distorcer radicalmente o modelamento das coberturas se não forem descartados. Além disso, as amostras também podem conter anomalias advindas de falhas de projeto, como células em *overshooting* (sobrecobertura) ou *undershooting* (subcobertura), sendo que essas anomalias devem ser corretamente representadas no modelo. O segundo problema é, após mapeado o ambiente de RF, o de se determinar de forma automática um conjunto de ajustes de parâmetros de rede de forma a se obter uma melhor eficiência de cobertura com qualidade – ou seja, cobertura da portadora com dominância efetiva sobre interferência – nesse sistema. Neste contexto, propõe-se nesta tese uma metodologia que abranja todos as etapas de um processo de melhoria da cobertura da rede que seja auto-contida no contexto de rede e com uma abordagem orientada a dados (*data-driven*), ou seja, que não dependa de modelagens matemáticas (por exemplo, modelos de propagação de rádio-frequências). Essa metodologia será proposta de maneira fim-a-fim, iniciando pelo exame da etapa de aquisição de dados (focado em uma implementação prática de *crowdsourcing* como alternativa ao *drive test*), passando pelo modelamento do ambiente de RF através de métodos para extrapolação dos dados que excluam *outliers* mas preservem as representações de anomalias de rede e, por fim, propondo um método de melhoria da cobertura do *cluster* que resulte em uma configuração de rede com desempenho superior com relação à configuração inicial. Os resultados mostram que, tanto a metodologia usada no modelamento é adequada ao problema proposto quanto o método para aprimoramento da cobertura mostrou um ganho significativo de qualidade de cobertura, com o ajuste (aumento ou redução) de potências das células do *cluster*, melhorando tanto a área coberta quanto a área com dominância de uma única servidora, situação ideal para redução da interferência de um sistema celular.

Palavras-chave: RAN, celular, redes móveis, melhoria de cobertura, expansão de dados, processos gaussianos, redes neurais, CMA-ES, LTE, 4G, 5G, 6G.

Abstract

The evolution of mobile networks and the increasing complexity of their management have made fine-tuning the RF environment progressively more challenging. As a result, standardization bodies and industry players have been investing in the automation of labor-intensive and complex processes, particularly the improvement of RAN coverage in cellular systems. Some proposals, such as the MDT functionality introduced in recent 3GPP releases, aim to collect raw network data to enable the development of tools that automate tasks previously carried out by specialized engineers. However, as observed in practice, RF coverage data are often sparse and do not by themselves provide comprehensive information about cell coverage, making systematic and thorough analysis difficult. The main problems leading to coverage efficiency loss are known as overshooting, where a cell covers more area than intended, interfering with and reducing the capacity of neighboring cells, and undershooting, where coverage falls well short of the intended area, causing service gaps and underutilization of network resources. This work proposes the Adaptative Radio Coverage Analysis through Data-driven Evolution (ARCADE) methodology, which, drawing on RF data obtained by any means (such as MDT, drive tests, or crowdsourcing), enables the determination of an improved network coverage configuration through an evolutionary process, designed to be eventually automated within the system itself. The proposal also assumes that no design data are available (such as tower heights, antenna orientation, or antenna models, or terrain databases), allowing a self-contained implementation within the cellular system itself. This eliminates dependence on the reliability and accuracy of such data and the need to acquire and integrate these databases, simplifying solution deployment, while also avoiding the deviations inherent in the various theoretical mathematical propagation prediction models commonly used today. This work can be divided into two distinct problems to be investigated sequentially. The first arises from the need for adequate modeling of the RF environment based on data that are often sparse, depending on the volume of users participating in the feature or application. It must also be considered that the data will contain outlier points, stemming from reading or processing errors, or from highly sporadic measurements outside the intended

context (such as those from user devices located on high rooftops), which can radically distort coverage modeling if not discarded. Furthermore, the samples may also contain anomalies arising from design faults, such as overshooting or undershooting cells, and these anomalies must be correctly represented in the model. The second problem, after the RF environment has been mapped, is the determination of a set of network parameter adjustments in order to achieve better coverage efficiency with quality – that is, carrier coverage with effective dominance over interference – in the system. In this context, this thesis proposes a methodology encompassing all stages of a network coverage improvement process that is self-contained within the network context and follows a data-driven approach – that is, one that does not rely on mathematical modeling (e.g., radio frequency propagation models). This methodology is proposed end-to-end, beginning with an examination of the data acquisition stage (focused on a practical crowdsourcing implementation as an alternative to *drive test*), proceeding through RF environment modeling via data extrapolation methods that exclude outliers while preserving representations of network anomalies, and finally proposing a cluster coverage improvement method that yields a network configuration with superior performance relative to the initial configuration. Results show that both the modeling methodology is well-suited to the proposed problem and the coverage improvement method achieved significant gains in coverage quality, with cell power adjustments (increases or reductions) within the cluster improving both the covered area and the area with single-server dominance – the ideal condition for interference reduction in a cellular system.

Keywords: RAN, cellular, mobile networks, coverage improvement, data expansion, gaussian processes, neural networks, CMA-ES, LTE, 4G, 5G, 6G.

Lista de ilustrações

Figura 1.1 – Representação da estrutura da tese e respostas às questões de pesquisa. Figura elaborada pelo autor.	35
Figura 2.1 – Componentes básicos de uma rede móvel e terminologias. A terminologia geral da primeira coluna será utilizada neste trabalho. Figura elaborada pelo autor.	39
Figura 3.1 – Esquemático da metodologia ARCADE para aprimoramento da RAN. Figura elaborada pelo autor.	54
Figura 3.2 – Grafo representando dois <i>clusters</i> independentes para efeito de exemplificação de conceito de definição de escopo de análise. Figura elaborada pelo autor.	57
Figura 3.3 – Exemplo de amostras de RF de uma única célula dispostas em uma determinada região geográfica. Figura elaborada pelo autor.	59
Figura 3.4 – Estrutura de grade georreferenciada e informações por elemento da grade. Figura elaborada pelo autor.	60
Figura 3.5 – Representação da amostragem após o aumento de dados com GPSK.	65
Figura 4.1 – Arquitetura simplificada da solução para aquisição de dados em <i>crowdsourcing</i> . Figura elaborada pelo autor.	77
Figura 4.2 – Foto do equipamento PCTEL IBflex. Fonte: (PCTEL, 2020).	82
Figura 4.3 – Mapa indicando a posição das estações componentes do <i>cluster</i> analisado. Figura elaborada pelo autor, usando mapa obtido no software Google Earth.	83
Figura 4.4 – Variação de R^2 e <i>Root Mean Squared Error</i> (RMSE) com α . Eixo X não é linear. Figura elaborada pelo autor	87
Figura 4.5 – Mapas de cobertura usando $\alpha = 0,00050$; $0,00100$; $0,00145$; $0,00170$; $0,00500$ e $0,05000$. Figura elaborada pelo autor.	87
Figura 4.6 – Mapas mostrando dados disponíveis e cobertura com níveis preditos de RSRP. Figura elaborada pelo autor.	89

Figura 4.7 – Duas representações geográficas das amostras do <i>cluster</i> , a primeira caracterizada pelo PCI, a segunda por classificação de normalidade. Figura elaborada pelo autor.	91
Figura 4.8 – Nível de sinal (RSRP) por amostra na geografia de análise. Figura elaborada pelo autor.	91
Figura 4.9 – Amostras por PCI individual, com representação de RSRP. Figura elaborada pelo autor.	92
Figura 4.10 – Variação de RMSE e R^2 com α . Figura elaborada pelo autor.	93
Figura 4.11 – Mapas de RSRP por PCI após aumento de dados por GPSK. Figura elaborada pelo autor.	94
Figura 4.12 – Áreas de Normalidade por PCI. Os pontos representam os baricentros das respectivas células e o retângulo, as respectivas áreas de normalidade. Figura elaborada pelo autor	95
Figura 4.13 – Mapas de RSRP por PCI após aumento de dados por RNA. Figura elaborada pelo autor.	98
Figura 4.14 – Processo de refinamento da calibração dos pesos de $\tilde{J}(P)$. Figura elaborada pelo autor	102
Figura 4.15 – Distribuição dos ajustes de potência por PCI. Figura elaborada pelo autor	107
Figura 4.16 – Mapas de cobertura e interferência antes e após o aprimoramento da cobertura baseado em <i>Covariance Matrix Adaptation Evolution Strategy</i> (CMA-ES). Regiões em branco indicam ausência de cobertura (RSRP < -105 dBm). Regiões em vermelho correspondem a pontos cobertos com dominância insuficiente (dominância < 9 dB), enquanto regiões em azul indicam pontos cobertos com dominância adequada (dominância \geq 9 dB).	109
Figura 4.17 – Mapa de variação do <i>Reference Signal Received Power</i> (RSRP) após a correção das potências obtidas com CMA-ES. Regiões em verde indicam aumento do RSRP da célula servidora, regiões em amarelo indicam redução, e regiões em azul indicam ausência de variação em relação à configuração de referência.	110
Figura 4.18 – Diagrama preliminar comparativo das etapas do processo de <i>tuning</i> de RAN. Figura elaborada pelo autor.	115
Figura 4.19 – Diagrama final comparativo das etapas do processo de <i>tuning</i> de RAN. Figura elaborada pelo autor.	120

Lista de tabelas

Tabela 2.1 – Resumo comparativo dos trabalhos relacionados em modelamento de RAN. Círculos preenchidos representam critérios aderentes, os não preenchidos, critérios não aderentes e os pontos, critérios não aplicáveis.	47
Tabela 2.2 – Resumo comparativo dos trabalhos relacionados em auto-organização e melhoria de cobertura de RAN. Círculos preenchidos representam critérios aderentes e não preenchidos, critérios não aderentes.	51
Tabela 4.1 – Exemplos de amostras obtidas por meio do aplicativo para <i>crowdsourcing</i> com dados anonimizados.	80
Tabela 4.2 – Configuração do <i>hardware</i> para processamento dos dados.	84
Tabela 4.3 – Quantidade de amostras, anomalias e <i>outliers</i> por PCI	90
Tabela 4.4 – Distâncias mínimas associadas às três anomalias mais próximas à A_N por PCI	96
Tabela 4.5 – Desempenho da RNA por PCI em termos de tempo de processamento e métricas de erro	99
Tabela 4.6 – Vetor de ajustes de potência na melhor solução por PCI	106
Tabela 4.7 – Análise de robustez considerando 1.000 execuções independentes do CMA-ES	107
Tabela 4.8 – Distribuição de cobertura e dominância antes e após as adequações de potência	110
Tabela 4.9 – Médias, dispersão e intervalos de confiança obtidos após 1.000 execuções.	111
Tabela 4.10–Desempenho computacional ao longo de 1.000 execuções independentes do CMA-ES	112
Tabela 4.11–Comparativo de melhoria de cobertura e qualidade entre o ARCADE e o modelo proposto em (FAGEN; VICHARELLI; WEITZEN, 2008).	113
Tabela 4.12–Consumo energético e emissão de carbono associados ao processamento das RNAs por PCI	118
Tabela A.1–Resultados para $\lambda = 0,1$	143
Tabela A.2–Resultados para $\lambda = 0,3$	144

Tabela A.3–Resultados para $\lambda = 1$	144
Tabela A.4–Resultados para $\lambda = 0,1$	145
Tabela A.5–Resultados para $\lambda = 0,3$	146
Tabela A.6–Resultados para $\lambda = 0,1$	147
Tabela A.7–Resultados para $\lambda = 0,3$	148
Tabela A.8–Resultados da Fase 4	148

Lista de siglas

3GPP *Third Generation Partnership Project*

API *Application Programming Interface*

APN *Access Point Name*

ARCADE *Adaptative Radio Coverage Analysis through Data-driven Evolution*

AWS *Amazon Web Services*

CMA-ES *Covariance Matrix Adaptation Evolution Strategy*

CN *Core Network*

CI *Cell Identity*

CRS *Cell-Specific Reference Signal*

CCO *Coverage and Capacity Optimization*

CGI *Cell Global Identity*

CSV *Comma-Separated Values*

COM *Cluster Optimization Module*

DAM *Data Acquisition Module*

DE *Differential Evolution*

DPM *Data Preprocessing Module*

DRL *Deep Reinforcement Learning*

DT *drive test*

dBm *decibéis-miliwatt*

dB *decibel*

DBSCAN *Density-Based Spatial Clustering of Applications with Noise*

DDPG *Deep Deterministic Policy Gradient*

EMS *Element Management System*

E-UTRAN *Evolved Universal Terrestrial Radio Access Network*

EIRP *Effective Isotropic Radiated Power*

EPS *Evolved Packet System*

EC2 *Elastic Compute Cloud*

EARFCN *E-UTRA Absolute Radio Frequency Channel Number*

FR1 *Frequency Range 1*

FR2 *Frequency Range 2*

FDD *Frequency Division Duplex*

FSPL *Free-Space Path Loss*

GPU *Graphics Processing Unit*

GPSK *Gaussian Process with Spatial Kernel*

GNSS *Global Navigation Satellite System*

GERAN *GSM EDGE Radio Access Network*

GA *Genetic Algorithm*

GPS *Global Positioning System*

IoT *Internet of Things*

IA *Inteligência Artificial*

KPI *Key Performance Indicator*

LTE *Long Term Evolution*

MDT *Minimization of Drive Tests*

MR *Measurement Report*

MSE *Mean Squared Error*

MAE *Mean Absolute Error*

MIMO *Multiple Input Multiple Output*

mMIMO *Massive Multiple Input Multiple Output*

MTGM *Model Training and Generalization Module*

MCC *Mobile Country Code*

MNC *Mobile Network Code*

MCTI *Ministério da Ciência, Tecnologia e Inovação do Brasil*

NMS *Network Management System*

NG-RAN *Next-Generation Radio Access Network*

NDT *Network Digital Twin*

O-RAN *Open RAN*

PDU *Packet Data Unit*

PCI *Physical Cell ID*

PCA *Principal Component Analysis*

PSO *Particle Swarm Optimization*

PCRF *Policy and Charging Rules Function*

RAN *Radio Access Network*

RBF *Radial Basis Function*

RE *Resource Element*

RIC *RAN Intelligent Controller*

RF *Radiofrequência*

RMSE *Root Mean Squared Error*

RSRP *Reference Signal Received Power*

RSRQ *Reference Signal Received Quality*

RRC *Radio Resource Control*

RSSI *Received Signal Strength Indicator*

RNA *Rede Neural Artificial*

RBS *Radio Base Station*

ReLU *Rectifier Linear Unit*

SO *sistema operacional*

SON *Self-Organizing Networks*

SINR *Signal-to-Interference-plus-Noise Ratio*

SAM *Samples Augmentation Module*

SDI *Spatial Distribution Index*

TAC *Tracking Area Code*

TTW *Tank-to-Wheel*

TabPFN *Tabular Prior-data Fitted Network*

UE *User Equipment*

UTRAN *Universal Terrestrial Radio Access Network*

WTW *Well-to-Wheel*

Sumário

1	INTRODUÇÃO	29
1.1	Motivação	30
1.2	Objetivos	31
1.3	Hipóteses de Pesquisa	32
1.4	Contribuições	33
1.5	Organização da Tese	34
2	FUNDAMENTAÇÃO TEÓRICA	37
2.1	Introdução	37
2.2	Conceitos Básicos	38
2.2.1	Arquitetura de Sistemas Móveis 3GPP	38
2.2.2	Cobertura e Interferência em Redes LTE e 5G	38
2.2.3	Obtenção de Dados de Radiofrequência	40
2.2.4	Processos Gaussianos	42
2.2.5	Redes Neurais Artificiais	43
2.2.6	CMA-ES	43
2.3	Trabalhos Relacionados	44
2.3.1	Trabalhos Relacionados à Modelagem de RAN	44
2.3.2	Trabalhos Relacionados ao Aprimoramento de Cobertura Celular	46
3	ARCADE: ADAPTATIVE RADIO COVERAGE ANALYSIS THROUGH DATA-DRIVEN EVOLUTION	53
3.1	Metodologia ARCADE	53
3.1.1	<i>Data Acquisition Module (DAM)</i>	53
3.1.2	<i>Data Preprocessing Module (DPM)</i>	54
3.1.3	<i>Samples Augmentation Module (SAM)</i>	55
3.1.4	<i>Model Training and Generalization Module (MTGM)</i>	55
3.1.5	<i>Cluster Optimization Module (COM)</i>	55

3.2	Definição do <i>Cluster</i> de Análise	56
3.3	Aquisição de Dados via <i>Crowdsourcing</i> (DAM)	58
3.4	Modelagem dos Dados de Entrada (DPM)	58
3.4.1	Classificação de Amostras	61
3.4.2	Classificação de Áreas	62
3.5	Processos Gaussianos para Aumento de Dados (SAM)	64
3.5.1	Aumento de Amostragem na Área de Anormalidade	64
3.6	Modelagem de Cobertura Usando Rede Neural Artificial (RNA) (MTGM)	66
3.7	Método para Aprimoramento de Cobertura (COM)	66
3.7.1	Formulação da Função Objetivo	67
3.7.2	Definição da Função Objetivo Básica $J(P)$	71
3.7.3	Resolução do Problema Utilizando-se CMA-ES	72
4	AVALIAÇÃO EXPERIMENTAL	75
4.1	Aquisição de Dados Via <i>Crowdsourcing</i>	76
4.1.1	Arquitetura da Solução	76
4.1.2	Informações Coletadas e Disponibilizadas	76
4.1.3	Desafios na Implementação	78
4.1.4	Resultados Obtidos	79
4.2	Cenário Real Analisado e Aquisição dos Dados de Cobertura .	80
4.2.1	Equipamentos de Coleta de Amostras	81
4.2.2	<i>Cluster</i> de Células Analisado	82
4.2.3	Hardware para Processamento dos Dados	84
4.3	Avaliação Experimental de Predição de Cobertura	84
4.3.1	Aumento da Amostragem com GPSK	84
4.3.2	Generalização com RNA (Módulo MTGM)	88
4.4	Aumento de Dados para o <i>Cluster</i>	89
4.4.1	Extrapolação de Dados com <i>Gaussian Process with Spatial Kernel</i> (GPSK) .	93
4.4.2	Obtenção de Cobertura com RNA	97
4.5	Aprimoramento da Qualidade de Cobertura do <i>Cluster</i>	97
4.5.1	Definição de Parâmetros e Calibração de Pesos de $\tilde{J}(P)$	100
4.5.2	Considerações Finais sobre a Calibração	104
4.5.3	Execução do Aprimoramento de Cobertura com os Pesos Definidos . . .	105
4.6	Resultados e Discussão	107
4.6.1	Avaliação da Robustez do Método	107
4.6.2	Melhorias de Cobertura e Interferência	108
4.6.3	Ganhos Quantitativos de Desempenho	110
4.6.4	Aspectos Computacionais e Considerações de Implantação	111
4.6.5	Comparação dos Resultados Obtidos com Trabalhos Relacionados . . .	112

4.7	Ponderações sobre Eficiência Energética e Emissão de Carbono	114
4.7.1	Metodologia para Comparação de Emissão de Carbono	114
4.7.2	Emissão de Carbono na Etapa de Aquisição de Amostras	115
4.7.3	Emissão de Carbono na Etapa de Extrapolação de Dados	116
4.7.4	Emissão de Carbono na Etapa de <i>Tuning</i> da RAN	119
4.7.5	Conclusão sobre Impacto da Metodologia ARCADE na Emissão de Carbono	119
4.8	Respostas às Questões de Pesquisa	121
5	CONSIDERAÇÕES FINAIS	127
5.1	Principais Contribuições	128
5.2	Contribuições em Produção Bibliográfica	129
5.3	Trabalhos Futuros	130
	REFERÊNCIAS	133

APÊNDICES **141**

APÊNDICE A	– RESULTADOS COMPLETOS DO PROCESSO DE CALIBRAÇÃO DOS PESOS	143
A.1	Fase 1: Análise Preliminar	143
A.2	Fase 2: Primeiro Refinamento	144
A.3	Fase 4: Validação Final e Consolidação dos Pesos	148

Introdução

Nos últimos anos, as redes celulares tornaram-se um componente indispensável e essencial da sociedade contemporânea, especialmente em uma comunidade digitalmente globalizada que depende de serviços ubíquos e confiáveis. Até 2027, espera-se que o 5G se torne a tecnologia dominante no mercado global e, até 2030, o número de assinantes deve atingir 9,5 bilhões, com 6,3 bilhões utilizando 5G e alguma penetração incipiente de 6G (ERICSSON, 2025). Esse aumento exponencial no uso de dispositivos conectados tornou as *Radio Access Networks* (RANs) extremamente complexas, especialmente considerando a coexistência de múltiplas gerações de tecnologia, nomeadamente 2G, 3G, 4G e 5G, e a disseminação de dispositivos do tipo *Internet of Things* (IoT). Nesse contexto, há uma demanda crescente por soluções que simplifiquem o aprimoramento da qualidade da RAN, particularmente porque, em sistemas como 4G e 5G, a autointerferência da rede está inversamente correlacionada com a eficiência e a capacidade de tráfego. Os métodos tradicionais de melhoria de rede, baseados em processos manuais e em conhecimentos técnicos altamente especializados, estão se tornando cada vez menos eficazes à medida que a capacidade e a complexidade do sistema se expandem.

Para enfrentar esses desafios, há uma tendência clara em direção ao desenvolvimento de métodos automatizáveis para aprimoramento da RAN. Esse conjunto de soluções é denominado *Self-Organizing Networks* (SON). As técnicas de SON baseiam-se em um misto de padronizações (3GPP, 2024d) e soluções proprietárias, frequentemente utilizando inteligência artificial e aprendizado de máquina para se aprimorar configurações e ajustes da RAN.

Um dos desafios ao trabalhar com conjuntos de dados obtidos da RAN advém de suas esparsidades, que dependem do método de aquisição dos dados e de fatores inerentes a cada um deles, como a densidade da coleta dos *drive tests* (DTs), ou nível de adesão dos dispositivos móveis à funcionalidade *Minimization of Drive Tests* (MDT), para citar alguns exemplos. Outro desafio no uso de métodos convencionais de interpolação é a presença de anomalias de cobertura, frequentemente causadas por *overshooting* ou *undershooting* de células, resultantes de deficiências na configuração da RAN e a presença

de *outliers*, causados por erros de medição ou amostragens atípicas (por exemplo, medições de usuários em terraços elevados). Essas anomalias tendem a fazer com que métodos simples de interpolação produzam modelos imprecisos, tipicamente exagerando anormalidades e criando artefatos, devido ao sobreajuste do modelo e à ausência de condições de contorno adequadas, enquanto os *outliers* comprometem a capacidade de generalização de um modelo, introduzindo distorções que não representam a realidade.

Porém, uma vez que os dados sejam tratados e uma cobertura “contínua” seja modelada em uma base de dados que represente bem a realidade da rede, ainda resta o problema de aprimoramento da rede com uma metodologia que seja automatizável.

O desafio proposto nesta tese consiste, além da ausência de dados de projeto, também na restrição do uso de modelos matemáticos de predição. Isso quer dizer que o modelo proposto neste trabalho não deve depender nem de uma base de dados de posicionamento e configuração de *Radio Base Stations* (RBSs), nem de uma base topológica e morfológica, necessária para cálculos teóricos de predição de cobertura baseados em modelos clássicos como o de Okumura, Hata ou COST-231 (SINGH, 2012). Em outras palavras, a solução deve, idealmente, ser totalmente orientada a dados (*data-driven*) e independente de informações externas ao próprio sistema celular.

Desta forma, será proposta nesta tese uma metodologia denominada *Adaptative Radio Coverage Analysis through Data-driven Evolution* (ARCADE), que compreende uma abordagem para modelamento do ambiente de Radiofrequência (RF) e posterior aprimoramento de cobertura de redes móveis. A proposta tem como objetivo final obter, sem informações externas ao sistema, uma melhoria significativa da quantidade e qualidade de cobertura de um *cluster* de RBSs, viabilizando uma futura possibilidade de automação de processo de aprimoramento da RAN. A metodologia será avaliada através da implementação de uma prova de conceito experimental, utilizando-se dados reais obtidos de uma operadora comercial por meio de DT.

No entanto, é importante ressaltar que a proposta ARCADE não é um substituto dos processos de *tuning* já praticados pelas operadoras, mas um complemento ao mesmo. É essencial que uma rede já se encontre com seus parâmetros físicos calibrados, por exemplo, com bons posicionamentos geográficos das estações e correto direcionamento das antenas, antes que a metodologia proposta seja implementada. Ou seja, trata-se de uma metodologia complementar e não substituível do processo de *tuning* de rede tradicional.

1.1 Motivação

A natureza complexa e trabalhosa do processo de melhoria de uma RAN, o crescente incremento de tecnologias disponíveis (enquanto as legadas se mantêm ativas) e o contínuo adensamento das redes requerem um grande grau de esforço para que essa atividade seja desenvolvida a contento. Uma rede com menor interferência entre suas próprias células

possui uma eficiência de capacidade notadamente maior, de forma que uma maior qualidade de cobertura implica diretamente em menor impacto de CapEx e OpEx por parte das operadoras. A melhoria do desempenho financeiro em uma operação mais eficiente é naturalmente um grande motivador para a busca de soluções que tentem aumentar essa eficiência.

Porém, alguns desafios se apresentam nesse sentido. O primeiro é que as padronizações hoje abordam apenas aspectos mais gerais para as SONS (3GPP, 2024d), deixando as implementações práticas a cargo de fornecedores de tecnologia, normalmente as grandes indústrias do setor (NOKIA, 2025) (SAMSUNG, 2025). Naturalmente, essas implementações são proprietárias e suas especificações não são públicas.

Um segundo desafio advém da quantidade de informações e bases de dados necessárias para avaliação da eficiência de uma cobertura celular. Muitas das plataformas usam modelos matemáticos para o cálculo de perda de propagação. Além disso, os sistemas de predição e tratativas tradicionalmente requerem *Key Performance Indicators* (KPIs) de sistemas de gerência de rede, sejam eles de níveis *Element Management System* (EMS) ou *Network Management System* (NMS), bases topológicas e morfológicas da geografia analisada, além de informações sobre o projeto da rede, que incluem posicionamento de RBSs, azimute, *downtilt* e modelo de antenas, informações sobre vizinhanças de células, dentre outros (WANG; HANDURUKANDE, 2018) (SKOCAJ et al., 2022) (DREIFUERST et al., 2021) (OJO; IMOIZE; ALIENYI, 2021). O levantamento dessas informações demanda integrações de sistemas EMS/NMS, além da aquisição de bases geográficas e depende de se ter uma base com o projeto da rede que represente fielmente o fisicamente implantado, o que demanda esforço e custos extras no processo. Esses desafios, em si, são grandes motivadores para a descrição de um método que simplifique o processo de *tuning* de rede e contribua para uma futura automação desse processo.

Por fim, um último motivador é a complexidade crescente das redes celulares. O adensamento progressivo de estações aumenta a probabilidade de ocorrência de anomalias de cobertura, como áreas de sombra ou de superposição de células. Quando o ambiente se torna muito intrincado, a simples inspeção visual de dados tende a apresentar-se caótica e pouco intuitiva, tornando o processo de aprimoramento da rede demasiadamente complexo. Uma metodologia de melhoria estruturada, com processos claramente definidos e arquitetura auto-contida, isto é, que não dependa de informações externas nem de integração com sistemas de terceiros, tem um grande potencial para adoção prática e aplicabilidade direta na indústria.

1.2 Objetivos

O objetivo geral desta tese é propor uma metodologia fim a fim para viabilizar o aprimoramento de cobertura celular de forma automatizável (ainda que a solução expe-

rimental apresentada neste trabalho não opere, por ora, de forma autônoma), contando apenas com dados georreferenciados de níveis de cobertura por célula (*Reference Signal Received Power* (RSRP)), sem o auxílio de informações de projetos de RBSs, KPIs de sistemas de gerência ou bases de dados geográficos ou morfológicos. Isso também implica na não utilização de modelos empíricos de predição matemática de cobertura, de forma que o modelo deve se basear apenas em dados reais. Além disso, os dados de cobertura podem ser esparsos e não abranger totalmente a área analisada, ou seja, há necessidade de extrapolação da cobertura com base nos dados informados. Esses dados podem também conter anomalias e *outliers*, que impõem desafios extras nessa extrapolação de cobertura.

Como forma de atingir esse objetivo geral, os seguintes objetivos específicos são definidos:

- Propor um método para modelamento dos dados de RF coletados de modo a refletir eventuais anomalias de cobertura, representadas por dados isolados, de forma que não haja um exagero na estimativa de cobertura nem um descarte de dados relevantes, porém que também leve em consideração a presença de eventuais *outliers* que possam distorcer esse modelamento, e que devem ser desprezados;
- Propor um método para o aprimoramento da rede de acesso de rádio usando apenas as informações extrapoladas obtidas das medições, fazendo com que um processo auto-contido consiga ajustar suas configurações de modo tanto a maximizar a cobertura da área analisada quanto minimizar a interferência entre as células do *cluster*.

1.3 Hipóteses de Pesquisa

As hipóteses da pesquisa são baseadas nas viabilidades de demonstração dos objetivos específicos da tese, ou seja, com a demonstração de que existem métodos viáveis para o atingimento das propostas apresentadas.

A seguir são apresentadas as hipóteses definidas nessa tese:

- **Hipótese 1:** *A extrapolação contínua das coberturas das células em uma geografia, a partir de dados de RF esparsos adquiridos de uma rede celular operacional, é condicionada à adoção de um método que trate diferenciadamente as amostras presentes nesses dados: preservando na extrapolação aquelas que representam anomalias de cobertura e descartando as que constituem outliers.*

As seguintes questões de pesquisa estão associadas à Hipótese 1.

- **Q1:** Como representar as amostras na geografia a ser analisada de uma forma contínua e organizada, de modo que caracterizem o ambiente de RF de forma acurada?

- **Q2:** Como, célula a célula, determinar-se quais amostras podem ser consideradas normais, quais são anomalias e quais correspondem a *outliers*?
 - **Q3:** Quais métodos são eficazes para se realizar a extrapolação dos dados, inferindo-se valores ausentes em coordenadas onde não há amostras?
 - **Q4:** Como fazer com que as amostras anômalas não interfiram no modelo de extrapolação, exagerando na representação da cobertura, sem que sejam simplesmente descartadas?
- **Hipótese 2:** *A quantificação consistente da qualidade de cobertura de um cluster de células quanto ao grau de sua eficiência é condicionada exclusivamente aos dados georreferenciados de RF, dispensando quaisquer informações externas ao sistema.*

A questão de pesquisa associada à Hipótese 2 é:

- **Q5:** Dado o mapa de cobertura contínuo do *cluster* de células, na forma de um banco de dados estruturado que representa a configuração do sistema, como quantificar a qualidade dessa cobertura de forma a se poder avaliar seu grau de eficiência?
- **Hipótese 3:** *A melhoria do desempenho do sistema, mensurada pelo aumento da qualidade de cobertura na área analisada, é condicionada à existência de um método matemático formal de ajuste dos parâmetros das células do cluster, cuja formulação pressupõe a validade das Hipóteses 1 e 2.*

A questão de pesquisa associada à Hipótese 3 é:

- **Q6:** Qual abordagem de modelagem e avaliação de desempenho é mais adequada para permitir que o sistema determine ajustes em seus parâmetros, de modo a melhorar a qualidade da cobertura na área atendida pelo cluster de células?

1.4 Contribuições

A proposta de uma metodologia formal que permita o aprimoramento da cobertura de um sistema baseado puramente em dados coletados do ambiente de rádio (RAN) pode simplificar bastante um processo que, na prática, é recorrente nas operadoras móveis, permitindo a melhoria constante da eficiência da rede baseada em regras claras e bem definidas. A automação desse processo sem dúvida contribui significativamente para auxiliar operacionalmente o processo de *tuning* de rede, em adição ao processo de calibração de parâmetros físicos.

Desta forma, as principais contribuições do presente trabalho são:

- ❑ Uma metodologia para extrapolação de dados esparsos de forma que não haja uma superestimativa de anomalias, mas também não se percam informações relevantes de inconsistências de rede.
- ❑ Um método para que o próprio sistema consiga potencialmente se avaliar e propor alterações em seus parâmetros de funcionamento de forma a maximizar seu desempenho no ambiente de RF.

1.5 Organização da Tese

O restante desta tese está organizado da seguinte maneira.

O Capítulo 2 apresenta alguns conceitos básicos para o entendimento do trabalho e uma visão geral do mesmo. São introduzidas visões dos sistemas de telecomunicações móveis, além de fundamentos das ferramentas matemáticas e computacionais a serem utilizadas na metodologia.

Ainda no Capítulo 2 será apresentada uma revisão bibliográfica sistemática de trabalhos relacionados ao tema da tese. Esses trabalhos são classificados de acordo com os objetivos da tese, englobando o processo de modelamento de ambiente de rádio de sistemas móveis e os métodos de aprimoramento de RAN.

O Capítulo 3 apresenta o ARCADE e fornece uma descrição detalhada da metodologia proposta de forma didaticamente estruturada em módulos, juntamente com uma proposta de estratégia racional para escolha de um *cluster* de células e geografia de análise. Além da apresentação do ARCADE, ele será dividido em duas seções distintas. A primeira descreve o método usado para extrapolação de dados de forma a descrever-se continuamente o ambiente de rádio. A segunda propõe a abordagem para a melhoria do *cluster* de células através da definição de uma função objetivo e sua avaliação e aprimoramento.

Já o Capítulo 4 descreve os experimentos realizados para a validação prática das propostas apresentadas no Capítulo 3. Detalha-se a metodologia usada para essa validação, bem como análises quantitativas das inferências obtidas a partir de dados reais advindos de uma operação móvel em ambiente de produção.

Finalmente, no Capítulo 5, são apresentadas as considerações finais sobre a tese, abordando as limitações da proposta e sugerindo possíveis direções para trabalhos futuros. Esse capítulo responde às questões de pesquisa levantadas na Seção 1.3, revisitando as principais contribuições do estudo e listando as publicações mais relevantes originadas a partir desta pesquisa.

As questões de pesquisa serão endereçadas e respondidas nas seções e capítulos de acordo com o diagrama mostrado na Figura 1.1.

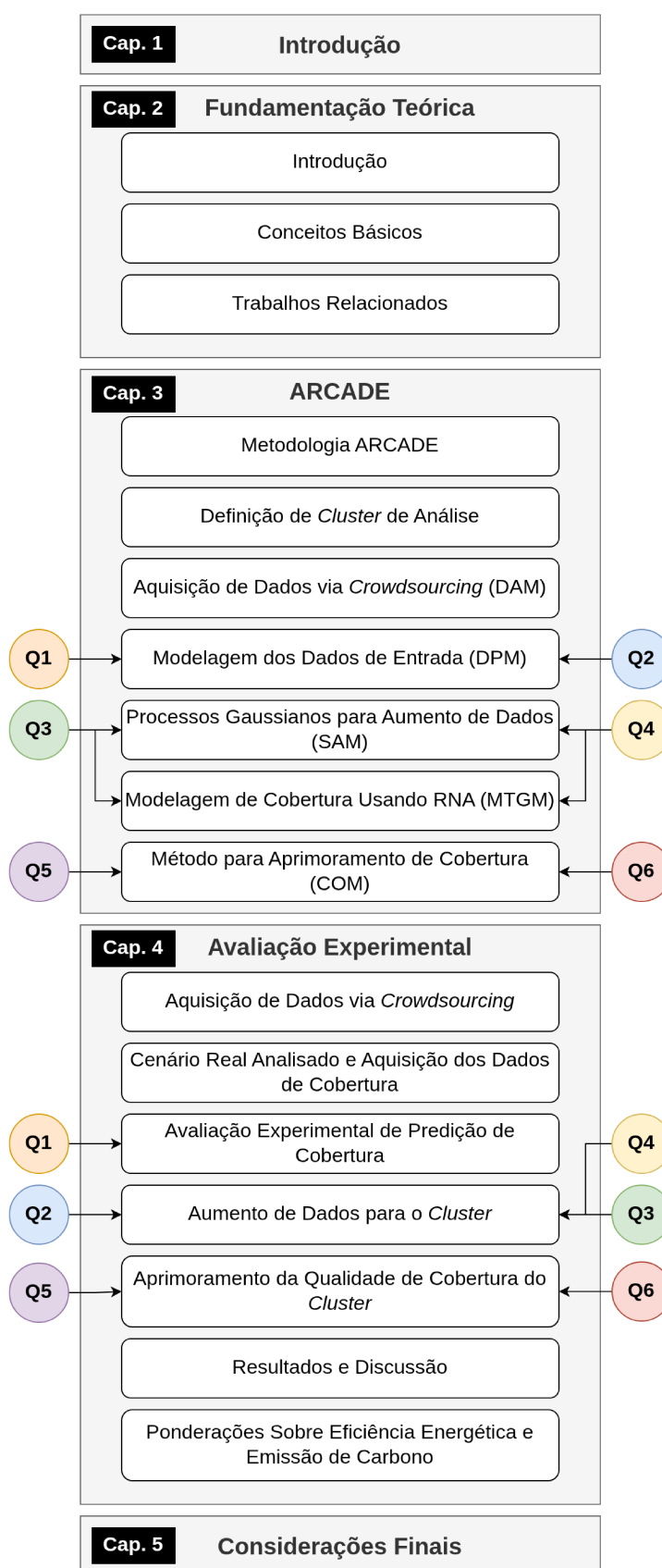


Figura 1.1 – Representação da estrutura da tese e respostas às questões de pesquisa. Figura elaborada pelo autor.

Fundamentação Teórica

2.1 Introdução

A partir do momento em que os sistemas celulares começaram a se popularizar – ao contrário do período inicial, quando aparelhos celulares eram vistos como artigos de luxo –, a complexidade da infraestrutura de rede aumentou significativamente, visando suportar não apenas o crescente volume de tráfego de voz, mas também a demanda emergente por serviços de dados. Houve sobretudo a necessidade do adensamento de RBSs com o propósito de se colocar mais usuários simultâneos em uma mesma área geográfica, mantendo-se os recursos de espectro. Esse processo de adensamento de estações demandou das operadoras uma necessidade de se ajustar continuamente os parâmetros de rede, principalmente para conter coberturas de células que anteriormente deveriam cobrir grandes áreas. Essa atividade é denominada *tuning* de rede, é hoje um trabalho contínuo e essencial na operação móvel, e é normalmente desenvolvido por engenheiros qualificados nas próprias operadoras, empresas terceiras especializadas ou até grandes fornecedores de tecnologia que incorporam esse serviço em suas soluções de expansão de rede.

Há algum tempo, a indústria e órgãos de padronização se esforçam para automatizar as atividades de melhoria de rede, com resultados mistos. Mesmo as padronizações que definem o conceito de SON têm avançado timidamente quando comparadas com as demandas atuais (3GPP, 2024d). Soluções proprietárias também têm ganhado impulso, mas normalmente são caras e demandam um esforço significativo para serem efetivas em seus propósitos (NOKIA, 2025) (SAMSUNG, 2025).

A solução proposta nesta tese envolve o desenvolvimento de um modelo que permita que um sistema celular já implementado com boas configurações físicas seja avaliado e aprimorado com base apenas em dados de RF, que podem ser adquiridos de diversas formas, como será apresentado na Seção 2.2.3. Ao conjunto de procedimentos que compõe essa metodologia, foi dado o nome de ARCADE.

Esse capítulo é dividido em duas partes: na Seção 2.2 são apresentados conceitos básicos para o entendimento dos problemas e das soluções propostas e, na Seção 2.3,

é realizada uma revisão da literatura, com trabalhos já desenvolvidos em abordagens similares, complementares ou alternativas aos métodos apresentados.

2.2 Conceitos Básicos

A seguir serão apresentados alguns conceitos básicos da arquitetura de sistemas que utilizam o padrão *Third Generation Partnership Project* (3GPP), da cobertura e interferência em sistemas celulares modernos, dos processos envolvidos na obtenção de dados de RF, além das ferramentas matemáticas e computacionais usadas na abordagem do problema de modelamento de cobertura e melhoria da rede.

2.2.1 Arquitetura de Sistemas Móveis 3GPP

Em todas as arquiteturas de redes móveis, independente da geração, é possível dividir um sistema celular em macro componentes bem distintos, que integram o caminho fim a fim da informação. Esses componentes são interconectados por meio de interfaces padronizadas, de forma a permitir a interoperabilidade entre fornecedores de tecnologia (SESIA; TOUFIK; BAKER, 2009). A denominação formal de cada componente varia de geração a geração. Por exemplo, a RAN é denominada *GSM EDGE Radio Access Network* (GERAN) no 2G, *Universal Terrestrial Radio Access Network* (UTRAN) no 3G, *Evolved Universal Terrestrial Radio Access Network* (E-UTRAN) no 4G e *Next-Generation Radio Access Network* (NG-RAN) no 5G (3GPP, 2024f). Para não haver dúvidas, neste trabalho será usada a terminologia geral, que é usualmente empregada nos meios da indústria e academia: *User Equipment* (UE) para designar o dispositivo do usuário, RBS na designação da estação rádio-base (cujo conjunto é chamado RAN) e *Core Network* (CN) para o *core network*. A Figura 2.1 ilustra esses componentes, juntamente com as terminologias específicas para o 4G e 5G.

2.2.2 Cobertura e Interferência em Redes LTE e 5G

O correto dimensionamento de cobertura e interferência das células na RAN é um imprescindível aspecto a ser considerado no projeto de uma rede móvel, de modo a se ter o máximo de eficiência de recursos caros e limitados, como o espectro de radiofrequência e os ativos que compõem as RBSs. Um bom trabalho preliminar tanto de projeto quanto de posterior *tuning* da RAN é essencial antes de qualquer outra camada visando a melhoria de qualidade de cobertura seja implementada.

A área de cobertura de um sistema celular é dividida em células, sendo essas logicamente implementadas por uma RBS. A cobertura de uma célula pode ser ajustada com relação à sua abrangência em uma área geográfica, usualmente alterando-se parâmetros como a inclinação (*downtilt*) da antena (angulação física da antena em seu eixo

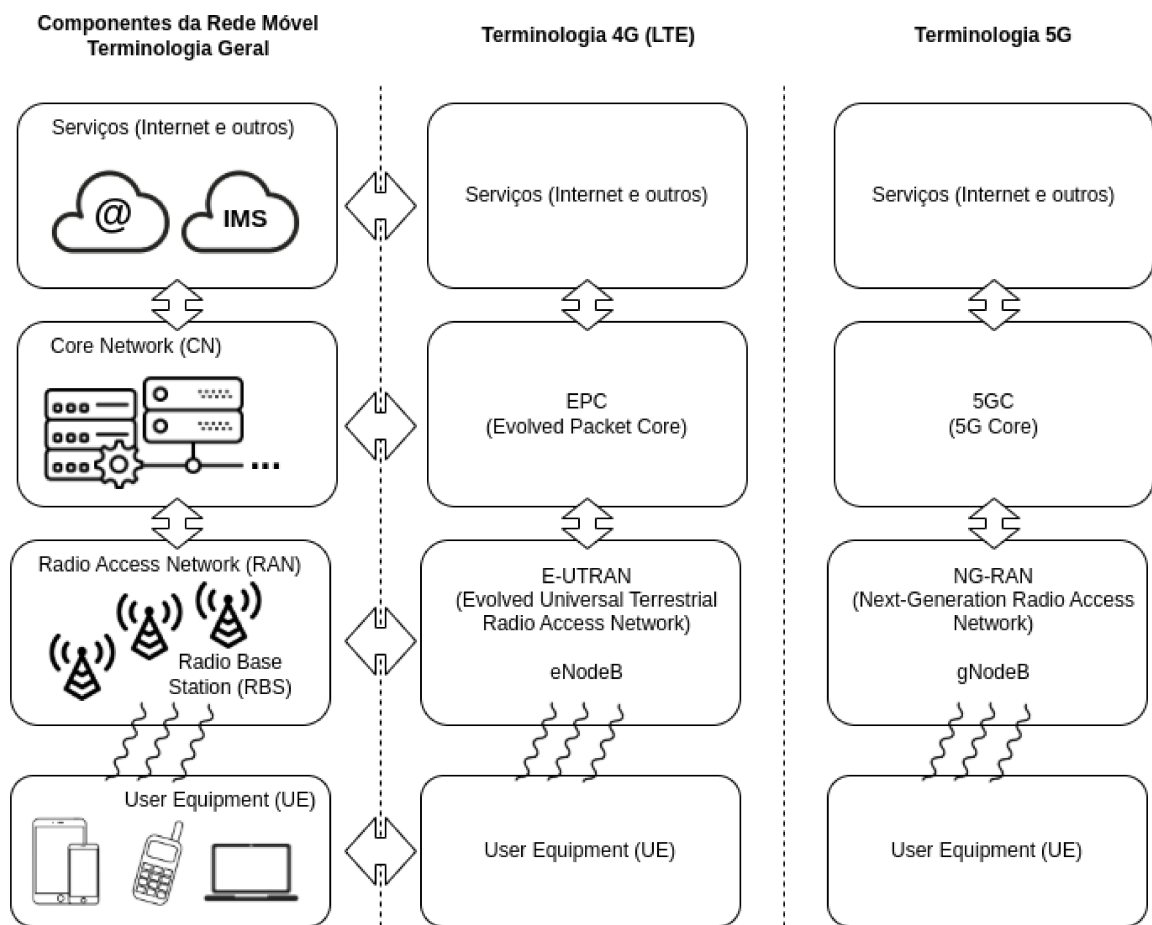


Figura 2.1 – Componentes básicos de uma rede móvel e terminologias. A terminologia geral da primeira coluna será utilizada neste trabalho. Figura elaborada pelo autor.

vertical), azimute (angulação em seu eixo horizontal), ou *Effective Isotropic Radiated Power* (EIRP), mediante ajuste da potência de transmissão do rádio. Em *Long Term Evolution* (LTE), a potência associada aos sinais de referência, ou *Cell-Specific Reference Signals* (CRSs), utilizados pelas UEs para medições de RSRP e decisões de mobilidade, pode ser configurada por meio do parâmetro de potência de referência do sinal (*referenceSignalPower*), influenciando diretamente a percepção de cobertura, o que será especialmente importante neste trabalho (GHOSH, 2017).

Com relação ao mecanismo de mitigação de interferência no LTE e no 5G, as especificações foram baseadas em técnicas avançadas de modulação e processamento de sinal que permitem que uma mesma faixa de frequência seja reutilizada em todas as células (inclusive as vizinhas), sem a necessidade de subdivisão do espectro em canais menores e designação de um plano de frequências.

Para a nomenclatura utilizada nesta tese, é importante a definição formal de conceitos básicos e terminologias que serão utilizados durante a evolução da mesma. As seguintes definições foram obtidas de (3GPP, 2024f; 3GPP, 2024a; 3GPP, 2025; 3GPP, 2024b; 3GPP, 2024e; HOLMA; TOSKALA, 2009; SESIA; TOUFIK; BAKER, 2009; KHAN,

2009):

- *Physical Cell ID* (PCI): Identidade única atribuída a cada célula em uma rede LTE e 5G para diferenciar transmissões de células vizinhas. É um valor inteiro no intervalo $[0, 503]$ no LTE e $[0, 1007]$ no 5G.
- *Cell-Specific Reference Signal* (CRS): Em sistemas LTE, os CRSs consistem em sinais de referência transmitidos periodicamente pela RBS, mapeados em elementos físicos específicos no domínio tempo-frequência, cuja potência é configurada por célula. Os CRSs são utilizados pelas UEs para a estimação do estado do canal, medições de qualidade do enlace e procedimentos de mobilidade, tais como resseleção e *handover* entre células.
- *Reference Signal Received Power* (RSRP): no contexto do LTE, corresponde à potência média recebida nos *Resource Elements* (REs) que transportam os CRSs da célula servidora, sendo amplamente utilizado em medições de cobertura e mobilidade.

Além disso, no contexto deste trabalho, também se define o conceito de *dominância*. Embora essa métrica seja amplamente utilizada na prática para efeito de planejamento e *tuning* de redes RAN, não existe uma definição formal nas especificações 3GPP nem uma equação padronizada na literatura acadêmica. Neste trabalho, adota-se a seguinte definição: a dominância em um ponto i em uma área analisada é a diferença, em decibel (dB), entre a potência RSRP (em decibéis-miliwatt (dBm)) recebida da célula nesse ponto e a potência RSRP (também em dBm) recebida da célula interferente dominante, ou seja, a segunda célula com maior potência recebida nesse ponto:

$$D(i) = \text{RSRP}_{s,i} - \text{RSRP}_{d,i} \quad (1)$$

onde $D(i)$ é a dominância no ponto i , em dB, $\text{RSRP}_{s,i}$ é a potência RSRP recebida da célula servidora no ponto i , e $\text{RSRP}_{d,i}$ é a potência RSRP recebida da célula interferente dominante no ponto i (todas as potências em dBm). Essa formulação define, de forma pontual (i.e., em um ponto específico em uma região) e quantitativa, o princípio de que a célula servidora deve apresentar clara superioridade sobre o principal interferente. O conceito de dominância é muito utilizado na indústria, especialmente nos processos de planejamento e aprimoramento de redes móveis.

2.2.3 Obtenção de Dados de Radiofrequência

Um desafio para o processo de melhoria da qualidade de cobertura da RAN consiste na obtenção dos dados que permitam um modelamento do ambiente de RF adequado que sirva como base para essa atividade. Conforme mencionado no Capítulo 1, esses dados

podem ser adquiridos de diversas maneiras, das quais se destacam o método baseado em *drive tests*, via *crowdsourcing* e via MDT, sendo que os três serão brevemente descritos a seguir.

A primeira alternativa – e a mais tradicionalmente usada – para obtenção de informações sobre o ambiente de RF é a realização de *drive tests*, ou DTs. Essa é uma abordagem muito utilizada desde as primeiras gerações de sistemas celulares, sendo ainda empregada no presente, e o objetivo final é a avaliação e melhoria da RAN, consistindo em uma coleta de dados de rádio através de equipamentos especializados (normalmente UEs comuns com software modificado) e um sistema de georreferenciamento, normalmente *Global Positioning System* (GPS), com o qual se posicionam as medições de desempenho e outras informações na geografia onde foram coletadas. São coletadas especialmente métricas georreferenciadas de RSRP, *Reference Signal Received Quality* (RSRQ) e demais informações de camada *Radio Resource Control* (RRC).

Uma segunda solução para o levantamento de dados de RAN é uma funcionalidade padronizada pelo 3GPP, denominada MDT. Ela permite a coleta automatizada das informações que normalmente seriam obtidas pelos DTs, porém diretamente dos terminais dos usuários, que hoje já possuem capacidades de georreferenciamento embutidas, como *Global Navigation Satellite System* (GNSS) ou outros sistemas de localização. O MDT foi introduzido pelo 3GPP em sua *Release 10* e continua evoluindo até as versões mais recentes (3GPP, 2024c). Na prática, os sistemas de MDT são implementados de forma proprietária por cada fornecedor de tecnologia, como em (SAMSUNG, 2025) e sua utilização não tem sido ampla, embora haja registro de casos isolados de uso em operadoras (NOKIA, 2018). Parte da baixa adoção deve-se à complexidade e custo desses sistemas proprietários. Logo, embora esse recurso seja uma excelente alternativa ao processo de DT, a adoção da funcionalidade pela indústria, em termos práticos, ainda não é ampla ou consolidada.

Por fim, é importante citar a alternativa de *crowdsourcing*. A aquisição dos dados georreferenciados de RF por meio desse método é tecnicamente bem estabelecida e de maior eficiência quando comparada com o caro e trabalhoso DT. No contexto de redes móveis, o *crowdsourcing* é a técnica na qual se obtêm as informações do ambiente RAN através dos próprios dispositivos UE na rede que, por meio de aplicações específicas ou *applets* embarcados em apps maiores, coletam e enviam esses dados de forma anonimizada e normalmente sem a participação ativa do usuário e transparente aos subsistemas da rede móvel. Os dados via *crowdsourcing* podem ser obtidos tanto pela própria operadora (sendo restritos aos dados dos próprios clientes), quanto podem ser contratados por fornecedores desse serviço. Um maior detalhamento sobre o método é abordado na Seção 3.3 e uma implementação prática será demonstrada na Seção 4.1.

2.2.4 Processos Gaussianos

Um processo gaussiano é definido por um conjunto de variáveis aleatórias que segue um padrão de distribuição gaussiana multivariada e é especificado por uma função de média $m(x)$ e uma função de covariância $k(x, x')$, também conhecida como núcleo, ou *kernel*, o qual define a relação estatística entre essas variáveis (SCHÖLKOPF, 2002). Técnicas baseadas em processos gaussianos são, por vezes, empregadas para a expansão de dados, a fim de delinear condições de contorno e proporcionar maior coerência a modelos baseados em aprendizado de máquina (TANG et al., 2023).

Também é possível definir processo gaussiano como uma distribuição sobre funções $f(x)$ onde qualquer conjunto finito de pontos segue uma distribuição normal multivariada (RASMUSSEN; WILLIAMS, 2008):

$$f(x) \sim GP(m(x), k(x, x')) \quad (2)$$

Onde $GP(\cdot)$ denota um processo gaussiano, isto é, uma distribuição de probabilidade sobre funções, $m(x)$ é a função de média, geralmente assumida como zero e, $k(x, x')$ é a função de covariância (ou *kernel*), que define a relação entre os pontos no espaço de entrada.

O objetivo geral da utilização dos processos gaussianos é estimar valores em um conjunto de dados não observados a partir de amostras disponíveis. Esses padrões são definidos a partir do *kernel* utilizado, de forma que, a cada caso de uso, é essencial a identificação da modelagem mais adequada.

Para casos de uso como o apresentado nesse trabalho, o *kernel Radial Basis Function* (RBF) é bastante adequado, uma vez que uma das suas propriedades é a invariância sob a ótica de operações de rotação ou espelhamentos. Isso significa que os resultados do processo de aprendizagem não dependem de qual sistema de coordenadas ortogonais é utilizado para representar os dados de entrada (SCHÖLKOPF, 2002). O *kernel* RBF é dado pela expressão:

$$k(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \quad (3)$$

onde σ^2 é a variância do processo, l é o parâmetro de escala (ou comprimento de correlação) que define o quão rápido a função muda, e $\|x - x'\|^2$ é a distância euclidiana entre os pontos de entrada.

Os processos gaussianos apresentam um método bem fundamentado em termos probabilísticos para a estimativa de valores e aumento de amostragem, o que será importante como técnica para evitar o efeito de *overfitting* em Rede Neural Artificiais (RNAs) (YING, 2019), como será mostrado na Seção 2.2.5 .

2.2.5 Redes Neurais Artificiais

A RNA, ou simplesmente rede neural, talvez seja uma das ferramentas habilitadoras mais amplamente utilizadas hoje em dia no campo da inteligência artificial. (HAYKIN, 2009) define uma rede neural como um processador distribuído, massivamente paralelo, composto de unidades de processamento simples e que possui uma propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso.

As características de uma RNA as tornam especialmente boas para resolver problemas de regressão, como os que se apresentam nesta tese, quando há alguma informação sobre um sistema e deseja-se inferir valores fora do universo da amostragem (ZHANG; PATUWO; HU, 1998).

A RNA deve ser modelada de acordo com o problema a ser resolvido. A modelagem inclui a definição de diversos parâmetros, dentre eles (HAYKIN, 2009): quantidade de nós de entrada e de saída, número de camadas ocultas e suas quantidades de nós, função de ativação por camada, dentre outros.

Um desafio comum no treinamento das RNAs é o chamado sobreajuste (mais conhecido pela sua designação em inglês *overfitting*) que ocorre quando a rede aprende excessivamente os detalhes e ruídos do conjunto de treinamento, em vez de generalizar adequadamente. Isso faz com que o desempenho da rede seja alto nos dados de treinamento, mas ruim em dados fora desse conjunto, comprometendo sua capacidade de generalização. Há várias razões para o *overfitting*, incluindo uma excessiva complexidade da RNA, dados esparsos ou presença de *outliers*. Para evitar esse efeito, podem ser usadas várias técnicas complementares, filtragem de *outliers* ou aumento de dados (*data augmentation*), técnicas que serão empregadas nesta tese. (YING, 2019). No caso específico em que se aplica o aumento de dados, os processos gaussianos podem se apresentar como um bom método para estimativa de dados baseados em regressão, desde que se conheça bem o problema e a abordagem adequada para seu modelamento estatístico (TANG et al., 2023).

2.2.6 CMA-ES

A técnica *Covariance Matrix Adaptation Evolution Strategy* (CMA-ES), proposta inicialmente por (HANSEN; OSTERMEIER, 2001), é um algoritmo estocástico de busca que aborda o seguinte problema de otimização: minimizar uma função objetivo não-linear que realiza um mapeamento do espaço de busca $S \subset \mathbb{R}^n$ para \mathbb{R} , onde \mathbb{R}^n denota o espaço vetorial real de dimensão n , correspondente ao espaço de busca contínuo do problema. Um tutorial mais prático foi apresentado posteriormente em (HANSEN, 2023). Trata-se de uma classe de algoritmos evolutivos desenvolvida com o objetivo de otimizar funções de custo do tipo “caixa-preta”, nas quais não se dispõe de informações analíticas, como gradientes, sendo o acesso restrito apenas à avaliação da função objetivo.

O CMA-ES destaca-se por sua robustez, eficiência e capacidade de adaptação auto-

mática da geometria do espaço de busca (HANSEN; OSTERMEIER, 2001). Isso o torna muito adequado a problemas complexos de otimização, como o caso de calibração de parâmetros para melhorias de desempenho do ambiente de RAN em redes móveis celulares (SEMENCHIKOV et al., 2021), nos quais o comportamento do sistema como um todo é influenciado pela alteração em cada elemento.

2.3 Trabalhos Relacionados

Nesta seção, são apresentados os trabalhos relacionados ao tema desta tese. Como o problema é abordado em duas etapas, são apresentados os trabalhos relacionados a cada uma delas de forma independente, sendo a primeira subseção relativa aos trabalhos correlatos à modelagem de ambientes de RF de redes móveis e interpolação e generalização de dados, e a segunda citando as referências a trabalhos de relacionados ao processo de *tuning* de RAN.

2.3.1 Trabalhos Relacionados à Modelagem de RAN

Alguns esforços visam propor modelagens de *digital twin* para redes móveis e ambientes de RF, como os apresentados em (DENG et al., 2021) e (NGUYEN et al., 2021). No entanto, essas abordagens não são exclusivamente focadas na RAN e não têm como objetivo modelar a cobertura celular.

Em (TANG et al., 2023), propõe-se o aumento de dados (*data augmentation*) como estratégia para contrapor o *overfitting* por meio de processos gaussianos para aprimorar a localização de dispositivos em ambientes Wi-Fi internos. Um trabalho mais próximo ao aqui apresentado, que trata da melhoria da cobertura e capacidade de redes móveis utilizando dados de MDT e aprendizado profundo, é apresentado em (SKOCAJ et al., 2022). Entretanto, essa abordagem depende tanto de dados de projeto quanto de modelos matemáticos analíticos para estimar cobertura e capacidade da rede; é baseada em episódios e não tem como objetivo gerar um mapa de cobertura a partir de dados esparsos (embora essa limitação seja mencionada no estudo). Já (CHEERLA; RATNAM; BORRA, 2018) faz uma proposta de modelo de predição de perda de percurso (*path loss*) baseado em RNAs, treinado a partir de medições reais de sinais nas bandas de 800 e 1.800 MHz em uma pequena área geográfica urbana, com o objetivo de corrigir limitações de modelos de propagação como o COST-231 Walfisch-Ikegami, não objetivando modelar o ambiente de RF de forma puramente orientada a dados, mas sim propor correções e aprimoramentos a modelos teóricos de predição de perda de percurso, a partir das medições realizadas.

Um outro trabalho relevante é apresentado em (DREIFUERST et al., 2021), que investiga a melhoria de cobertura e capacidade em conjunto em redes LTE por meio de técnicas orientadas a dados, usando *Deep Deterministic Policy Gradient* (DDPG), otimização bayesiana multiobjetivo e *random search*. São usados dados simulados baseados

em modelos de propagação, e não coletados em campo, e o artigo não constrói um modelamento do ambiente de RF ou estima RSRP diretamente a partir de medições, mas foca no aprimoramento de parâmetros de rede como potência de transmissão das células e *downtilt* da antena, discutindo resultados obtidos pelas três técnicas. Finalmente, (OJO; IMOIZE; ALIENYI, 2021) propõe modelos de predição de perda de percurso baseados em RNAs treinadas com medições reais de campo em redes LTE com uma RBS tri-setorizada, apresentando um desempenho superior com relação a modelos clássicos como COST-231 e outros. O objetivo do trabalho é modelar a perda de percurso com foco em melhorar a acurácia de modelos empíricos para fins de planejamento de propagação, e não objetiva a extrapolação espacial de uma célula.

No âmbito dos estudos mais recentes, o trabalho apresentado em (REKKAS et al., 2025) propõe o uso de técnicas de aprendizado de máquina baseadas em processos gaussianos – como o *GPBoost* – sobre dados reais de medições para construir um mapa de ambiente de rádio, que inclui parâmetros como RSRP e RSRQ em redes LTE e 5G. O método é orientado a dados e a acurácia do método é relativamente boa, da ordem de 8 a 9 dB de *Root Mean Squared Error* (RMSE) em LTE. No entanto, o treinamento do modelo depende de dados do projeto e informações geoespaciais, como o cenário topográfico e atributos de posicionamento das estações e de vizinhanças entre as mesmas. Em (QIU; BOSE, 2025), é proposto um conjunto de modelos de aprendizado de máquina para predição de métricas de rádio em redes LTE, utilizando dados obtidos via *crowdsourcing*. Apresenta uma abordagem híbrida com um modelo que usa predição matemática associada a uma RNA, obtendo-se uma acurácia da ordem de 10 dB de RMSE para RSRP, avaliando mais de 300.000 amostras em ambientes urbanos. Todavia, é utilizado um modelo de predição de propagação baseado em *Free-Space Path Loss* (FSPL), com correção via RNA. Além disso, o método depende da entrada de um conjunto de informações denominadas *Engineered Features*, que inclui a distância entre a estação e a UE, potência de transmissão da estação e dados de morfologia 3D, dentre outras.

O estudo (TSOULOS; TSOULOS; ATHANASIADOU, 2025) usa especificamente dados de DT para fazer predições de RSRP, RSRQ e *Signal-to-Interference-plus-Noise Ratio* (SINR) de sistemas celulares LTE e 5G, avaliando quatro modelos de aprendizado de máquina com relação ao Erro Absoluto Médio, ou *Mean Absolute Error* (MAE), desses KPIs. Foram realizadas medições reais com sete dispositivos em cinco posições diferentes dentro do veículo usado no DT, onde foram coletadas amostras em um percurso linear de 100 km na Grécia. O método aplica-se somente ao caso específico de medições *in-car*, sendo que o objetivo do trabalho é a predição de KPIs em posições não medidas dentro de veículos, ou estimativas de indicadores quando há dados faltantes. Já o trabalho apresentado em (ZHANG et al., 2025) demonstra um *framework* para predição precisa de potência recebida de uma estação celular (atingindo um RMSE de 2,54 dB para RSRP) para uma determinada área, usando uma rede neural convolucional do tipo

Residual Network com 18 camadas (ResNet-18). Porém o método depende tanto de informações do projeto quanto de dados geográficos da área, incluindo base de dados de altura de edificações, morfologia e elevação de terreno.

Um trabalho de especial aderência é descrito em (JIANG et al., 2025), que propõe o uso de um modelo chamado *Tabular Prior-data Fitted Network* (TabPFN) como solução para a predição de cobertura (RSRP) em redes móveis. O modelo foi treinado com um conjunto de dados de DT em rede LTE e apresenta bons resultados para RMSE (2,46 dB) e R^2 (0,94). O método é orientado a dados, não dependendo de nenhum modelamento matemático para a predição e descobre automaticamente estratégias para lidar com valores faltantes ou *outliers*. Entretanto, o TabPFN recebe explicitamente como entrada sete parâmetros relativos à rede e ao ambiente, como a distância entre a RBS e o UE, o ângulo de azimute e de *downtilt* relativos a esses dois nós, a morfologia do terreno e o tipo de propagação do sinal (linha de visada ou em obstrução); logo, é altamente dependente de informações externas de projeto de rede.

Outros trabalhos de analisados incluem (HUANG et al., 2026), que realiza um modelamento *data-driven* da RAN em um *digital twin*, permitindo assim predições de cobertura de uma área, mas também sendo altamente dependente de dados do projeto, e (JESKE et al., 2025), que propõe um modelo de aprendizado de máquina baseado em *random forest* para predição precisa de RSRP em cenários urbanos e suburbanos, com um RMSE de aproximadamente 3,5 dB. Para tal, o modelo usa 64 *features*, incluindo parâmetros de antenas, estatísticas de terreno, modelos de propagação, dentre outros.

Por fim, contrasta-se este trabalho na Tabela 2.1, destacando características essenciais e suas diferenças. Os itens com círculos preenchidos representam critérios aderentes ao trabalho; círculos sem preenchimento mostram critérios que não são aderentes ou abor-dados, e os pontos pequenos indicam que o critério não se aplica ao respectivo trabalho.

2.3.2 Trabalhos Relacionados ao Aprimoramento de Cobertura Celular

Há trabalhos específicos que abordam a melhoria da cobertura de uma RAN, muitos deles de forma automatizada ou automatizável, além da detecção de anomalias e auto-cura (*self-healing*) de redes. Porém, no contexto específico a ser descrito neste trabalho, não há um desenvolvimento que aborde completamente o problema de forma direta e dentro do escopo aqui proposto. O que se percebe, em especial, é que os trabalhos, em sua maioria, utilizam dados de rede externos à informação disponível na própria operação de um sistema móvel celular padrão 3GPP.

O *survey* (KLAINÉ et al., 2017) aborda técnicas de aprendizado de máquina aplicadas à auto-organização de redes móveis, com alguns objetivos específicos relacionados ao crescimento de cobertura e capacidade, melhoria de qualidade de serviço e de experiência

Tabela 2.1 – Resumo comparativo dos trabalhos relacionados em modelamento de RAN. Círculos preenchidos representam critérios aderentes, os não preenchidos, critérios não aderentes e os pontos, critérios não aplicáveis.

Referência	Rede baseada em RNA	Independência de dados de projeto	Independência de predição analítica	Suporta anomalias de RF	Independência de KPIs de sistema	Aplica-se a sistemas 3GPP (RAN)
Juan Deng <i>et al.</i> 2021 (DENG <i>et al.</i> , 2021)	○	○	○	○	○	●
Skocaj <i>et al.</i> 2022 (SKOCAJ <i>et al.</i> , 2022)	●	○	○	●	○	●
Tang, 2023 (TANG <i>et al.</i> , 2023)	●	○	●	●	●	○
Cheerla, 2018 (CHEERLA; RATNAM; BORRA, 2018)	●	○	○	●	●	●
Dreifuerst, 2021 (DREIFUERST <i>et al.</i> , 2021)	●	○	●	●	●	●
Ojo, 2020 (OJO; IMOIZE; ALIENYI, 2021)	●	○	●	●	●	●
Rekkas, 2025 (REKKAS <i>et al.</i> , 2025)	○	○	●	●	●	●
Qiu, 2025 (QIU; BOSE, 2025)	●	○	○	●	●	●
Tsoulos, 2025 (TSOULOS; TSOULOS; ATHANASIADOU, 2025)	○	●	●	●	●	●
Zhang, 2025 (ZHANG <i>et al.</i> , 2025)	●	○	●	○	●	●
Jiang, 2025 (JIANG <i>et al.</i> , 2025)	●	○	●	●	●	●
Huang, 2025 (HUANG <i>et al.</i> , 2026)	●	○	●	●	○	●
Jeske, 2025 (JESKE <i>et al.</i> , 2025)	○	○	○	○	●	●
Este trabalho	●	●	●	●	●	●

e prover maior flexibilidade e inteligência na rede. Embora aponte as RNAs e algoritmos heurísticos como soluções para os problemas acima, as aplicações citadas são relacionadas ao gerenciamento de mobilidade e posicionamento, alocação e balanceamento de recursos, melhoria de eficiência de *handover* e falhas de serviço. Uma abordagem à melhoria de cobertura se refere ao caso específico de *femtocells*.

Um outro *survey* apresentado em (MOYSEN; GIUPPONI, 2018) apresenta uma visão abrangente e sistemática do então estado arte em SONs e gerenciamento de redes móveis, com ênfase no papel do aprendizado de máquina como ferramenta habilitadora de automação fim a fim. Ele organiza e classifica conceitos, arquiteturas, casos de uso e técnicas de aprendizado aplicáveis ao gerenciamento de redes, abordando também aspectos de padronização. O *survey* menciona o MDT como um dos habilitadores do caso de uso *Coverage and Capacity Optimization* (CCO), citando também esforços de *crowdsourcing* como fontes de dados. Cita especificamente quatro trabalhos no problema específico do CCO: (RAZAVI; KLEIN; CLAUSSEN, 2010; ISLAM; MITSCHELE-THIEL, 2012; LI *et*

al., 2012; FAN; TIAN; SENGUL, 2014), todos baseados em técnicas de *fuzzy Q-learning*, nenhum deles com abordagem similar a esta tese.

Uma abordagem particularmente relevante de diagnóstico e tratativa de problemas de cobertura e interferência de células baseada em ajuste de potência é descrita no trabalho de (FAGEN; VICHARELLI; WEITZEN, 2008). Nele, são introduzidos conceitos básicos de indicadores de cobertura, sobreposição e a relação entre essas grandezas, os quais são parcialmente adotados neste trabalho. Para caracterizar e controlar a sobreposição de cobertura entre RBSs, é utilizada uma função objetivo com a métrica denominada *coeficiente de cobertura* para a i -ésima estação, definida como:

$$c_i = \frac{\beta_i}{\alpha_i} \quad (4)$$

onde β_i representa a área de cobertura da estação no cenário com a presença de vizinhas (melhor servidor), e α_i corresponde à área de cobertura em isolamento, sendo $\alpha_i \geq \beta_i$. O coeficiente de cobertura varia entre $0 \leq c_i \leq 1$, onde o valor $c_i = 1$ indicaria uma ausência de sobreposição de cobertura, ou seja, a área atendida pela estação em rede coincide com sua área de cobertura isolada. Então formula um vetor de potências de transmissão das células p , que é ajustado iterativamente de modo a maximizar a cobertura e, simultaneamente, minimizar a interferência intercelular, com base em técnicas clássicas de otimização (Newton–Raphson), com ajustes lineares (aumento ou redução) da potência das células de modo a se maximizar c_i . Um caso real utilizando-se 12 RBSs é apresentado, sendo que as coberturas são obtidas via predição computacional. O resultado da aplicação do algoritmo mostrou uma redução de aproximadamente 47% do coeficiente de cobertura, com uma redução discreta da cobertura total de apenas 3%. O trabalho usa predições computacionais de cobertura e não dados esparsos coletados, e são informados os dados de projeto, sobretudo as coordenadas das estações e potências dos transmissores, essas últimas ajustadas de acordo com interações no processo apresentado no algoritmo. Além disso, presume uma relação relativamente homogênea e estável de morfologia e projeto. No entanto, o trabalho é especialmente aderente a esta tese na proposta de um vetor de correção de potência das células e métricas de cobertura e sobreposição para se avaliar a cobertura das células e interação entre as mesmas.

O trabalho apresentado em (NWELIH; ISABONA; IMOIZE, 2022) propõe um método de *auto-tuning* baseado em dados reais de rede e algoritmos genéticos; entretanto, o ajuste fino dos parâmetros é realizado utilizando o modelo matemático de propagação COST-231 Hata. Em (SKOCAJ et al., 2022), é empregada uma abordagem baseada em dados de MDT e *Deep Reinforcement Learning* (DRL) para ajustes do *downtilt* de antenas em uma rede real, fazendo uso de dados de planejamento, KPIs e simulações eletromagnéticas para o treinamento do agente.

Outros trabalhos são focados na detecção de anomalias em redes celulares. (WANG; HANDURUKANDE, 2018) propõe um método para detecção de anomalias de cobertura,

porém é focado em análises temporais, ou seja, há a necessidade de coleta de indicadores em um fluxo de dados contínuo e as anomalias são avaliadas na variação desses indicadores no tempo sendo atribuídos a elas uma pontuação e probabilidade em intervalos discretos nos quais os KPI são analisados. Da mesma maneira, em um trabalho anterior, (CIOCARLIE et al., 2013) apresentava um método de detecção de anomalias através de um método composto de análises de indicadores de desempenho, avaliando-se a probabilidade de se ter uma anomalia de rede em cada algoritmo, combinando os resultados e mapeando-se essas probabilidades finais para valores binários, representando 0 (normal) ou 1 (anomalia). Porém, o conceito de análise temporal também se aplica a esse trabalho.

A dissertação apresentada por (CORRÊA, 2020) também foca na análise e detecção de anomalias, empregando as técnicas *Principal Component Analysis* (PCA) e *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) aplicados a KPIs como SINR e *Received Signal Strength Indicator* (RSSI), além de número médio de usuários ativos e taxa de dados média no *downlink*, tendo como classificação final das células os rótulos “Normal”, “*Undershooting*” e “*Overshooting*”, dentre outros. No entanto, o trabalho também é baseado em análises temporais e não utiliza dados de RF georreferenciados como informação de entrada e, embora seja fornecido um diagnóstico da cobertura, o método não propõe uma solução para a ajustes de rede.

Além disso, alguns trabalhos não propõem técnicas de melhoria ou detecção de anomalias de cobertura, porém demonstram a aplicabilidade das técnicas usadas nesta tese. (AKBARI et al., 2016) demonstra que, apesar de imprecisões relacionadas ao posicionamento, os dados de MDT constituem uma fonte altamente confiável de informações da RAN para a estimativa de cobertura. O estudo apresentado em (SEMENCHIKOV et al., 2021) avalia múltiplos algoritmos metaheurísticos para a aprimoramento da RAN utilizando modelos de propagação e demonstra que o CMA-ES e suas variantes apresentam desempenho superior em vários aspectos quando comparados às técnicas *Particle Swarm Optimization* (PSO), *Differential Evolution* (DE) e a outras avaliadas, sendo uma referência especialmente importante na escolha da abordagem na etapa de *auto-tuning*.

Os trabalhos mais recentes da área dão muito foco às visões para evolução ao 6G, usando *agentic AI*, *Network Digital Twins* (NDTs) e *reinforcement learning*. O trabalho de (LIU et al., 2025) propõe um *framework* completo para RAN 6G, através de Inteligência Artificial (IA) nativa e o uso de NDT, com o objetivo de viabilizar quatro propriedades de self-* da rede (*self-configuration*, *self-optimization*, *self-healing* e *self-evolution*). Embora seja um *framework* totalmente orientado a dados e não utilizar modelamento matemático para predições, ele não foca exclusivamente na melhoria da cobertura e na minimização da interferência, e necessita coletar diversas informações de KPIs de rede, com o objetivo de construção do modelo do NDT. O modelo também não é aplicável a redes LTE ou 5G atuais, sendo uma visão específica para o 6G.

Já em (HU et al., 2025), é proposto o primeiro *framework* de *auto-tuning* para RAN

6G baseado em *agentic AI*, integrando quatro agentes especializados em um *loop* fechado que lhes permite iterar de forma autônoma, possibilitando que o sistema se auto-corrija. Experimentos em cenários de *clusters*, avaliando a interferência inter-células, mostraram ganhos de 17,1% de *throughput*, dentre outros. Porém, o modelo não é aplicável aos atuais sistemas LTE e 5G, sendo uma proposta para a evolução 6G. O agente de simulação usa modelos analíticos de propagação para gerar KPIs da rede, além de não ter sido aplicado a uma rede real (foram usadas plataformas de simulação, baseadas no método de Monte Carlo).

Finalmente, (ABOUELMAATI et al., 2025) apresenta um *survey* abrangente que analisa a sinergia entre *Open RAN* (O-RAN) e SON para habilitar autonomia via IA em redes 5G e 6G. O artigo destaca técnicas de aprendizado de máquina como as principais ferramentas para *auto-tuning*, centralizando-as no *RAN Intelligent Controller* (RIC) (um conceito inerente da arquitetura O-RAN), que é o componente central de inteligência e automação da RAN. O artigo cita vários casos e testes comerciais, todos dependentes da parametrização da rede, dados de topologia e morfologia ou KPIs em tempo real e, sendo um *survey*, não apresenta uma proposta específica para melhoria de RAN, mas um compilado das diversas técnicas abordadas no trabalho.

A Tabela 2.2 apresenta um comparativo dos trabalhos citados, comparando-os com a tese aqui apresentada. Novamente, os itens com círculos preenchidos representam critérios aderentes, enquanto os círculos sem preenchimento representam critérios não aderentes ou não abordados.

Em resumo, percebe-se que os trabalhos de modelagem com melhor acurácia dependem fortemente de informações externas – dados de projeto, morfologia do terreno ou *features* de engenharia – enquanto os métodos puramente empíricos não buscam generalizar a cobertura de uma célula a partir de dados esparsos

Do lado do aprimoramento de cobertura, as soluções existentes recorrem a modelos analíticos de propagação, limitam-se a detectar anomalias sem propor correções, ou são propostas voltadas apenas ao 6G. Não foi encontrado um trabalho que trate o problema de ponta a ponta: modelar o ambiente de RF usando somente as medições coletadas pela própria rede e, a partir desse modelo, propor ajustes de seus parâmetros.

A metodologia ARCADE se propõe a preencher essa lacuna, com a aderência a todos os critérios, conforme mostrado nas Tabelas 2.1 e 2.2.

Tabela 2.2 – Resumo comparativo dos trabalhos relacionados em auto-organização e melhoria de cobertura de RAN. Círculos preenchidos representam critérios aderentes e não preenchidos, critérios não aderentes.

Referência	Técnica utilizada	Independência de dados de projeto	Independência de modelos analíticos	Caso de uso em rede real	Aprimoramento de RAN
Razavi <i>et al.</i> (RAZAVI; KLEIN; CLAUSSEN, 2010)	<i>Fuzzy Q-Learning</i>	○	●	○	●
Islam <i>et al.</i> (ISLAM; MITSCHLE-THIEL, 2012)	<i>Fuzzy Q-Learning</i>	○	●	○	●
Fan <i>et al.</i> (FAN; TIAN; SENGUL, 2014)	<i>Fuzzy Q-Learning</i>	○	●	○	●
Fagen <i>et al.</i> (FAGEN; VICHARELLI; WEITZEN, 2008)	Método analítico (Newton–Raphson)	○	○	●	●
Nwelih <i>et al.</i> (NWELIH; ISABONA; IMOIZE, 2022)	GA	○	○	●	●
Skocaj <i>et al.</i> (SKOCAJ <i>et al.</i> , 2022)	DRL	○	○	●	●
Wang <i>et al.</i> (WANG; HANDURUKANDE, 2018)	Heurísticas e análise temporal	●	●	●	○ (identifica anomalias)
Ciocarlie <i>et al.</i> (CIOCARLIE <i>et al.</i> , 2013)	Heurísticas e análise temporal	●	●	●	○ (identifica anomalias)
Correa <i>et al.</i> (CORRÊA, 2020)	PCA + DBSCAN	●	●	●	○ (identifica anomalias)
Liu <i>et al.</i> (LIU <i>et al.</i> , 2025)	NDT	●	●	○	● (apenas 6G)
Hu <i>et al.</i> (HU <i>et al.</i> , 2025)	<i>Agentic IA</i>	●	○	○	● (apenas 6G)
Abouelmaati <i>et al.</i> (ABOUELMAATI <i>et al.</i> , 2025)	ML/AI	○	●	●	● (survey)
Este trabalho	CMA-ES	●	●	●	●

ARCADE: *Adaptative Radio Coverage Analysis through Data-driven Evolution*

Este capítulo apresenta a proposta desenvolvida para a avaliação da cobertura de um sistema celular através do processo denominado *Adaptative Radio Coverage Analysis through Data-driven Evolution* (ARCADE). O capítulo é estruturado da seguinte forma: a Seção 3.1 apresenta uma visão condensada da metodologia ARCADE, organizada didaticamente em módulos, e apresenta um método para a definição formal de um *cluster*. Em seguida, a Seção 3.3 propõe uma visão de *crowdsourcing* como uma alternativa à coleta de dados via DT, sendo que essa implementação será demonstrada na prática no Capítulo 4. Na sequência, a Seção 3.4 detalha a metodologia para organização, extrapolação e estruturação dos dados utilizados na análise. Logo a seguir, a Seção 3.7 apresenta as técnicas empregadas para a realização do aprimoramento da cobertura de um determinado *cluster* de células a partir dos dados extrapolados e estruturados.

3.1 Metodologia ARCADE

A Figura 3.1 apresenta, de forma resumida, um diagrama com a metodologia ARCADE completa, desde a aquisição de amostras até a obtenção de um *cluster* de células com suas potências ajustadas de forma a melhorar o desempenho de cobertura. Para fins didáticos, as atividades e processos foram agrupados em módulos componentes da metodologia, que são executados sequencialmente e serão apresentados com mais detalhes a seguir.

3.1.1 *Data Acquisition Module (DAM)*

O DAM é responsável pela aquisição de dados brutos necessários para o processo de modelagem. Este módulo recebe ou coleta informações provenientes de diferentes fontes, como MDT, *crowdsourcing* e/ou DTs, assegurando que os dados sejam adequados para as etapas subsequentes do sistema. Além disso, o DAM consolida as amostras, que podem

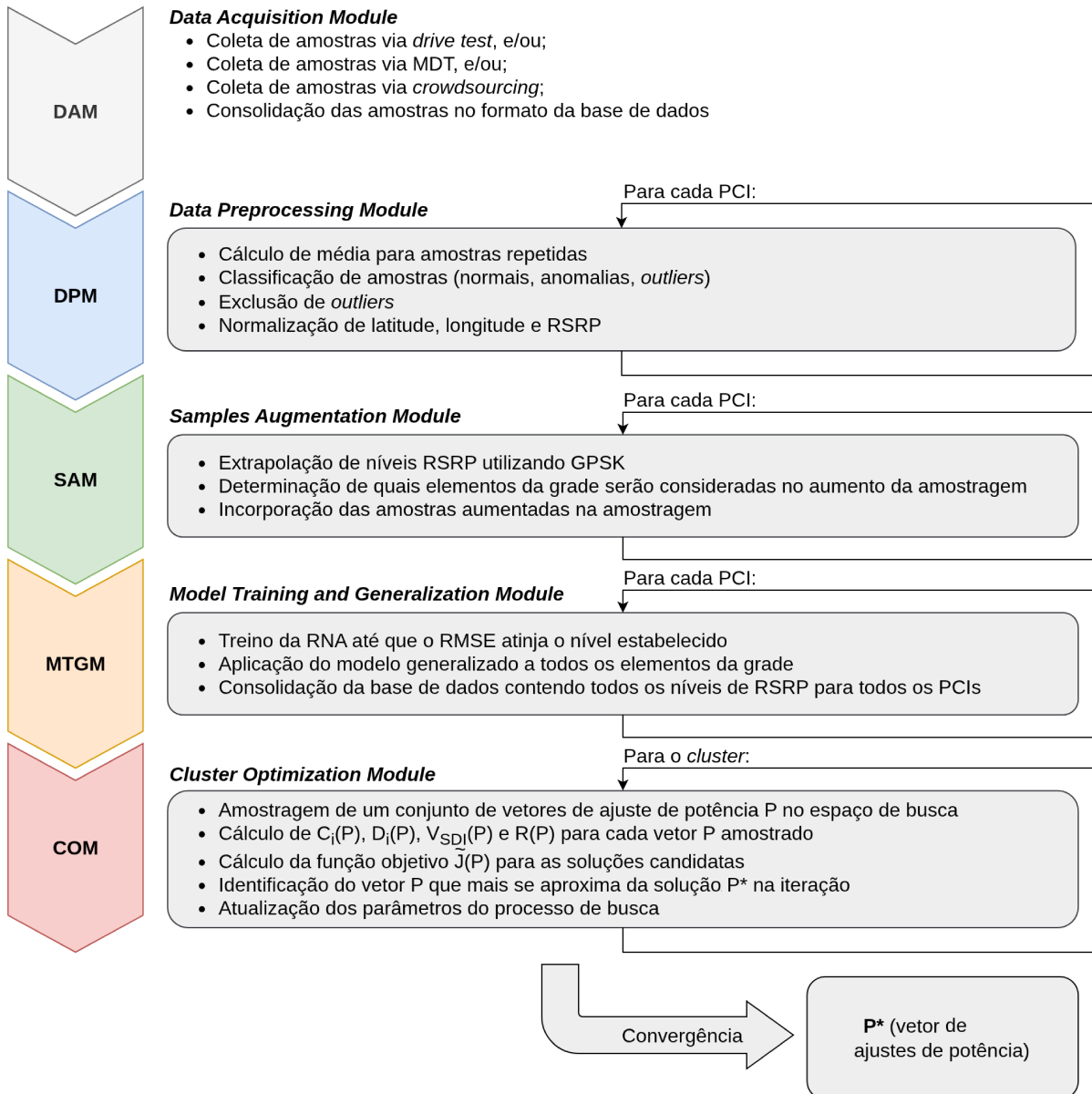


Figura 3.1 – Esquemático da metodologia ARCADE para aprimoramento da RAN. Figura elaborada pelo autor.

ser provenientes de mais de uma fonte, em uma base de dados única. Uma visão sobre a implementação do DAM na forma de *crowdsourcing* será apresentada na Seção 3.3.

3.1.2 Data Preprocessing Module (DPM)

O DPM realiza o pré-processamento dos dados adquiridos, preparando-os para a análise. As principais atividades incluem limpeza de dados, consolidando amostras repetidas em uma única (com valor de RSRP médio das mesmas), exclusão de registros com campos ausentes ou com erro, classificação das amostras (normais, anomalias e *outliers*), exclusão das amostras *outliers* e normalização da latitude, longitude e RSRP. Essa normalização é feita por último, preservando-se as informações originais, para flexibilizar os cálculos

relativos à classificação das amostras – por exemplo, que o cálculo da distância da amostra ao centróide correspondente à cobertura da célula possa ser realizado em unidades de metros. Esse módulo DPM será apresentado com mais detalhes na Seção 3.4.

3.1.3 *Samples Augmentation Module (SAM)*

O SAM tem como função a ampliação das amostras de dados disponíveis, aplicando as técnicas de aumento de dados a serem apresentadas na Seção 3.5, modelando o espaço analisado através da técnica *Gaussian Process with Spatial Kernel* (GPSK). Além disso, o módulo é responsável por determinar a área de anomalias A_A e incorporar as amostras enriquecidas à mesma. O módulo SAM irá ser descrito com mais detalhes na Seção 3.5.

3.1.4 *Model Training and Generalization Module (MTGM)*

O MTGM é dedicado ao treinamento dos modelos e à determinação de sua generalização. Nesta etapa, uma RNA é treinada, visando encontrar padrões nos dados de cobertura de cada PCI, incluindo as amostras coletadas e as aumentadas obtidas através do SAM. Utiliza técnicas de validação cruzada e ajuste de hiperparâmetros para maximizar o desempenho preditivo e minimizar a *overfitting*. Ao final, o módulo produz, para cada célula do *cluster*, um mapa de RSRP na sua área total A_T , que alimenta o processo subsequente de aprimoramento de cobertura da RAN. O módulo MTGM é apresentado na Seção 3.6.

3.1.5 *Cluster Optimization Module (COM)*

O COM é responsável pelo processo de melhoria da qualidade da cobertura do *cluster* de células, a partir do resultado obtido na etapa anterior. É neste módulo que são avaliadas a cobertura, dominância e equilíbrio entre as células do *cluster* e propostos os ajustes de potência de modo a se ter uma configuração melhorada do sistema, conforme será descrito na Seção 3.7. Nesse módulo, os parâmetros de controle das células – em particular os ajustes de potência – são tratados de forma coordenada, levando em conta as interdependências existentes entre células adjacentes e os efeitos de interferência mútua. O objetivo do COM é a busca por uma configuração melhorada do *cluster* como um todo, que resulte em um equilíbrio adequado entre cobertura, dominância servidor/interferente e homogeneidade espacial das células. Ele constitui a etapa final do escopo do processo, sendo o responsável por transformar todos os dados obtidos e calculados em propostas concretas de aprimoramento da RAN, orientadas pela função objetivo $\tilde{J}(P)$ e resolvidas por meio de técnicas de otimização estocástica. O detalhamento do módulo COM será apresentado na Seção 3.7.

3.2 Definição do *Cluster* de Análise

É importante ressaltar que, antes que a metodologia possa ser colocada em prática, há a necessidade de delimitação do escopo da análise em termos de *cluster* de células, dentro do universo de elementos disponíveis. Uma operadora pode ter centenas, milhares ou eventualmente milhões de células em sua rede, de forma que a escolha do escopo da análise deve seguir certos critérios. Essa seleção pode ser feita intuitivamente – como, por exemplo, incluir todas as células no perímetro urbano de uma pequena cidade –, porém, quando o espaço amostral potencial de elementos é muito grande, é recomendável definir-se critérios bem fundamentados para essa seleção. Essa delimitação do escopo é uma etapa importante e deve ser realizada de forma racional: um *cluster* deve ter uma quantidade razoável de células que interagem entre si, ou seja, apresente interação ou superposição de cobertura e, idealmente, possa ser representado em um único grafo conexo, cujos vértices são as células e as arestas, as relações bilaterais de vizinhanças. Ou seja, se o *cluster* escolhido for representado por mais de um grafo conexo, recomenda-se que o escopo seja quebrado em análises distintas.

Além disso, é importante ressaltar que a modelagem proposta neste trabalho considera um cluster de células pertencentes a uma única camada tecnológica homogênea, não contemplando cenários de redes heterogêneas operando de forma sobreposta. Em arquiteturas heterogêneas, células micro, pico e femtocélulas são tipicamente implantadas de maneira hierárquica e intencionalmente sobrepostas à cobertura de macrocélulas, de forma que não se aplicam as definições de dominância de células da forma em que foram formuladas neste trabalho. Na prática, para o ARCADE, os PCIs de células em camadas mais baixas devem ser desconsideradas da análise.

De forma racional, propõe-se as seguintes definições abaixo. Seja:

- \mathcal{C} o universo de células disponíveis na rede de uma operadora que potencialmente poderia fazer parte da análise, com $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$ e $n = |\mathcal{C}|$;
- $\mathcal{N} \subseteq \mathcal{C} \times \mathcal{C}$ o conjunto de pares de células com relações bilaterais de vizinhança, sendo \mathcal{N} simétrico, ou seja, $(c_i, c_j) \in \mathcal{N} \Rightarrow (c_j, c_i) \in \mathcal{N}$;
- $G = (\mathcal{C}, \mathcal{N})$ o grafo de vizinhança global da rede.

Um *cluster* de análise pode ser definido como um subgrafo conexo $G' = (\mathcal{C}', \mathcal{N}') \subseteq G$, tal que:

$$\begin{aligned} \mathcal{C}' &\subseteq \mathcal{C} \\ \mathcal{N}' &= \mathcal{N} \cap (\mathcal{C}' \times \mathcal{C}') \\ G' &\text{ é conexo} \end{aligned}$$

Se o conjunto $\mathcal{C}'' \subseteq \mathcal{C}$ induzir um subgrafo $G'' = (\mathcal{C}'', \mathcal{N}'')$, com $\mathcal{N}'' = \mathcal{N} \cap (\mathcal{C}'' \times \mathcal{C}'')$, que seja desconexo, então:

$$G'' = G_1 \cup G_2 \cup \dots \cup G_k \quad (5)$$

onde $G_i = (\mathcal{C}_i, \mathcal{N}_i)$ são os componentes conexos de G'' . Neste caso, cada componente G_i deve ser tratado como um *cluster* de análise independente.

Como exemplificação desse conceito, considere-se quinze células, representadas da seguinte forma. Seja o conjunto de células (vértices) do grafo de vizinhanças:

$$\mathcal{C}' = \{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O\} \quad (6)$$

Seja o conjunto de arestas representando as vizinhanças entre as células, sendo as relações bidirecionais:

$$\begin{aligned} \mathcal{N}' = \{ & (A, B), (A, C), (A, D), (B, C), (C, D), (D, E), \\ & (E, F), (F, G), (G, H), (H, E), (B, F), \\ & (I, J), (I, K), (J, K), (K, L), (L, M), \\ & (M, N), (N, O), (O, I), (J, M), (L, N)\} \end{aligned} \quad (7)$$

O diagrama que representa esse grafo é ilustrado na Figura 3.2.

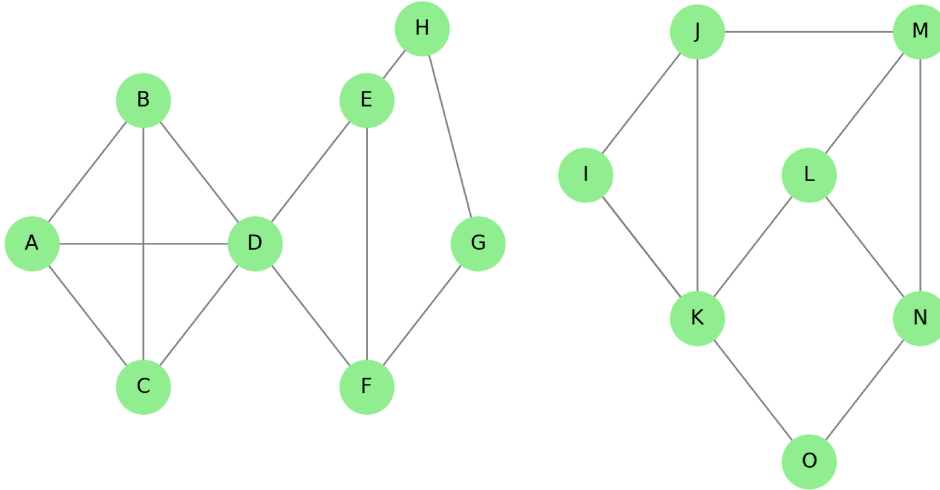


Figura 3.2 – Grafo representando dois *clusters* independentes para efeito de exemplificação de conceito de definição de escopo de análise. Figura elaborada pelo autor.

Neste caso, em vez de se realizar uma análise conjunta do *cluster* $\{A, B, \dots, O\}$, o mais adequado é segmentar a análise em duas partes distintas, correspondentes aos subconjuntos $\{A, B, \dots, H\}$ e $\{I, J, \dots, O\}$, uma vez que não há interação entre as células desses dois *clusters*.

A relação das vizinhanças entre células é muito bem definida na configuração do sistema e pode ser utilizada como critério para determinação dos *clusters*. Embora já tenha

sido definido como premissa que os dados de configuração do sistema não serão usados no ARCADE, essa etapa de definição de células e geografia não faz parte do escopo da metodologia e não necessariamente tem que seguir critérios rígidos, sendo essa proposta meramente uma sugestão formal para essa definição do escopo de análise.

3.3 Aquisição de Dados via *Crowdsourcing* (DAM)

Uma abordagem para aquisição de dados e mapeamento do ambiente de RF que já está disponível e é bastante utilizada na indústria atualmente é o *crowdsourcing*. Essa alternativa é bastante interessante, sobretudo levando-se em consideração que boa parte das operadoras não têm a funcionalidade MDT disponível e tendo em conta o grande esforço e custo para coleta e processamento de amostras via DTs. Há várias empresas que disponibilizam dados de medições georreferenciadas de RF, obtidas via *crowdsourcing*. Dentre elas, as mais famosas são a Opensignal (OpenSignal, 2025) e a Ookla (Ookla, 2025). Ambas as empresas realizam coletas de dados através de seus aplicativos móveis e websites, além de contratos e parcerias com diversos fornecedores de *apps* através de uma *Application Programming Interface* (API) aberta que coleta dados anonimizados de usuários durante suas execuções (Opensignal, 2026) (Ookla, 2026).

Além da contratação de dados de fornecedores externos, há ainda a possibilidade da operadora implementar *applets* de coleta de dados em sua própria aplicação. Uma vantagem dessa abordagem é o custo recorrente significativamente menor para a obtenção desses dados e um maior controle no processo como um todo. Em contraponto, uma desvantagem é a de não se ter informações de outras operadoras, inclusive as concorrentes, que viabilizariam o *benchmarking* em áreas geográficas específicas, sendo que, nesse caso, o uso se restringe basicamente a aplicações de engenharia ou de mercado com muito menor aplicabilidade.

De qualquer forma, a aquisição de dados através de técnicas de *crowdsourcing* é uma alternativa válida e muito interessante para a obtenção da informação de cobertura. Com o propósito de demonstrar uma solução fim a fim na proposta desta tese, será apresentada na Seção 4.1 uma abordagem prática de implementação de um serviço *crowdsourcing*.

3.4 Modelagem dos Dados de Entrada (DPM)

Uma vez resolvido o problema da obtenção de dados para alimentar o ARCADE, o próximo passo consiste em, com as informações do ambiente de RF em mãos, classificá-las e extrapolá-las, de modo a obter um retrato completo da cobertura em uma determinada área geográfica. O trabalho com os dados “crus” torna-se difícil devido à heterogeneidade e esparsidade das amostras, sobretudo nas regiões de borda e sujeitas a anomalias de cobertura. A Figura 3.3 ilustra, na prática, como tipicamente se distribuem os dados de

RF de uma única célula em uma área geográfica a ser analisada, onde diferentes cores representam diferentes níveis de RSRP.

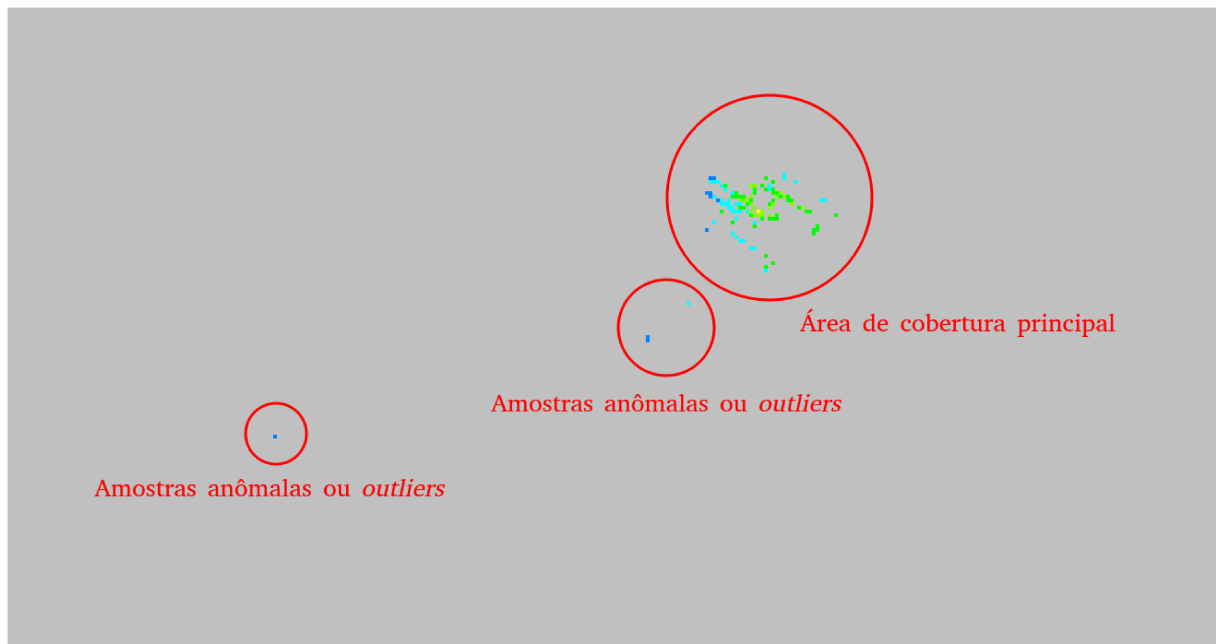


Figura 3.3 – Exemplo de amostras de RF de uma única célula dispostas em uma determinada região geográfica. Figura elaborada pelo autor.

Desta forma, é essencial estabelecer-se uma metodologia para a estruturação das amostras e modelamento da informação de entrada de forma que uma análise sistemática seja possível. O objetivo final é obter-se a cobertura “contínua” de cada célula analisada contida em um *cluster* de RBSs dentro de uma área geográfica pré-determinada. A proposta deste trabalho é que isso seja realizado dividindo-se a área total em uma grade georreferenciada, sendo que, para cada elemento da mesma, haverá uma (e apenas uma) amostra de RSRP associada a cada PCI dentro do *cluster*. As dimensões da área geográfica representada e do elemento da grade podem ser especificadas de acordo com a necessidade da aplicação, dependendo, por exemplo, do tamanho da área total e da quantidade de estações e células a serem analisadas no *cluster*. A Figura 3.4 ilustra esse conceito.

Considerando-se a esparsidade das amostras de RFs, quer sejam elas adquiridas por DT, *crowdsourcing* ou por MDT, provavelmente a maior parte dos elementos da grade ficaria vazia e, naturalmente, não haveria informações de todos os PCIs em todos os elementos. As anomalias de cobertura não necessariamente ficariam caracterizadas, pois a dispersão e escassez das amostras poderiam dissimulá-las. Além disso, há uma grande possibilidade de que um mesmo elemento da grade tenha mais de uma amostra referente ao mesmo PCI.

Há vários desafios na obtenção desse mapeamento em grade, sobretudo os relativos às características dos dados em si. A principal contribuição desta seção é propor uma abordagem de modelagem para a cobertura celular na presença de anomalias, *outliers* e

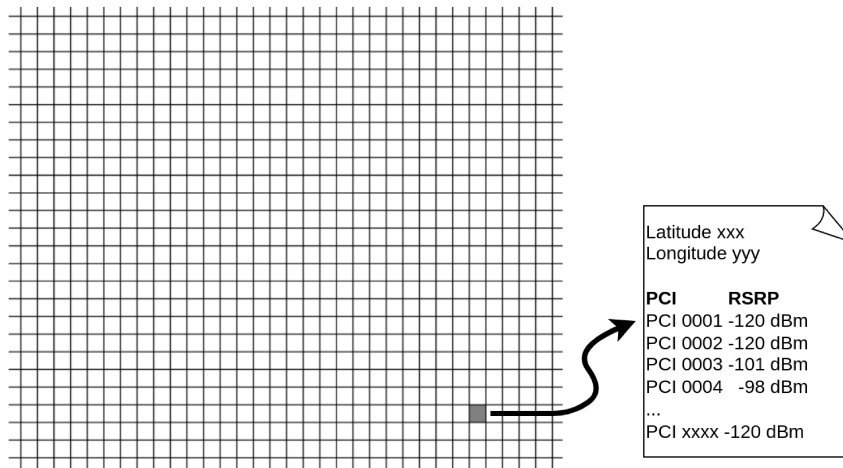


Figura 3.4 – Estrutura de grade georreferenciada e informações por elemento da grade. Figura elaborada pelo autor.

dados esparsos, mantendo, ao mesmo tempo, uma descrição coerente do ambiente de RF.

Assim, esta seção aborda os problemas e desafios da obtenção de um mapa contínuo de cobertura para sistemas celulares utilizando dados esparsos e amostras dispersas quando técnicas de predição baseadas em redes neurais artificiais são empregadas. Também apresenta uma proposta para mitigar questões decorrentes do *overfitting* dos modelos por meio da utilização de técnicas de expansão dos dados de treinamento, adicionando condições de contorno essenciais ao espaço analisado, de modo que o modelo se ajuste adequadamente às condições físicas reais observadas na prática – sem, no entanto, recorrer a modelos empíricos (como Okumura-Hata ou COST 231-Hata), semi-empíricos (como Walfisch-Ikegami) ou determinísticos (como *ray tracing*); todos eles baseados em predições matemáticas que simulam cobertura em topologias e morfologias específicas com base em modelos de propagação, como é comumente praticado em softwares de predição de cobertura.

No aprendizado de máquina, tanto o algoritmo utilizado quanto a estrutura da RNA, juntamente com seus parâmetros, são características essenciais para um bom desempenho em termos de acurácia e precisão de sua saída; no entanto, esses fatores não são os únicos que influenciam esses resultados. O desempenho do sistema é fortemente afetado pela quantidade e qualidade dos dados disponíveis para treinamento e teste (YING, 2019). Conforme já mencionado na Seção 2.2.5, na presença de ruídos ou ao lidar com dados muito esparsos que não representam casos genéricos, o modelo pode apresentar bom desempenho no conjunto de treinamento, mas falhar em generalizar para o conjunto de teste, o que é conhecido como *overfitting*, e há várias razões para sua ocorrência: (a) a presença de dados ruidosos no conjunto de treinamento; (b) complexidade excessiva da hipótese (muitas entradas), e; (c) problemas inerentes ao processo de tomada de decisão das funções de avaliação, que se baseiam em uma pontuação máxima (YING, 2019). No caso apresentado nesta tese, a primeira razão é a principal causa de sobreajuste no modelo,

uma vez que sua complexidade não é elevada, e assume-se que a modelagem é adequada à aplicação. Logo, dada a natureza desta investigação, uma abordagem adequada é a utilização de uma técnica de expansão dos dados de treinamento.

No método proposto, classifica-se as amostras de acordo com seu agrupamento e coerência com a cobertura da célula e, em seguida, categoriza-se as áreas de análise com base em sua prevalência, conforme será apresentado a seguir.

3.4.1 Classificação de Amostras

Como mencionado anteriormente, a confusão de anomalias de cobertura decorrentes de configurações subótimas de sistemas celulares (por exemplo, posicionamento inadequado de antenas) com a possível existência de amostras consideradas *outliers* causadas por erros de leitura ou situações anômalas, representa um desafio significativo para a modelagem de cobertura. Uma RNA que use os dados diretamente pode interpretar essas amostras anômalas como padrões a serem seguidos, resultando em modelos incorretos, incluindo, por exemplo, “rajadas de cobertura” exageradas em direção às anomalias.

Nesta investigação, o primeiro passo consiste em propor um método de categorização através do qual as amostras possam ser classificadas de acordo com suas características de dispersão espacial. As definições relativas à categorização das amostras e à definição formal de cada categoria são apresentadas abaixo.

Sejam definidos:

- $S = \{s_i(m, n) \mid i \in \mathbb{N}\}$: o conjunto de amostras s_i de uma célula, com coordenadas m e n ;
- $N = |S| < \infty$: o número finito de amostras da célula;
- $c = (m_c, n_c)$: o centróide espacial do conjunto S , dado por:

$$m_c = \frac{1}{N} \sum_{i=1}^N m_i, \quad n_c = \frac{1}{N} \sum_{i=1}^N n_i \quad (8)$$

- $d(s_i, c)$: a distância da amostra s_i ao centróide c , definido como:

$$d(s_i, c) = \sqrt{(m_i - m_c)^2 + (n_i - n_c)^2} \quad (9)$$

- μ_d : a média de todas as distâncias $d(s_i, c)$:

$$\mu_d = \frac{1}{N} \sum_{i=1}^N d(s_i, c) \quad (10)$$

□ σ_d : o desvio-padrão das distâncias $d(s_i, c)$:

$$\sigma_d = \sqrt{\frac{1}{N} \sum_{i=1}^N (d(s_i, c) - \mu_d)^2} \quad (11)$$

Usando as definições acima, pode-se então especificar os conjuntos da seguinte forma:

□ **Conjunto de amostras normais** (S_N): Essas são amostras localizadas em áreas tipicamente cobertas pela célula, mais próximas ao centróide de sua região de cobertura:

$$S_N = \{s_i \in S \mid d(s_i, c) < \mu_d + 3\sigma_d\} \quad (12)$$

□ **Conjunto de amostras anômalas (anomalias)** (S_A): Essas são amostras que se desviam significativamente de área de cobertura normal da célula, tipicamente representando problemas de cobertura (por exemplo, *overshooting*):

$$S_A = \{s_i \in S \mid \mu_d + 3\sigma_d \leq d(s_i, c) < \mu_d + 6\sigma_d\} \quad (13)$$

□ **Conjunto de amostras *outliers*** (S_O): Essas são amostras oriundas de erros de medição, tanto devido a problemas técnicos com as UEs, ou posicionamentos atípicos, como UEs em topos de edifícios altos medindo células muito longe de suas áreas de cobertura. Nas análises aqui apresentadas, as amostras *outliers* S_O serão descartadas do conjunto S , pois são consideradas ruído que distorceriam o modelo:

$$S_O = \{s_i \in S \mid d(s_i, c) \geq \mu_d + 6\sigma_d\} \quad (14)$$

O conjunto total S seria a união dos três subconjuntos, mas como as amostras S_O são descartadas, o conjunto S passa a ser:

$$S = S_N \cup S_A \quad (15)$$

E a interseção entre S_N , S_A , e S_O são conjuntos vazios:

$$S_N \cap S_A = \emptyset, \quad S_N \cap S_O = \emptyset, \quad S_A \cap S_O = \emptyset \quad (16)$$

3.4.2 Classificação de Áreas

Agora é possível classificar a área total (A_T) entre a Área de Normalidade (A_N) e a Área de Anormalidade (A_A) da seguinte maneira:

- **Área Total** (A_T): Corresponde à área total considerada na análise, previamente delimitada como a região completa de interesse para o processo de aprimoramento da cobertura celular de um *cluster*.
- **Área de Normalidade** (A_N): É o retângulo contido em A_T , definido pelas coordenadas mínimas e máximas das amostras em S_N , sendo m_i e n_i correspondentes, respectivamente, à latitude e longitude da amostra s_i :

$$m_{\min} = \min\{m_i \mid s_i \in S_N\}, \quad (\text{mínima latitude dos pontos em } S_N) \quad (17)$$

$$m_{\max} = \max\{m_i \mid s_i \in S_N\}, \quad (\text{máxima latitude dos pontos em } S_N) \quad (18)$$

$$n_{\min} = \min\{n_i \mid s_i \in S_N\}, \quad (\text{mínima longitude dos pontos em } S_N) \quad (19)$$

$$n_{\max} = \max\{n_i \mid s_i \in S_N\}, \quad (\text{máxima longitude dos pontos em } S_N) \quad (20)$$

A Área de Normalidade (A_N) então compreende o conjunto de amostras dentro do retângulo definido por:

$$A_N = \{(m, n) \mid m_{\min} \leq m \leq m_{\max}, n_{\min} \leq n \leq n_{\max}\}. \quad (21)$$

- **Área de Anormalidade** (A_A): É a área definida pelo complemento da área total (A_T), após a exclusão da Área de Normalidade (A_N), ou:

$$A_A = A_T \setminus A_N, \quad (22)$$

Desta forma, todos os conjuntos que representam as amostras e áreas estão definidos formalmente para o prosseguimento da análise.

Após todas as definições, ressalta-se que as amostras serão dispostas de forma discreta na estrutura de grade, tendo-se assim uma estrutura como mostrada na Figura 3.4. Se no mesmo elemento (que representa um ponto de coordenada geográfica) houver múltiplas amostras, denotadas por $s'_j(m, n)$, sendo essa a j -ésima amostra para a mesma coordenada (m, n) , definimos então que o valor de $s_i(m, n)$ será a média das amostras $s'_j(m, n)$, conforme a seguinte fórmula:

$$s_i(m, n) = \frac{1}{|\mathcal{T}(m, n)|} \sum_{s'_j \in \mathcal{T}(m, n)} s'_j(m, n) \quad (23)$$

onde:

- $\mathcal{T}(m, n)$ representa o conjunto das amostras coletadas na mesma coordenada (m, n) ;
- $|\mathcal{T}(m, n)|$ é o número de amostras em (m, n) .

Caso haja apenas uma única amostra $s'_j(m, n)$, então $s_i(m, n) = s'_j(m, n)$.

De forma simplificada, $s_i(m, n)$ representa a média aritmética simples das amostras $s'_j(m, n)$ quando há múltiplos valores de medição em uma mesma coordenada. Desta forma, para cada elemento de grade, haverá apenas um único valor de RSRP associado a cada PCI dentro do escopo da análise.

3.5 Processos Gaussianos para Aumento de Dados (SAM)

O problema ao tentar interpolar os dados concentrados em A_N na presença de anomalias em A_A (que devem ser incorporadas ao modelo), é que a escassez de informações em A_A leva o modelo da RNA a gerar inconsistências nessa região. A solução encontrada foi aumentar a densidade dos dados em A_A de maneira que se mantenha a coerência com o que é verificado na prática. A abordagem adotada neste trabalho baseia-se na extrapolação de dados utilizando processos gaussianos com *kernel* espacial, ou GPSKs. Técnicas de processos gaussianos estão entre as utilizadas para o aumento do número de amostras a fim de delinear condições de contorno e proporcionar maior coerência a modelos baseados em aprendizado de máquina (TANG et al., 2023).

No caso específico deste trabalho, será utilizado o *kernel* RBF, que, conforme já explicado na Seção 2.2.4, é particularmente adequado quando a função de covariância representa uma dependência espacialmente radial.

Para cada PCI e para cada elemento da grade, é realizada a extrapolação dos dados, embora nem todas elas sejam utilizadas no aumento das amostras, como será descrito a seguir.

3.5.1 Aumento de Amostragem na Área de Anormalidade

Conforme explicado, é necessária a aplicação de uma técnica para aumento de amostras para uma boa convergência de uma RNA, o que será feito com GPSK. No caso do problema abordado neste trabalho, as amostras adicionais devem ser agregadas sobretudo onde há esparsidade de dados, no intuito tanto de ressaltar as anomalias identificadas por amostras isoladas, quanto de estabelecer condições de contorno que apontam áreas sem cobertura a uma certa distância da área de cobertura normal da célula (A_N).

Desta forma, apenas a área com menor densidade amostral, A_A , terá aumento de dados por meio de processos gaussianos. Isso se justifica pelo fato de que as regiões de maior interesse, A_N , já contêm dados reais suficientes e, portanto, não necessitam de extrapolação adicional.

O espaçamento entre as amostras a serem incorporadas ao modelo depende da proximidade das anomalias em relação a A_N . Para cada amostra anômala $s_i \in S_A$, define-se

$d_N(s_i)$ como a menor distância, ao longo dos eixos de latitude ou longitude, entre s_i e a fronteira mais próxima do retângulo A_N , isto é, seus limites norte, sul, leste e oeste. A menor dessas distâncias, considerando todas as amostras em S_A , é denotada por d_{\min} . Assim, o intervalo entre as amostras obtidas por GPSK a serem consideradas no treinamento do modelo é dado pela distância amostral d_S :

$$d_N(s_i) = \min(|m_i - m_{\min}|, |m_i - m_{\max}|, |n_i - n_{\min}|, |n_i - n_{\max}|) \quad (24)$$

$$d_{\min} = \min_{s_i \in S_A} d_N(s_i), \quad d_S = \frac{d_{\min}}{2} \quad (25)$$

Dessa forma, o espaçamento entre as amostras corresponde à metade da menor distância entre uma amostra anômala e a fronteira de A_N . Essa definição garante que as anomalias serão bem representadas no conjunto de dados aumentados. A Figura 3.5 mostra uma representação gráfica do aumento de dados obtido com o método especificado. Como é uma imagem ilustrativa, as dimensões não estão em escala.

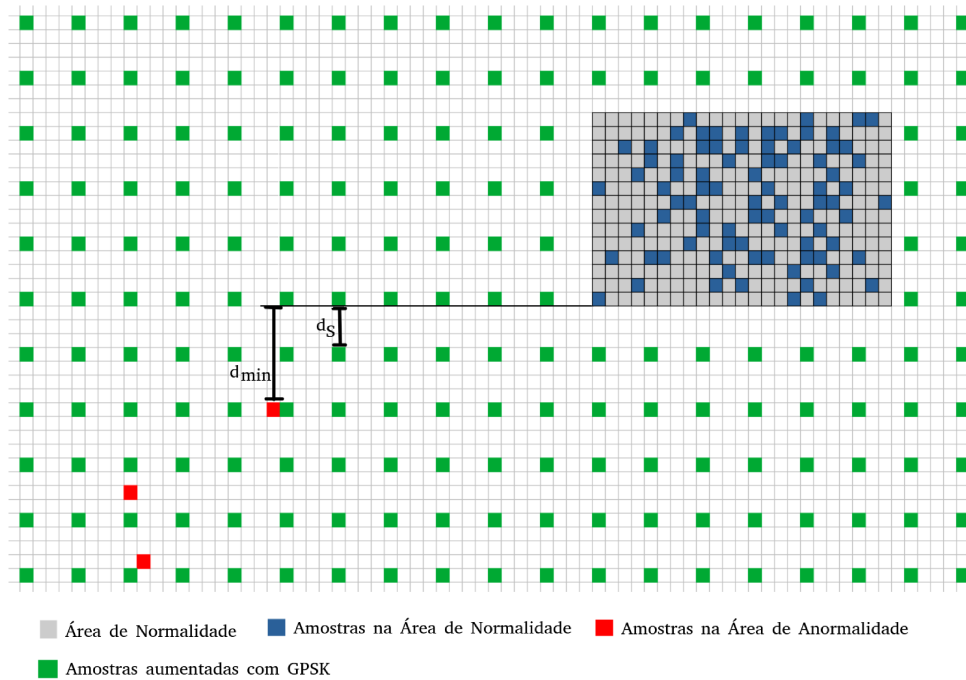


Figura 3.5 – Representação da amostragem após o aumento de dados com GPSK.

O conjunto de amostras agregadas através de GPSK será denominado S_G (pois corresponde a amostras geradas a partir de um modelo gaussiano). Assim, obtém-se o conjunto total de amostras (S_T), lembrando que o conjunto de amostras *outliers* S_O será descartado:

$$S_T = S_N \cup S_A \cup S_G \quad (26)$$

Ao final, S_T será o conjunto de amostras que alimentará a RNA responsável por generalizar o modelo para A_T .

3.6 Modelagem de Cobertura Usando RNA (MTGM)

Finalmente, uma vez que haja amostras suficientes que permitam a representação dos elementos em toda a área estudada, isto é, haja amostras capazes de caracterizar as regiões de anomalias e estabelecer condições de contorno em regiões longe da cobertura principal da célula, é proposto que uma RNA seja treinada com esses dados e modele, de uma maneira genérica, a cobertura para todos os elementos da grade na Área Total (A_T).

A estrutura da RNA deve ser determinada em experimentos que meçam sobretudo sua eficiência, precisão e acurácia na generalização do modelo. Porém, a estrutura geral da RNA para generalizar a cobertura de uma célula na área (A_T), conforme o contexto deste trabalho, deve necessariamente levar em conta:

- ❑ Camada de entrada: deverá conter dois nós, representando as coordenadas (latitude e longitude) normalizadas de cada elemento da grade que representa (A_T);
- ❑ Camada de saída: deverá ter um nó, representando o nível RSRP normalizado correspondente às coordenadas de entrada;

A determinação da quantidade de camadas ocultas, nós por camada, funções de ativação e demais parâmetros da rede será demonstrada no capítulo 4.

O treinamento da rede é realizado célula a célula, isto é, o treinamento se dará para cada célula na área total. Após esse treinamento o nível de RSRP para todos os elementos da estrutura da grade referenciada – conforme mostrado na Figura 3.4 – será inferido pelo modelo. Isso significa também que, para cada célula no *cluster*, será repetido todo o processo de classificação das amostras, classificação das áreas, aumento de dados, treinamento da RNA e, finalmente, predição do RSRP na (A_T).

3.7 Método para Aprimoramento de Cobertura (COM)

Finalmente, uma vez obtida a base de dados final a ser avaliada, contendo as informações mostradas na Figura 3.4, inicia-se a etapa de avaliação da quantidade, qualidade e equilíbrio da cobertura, de modo a viabilizar uma função objetivo que reflita as condições gerais do *cluster* em termos de desempenho e eficiência na cobertura.

Em redes móveis, a avaliação do desempenho de uma célula não pode ser realizada de forma isolada, uma vez que há um alto grau de interação entre células em um mesmo *cluster*. Em ambientes reais, o desempenho de uma célula está intrinsecamente ligado ao contexto na qual ela está inserida, que depende de vários fatores, como densidade da região de cobertura, topologia e morfologia do terreno, dentre outros. Porém, mesmo em um sistema de altíssima eficiência de cobertura, a interferência entre células do mesmo *cluster* é inevitável. Existe um nível razoável e aceitável de sobreposição das células, no qual se atinge um bom equilíbrio entre cobertura e interferência, e além do qual ajustes

de potência com o intuito de diminuir a interferência ou aumentar a cobertura acabam por reduzir a qualidade do serviço. Desta forma, é necessária a definição de grandezas mensuráveis para a avaliação tanto da cobertura de uma célula no contexto de um *cluster* quanto da interferência entre as mesmas.

Para compor a função objetivo, que deve ser minimizada para se obter o melhor desempenho e eficiência do cluster, são avaliados três aspectos distintos e intrínsecos ao *cluster*, que serão formalmente definidos na Seção 3.7.1: cobertura, dominância e equilíbrio no tamanho das células.

A partir desses parâmetros, será formulada a função objetivo $\tilde{J}(P)$, que incorpora os três aspectos descritos. Ressalta-se que existe um compromisso mútuo (*trade-off*) entre cobertura e dominância, uma vez que o aumento da cobertura de uma célula tende a intensificar a interferência sobre células adjacentes, enquanto estratégias voltadas ao aumento da dominância, como a redução da potência de células interferentes, podem resultar em perdas de cobertura.

A seguir, é demonstrada a formulação da função objetivo considerada neste trabalho, e que orienta o processo de busca por soluções aprimoradas utilizando o algoritmo CMA-ES.

3.7.1 Formulação da Função Objetivo

Considere-se um conjunto de N pontos de amostragem e um cluster composto por K células. Seja:

$$P = (\Delta P_1, \Delta P_2, \dots, \Delta P_K) \in \mathbb{R}^K \quad (27)$$

o vetor de ajustes de potência, em dB, aplicados ao canal de referência CRS cada célula do *cluster*, lembrando que essa é a potência efetivamente medida pelas UEs para determinação da intensidade do sinal da célula e de sua qualidade.

Seja $RSRP_{i,k}$ o valor original do RSRP da célula k no ponto i , sendo que esse ponto representa um elemento na matriz que mapeia a A_T . O modelo de potência ajustada é dado por:

$$RSRP'_{i,k}(P) = RSRP_{i,k} + \Delta P_k, \quad i = 1, \dots, N, \quad k = 1, \dots, K. \quad (28)$$

Para cada ponto i , define-se o índice da célula servidora por:

$$k^*(i) = \arg \max_{k \in \{1, \dots, K\}} RSRP'_{i,k}(P), \quad (29)$$

e o valor do servidor com a potência ajustada é:

$$S_i(P) = RSRP'_{i,k^*(i)}(P). \quad (30)$$

Para o entendimento claro, o termo $S_i(P)$ representa o nível de potência do sinal recebido da célula servidora no ponto i , considerando os ajustes de potência definidos pelo vetor P . Em outras palavras, $S_i(P)$ corresponde ao maior valor de RSRP ajustado entre todas as células do *cluster* no ponto i . No caso particular em que não são aplicados ajustes de potência às células, isto é, $\Delta P_k = 0$ para todo k , o valor de $S_i(P)$ coincide com o nível original de RSRP da célula que apresenta o maior sinal recebido no ponto i , caracterizando a célula servidora original desse ponto.

O interferente dominante (segundo maior RSRP) é definido como:

$$I_i(P) = \max_{k \in \{1, \dots, K\} \setminus \{k^*(i)\}} RSRP'_{i,k}(P). \quad (31)$$

Na prática, $I_i(P)$ corresponde à potência recebida da segunda melhor servidora no ponto i , sendo considerada interferência nesse ponto.

Define-se ainda a diferença servidor–interferente, ou simplesmente dominância, como a diferença (em dB) entre a potência recebida da célula melhor servidora e da célula interferente dominante, ou seja:

$$\Delta_i(P) = S_i(P) - I_i(P). \quad (32)$$

3.7.1.1 Definição do Termo Índice de Cobertura $C_i(P)$

Sejam os limiares S_{\min} o nível mínimo de RSRP válido para que o UE tenha serviço com alguma qualidade e S_{\max} o nível de RSRP acima do qual não há mais ganho de qualidade na ausência de interferência.

O *índice de cobertura* no ponto i $C_i(P)$ é definido por:

$$C_i(P) = \begin{cases} 0, & \text{se } S_i(P) \leq S_{\min}, \\ \frac{S_i(P) - S_{\min}}{S_{\max} - S_{\min}}, & \text{se } S_{\min} < S_i(P) < S_{\max}, \\ 1, & \text{se } S_i(P) \geq S_{\max}. \end{cases} \quad (33)$$

Esse termo quantifica de forma contínua a quantidade de cobertura no ponto i , refletindo o nível efetivo de potência do sinal recebido pela célula servidora. Valores de $C_i(P)$ iguais ou próximos a zero indicam regiões sem cobertura adequada, enquanto valores iguais ou próximos a um caracterizam áreas em que o nível de RSRP em si é suficiente para garantir o serviço, não havendo ganhos adicionais com o aumento de potência. Ressalta-se também que $C_i(P)$ não leva em consideração a interferência, de forma que não garante por si só uma boa qualidade na conexão.

3.7.1.2 Definição do Termo Índice de Dominância $D_i(P)$

Sejam os limiares Δ_{\min} a mínima diferença possível do nível servidor–interferente (dominância) e Δ_{ref} a mínima condição onde há serviço pleno em uma rede celular. O Δ_{\min} é zero por definição – uma vez que, se a célula interferente dominante tiver potência maior que a servidora, ela em si passa a ser servidora –, enquanto o Δ_{ref} é definido como a relação portadora–interferência mínima para a manutenção de uma boa percepção do serviço.

Define-se então o índice de dominância como:

$$D_i(P) = \begin{cases} 0, & \text{se } S_i(P) \leq S_{\min}, \\ \frac{\Delta_i(P) - \Delta_{\min}}{\Delta_{\text{ref}} - \Delta_{\min}}, & \text{se } S_i(P) > S_{\min} \text{ e } \Delta_{\min} < \Delta_i(P) < \Delta_{\text{ref}}, \\ 1, & \text{se } S_i(P) > S_{\min} \text{ e } \Delta_i(P) \geq \Delta_{\text{ref}}. \end{cases} \quad (34)$$

Esse índice quantifica a predominância da célula servidora em relação ao interferente dominante (segunda melhor servidora) no ponto i . Valores de $D_i(P)$ iguais ou próximos a zero indicam condições de forte interferência, enquanto valores iguais ou próximos à unidade caracterizam cenários em que o sinal do servidor é suficientemente dominante neste ponto.

3.7.1.3 Definição do Índice de Distribuição Espacial *Spatial Distribution Index* (SDI) e sua Variância Entre Células

Finalmente, define-se o SDI, que mede o quão similares são as coberturas das células, característica desejável em um sistema móvel, para efeito de distribuição mais uniforme e coerente do tráfego de rede. Para cada célula k do *cluster* define-se o conjunto de pontos relevantes de cobertura da forma:

$$\mathcal{S}_k(P) = \{(x_i, y_i, w_i(P)) \mid RSRP'_k(x_i, y_i; P) \geq RSRP_{\min}\}, \quad (35)$$

onde $w_i(P)$ é o peso atribuído ao ponto i , proporcional ao valor de RSRP normalizado da célula nesse ponto e $RSRP'_k(x_i, y_i; P)$ representa o nível de potência do sinal de referência recebido no ponto (x_i, y_i) proveniente da célula k , após a aplicação do vetor de ajustes de potência P . Pontos com RSRP menores que $RSRP_{\min}$ são descartados desse conjunto de pontos relevantes. Assim, pontos com maior potência de sinal recebido de uma determinada célula possuem maior influência no posicionamento do centróide ponderado desta célula, conforme será definido a seguir.

O centróide ponderado da célula é dado por:

$$\bar{x}_k(P) = \frac{\sum_i w_i(P)x_i}{\sum_i w_i(P)}, \quad \bar{y}_k(P) = \frac{\sum_i w_i(P)y_i}{\sum_i w_i(P)}. \quad (36)$$

Então, define-se o SDI de uma célula k como o raio médio ponderado dessa célula, ou, mais precisamente, a média ponderada das distâncias de suas amostras ao seu centróide:

$$SDI_k(P) = \frac{\sum_i w_i(P) d((x_i, y_i), (\bar{x}_k(P), \bar{y}_k(P)))}{\sum_i w_i(P)} \quad (37)$$

onde $d(\cdot)$ representa a distância geodésica entre os pontos considerados.

Define-se então a média dos SDIs das células através da equação:

$$\mu_{SDI}(P) = \frac{1}{K} \sum_{k=1}^K SDI_k(P) \quad (38)$$

Define-se também o valor médio de referência $\mu_{SDI}(P_0)$, onde P_0 é o vetor de ajustes de potência inicial (configuração nominal do sistema, sem correção nas potências). A variância normalizada entre os SDIs das células é então dada por:

$$V_{SDI}(P) = \frac{1}{K \mu_{SDI}^2(P_0)} \sum_{k=1}^K \left[SDI_k(P) - \mu_{SDI}(P) \right]^2 \quad (39)$$

A normalização por $\mu_{SDI}^2(P_0)$ – que é uma constante e não é atualizada nas iterações do CMA-ES – foi adotada para tornar o termo adimensional e relativo à configuração inicial do sistema.

Fisicamente, a variância dos valores de SDI entre as células expressa o grau relativo de diferença espacial das suas coberturas no *cluster*, com relação ao raio médio das células na configuração original. Valores elevados de $V_{SDI}(P)$ sugerem que existem células com áreas efetivas significativamente distintas proporcionalmente à configuração nominal, indicando cenários indesejados em que algumas células concentram grande parte da cobertura, enquanto outras apresentam abrangência reduzida. Por outro lado, valores baixos dessa variância refletem uma distribuição espacial mais equilibrada, na qual as células possuem dimensões geométricas semelhantes, favorecendo uma ocupação mais uniforme do espaço e uma distribuição mais homogênea do tráfego de rede.

3.7.1.4 Regularização dos Ajustes de Potência

Em problemas de busca por otimização envolvendo muitos parâmetros de controle, é comum a introdução de termos de regularização para se restringir o espaço de soluções admissíveis e evitar ajustes excessivos que, embora matematicamente possam ser ótimos, fisicamente são de pouca aplicação prática (MORADI; BERANGI; MINAEI, 2020). No contexto do *tuning* de potências de células de um *cluster*, ajustes extremos podem resultar em soluções tecnicamente inviáveis ou indesejáveis, como aumentos exagerados de potência em algumas células e redução a potências mínimas em outras. Desta forma, um termo de regularização que controla os passos de ajustes de potência é proposto de forma a mitigar esse efeito.

Com $\sigma_P > 0$ representando uma escala típica de variação admissível (em dB) dos ajustes de potência e o ΔP_k o ajuste de potência na célula k , define-se o termo de regularização como:

$$R(P) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\Delta P_k}{\sigma_P} \right)^2 \quad (40)$$

Esse termo de regularização enquadra-se na classe de penalizações de norma L2, também conhecidas como *weight decay* e é um dos mais comumente usados para modificações de funções objetivo do tipo $J'(\theta; X, y) = J(\theta; X, y) + \alpha\Omega(\theta)$ (MORADI; BERANGI; MINAEI, 2020). No contexto deste trabalho, o termo penaliza desvios elevados nos ajustes de potência individuais, favorecendo soluções em que as variações de potência permanecem dentro de uma faixa considerada razoável do ponto de vista físico e operacional. Desta forma, como será apresentado a seguir, a função objetivo $\tilde{J}(P)$ passará a incorporar não apenas o desempenho em termos de cobertura, dominância e equilíbrio no tamanho das células, mas também a suavidade da solução.

3.7.2 Definição da Função Objetivo Básica $J(P)$

Antes de apresentar a formulação final da função objetivo $\tilde{J}(P)$, será definida a função objetivo base $J(P)$, do qual o primeiro é uma extensão regularizada da segunda, tanto pela introdução do SDI quanto pelo termo de regularização $R(P)$.

Considere-se que o desempenho global da RAN de sistema celular depende, fundamentalmente, de dois critérios: i) a qualidade de cobertura nos pontos da área de amostragem e ii) a dominância do servidor frente ao interferente dominante. Assim, define-se a função básica $J(P)$ como:

$$J(P) = \frac{1}{N} \sum_{i=1}^N \left[w_{\text{cov}} C_i(P) + w_{\text{dom}} D_i(P) \right], \quad (41)$$

onde N é a quantidade de elementos da matriz que representa A_T , $C_i(P)$ é o índice de cobertura no ponto i considerando o ajuste de acordo com o vetor de potências P , e $D_i(P)$ é o índice de dominância servidor-interferente também com o ajuste com o vetor P , ambos já definidos anteriormente. Os parâmetros $w_{\text{cov}} > 0$ e $w_{\text{dom}} > 0$ são pesos que ajustam a relevância relativa de cada critério e que serão empiricamente obtidos, conforme será descrito no Capítulo 4.

A função $J(P)$ avalia o desempenho médio do cluster em termos de cobertura e dominância, mas ainda não incorpora penalizações por ajustes excessivos de potência ou desigualdades de abrangências geográficas entre células. Esses efeitos adicionais são introduzidos na função estendida $\tilde{J}(P)$, apresentada a seguir.

3.7.2.1 Função Objetivo Final

À função $J(P)$, serão acrescidos, finalmente, dois termos: um de regularização para mitigar grandes ajustes de potência e outro de penalização por desbalanceamento de SDI. Em termos didáticos:

$$\tilde{J}(P) = J(P) - \text{penalização espalhamento SDI} - \text{regularização} \quad (42)$$

A função a ser maximizada é dada por

$$\tilde{J}(P) = \frac{1}{N} \sum_{i=1}^N \left[w_{\text{cov}} C_i(P) + w_{\text{dom}} D_i(P) \right] - w_{\text{SDI}} V_{\text{SDI}}(P) - \lambda R(P) \quad (43)$$

onde w_{cov} , w_{dom} , w_{SDI} , $\lambda > 0$ são pesos dos parâmetros do modelo. Conforme já mencionado, todos esses pesos serão determinados empiricamente no Capítulo 4.

O problema consiste então em encontrar o vetor P^* que resulta no melhor resultado possível – observando-se a estocasticidade do processo – da função objetivo $\tilde{J}(P)$, ou, de maneira formal:

$$P^* = \arg \max_{P \in \mathbb{R}^K} \tilde{J}(P) \quad (44)$$

Essa categoria de problema é tipicamente resolvida por um método de otimização estocástica, que será definido na próxima seção.

3.7.3 Resolução do Problema Utilizando-se CMA-ES

Conforme definido na seção anterior, o problema torna-se encontrar o valor do vetor de ajuste de potências P^* que maximiza o valor de $\tilde{J}(P)$. Essa solução não pode ser obtida de forma analítica nem trivial e, nas condições usuais da quantidade de células e da granularidade dos ajustes do vetor P , a quantidade de soluções possíveis é muito grande. Isso porque a função objetivo não é diferenciável e por isso não pode ser resolvida com métodos clássicos baseados em gradiente, pois o modelo de avaliação de cobertura e interferência está fundamentado em dados empíricos discretizados no espaço e não em equações de predição.

Desta forma, a resolução do problema é abordada por meio de um método de otimização estocástica, especificamente o CMA-ES (HANSEN; OSTERMEIER, 2001). Conforme apresentado na Seção 2.2.6, o CMA-ES é particularmente adequado para problemas de otimização estocástica contínua em espaços reais de alta dimensionalidade, nos quais a função objetivo é uma “caixa-preta”, isto é, avaliada unicamente de forma numérica, sem uma estrutura analítica diferenciável.

Neste trabalho, a aplicação desse método consiste em encontrar um vetor de ajustes de potência P que resulte em um aprimoramento real da cobertura do *cluster*, através de um processo iterativo, no qual a cada iteração um conjunto de soluções candidatas

é amostrado a partir de uma distribuição normal, sendo cada amostra avaliada através da função objetivo $\tilde{J}(P)$. Com base no desempenho dessas soluções, os parâmetros estatísticos da distribuição são adaptados de modo a direcionar a busca para regiões mais promissoras do universo de soluções. Essa estratégia apresenta vantagens importantes no contexto de melhoria de cobertura da RAN em sistemas móveis quando comparada com outros métodos (SEMENCHIKOV et al., 2021).

Desta forma, o CMA-ES foi escolhido por ser apropriado em termos de estratégia adaptativa de amostragem, fazendo com que a busca por soluções aprimoradas seja viável com um número relativamente reduzido de avaliações da função objetivo.

Avaliação Experimental

Esta tese propõe o ARCADE, uma metodologia para o aprimoramento da qualidade da cobertura celular de um *cluster* específico com base em informações esparsas do ambiente de RF. Conforme já visto no Capítulo 3, há uma sequência de processos a serem seguidos para se chegar às conclusões pretendidas. Este capítulo irá detalhar os experimentos realizados com base em dados reais coletados de uma rede móvel operacional, comparando o desempenho de uma RAN em suas configurações originais de rede com o desempenho após implementadas soluções propostas pelo processo automatizável sugerido neste trabalho. Essa comparação tem como objetivo evidenciar a eficácia da metodologia ARCADE na potencial automatização de processos intrinsecamente realizados por engenheiros especialistas, como a melhoria na qualidade de cobertura de um ambiente RAN.

O Capítulo 4 será dividido da seguinte forma: a Seção 4.1 descreve uma implementação experimental do serviço de aquisição de dados via *crowdsourcing*, incluindo uma prova de conceito testada em uma operadora real. A Seção 4.2 explica a origem da informação de entrada, a rede e o *cluster* a ser analisado, além dos recursos computacionais usados no processamento das informações. A Seção 4.3 demonstra detalhadamente como é feito o aumento de dados usando um caso de uma única célula como exemplo, sendo demonstrado o método GPSK para aumento de amostragem e posterior generalização com a RNA, e comparando o resultado dos pontos de amostra com a cobertura generalizada da mesma. A Seção 4.4 mostra como foi realizado o aumento das amostras e generalização de cobertura para todo o cluster de 24 células, incluindo a aplicação do módulo SAM e do módulo MTGM. Então, na Seção 4.5, o processo de aprimoramento da cobertura do *cluster* é realizado, demonstrando a aplicação da metodologia proposta e o resultado obtido. A Seção 4.6 apresenta uma discussão sobre os resultados encontrados, apresentando o comparativo dos cenários antes e depois de aplicado o processo, através de mapas e tabelas, além de discutir sobre o esforço computacional no mesmo. Na Seção 4.7 é apresentado um comparativo da eficiência energética e pegada de carbono entre a solução proposta em relação aos processos atualmente mais usados na prática. Finalmente, na Seção 4.8,

são respondidas as questões de pesquisa propostas neste trabalho.

4.1 Aquisição de Dados Via *Crowdsourcing*

Nesta seção, será brevemente apresentada a abordagem para o projeto e desenvolvimento de um sistema para aquisição de dados via *crowdsourcing* e apresentado um protótipo funcional do mesmo, implementado em uma operadora real. Como já comentado no Capítulo 3, a coleta de dados nessa modalidade é uma alternativa viável e interessante para o mapeamento do perfil de cobertura de células em uma determinada geografia.

4.1.1 Arquitetura da Solução

A proposta é o desenvolvimento de um *applet* para dispositivos com sistema operacional (SO) Android, que pode ser embarcado em um aplicativo instalado nos celulares que possuam esse SO, inclusive nos próprios apps das operadoras, que já têm uma aderência significativa dentro de suas bases de clientes. Esse *applet* roda em plano de fundo, de modo a não influenciar na operação do dispositivo e não ser percebido pelo usuário. Além disso, é necessário estruturar os servidores e bases de dados que irão receber, consolidar e organizar as informações, tornando-as prontas para serem publicadas. Para utilização direta pela operadora, ainda é necessário o desenvolvimento de um programa que permita o acesso aos dados pelos usuários finais e publicação na forma de tabelas e mapas, porém essa última etapa não será abordada neste trabalho, por não ser escopo do mesmo.

Para validação da prova de conceito, optou-se pelo desenvolvimento de um aplicativo completo, em vez da abordagem baseada em *applet*. O aplicativo foi desenvolvido no *framework React Native*, usando *TypeScript* para a camada de aplicação multiplataforma e *Kotlin* para a implementação de funcionalidades nativas no ambiente Android. A base de dados utilizada foi o PostgreSQL, escolhido por ser um sistema gerenciador de banco de dados relacional maduro, amplamente usado na indústria, de fácil utilização, com ampla adoção, alto desempenho, robustez e suporte nativo a extensões avançadas para manipulação de dados espaciais e temporais. Além disso, trata-se de um software de código aberto, com amplo suporte da comunidade. O servidor foi implantado como serviço em nuvem na *Amazon Web Services (AWS)*, utilizando instâncias do serviço *Elastic Compute Cloud (EC2)*. A Figura 4.1 demonstra a arquitetura simplificada da solução.

4.1.2 Informações Coletadas e Disponibilizadas

Durante sua execução, o aplicativo coleta blocos de 96 amostras de RF georreferenciadas e com identificação da célula, por meio da leitura de medições de sinal do celular, dados de geolocalização do GPS (latitude e longitude) e identificação da célula, através de seu *Cell Global Identity (CGI)*. Ressalta-se que o Android não expõe diretamente as

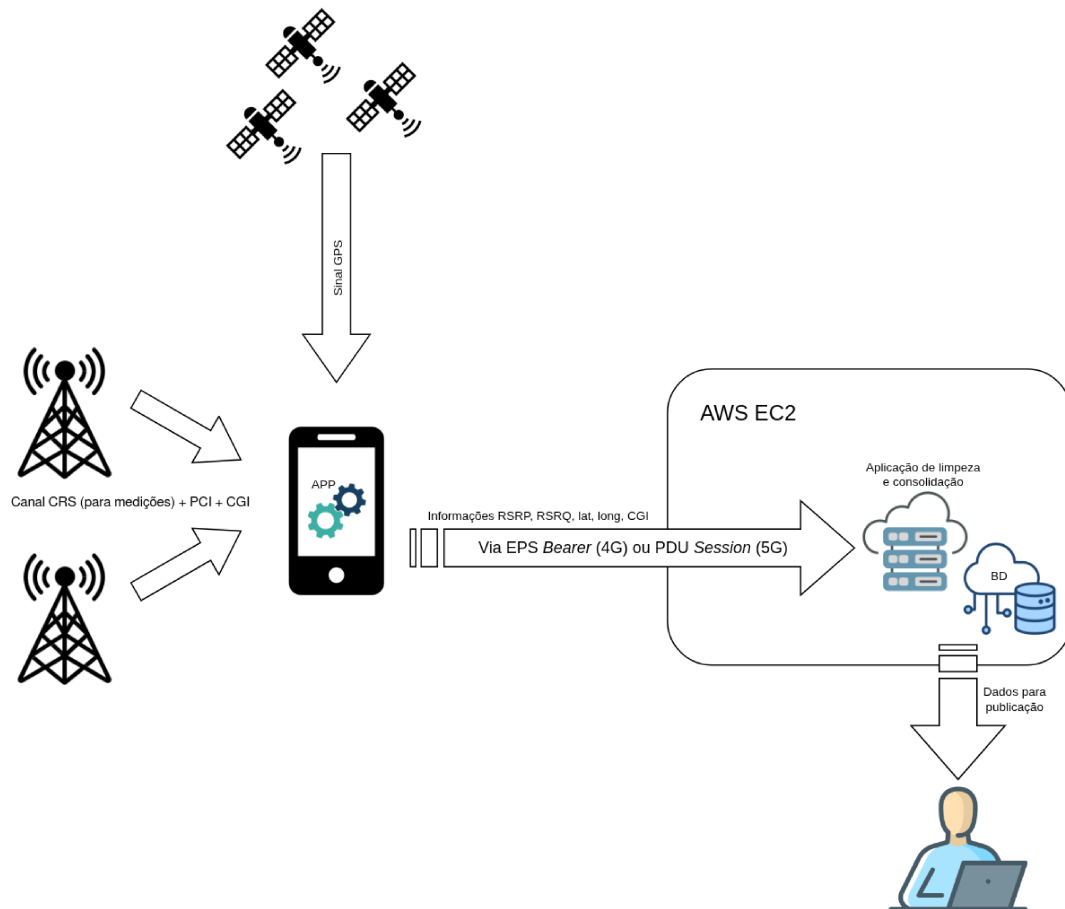


Figura 4.1 – Arquitetura simplificada da solução para aquisição de dados em *crowdsourcing*. Figura elaborada pelo autor.

mensagens de sinalização da camada RRC, tais como os *Measurement Reports* (MRs) definidos pelo 3GPP. Em vez disso, o sistema operacional fornece “fotos” instantâneas das condições da célula servidora e, em alguns casos, de células vizinhas, por meio de APIs de medições da camada de rádio. Essas informações foram obtidas da forma descrita a seguir.

As informações relacionadas à identificação da célula LTE são coletadas por meio da API de telefonia do Android, utilizando a classe `TelephonyManager` e o método `getAllCellInfo()`, que retorna uma lista de objetos do tipo `CellInfo`. Para células LTE, os parâmetros são acessados por instâncias da classe `CellInfoLte`, a partir das quais é possível extrair informações de identidade e configuração da célula servidora.

A identidade da célula é obtida por meio da classe `CellIdentityLte`, permitindo o acesso aos seguintes parâmetros:

- ❑ *Mobile Country Code* (MCC) e *Mobile Network Code* (MNC), que compõem o identificador da rede móvel, correspondendo ao código de identificação do país e da operadora, respectivamente;
- ❑ *Tracking Area Code* (TAC), que identifica a área de rastreamento da célula;

- ❑ *Cell Identity* (CI), correspondente ao identificador da célula LTE;
- ❑ PCI, que caracteriza a identificação física da célula;
- ❑ *E-UTRA Absolute Radio Frequency Channel Number* (EARFCN), que indica o canal de radiofrequência utilizado na interface LTE.

Esses parâmetros permitem a construção do identificador global da célula (CGI), além de obter outras informações, como seu PCI e a faixa de frequência em que ela opera.

Os indicadores de qualidade do sinal LTE são obtidos a partir da classe `CellSignalStrengthLte`, associada ao objeto `CellInfoLte`. Essa classe disponibiliza métodos específicos para acesso direto aos principais indicadores de desempenho da camada física, incluindo RSRP e RSRQ.

Finalmente, as coordenadas geográficas associadas a cada medição são obtidas por meio das APIs de localização do Android, que permitem o acesso à latitude e longitude do UE. A aquisição da informação é realizada utilizando os provedores de localização do sistema, como o `GPS_PROVIDER` ou outros mecanismos de obtenção de localização do dispositivo.

A coleta das medidas é realizada em cada dispositivo com o aplicativo instalado a cada 15 minutos, totalizando cerca de 96 coletas diárias por usuário, ficando esses dados gravados localmente na memória da UE. Após essas 96 coletas, o dispositivo envia as informações de forma consolidada em um arquivo *Comma-Separated Values* (CSV) via conexão de dados de usuário – normalmente *Evolved Packet System* (EPS) *bearer* no LTE ou *Packet Data Unit* (PDU) *session* no 5G – para o servidor AWS EC2 que consolida, organiza e armazena as informações na base de dados.

4.1.3 Desafios na Implementação

Durante o projeto da solução e a implementação da prova de conceito, alguns desafios foram identificados e tratados:

- ❑ Garantir que a coleta de dados seja contínua, mesmo após reinicializações dos dispositivos: para isso, o aplicativo foi configurado para ser inicializado automaticamente durante o processo de *boot* do sistema. Além disso, uma vez em execução, o aplicativo implementa um mecanismo de autoagendamento que assegura sua reinicialização automática caso o processo seja interrompido ou finalizado inesperadamente.
- ❑ Identificar quais informações de rede são acessíveis ao aplicativo: as informações são obtidas por meio de métodos de coleta implementados no aplicativo, os quais exploram interfaces e APIs disponibilizadas pelo sistema operacional móvel para acesso a parâmetros de localização, tempo e desempenho da rede celular. São coletados

RSRP, RSRQ, *timestamp* da coleta, posicionamento (latitude e longitude) do UE, EARFCN e CGI.

- ❑ Garantir que não haja impacto significativo no consumo de bateria: o aplicativo é ativado de forma periódica por meio de um mecanismo de agendamento, executa a coleta de dados uma única vez a cada ciclo e, em seguida, entra em estado de hibernação. Esse procedimento é repetido em intervalos regulares de 15 minutos, minimizando o uso de energia da bateria.
- ❑ Consumir o mínimo de armazenamento no dispositivo: os dados coletados são armazenados localmente em pacotes de pequeno porte, correspondentes a apenas 96 registros diários no formato CSV. Após esse período, o arquivo é transmitido para o servidor e apagado do dispositivo, evitando o acúmulo de dados locais.
- ❑ Implementar o aplicativo de modo que sua operação não consuma a cota de dados do cliente: para isso, é aplicada uma regra específica de *Access Point Name* (APN) na rede da operadora. Todo o tráfego de dados originado ou destinado ao endereço IP do servidor de coleta é tratado por uma política definida no elemento *Policy and Charging Rules Function* (PCRF), a qual isenta esse tráfego de tarifação.

É importante ressaltar que todas as informações coletadas são anônimas, ou seja, não são associadas a nenhuma forma de identificação do usuário. De qualquer forma, o acesso às informações descritas requer permissões pelo usuário, uma vez que o sistema Android impõe restrições de privacidade ao acesso a dados do dispositivo. Por fim, a disponibilidade de determinados parâmetros pode variar conforme a versão e personalização do SO e o fabricante do dispositivo, devendo tais limitações ser consideradas na análise e interpretação dos dados coletados.

4.1.4 Resultados Obtidos

O aplicativo foi instalado em nove celulares para os testes de prova de conceito. Foi observado que todos os desafios de implementação foram superados com as tratativas já apresentadas na Seção 4.1.3. Ao final, o formato da tabela de dados coletados é mostrado na Tabela 4.1. Para efeitos de anonimização dos dados, os campos CGI, EARFCN, latitude e longitude foram propositalmente alterados.

Durante o período de teste, foram coletadas em média 78 amostras diárias por usuário durante 14 dias, totalizando 9.828 amostras, um volume menor que o correspondente às 96 amostras diárias esperadas por usuário. Essa diferença deve-se a desligamentos pontuais dos dispositivos ou a períodos em que permaneceram sem bateria, além de amostras corrompidas (sem medições) que foram descartadas. A avaliação do consumo de bateria foi feita de forma indireta, por meio de entrevistas com os usuários. Em todos os casos, não houve percepção de redução do tempo útil da bateria ao longo do dia, o que era esperado,

Tabela 4.1 – Exemplos de amostras obtidas por meio do aplicativo para *crowdsourcing* com dados anonimizados.

timestamp	rsrp	rsrq	cgi	earfcn	technology	latitude	longitude
1759146800	-96	-10	7249901234567	1650	LTE	-15.922480	-31.948030
1759147702	-97	-11	7249901234567	1650	LTE	-15.921495	-31.949012
1759148604	-96	-11	7249901234567	1650	LTE	-15.921045	-31.949396
1759149506	-99	-13	7249901234567	1650	LTE	-15.920630	-31.949915
1759150408	-101	-14	7249901234567	1650	LTE	-15.920195	-31.950308
1759151310	-98	-12	7249901234567	1650	LTE	-15.919710	-31.950812
1759152212	-96	-11	7249901234567	1650	LTE	-15.919315	-31.951085
1759153114	-100	-13	7249901234567	1650	LTE	-15.918775	-31.951565
1759154016	-99	-12	7249901234567	1650	LTE	-15.918390	-31.952035

uma vez que a execução do aplicativo é realizada em *background*, não utiliza tempo de tela, não ativa nenhum recurso adicional no dispositivo e permanece em hibernação quando não está executando sua tarefa, a qual ocorre apenas uma vez a cada 15 minutos.

Ao final do período de testes, a prova de conceito foi conduzida com sucesso, atendendo aos objetivos propostos para avaliação da abordagem de aquisição de dados via *crowdsourcing*. As estratégias adotadas para a execução em segundo plano, controle do consumo de energia, limitação de consumo de armazenamento local e transmissão dos dados mostraram-se adequadas. Os testes mostraram a viabilidade prática da solução em um ambiente de rede real e em produção, incluindo a infraestrutura de *backend* implementada, que foi capaz de receber e armazenar de forma adequada o volume de dados gerado, embora esse volume tenha sido relativamente pequeno.

De uma forma geral, os testes demonstraram a efetividade da solução e que a abordagem proposta é tecnicamente adequada para a aquisição de medições georreferenciadas de intensidade e qualidade de sinal em redes móveis. Desta maneira, o método apresenta-se como uma alternativa prática viável à coleta baseada em *drive tests*.

4.2 Cenário Real Analisado e Aquisição dos Dados de Cobertura

Para a realização da avaliação experimental da metodologia ARCADE, é essencial que tenha-se dados reais de uma rede móvel. Isso porque, dada a complexidade e nuances características de um ambiente de RF de um sistema celular, que inclui propriedades físicas que se alteram de forma caótica como *fading* rápido e lento, imperfeições topológicas e morfológicas, reflexão e difração de sinal, dentre outras, o uso de dados simulados (ou baseados em previsões matemáticas) não confere uma legitimação da metodologia proposta. Uma vez que os dados de *crowdsourcing* apresentados na Seção 4.1 não estavam disponíveis pelo fato do sistema de coleta ainda estar restrito a poucos celulares (por se

tratar de uma prova de conceito não comercial), e pelo fato de a operadora em questão não possuir funcionalidade MDT implementada, foi necessária a coleta em campo de amostras da cobertura de um sistema real e operacional, o que foi feito através de um processo de DT minucioso.

O sistema analisado tem as seguintes características:

- ❑ Tecnologia: LTE Release 15
- ❑ Frequência: 3GPP banda B3, ou seja, faixa de 1.800 MHz em *Frequency Division Duplex* (FDD)
- ❑ Largura de banda: 10+10 MHz

Embora o sistema em si tenha outras faixas em utilização no 4G, além de outras tecnologias de RAN (incluindo GSM, UMTS e 5G), o estudo foi restrito ao LTE na faixa de 1.800 MHz FDD. Os dados dos demais sistemas e bandas foram desconsiderados. A razão é que faixas distintas têm características de cobertura diversas, exigindo parametrizações diferentes, de forma que, para manter o rigor da análise, uma única faixa foi escolhida para concentrar os estudos. Porém, nada impede que a mesma metodologia seja aplicada, também de forma isolada, a outras bandas ou mesmo outras tecnologias de acesso, tendo-se o cuidado, naturalmente, de se ajustar e calibrar a parametrização do processo. Análises de outras faixas ou tecnologias podem ser escopo para projetos futuros.

4.2.1 Equipamentos de Coleta de Amostras

O DT foi realizado com ajuda de terminais celulares categoria 20 (3GPP Cat-20) da marca Samsung com um software modificado que permite a coleta, georreferenciamento e armazenamento de dados de MRs. Para os dados úteis neste estudo, utilizou-se um terminal móvel em modo *RRC Connected Mode*, isto é, esse móvel estava todo o tempo fazendo *downloads* pela conexão celular.

Outra fonte de dados foi uma sonda do tipo *scanner* que mede o RSRP dos PCIs do ambiente de RF, limitado aos 16 níveis de RSRP mais fortes. Essa sonda utilizou uma antena externa que, calibrada com uma atenuação no circuito, tem um ganho de 0 dBi, conforme especificação 3GPP para dispositivos Cat-20. A especificação técnica do equipamento está disponível em (PCTEL, 2020), e inclui, referente à tecnologia LTE:

- ❑ Equipamento: PCTEL IBFlex© Scanning Receiver
- ❑ Suporte a todas as bandas 3GPP existentes;
- ❑ Larguras de banda: 1,4 / 3 / 5 / 10 / 15 / 20 MHz;
- ❑ Medidas LTE em SISO e MIMO 2x2 e 4x2;



Figura 4.2 – Foto do equipamento PCTEL IBflex. Fonte: (PCTEL, 2020).

- ❑ Taxa de medidas para LTE FDD: 50/s;
- ❑ Mínimo nível para detecção de RSRP: -140 dBm;
- ❑ Ganho de antena + atenuação no circuito: 0 dBi;
- ❑ Número máximo de PCIs: 16.

A foto da sonda *scanner* é mostrada na Figura 4.2.

Os dados dos MRs e os obtidos com a sonda, já devidamente georreferenciados em sua origem, foram agregados a uma mesma base de dados. Quando houve múltiplos valores para uma mesma coordenada geográfica, uma média dos mesmos foi utilizada, conforme já exposto na Seção 3.4.2.

4.2.2 *Cluster* de Células Analisado

O *cluster* a ser analisado corresponde a oito RBSs, cada uma composta de três células LTE em banda B3 (1.800 MHz FDD), num total de 24 células. A topologia é prevalentemente plana com inclinações suaves, com uma faixa de altitude que vai de um mínimo de 848 m de altitude acima do nível do mar até 769 m, ou seja, uma variação de 79 m entre os pontos mais baixo e mais alto. A morfologia predominante é suburbana, sendo as edificações casas ou pequenos edifícios em sua grande maioria, mas há casos isolados de prédios mais altos em alguns pontos dentro da área analisada. A Figura 4.3 mostra as estações que irão compor o *cluster* e a área total que serão analisados.

Observa-se que seis estações são mais externas no contexto do *cluster* enquanto que duas estão “envolvidas” por elas. Também há áreas onde não há amostras, pois trata-se de regiões não habitadas ou pouco densas. Este fato contribui para que o caso seja mais geral, ou seja, a generalização da cobertura das células deve extrapolar os dados de cobertura nas regiões sem amostra. A configuração é típica de uma cobertura padrão para bairros de morfologia urbana e suburbana de uma cidade de médio porte.

As RBSs têm as seguintes características:

- ❑ Fornecedor: Huawei;



Figura 4.3 – Mapa indicando a posição das estações componentes do *cluster* analisado. Figura elaborada pelo autor, usando mapa obtido no software Google Earth.

- ❑ Células por RBS: 3;
- ❑ Configuração *Multiple Input Multiple Output* (MIMO) 2T2R;
- ❑ Tipo de antenas: painéis verticais diretivos em configuração 2T2R para LTE B3, 65° de ângulo de meia-potência no plano horizontal e várias configurações de abertura e ganho para o plano vertical (são usadas diferentes antenas para diferentes células);
- ❑ Potência por rádio: 80 W por *branch*;
- ❑ Altura, azumite e configuração de *downtilt* elétrico ou mecânico das antenas são característicos de cada célula (não serão relevantes no contexto deste trabalho).

É bom reforçar que, para o escopo deste trabalho, os parâmetros dos sistemas irradiantes não são formalmente fornecidos. Logo, informações da potência de irradiação, ganhos

e altura de antenas e demais atributos da RAN, bem como características específicas de topologia e morfologia do terreno, não serão levadas em conta na avaliação experimental.

4.2.3 Hardware para Processamento dos Dados

Para a implementação do modelo na prática, incluindo os códigos, além da realização dos experimentos, utilizou-se um computador pessoal do tipo *laptop* da marca Hewlett-Packard (HP), modelo HP 15-FD0016NP, dedicado exclusivamente a este projeto. Embora não se trate de uma estação de alto desempenho, seus recursos computacionais mostraram-se suficientes para a execução dos códigos de análise, incluindo os baseados em RNAs. Apesar de não estar equipado com *hardware* dedicado para aceleração gráfica (como normalmente é desejável no treinamento de redes neurais), não houve comprometimento na efetividade do processamento dos dados, embora, como será mostrado, o tempo de processamento tenha sido significativo nessas etapas.

Porém, na aplicação prática envolvendo centenas ou milhares de células em geografias muito maiores, certamente haverá necessidade de se dimensionar o *hardware* de forma a se ter uma maior eficiência na execução dos processos.

A Tabela 4.2 apresenta brevemente as principais especificações do equipamento utilizado.

Tabela 4.2 – Configuração do *hardware* para processamento dos dados.

Parâmetro	Valor
CPU	13 th Gen. Intel Core i5-1335U (10 cores) 1,30 GHz
Sistema Operacional	Ubuntu 22.04.5 LTS 64 bits
Memória RAM	DDR4 (SO-DIMM) 16 GB
Disco Rígido	SSD 512 GB

4.3 Avaliação Experimental de Predição de Cobertura

Nesta seção, será demonstrado um caso específico e detalhado da aplicação dos módulos SAM (aumento de amostragem com GPSK) e MTGM (extrapolação dos dados com RNA) a uma única célula, com a finalidade de extrapolar-se sua cobertura sobre uma grande área geográfica.

4.3.1 Aumento da Amostragem com GPSK

Como forma de validação da metodologia de aumento de amostragem através de GPSK, foi realizado todo o processo para uma célula, de PCI 193. Porém, a coleta e análise de dados foram realizadas em uma área total de aproximadamente 558 km², medindo 22.145 metros na direção da latitude e 25.184 metros na direção da longitude. Essa

área engloba uma grande cidade, composta por subáreas com morfologias urbana densa, suburbana e rural. Nessa abordagem, dividiu-se a área total em uma grade de elementos com dimensões de 50 m x 50 m, resultando em 443 segmentos ao longo da latitude e 504 segmentos ao longo da longitude, totalizando aproximadamente 223.272 elementos de grade. Para converter coordenadas geográficas em distâncias em metros e alocar os elementos da grade, utilizou-se o método de Haversine (DAUNI et al., 2019). Cada elemento de grade pode estar associado a uma (e apenas uma) amostra de sinal RSRP por PCI, embora a maioria dos elementos não possua medições. Todo o trabalho foi baseado em dados provenientes de uma operadora real de rede móvel.

Para a célula analisada, foram coletadas 507 amostras de RF na área total, incluindo anomalias e outliers. Isso significa que há amostras para apenas 0,22% dos elementos da grade. Contudo, a maioria das amostras está localizada dentro da área de interesse para análise, especificamente na área de cobertura da célula (A_N). Por outro lado, fora dessa área (A_A), há apenas uma representação limitada, relacionada principalmente a anomalias de cobertura. De fato, para esta célula em particular, foram registradas apenas três anomalias e um outlier.

A extrapolação dos dados com GPSK e o cálculo dos valores de RSRP foi realizado para todos os elementos da grade na área total A_T , utilizando Python com a biblioteca *scikit-learn*. Os dados de amostragem foram lidos a partir de um arquivo CSV, já contendo entradas normalizadas para latitude, longitude e RSRP, com $lat_{min}/long_{min}$ definidos como 0, $lat_{max}/long_{max}$ como 1, e RSRP normalizado entre -120 dBm (equivalente a 0) e -50 dBm (equivalente a 1). Valores abaixo de -120 dBm foram considerados como 0, e valores acima de -50 dBm como 1, tanto para entradas quanto para saídas.

Utilizou-se 80% das amostras para treinamento, 10% para teste e 10% para validação, uma prática padrão em modelos de aprendizado de máquina. O kernel utilizado foi o RBF, com os seguintes parâmetros iniciais:

- C (Constante do *Kernel*): 0,1;
- Intervalo de ajuste para C : 10^{-4} a 10
- Comprimento de escala l (que regula o desvio padrão do *kernel* RBF): 0,1;
- $n_restarts_optimizer$ (número de reinicializações do otimizador com condições iniciais diferentes): 10.

A constante C , chamada “Constante do *Kernel*”, usada como uma constante de produto, é implementada pela biblioteca *scikit-learn* de modo que o valor do *kernel* é multiplicado por ela. Ela representa a amplitude do processo gaussiano, escalando a magnitude das variações esperadas na função modelada (scikit-learn developers, 2024a). Um valor inicial baixo de C é considerado conservador (assume-se que o RSRP não varia muito

antes de o otimizador calcular os dados). O otimizador calibra C dentro do intervalo de ajuste durante o treinamento, assim, o valor inicial tem impacto sobretudo no tempo de convergência, mas não no resultado final. O comprimento de escala l regula o alcance espacial da correlação: quanto maior a distância entre dois pontos em relação a l , menor o grau de influência de um sobre o outro. A escolha de 0,1 representa 10% da escala total do espaço, em coordenadas normalizadas. Novamente, o otimizador irá ajustar esse valor, sendo que a definição inicial apenas define o ponto de partida da busca (scikit-learn developers, 2024b).

Adicionalmente, o parâmetro *alpha* (α) controla o termo de regularização adicionado à matriz de covariância para modelagem do ruído nos dados, representando o nível de variância do ruído. Um α maior implica maior suavização do modelo, ao considerar maior incerteza nos dados (scikit-learn developers, 2023). Nesta implementação, assume-se que cada medição de RSRP é a soma do sinal verdadeiro com um ruído branco de variância α . Na prática, esse ruído simula imprecisões na coleta de dados, variações de desvanecimento rápido de multipercurso de rádio ou até erros na quantização do circuito do receptor. Na abordagem do *scikit-learn*, a matriz de covariância K é modificada pela adição do termo αI à sua diagonal (sendo I a matriz identidade), e K_α é então definida como:

$$K_\alpha = K + \alpha I \quad (45)$$

Essa matriz K_α é usada no fator K^{-1} , modificando $\mu(x')$ para a seguinte forma:

$$\mu(x') = K_*^T K_\alpha^{-1} Y \quad (46)$$

Foram realizados diversos experimentos, e exceto pelo parâmetro α , não houve necessidade significativa de ajuste dos demais parâmetros com relação aos adotados inicialmente. O parâmetro α , por sua vez, exigiu calibração para a obtenção de maior eficiência do modelo. Utilizou-se as métricas RMSE e R^2 para essa calibração, buscando maximizar esses indicadores nas saídas observadas para se encontrar o valor ótimo do parâmetro. O valor de R^2 representa o coeficiente de determinação e indica o quão bem os dados práticos observados são replicados pelo modelo, variando de 0 (o modelo não explica nenhuma variância dos dados de saída) a 1 (o modelo explica 100% da variância dos dados de saída). O RMSE representa a raiz quadrada do erro médio quadrático entre os valores previstos e os valores reais.

Os experimentos determinaram que $\alpha = 0,00145$ produziu os melhores resultados em termos de R^2 e RMSE, conforme ilustrado na Figura 4.10.

O valor alvo de R^2 é considerado arbitrário e fortemente dependente do contexto da aplicação; no entanto, o valor 0,67 – que indica que aproximadamente 67% da variância da variável dependente é explicada pelas variáveis independentes do modelo – é considerável como plenamente aceitável para os propósitos deste trabalho. De forma análoga, um

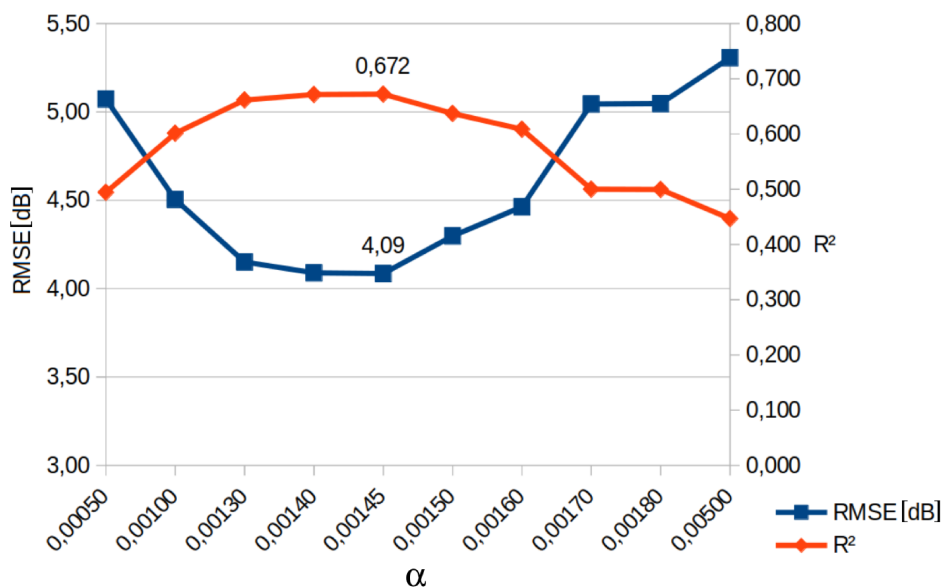


Figura 4.4 – Variação de R^2 e RMSE com α . Eixo X não é linear. Figura elaborada pelo autor

valor de RMSE de 4 dB é compatível com o erro de medição esperado para variações físicas de sinal devido ao desvanecimento (ou *fading*) rápido em ambientes urbanos e suburbanos (na faixa de 4 a 8 dB nessas áreas) (RAPPAPORT, 2009) (SIVERTSEN; LIOU; MICHELSON, 2010).

Na Figura 4.5, é possível verificar que o mapa gerado com α ajustado para 0,00145 apresenta coerência visual com os dados de saída esperados em comparação com outros valores. Para melhor visualização, a parte relevante da área A_T foi recortada nas imagens.

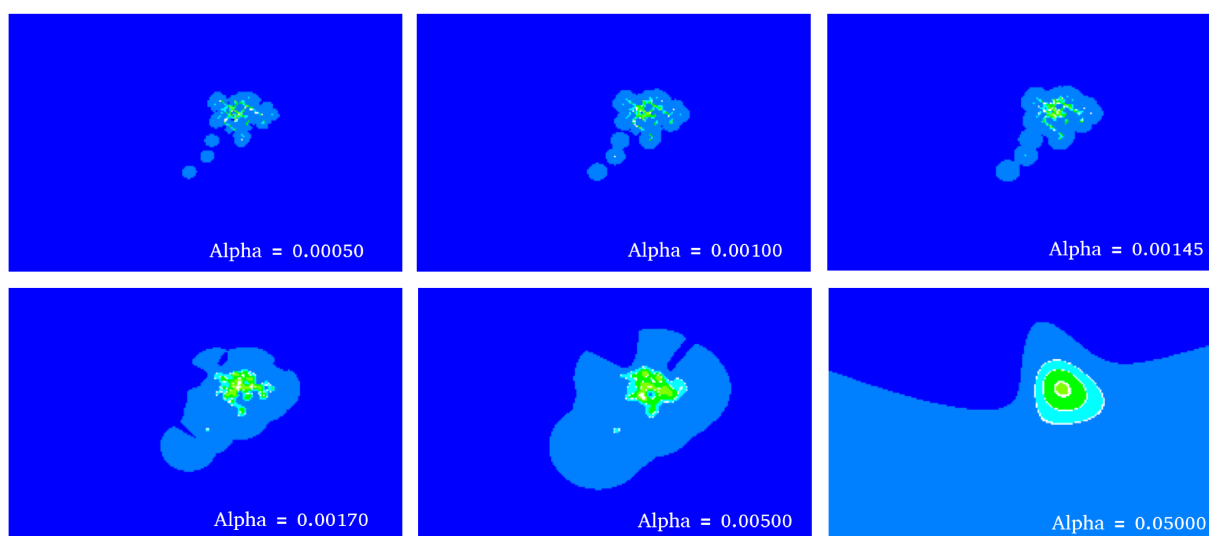


Figura 4.5 – Mapas de cobertura usando $\alpha = 0,00050$; $0,00100$; $0,00145$; $0,00170$; $0,00500$ e $0,05000$. Figura elaborada pelo autor.

4.3.2 Generalização com RNA (Módulo MTGM)

Por fim, a última etapa para se alcançar o mapeamento completo da cobertura da célula em toda a área de interesse A_T consiste na modelagem e no treinamento de uma RNA que represente adequadamente os dados coletados e extrapolados, denominado no ARCADE como módulo MTGM.

A RNA foi modelada experimentalmente em Python, utilizando a biblioteca *TensorFlow* com os seguintes hiperparâmetros:

- ❑ Dois neurônios de entrada (latitude e longitude normalizadas);
- ❑ Um neurônio de saída (valor de RSRP normalizado);
- ❑ Três camadas ocultas, cada uma com 20 neurônios;
- ❑ Função de ativação nas camadas de entrada e ocultas: *Rectifier Linear Unit* (ReLU);
- ❑ Função de ativação na camada de saída: linear;
- ❑ Particionamento do conjunto de dados em:
 - 80% para treinamento;
 - 10% para teste;
 - 10% para validação.
- ❑ Método para inicialização dos pesos: Glorot;
- ❑ Otimizador: Adam
- ❑ Taxa de aprendizado: 0,0001
- ❑ Função de perda: *Mean Squared Error* (MSE)

Esses hiperparâmetros foram definidos por meio de um processo sistemático, no qual se variaram a arquitetura da rede (número de camadas ocultas e de neurônios por camada), funções de ativação e demais parâmetros de treinamento, avaliando-se os resultados pelo RMSE e pelo coeficiente R^2 . Por não contribuir com o objeto principal desta tese, o detalhamento desse processo de calibração foi omitido, sendo apresentada apenas a configuração que produziu os melhores resultados.

Após a geração dos dados, o RMSE foi calculado para todas as amostras disponíveis, resultando em 4,02 dB para A_N e 0,31 dB para A_A . A avaliação foi realizada separadamente nessas áreas, uma vez que o RMSE em A_A deve ser significativamente inferior, dado que a grande maioria das amostras apresenta valor de -120 dB (ou 0, quando normalizado). Em A_N , o erro encontra-se dentro de margens aceitáveis, ainda considerando

a variabilidade na medição do sinal causada pelo desvanecimento rápido (RAPPAPORT, 2009) (SIVERTSEN; LIOU; MICHELSON, 2010).

Por fim, o mapa de cobertura previsto apresenta coerência visual com as medições reais, refletindo com precisão as anomalias detectadas nas três amostras indicadas em A_A , conforme ilustrado na Figura 4.6. Novamente, para melhor visualização, partes irrelevantes de A_T foram recortadas nas imagens.

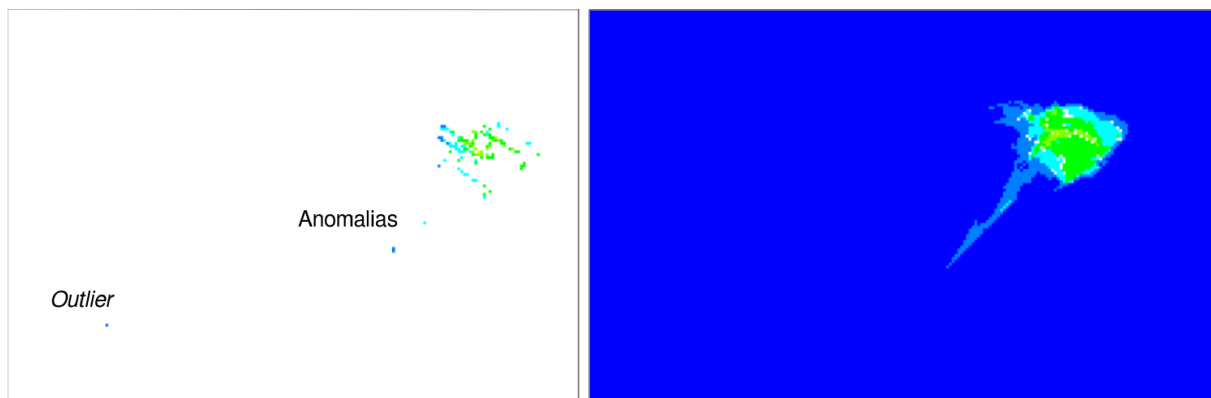


Figura 4.6 – Mapas mostrando dados disponíveis e cobertura com níveis preditos de RSRP. Figura elaborada pelo autor.

No mapa de predição gerado por meio da RNA, observa-se uma região caracterizada por uma anomalia da cobertura da célula, inferida pelas amostras anômalas. Embora, por motivo de clareza, o mapa não contenha informações sobre o arruamento da cidade, essa anomalia se apresenta em uma orientação que coincide precisamente com a direção das vias urbanas. Essa coincidência sugere fortemente uma coerência entre os resultados preditos pelo modelo e o comportamento esperado da propagação do sinal em ambiente urbano, evidenciando uma consistência do observado no ambiente real com o obtido pela metodologia ARCADE.

4.4 Aumento de Dados para o *Cluster*

Conforme mencionado na Subseção 4.2.2, a título de experimento e validação da metodologia, será analisado um *cluster* de oito RBSs, totalizando 24 células que cobrem uma região que engloba alguns bairros de características suburbanas, além de áreas rurais sem população ou amostras.

A quantidade de amostras, incluindo anomalias e *outliers* pode ser vista na Tabela 4.3

Tabela 4.3 – Quantidade de amostras, anomalias e *outliers* por PCI

PCI	Amostras (Sem <i>Outliers</i>)	Anomalias	<i>Outliers</i> (Removidos)
9	1.339	20	0
10	2.426	60	0
11	1.452	4	3
165	3.508	2	35
166	887	0	5
167	1.132	47	0
243	1.676	6	11
244	2.818	32	4
245	1.153	23	2
246	704	5	9
247	882	11	2
248	1.473	53	3
252	1.258	2	1
253	802	19	0
254	606	23	0
333	1.144	1	1
334	1.757	0	5
335	456	2	4
395	552	9	0
396	927	10	2
397	1.495	11	8
404	2.499	122	1
405	993	13	1
406	2.247	46	3
Total	34.186	521 1,52%	100 0,29%

As amostras podem ser visualizadas de forma georreferenciada na Figura 4.7, em duas representações: a primeira mostra a distribuição das amostras por PCI (onde cada cor representa uma célula distinta) e a segunda, por classificação da amostra, em normal, anomalia ou *outlier*.

O mapa de intensidade de cobertura (RSRP) pode ser visto na Figura 4.8, e mostra que a potência recebida não é homogênea em toda a área. Nessa visão, os *outliers* já foram excluídos.

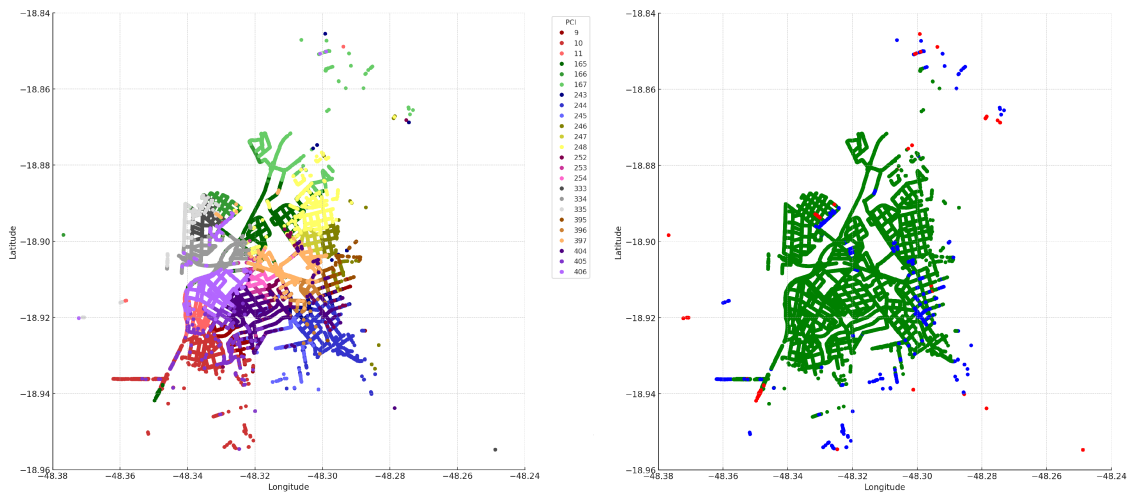


Figura 4.7 – Duas representações geográficas das amostras do *cluster*, a primeira caracterizada pelo PCI, a segunda por classificação de normalidade. Figura elaborada pelo autor.

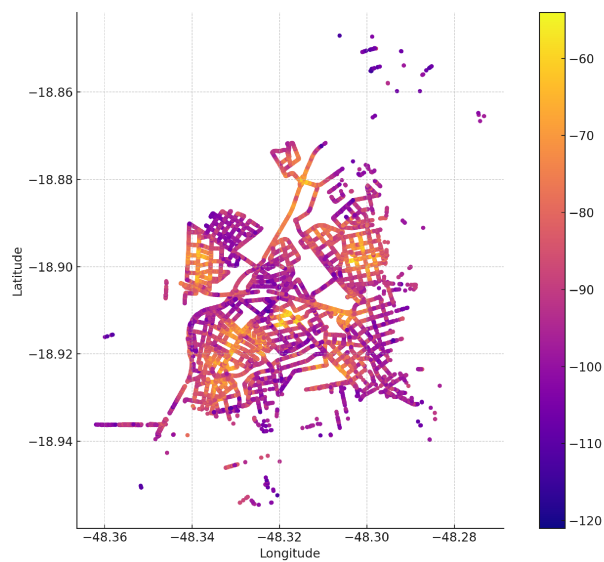


Figura 4.8 – Nível de sinal (RSRP) por amostra na geografia de análise. Figura elaborada pelo autor.

As amostras individuais por PCI podem ser vistas na Figura 4.9.

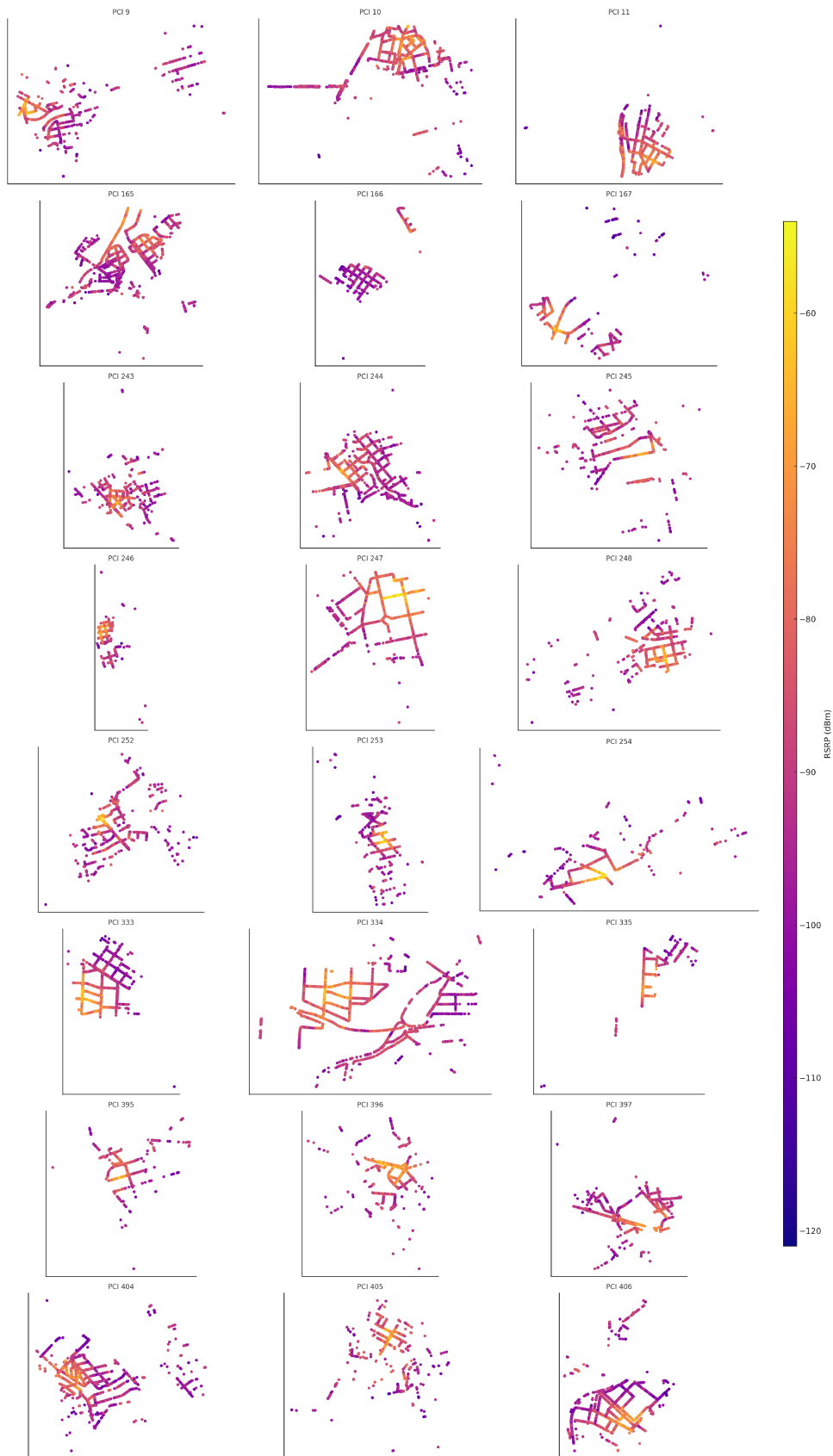


Figura 4.9 – Amostras por PCI individual, com representação de RSRP. Figura elaborada pelo autor.

4.4.1 Extrapolação de Dados com GPSK

As amostras mostradas foram submetidas ao aumento de dados, na mesma metodologia descrita na Subseção 4.3.1. No caso do *cluster*, foi utilizado um valor único de α para todas as extrapolações com GPSK. Isso é motivado pela necessidade de uniformidade desse processo de extrapolação. Caso cada célula utilizasse seu próprio α ótimo, as superfícies de RSRP apresentariam graus de suavização distintos entre si, podendo resultar em inconsistências em termos de resolução espacial. A determinação do α que maximiza o R^2 médio e minimiza o RMSE médio sobre todas as células foi um critério de melhoria global que equilibra a qualidade do ajuste do parâmetro com a homogeneidade do modelo em todo o *cluster*.

Para essa determinação do parâmetro α a ser usado, adotou-se um procedimento de avaliação de R^2 e RMSE para cada célula no intervalo de $\alpha = 0,001$ a $0,01$, com passo de $0,0002$, tendo sido calculados ambos os valores para cada PCI através da extrapolação pelo GPSK. Em seguida, calculou-se a média dessas métricas sobre todas as células, obtendo-se assim curvas de desempenho agregado em função de α , conforme ilustrado na Figura 4.10.

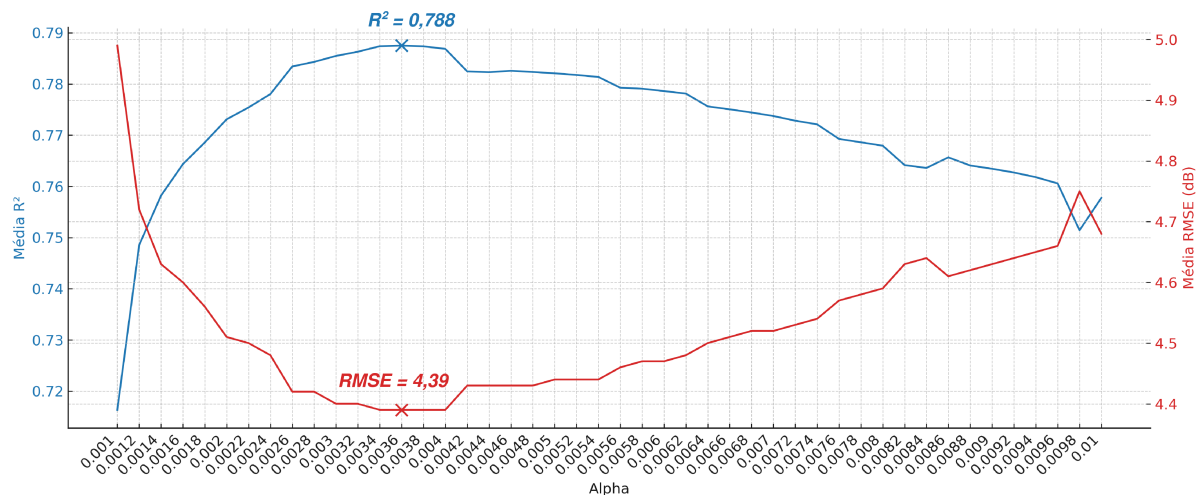


Figura 4.10 – Variação de RMSE e R^2 com α . Figura elaborada pelo autor.

A análise das curvas nessa figura mostra que, de forma consolidada, o α de melhor desempenho é igual a $0,0036$, onde encontra-se um R^2 máximo de $0,788$ e um RMSE mínimo de $4,39$ dBm (ou $0,06265$ quando normalizado). Desta forma, usando-se $\alpha = 0,0036$, os dados foram extrapolados de modo a se obter uma tabela completa de valores de RSRP por PCI em cada elemento da matriz que representa A_T . Os mapas por PCI podem ser vistos agora na Figura 4.11.

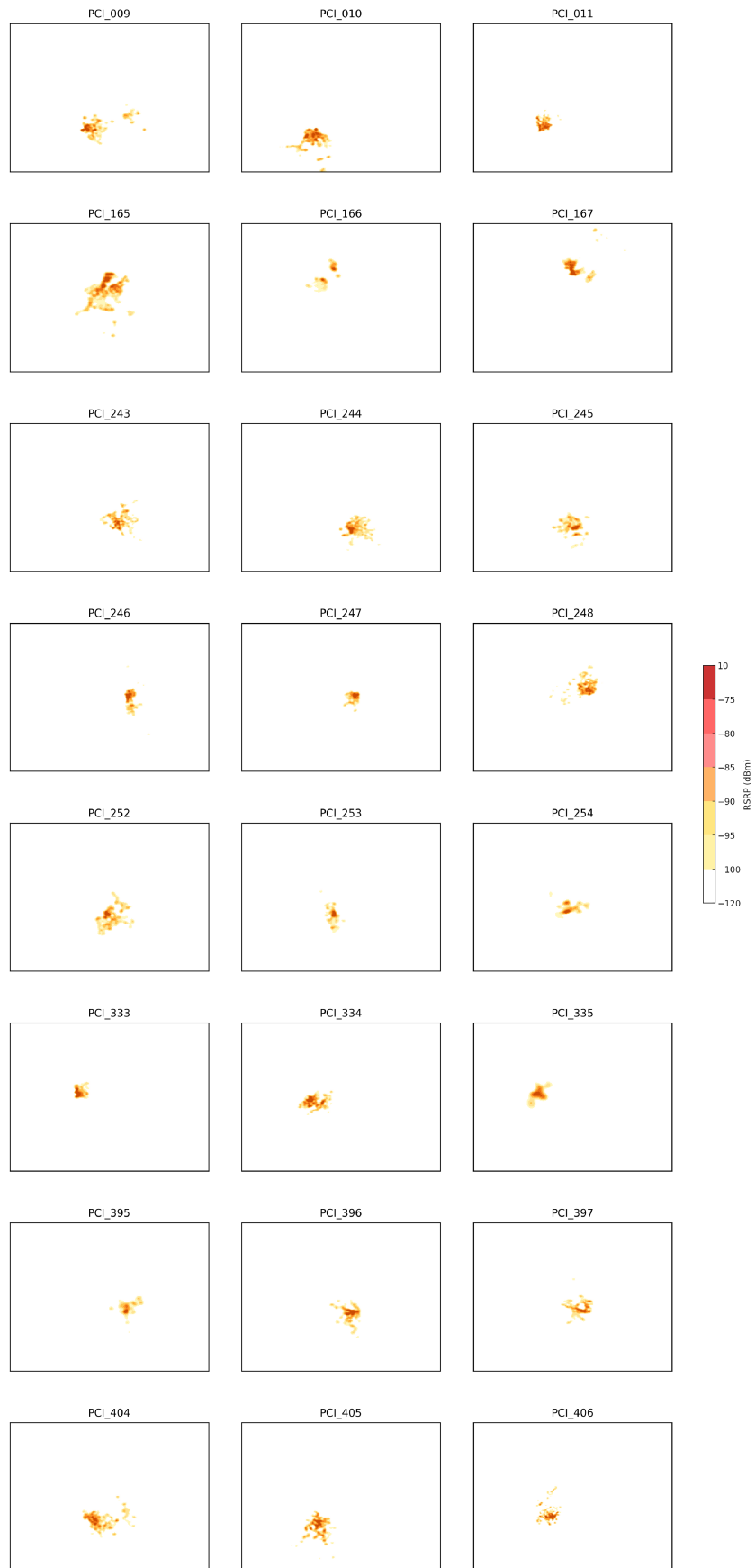


Figura 4.11 – Mapas de RSRP por PCI após aumento de dados por GPSK. Figura elaborada pelo autor.

O próximo passo para a obtenção dos mapas de cobertura para todas as células do *cluster* é o aumento dos dados de cada PCI, de forma a estabelecer condições de contorno e realçar regiões de anomalias para que as extrapolações de cobertura via RNA sejam condizentes com as condições reais observadas na prática. Conforme já descrito nas Subseções 3.4.1, 3.4.2 e 3.5.1, a metodologia consiste em identificar a área de normalidade A_N e de anormalidade A_A para cada célula, manter as amostras dentro das A_N e aumentar as amostras nas A_A . Esse aumento é realizado incorporando-se aos dados coletados amostras extras das coberturas extrapoladas via GPSK a intervalos correspondentes à metade da distância mínima (nas direções de latitude e longitude) entre a anomalia mais próxima da célula e o limite da A_N (d_{\min}). A Figura 4.12 demonstra as delimitações das áreas de normalidade (retângulos) e os baricentros (pontos) para cada PCI dentro do *cluster* analisado.

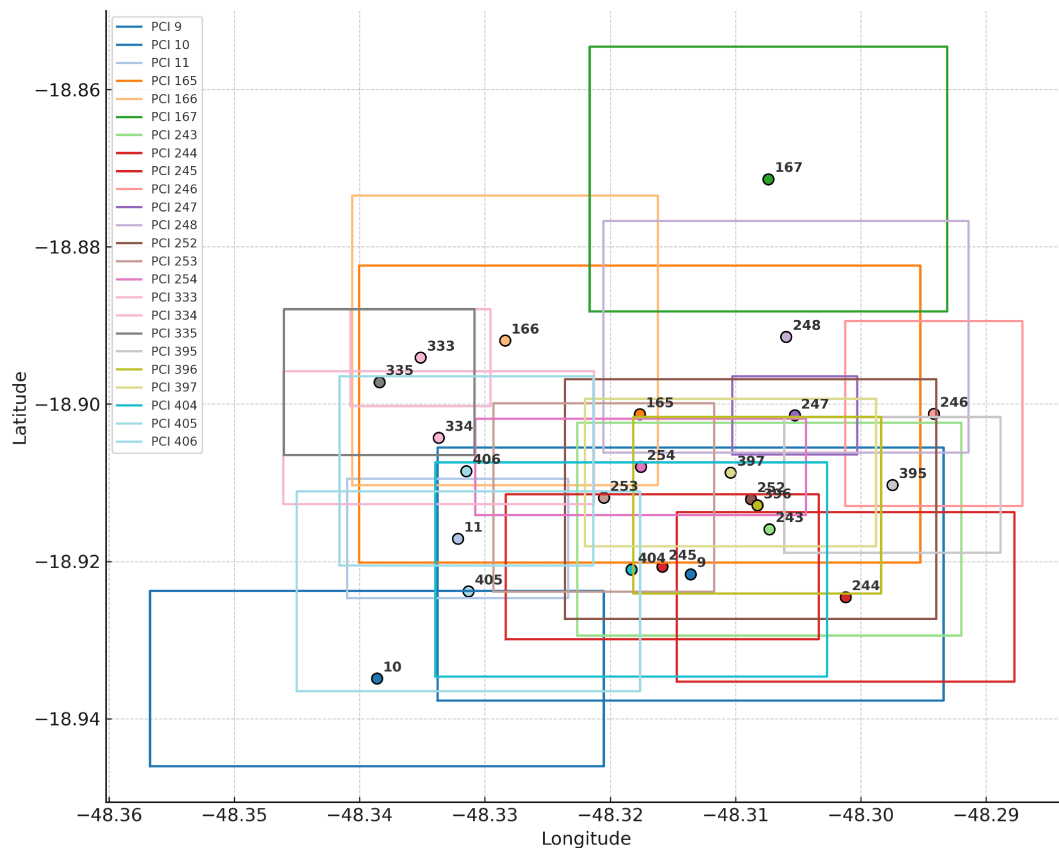


Figura 4.12 – Áreas de Normalidade por PCI. Os pontos representam os baricentros das respectivas células e o retângulo, as respectivas áreas de normalidade. Figura elaborada pelo autor

A distância d_{\min} foi então calculada de forma unificada para o *cluster*. Embora d_{\min} pudesse ser definido individualmente por célula, um valor global foi definido para todo o *cluster* – de forma análoga ao que foi feito com α –, para garantir uniformidade no processo de aumento de amostragem e evitar quaisquer vieses entre células, mantendo-se a mesma granularidade em todas elas. Para reforçar a robustez e a aplicabilidade da

metodologia no experimento, d_{\min} foi calculada para as três anomalias mais próximas em cada caso, $d_{\min1}$, $d_{\min2}$ e $d_{\min3}$, respectivamente, obtendo-se os valores mostrados na Tabela 4.4.

Tabela 4.4 – Distâncias mínimas associadas às três anomalias mais próximas à A_N por PCI

PCI	$d_{\min1}$ (m)	$d_{\min2}$ (m)	$d_{\min3}$ (m)
9	20,0	47,6	54,8
10	4,2	77,0	90,7
11	676,8	682,4	685,7
165	659,0	862,7	–
167	4,5	11,1	24,5
243	164,8	170,3	217,1
244	1,1	8,4	21,2
245	73,7	102,4	139,1
246	163,4	170,7	182,2
247	0,7	2,0	10,3
248	3,2	8,4	11,7
252	76,9	624,4	–
253	2,2	7,8	15,0
254	1,1	9,5	11,7
333	529,9	–	–
335	1056,4	1074,3	–
395	6,0	116,9	122,2
396	27,3	116,9	124,1
397	45,5	96,7	102,3
404	4,2	11,6	18,0
405	16,0	24,5	27,8
406	21,3	26,7	31,2

Como a granularidade do espaço amostral, ou seja, o tamanho do elemento na matriz de análise, é de 50 m por 50 m, distâncias de anomalias muito próximas às A_N devem ser descartadas, por não fazerem sentido no contexto de aumento de amostragem. Anomalias com d_{\min} de 100 m, por exemplo, resultariam em amostragens em elementos dentro ou adjacentes à área de normalidade, uma vez que esse valor é dividido por dois para se determinar a distância de amostragem. Desta forma, considerando que o elemento da grade tem uma resolução espacial de 50 m, valores de $d_{\min} < 4\Delta$ (onde $\Delta = 50$ m) foram descartados, pois resultariam em amostras redundantes dentro ou nas proximidades da região A_N . Após isso, foi calculada a mediana de d_{\min} , encontrando-se o valor de 680 m. Arredondando-se esse resultado, tem-se:

$$d_{\min} \approx 700 \text{ m} \quad (47)$$

$$\frac{d_{\min}}{2} \approx 350 \text{ m} \quad (48)$$

Ou seja, na área de anomalia de cada PCI, serão obtidas amostras extras a cada 49 elementos, uma vez que a distância de 350 m equivale linearmente a sete elementos.

4.4.2 Obtenção de Cobertura com RNA

Conforme o processo demonstrado na Seção 4.3.2, a metodologia baseada em RNA foi estendida a todas as 24 células pertencentes ao *cluster* do estudo. Os mapas são mostrados na Figura 4.13

Em comparação com os mapas mostrados na Figura 4.11, observa-se uma maior continuidade nos padrões de cobertura e uma suavização das transições entre regiões adjacentes, com a manutenção das características espaciais originais do sinal. Além disso, há uma redução dos artefatos advindos de anomalias de cobertura e interpolações puramente gaussianas.

A Tabela 4.5 mostra o desempenho geral do processamento em RNA de cada PCI em termos de tempo e métricas de erro.

Os tempos de processamento das RNAs são significativos, com uma média de 08 horas e 11 minutos por célula, usando o hardware descrito em 4.2.3 (considerado muito básico para esse tipo de aplicação), sendo muito superior ao tempo de processamento dos demais métodos empregados no ARCADE, como o GPSK e o CMA-ES.

O parâmetro R^2 foi geralmente bastante alto, sendo que o valor mínimo encontrado foi de 0,863, com uma média geral de 0,925 e baixo desvio padrão. Isso significa, na prática, que 92,5% da variabilidade dos valores reais de RSRP é explicada pelas predições do modelo da RNA, e indica que os padrões de atenuação e propagação presentes nos dados reais foi reproduzido de forma consistente.

O valor médio de RMSE foi de 4,73 dB. Conforme já mencionado, esse valor está muito condizente com margens de erro, considerando-se a presença de desvanecimento rápido de sinal intrínseco a qualquer forma de coleta de dados (RAPPAPORT, 2009) (SIVERTSEN; LIOU; MICHELSON, 2010).

4.5 Aprimoramento da Qualidade de Cobertura do *Cluster*

O processo de melhoria da qualidade de cobertura do *cluster* segue a metodologia apresentada na Seção 3.7, a partir da avaliação da função $\tilde{J}(P)$ através do método CMA-ES. A função objetivo, conforme já definida, é:

$$\tilde{J}(P) = \frac{1}{N} \sum_{i=1}^N \left[w_{\text{cov}} C_i(P) + w_{\text{dom}} D_i(P) \right] - w_{\text{SDI}} V_{\text{SDI}}(P) - \lambda R(P) \quad (49)$$

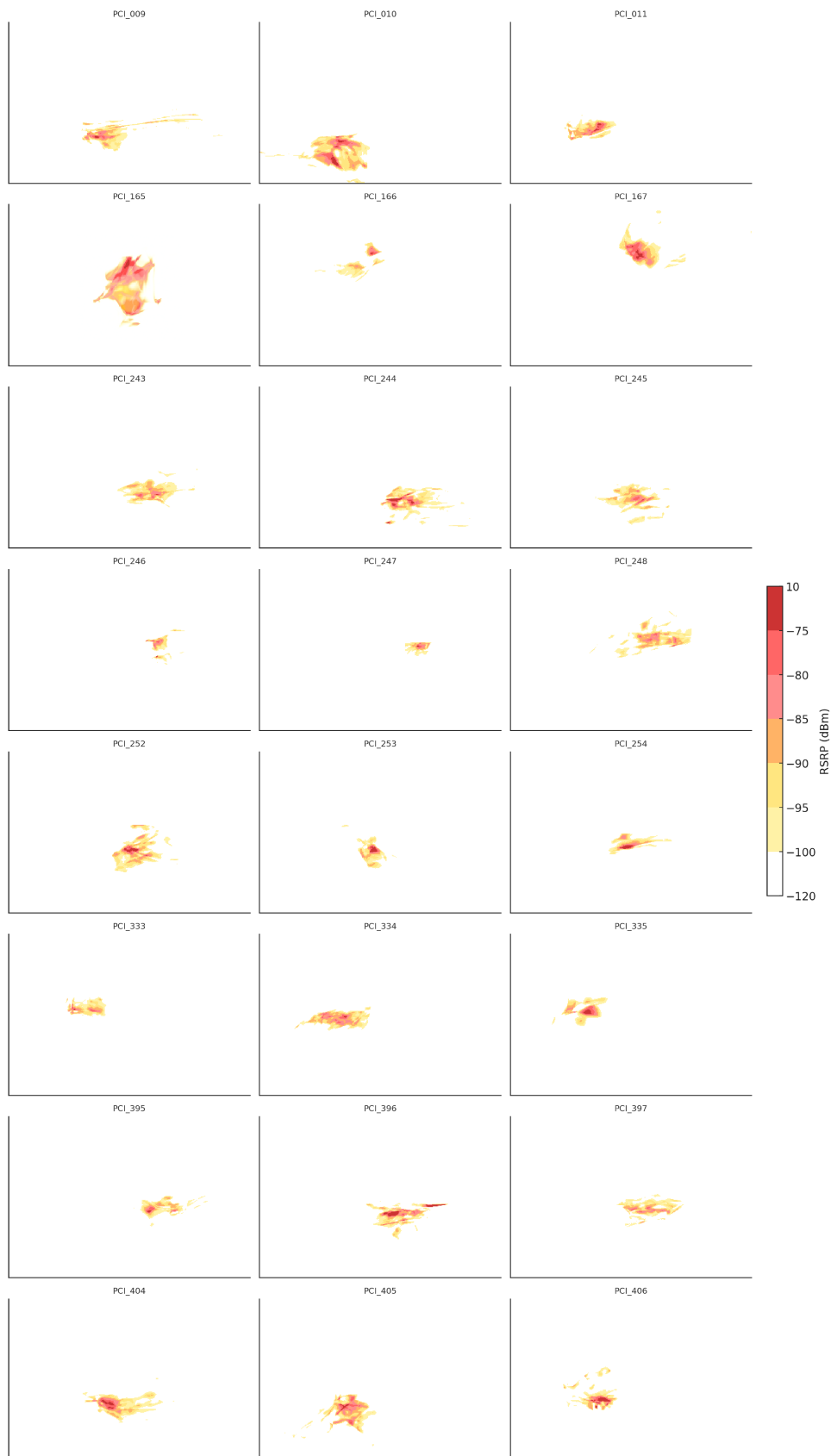


Figura 4.13 – Mapas de RSRP por PCI após aumento de dados por RNA. Figura elaborada pelo autor.

Tabela 4.5 – Desempenho da RNA por PCI em termos de tempo de processamento e métricas de erro

PCI	Tempo (hh:mm:ss)	R^2	MSE (normalizado)	RMSE (dB)
009	07:14:44	0,878	0,00862	6,50
010	10:50:56	0,939	0,00384	4,34
011	08:39:58	0,978	0,00186	3,02
165	10:24:54	0,939	0,00395	4,39
166	07:03:47	0,924	0,00278	3,69
167	06:32:19	0,961	0,00308	3,89
243	09:13:22	0,913	0,00650	5,65
244	12:04:26	0,892	0,00604	5,44
245	07:36:25	0,934	0,00392	4,38
246	06:32:04	0,863	0,00858	6,48
247	07:03:46	0,947	0,00393	4,39
248	08:40:05	0,937	0,00429	4,59
252	08:11:09	0,912	0,00587	5,37
253	06:51:55	0,948	0,00278	3,69
254	06:24:06	0,949	0,00264	3,60
333	07:59:00	0,954	0,00296	3,81
334	09:16:53	0,909	0,00605	5,45
335	06:07:26	0,906	0,00436	4,62
395	06:21:53	0,864	0,00561	5,24
396	07:14:24	0,938	0,00506	4,98
397	08:47:31	0,934	0,00392	4,39
404	09:58:26	0,926	0,00501	4,95
405	06:36:27	0,902	0,00786	6,21
406	10:46:55	0,941	0,00414	4,50
Média	08:11:22	0,925	0,00474	4,73
Desvio Padrão	01:41:07	0,029	0,00185	0,92

O primeiro passo para a avaliação de $\tilde{J}(P)$ consiste na determinação dos limiares de RSRP, do critério de dominância entre célula servidora e interferente, bem como dos pesos que compõem a estrutura da equação, notadamente w_{cov} , w_{dom} , w_{SDI} e λ . Esse processo é conduzido por meio de uma calibração contínua e sistemática desses parâmetros, conforme descrito na próxima subseção.

Também é importante ressaltar que o método CMA-ES é, por natureza, formulado como um algoritmo de minimização, enquanto, por definição, a função $\tilde{J}(P)$ deve ser maximizada. Para contornar essa diferença conceitual, na implementação computacional considera-se a minimização do negativo da função objetivo, isto é, minimiza-se $-\tilde{J}(P)$, o que é matematicamente equivalente à maximização de $\tilde{J}(P)$.

4.5.1 Definição de Parâmetros e Calibração de Pesos de $\tilde{J}(P)$

A seguir, são apresentados os parâmetros utilizados para a avaliação da função objetivo e a metodologia adotada para a calibração dos pesos. Em primeiro lugar, o *cluster* analisado consiste em 24 células, então, naturalmente, define-se $N = 24$.

Em seguida, são definidos os limiares mínimo e máximo de RSRP, para o cálculo de $C_i(P)$ e limiares de dominância para o cálculo de $D_i(P)$. Para os valores de limiares de RSRP, tanto observações práticas em campo quanto valores apresentados em (HOLMA; TOSKALA, 2011) convergem em níveis de $S_{\text{min}} = -105$ dBm e $S_{\text{max}} = -75$ dBm, respectivamente. Além disso, o limiar de dominância entre as células servidora e interferente dominante foi fixado em $\Delta_{\text{ref}} = 9$ dB, com referência empírica no valor de C/I associado à manutenção da modulação 16QAM, correspondente ao ponto médio do intervalo entre 6 e 12 dB, conforme (HOLMA; TOSKALA, 2011), sendo esse valor também muito utilizado na prática pela indústria. Δ_{min} é intuitivamente igual a 0 dB. A definição do parâmetro σ_P é arbitrária e foi fixada em 4 dB, sendo um valor considerado equilibrado que permite variações moderadas de potência e penaliza ajustes excessivamente elevados.

Para efeitos práticos, os valores dos elementos do vetor de ajuste de potência ΔP_k por célula foram limitados ao intervalo de -9 dB a +9 dB, compatível com *offsets* normalmente viáveis para ajustes da potência associada ao CRS de cada célula. Essa potência é a efetivamente refletida nas medições de RSRP realizadas pelas UEs e utilizadas para procedimentos de resseleção e *handover* entre células, e um aumento ou redução excessivos da mesma pode comprometer a aplicabilidade da solução encontrada. Na prática, operadores utilizam valores entre +12 e +22 dBm (tipicamente +18,2 dBm, ou ≈ 66 mW) para *referenceSignalPower*, sendo que o intervalo permitido pela padronização vai de -60 dBm a +50 dBm (3GPP, 2024e). Aplicando-se o intervalo de correção considerado ao valor típico, obtém-se os valores mínimos e máximos de +9,2 a +27,2 dBm ($\approx 8,3$ mW a 0,53 W), ainda muito dentro dos limites especificados pelo 3GPP. Os ajustes de potência são realizados em incrementos de $\pm 0,1$ dB, conforme especificado na documentação do fornecedor de tecnologia (Huawei Technologies Co., Ltd., 2018).

Estando essa parametrização estabelecida, o próximo passo é a calibração dos pesos w_{cov} , w_{dom} , w_{SDI} e λ , responsáveis por ajustar a contribuição relativa de cada métrica na composição de $\tilde{J}(P)$.

Para estabelecer uma referência para os demais pesos, foi definido $w_{\text{cov}} = 1$. Essa escolha tem como objetivo estabelecer uma base de normalização para os pesos, reduzindo o espaço de busca dimensional do problema e permitindo que os demais parâmetros sejam interpretados em termos relativos à componente de cobertura, que constitui o critério primário de viabilidade da rede, uma vez que representa a condição mínima para que qualquer comunicação possa ocorrer. Dessa forma, fixar w_{cov} em uma unidade preserva sua importância estrutural no modelo e transforma os demais pesos em fatores de amplificação ou atenuação relativos à sua referência. Isso também evita qualquer potencial redundância escalar, dado que a multiplicação simultânea de todos os pesos por uma constante não altera a estrutura relativa da função objetivo, mas apenas sua escala absoluta.

Então, o processo de determinação dos pesos foi conduzido de maneira iterativa e sistemática, dividido em quatro fases. Porém, alguns critérios na avaliação e escolha dos conjuntos de pesos foram determinados para refletir aspectos técnico-operacionais práticos e estabelecer condições gerais para a obtenção dos melhores pesos:

- Considerando $w_{\text{cov}} = 1$ como referência, definiu-se que os demais pesos não poderiam assumir valores inferiores a 0,1 (10% de w_{cov}) nem superiores a 10 (1.000% de w_{cov}), evitando assim altos vieses na contribuição relativa dos pesos;
- Adotou-se ainda um critério prático para evitar divergências associadas a aumentos excessivos de RSRP em grande número de células. Como limite arbitrário, definiu-se que no máximo 25% das células poderiam apresentar correções inferiores a -5 dB ou superiores a 5 dB. Tal restrição reflete o *trade-off* entre a melhoria de cobertura e dominância e o risco de redução ou aumento exagerado dos níveis de potência, o que poderia comprometer o balanceamento global da rede. Assim, manteve-se o limite de 25% como mecanismo de controle de estabilidade e realismo das soluções;
- A quantidade de execuções do algoritmo CMA-ES, com *seeds* fixas, bem como o número máximo de iterações, foram definidos de acordo com cada fase de refinamento. Para cada fase, calcularam-se e avaliaram-se as medianas das grandezas analisadas, em especial: (i) a quantidade de elementos da grade A_T com cobertura e boa dominância; e (ii) o número de *PCIs* cujos ajustes absolutos de potência superaram 5 dB;
- A cada nova fase do processo, aumentaram-se progressivamente a quantidade de *seeds*, o número máximo de iterações do CMA-ES e o conjunto de métricas consideradas na avaliação, caracterizando um procedimento incremental de refinamento da calibração. Por exemplo, a partir da Fase 2, outras grandezas como a quantidade

de elementos da grade com cobertura sem dominância e o percentual de *seeds* que geram soluções válidas também foram adicionadas à análise.

O objetivo final do processo de calibração dos pesos consiste em identificar o conjunto de parâmetros que maximize o desempenho da função objetivo sem violar os critérios e restrições anteriormente mencionados. Busca-se, assim, um ponto de equilíbrio entre performance e aderência às condições de estabilidade e realismo impostas ao modelo.

A adoção de uma abordagem estruturada em fases permitiu a exploração de um número maior de combinações admissíveis de pesos de maneira progressiva, sem comprometer a eficiência computacional. Na prática, foram avaliados múltiplos conjuntos candidatos com esforço computacional gradativo, refinando-se sucessivamente o espaço de busca até a etapa final, na qual os quatro melhores conjuntos foram submetidos a uma análise mais rigorosa, com maior número de *seeds* e limite máximo de iterações.

Esta estratégia é justificada pela enorme quantidade de possibilidades do espaço de busca e da grande quantidade de combinações possíveis entre os pesos e parâmetros envolvidos. A Figura 4.14 ilustra o processo gradual de calibração de pesos.

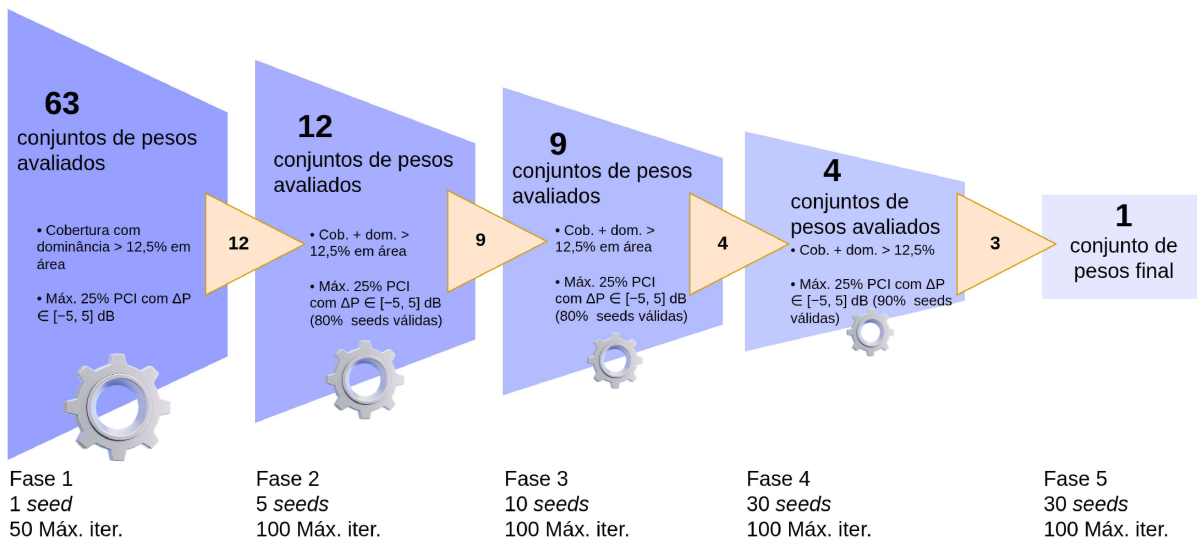


Figura 4.14 – Processo de refinamento da calibração dos pesos de $\tilde{J}(P)$. Figura elaborada pelo autor

As tabelas e valores encontrados na evolução dessas fases são mostrados com detalhes no Apêndice A, onde podem ser analisados de forma sistemática.

4.5.1.1 Fase 1: Análise Preliminar

Nesta etapa inicial, foram determinadas faixas de valores possíveis para cada peso (exceto w_{cov} , mantido na unidade) e analisadas as combinações entre eles, sendo cada combinação avaliada individualmente. Foram fixados três valores de λ , iguais a 0,1; 0,3 e 1,0 e, para cada um deles, foram avaliados w_{dom} nos valores 0,5; 1,0; 2,0; 3,5; 5,0; 7,0 e 10

e w_{SDI} nos valores 0,1; 0,3 e 1,0. No total, foram avaliados 63 conjuntos de parâmetros nos seguintes quesitos:

- Quantidade de elementos com cobertura com boa dominância;
- Quantidade de PCIs com ajuste de $|\Delta P| > 5$ dB (sobreajuste de potência).

Na Fase 1, os resultados foram avaliados com uma *seed* única e um número máximo de iterações igual a 50, o que permite maior agilidade na obtenção dos resultados, mantendo os critérios ainda menos restritivos. Os conjuntos que tiveram um sobreajuste de potência em mais que 25% dos PCIs foram descartados. Também foram descartados os resultados que apresentaram índices menores que 12,5% de área com cobertura com dominância. Dos 63 conjuntos de parâmetros analisados, 12 atenderam aos critérios estabelecidos e foram selecionados para análise na Fase 2.

4.5.1.2 Fase 2: Primeiro Refinamento

Durante a Fase 2, foram investigados os 12 melhores conjuntos de pesos obtidos da Fase 1. Nesta fase, os resultados foram obtidos executando-se 5 vezes o método CMA-ES, cada um com uma *seed* diferente. Foram mantidos os critérios de avaliação de quantidade de elementos com cobertura com dominância (mínimo de 12,5%) e sobreajuste de potência (máximo de 25%), sendo essas duas grandezas avaliadas como a mediana dos resultados obtidos das 5 execuções. A quantidade máxima de iterações foi aumentada para 100. Também foi adicionado um critério onde, para ser levado à próxima fase, um conjunto de pesos teria que ter no mínimo 80% das execuções individuais (*seeds*) levando a resultados sem sobreajuste de potência (máximo de 6 PCIs dos 24 com $|\Delta P| > 5$ dB).

Durante a Fase 1, todos os valores obtidos com $\lambda = 1,0$ mostraram-se com baixo desempenho no quesito de área de cobertura com dominância (todos ficaram com valores abaixo de 12,5% da área), de modo que foram analisados casos com λ iguais a 0,1 e 0,3, e para os valores de w_{dom} de 2,0; 3,5 e 7 e w_{SDI} de 0,1; 0,3 e 1,0.

Nessa fase foram registrados os valores das medianas de cobertura com dominância, cobertura sem dominância, quantidade de PCIs em sobreajustes, além da quantidade de *seeds* fora desse último critério.

Avaliados dessa forma, dos 12 conjuntos de parâmetros, 9 avançaram para a Fase 3, para uma análise mais detalhada.

4.5.1.3 Fase 3: Segundo Refinamento

Logo em seguida, na Fase 3, foram analisados os 9 conjuntos de pesos que levaram aos melhores resultados na Fase 2. O método adotado foi semelhante ao da Fase 2, porém a avaliação foi feita através do cálculo da mediana dos resultados obtidos com

10 execuções com *seeds* diferentes, mantendo-se os critérios anteriormente adotados em termos percentuais.

Todos os conjuntos avaliados foram válidos, sendo escolhidos os quatro melhores resultados de cobertura com dominância para a análise na próxima fase. Além disso, foram escolhidos casos com uma alta representatividade de diversidade dos pesos, nas quais foram selecionados casos com λ iguais a 0,1 e 0,3, w_{SDI} iguais a 0,1 e 1,0, e w_{dom} iguais a 3,5; 7,0 e 10,0.

4.5.1.4 Fase 4: Etapa Final e Definição dos Pesos

Por fim, chega-se à fase final do refinamento dos pesos. Nesta etapa, foram avaliados os quatro conjuntos de parâmetros previamente selecionados nas fases anteriores:

- ❑ Conjunto 1 ($\lambda = 0,3$, $w_{dom} = 10,0$, $w_{SDI} = 0,1$);
- ❑ Conjunto 2 ($\lambda = 0,3$, $w_{dom} = 10,0$, $w_{SDI} = 1,0$);
- ❑ Conjunto 3 ($\lambda = 0,1$, $w_{dom} = 3,5$, $w_{SDI} = 1$), e;
- ❑ Conjunto 4 ($\lambda = 0,3$, $w_{dom} = 7,0$, $w_{SDI} = 0,1$).

O Conjunto 2 foi desclassificado pois foi determinado que 90% das *seeds* deveriam corresponder a soluções válidas no critério de excesso de sobreajuste, tendo esse conjunto obtido 87% apenas. O Conjunto 3, foi considerado a melhor combinação dentre os três restantes, pois, embora o conjunto 1 tenha apresentado o maior valor de cobertura com dominância, a diferença relativa observada (inferior a 3%) caracteriza um empate técnico frente às incertezas inerentes ao processo de medição e à variabilidade estocástica entre *seeds*. Como critério adicional de decisão, foi priorizada uma configuração menos assimétrica entre os pesos. Desta forma, o Conjunto 3 apresentou melhor equilíbrio estrutural entre os termos da função e menor agressividade na ponderação dos pesos, sendo adotado como configuração recomendada.

Como já comentado, para não comprometer a fluidez do texto, as tabelas completas contendo todos os conjuntos avaliados com seus resultados em cada fase do processo de calibração são apresentadas em detalhes no Apêndice A, onde podem ser analisadas com maior profundidade.

4.5.2 Considerações Finais sobre a Calibração

O processo de determinação dos pesos foi guiado por um procedimento estruturado de calibração orientado por sensibilidade, estabilidade e coerência física. A calibração foi feita com base no desempenho real – cobertura com e sem dominância, além de critérios de viabilidade prática da solução – e não avaliando-se a função objetivo em si. Em outras

palavras, a escolha dos pesos foi realizada interativamente, avaliando-se processos no *outer loop* da aplicação. Dessa forma, os valores finais de w_{cov} , w_{dom} , w_{SDI} e λ correspondem a parâmetros que refletem um desempenho geral muito satisfatório do método.

Conforme mencionado, houve um bom equilíbrio entre os pesos, além de um adequado *trade-off* entre a melhor solução em termos de cobertura com dominância e a aplicabilidade prática dos ajustes de potência. Ao final, os pesos a serem usados na função objetivo são:

$$\square w_{\text{cov}} = 1,0;$$

$$\square w_{\text{dom}} = 3,5;$$

$$\square w_{\text{SDI}} = 1,0, \text{ e};$$

$$\square \lambda = 0,1.$$

Desta forma, a função objetivo utilizada no método de análise do *cluster* de células fica conforme mostrado na Equação 50.

$$\tilde{J}(P) = \frac{1}{N} \sum_{i=1}^N \left[1,0 C_i(P) + 3,5 D_i(P) \right] - 1,0 V_{\text{SDI}}(P) - 0,1 R(P) \quad (50)$$

Os próximos cálculos considerarão essa equação final como função objetivo para abordagem do problema de maximização de desempenho.

4.5.3 Execução do Aprimoramento de Cobertura com os Pesos Definidos

Devido à natureza estocástica do CMA-ES, cada experimento foi executado de forma independente 1.000 vezes, utilizando-se diferentes *seeds* e retendo-se a melhor solução obtida. O melhor resultado entre as *seeds* de 1 até 1.000 foi obtido com a *seed* 177. O vetor P obtido com essa *seed* é mostrado na Tabela 4.6.

Tabela 4.6 – Vetor de ajustes de potência na melhor solução por PCI

PCI	Ajuste de Potência (dB)
009	-2,5
010	8,9
011	7,6
165	8,5
166	-0,2
167	8,7
243	-2,2
244	9,0
245	1,5
246	2,9
247	2,7
248	0,3
252	-0,3
253	0,4
254	-0,6
333	1,0
334	-0,3
335	8,9
395	3,9
396	0,8
397	1,3
404	0,1
405	-1,4
406	0,7

A média de correção de potência proposta na melhor solução foi de 2,5 dB, indicando uma tendência de elevação da potência média no *cluster*. O desvio padrão foi relativamente alto, de aproximadamente 3,9 dB. Isso pode ser observado no fato de alguns PCIs terem grandes incrementos de potência, iguais ou próximos ao máximo de 9 dB, enquanto outros tiveram uma redução de até 2,5 dB.

A mediana dos ajustes foi de 0,9 dB (média dos valores centrais 0,8 e 1,0 dB), valor inferior à média, o que demonstra leve assimetria à direita na distribuição dos incrementos de potência, influenciada pelos valores mais elevados próximos ao limite superior estabelecido. O primeiro quartil (Q1) foi de -0,2 dB, enquanto o terceiro quartil (Q3) atingiu 5,8 dB, resultando em um intervalo interquartil (IQR) de 6,0 dB. Esse intervalo indica uma dispersão significativa mesmo na região central da distribuição. A amplitude total observada foi de 11,5 dB, evidenciando a grande variação entre os ajustes mínimos e máximos aplicados aos PCIs. Esses resultados mostram que, embora exista uma tendência média de elevação de potência no *cluster*, a solução ótima promove ajustes heterogêneos entre as células, compatíveis com as características espaciais de cobertura e dominância consideradas no processo de aprimoramento de cobertura.

A visualização dos dados acima é demonstrada na Figura 4.15.

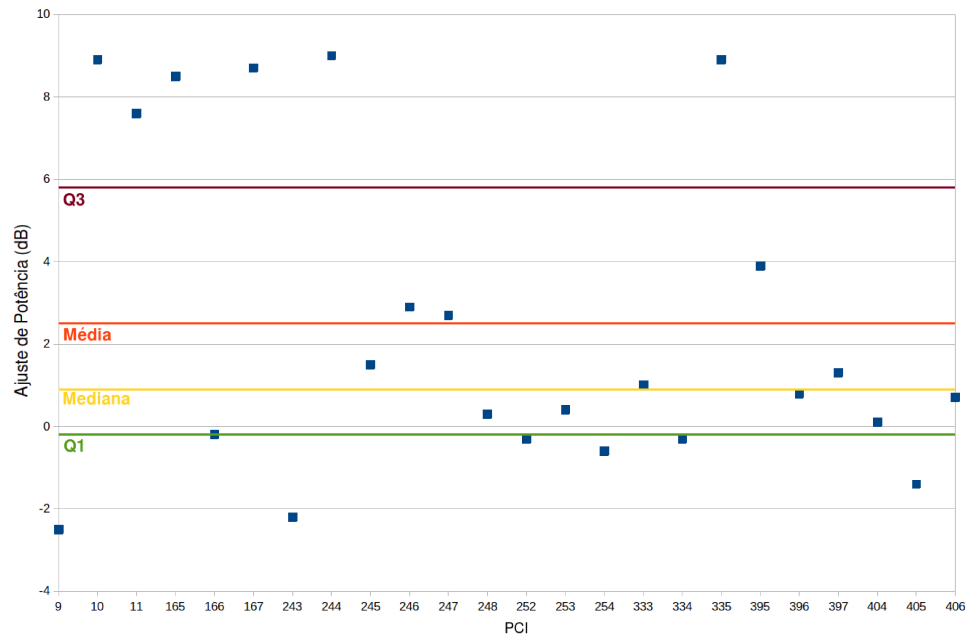


Figura 4.15 – Distribuição dos ajustes de potência por PCI. Figura elaborada pelo autor

4.6 Resultados e Discussão

Esta seção apresenta e discute os resultados obtidos com a aplicação do método de aprimoramento de cobertura proposta, com ênfase nas melhorias observadas em termos de cobertura e condições de interferência.

4.6.1 Avaliação da Robustez do Método

Para avaliar a robustez da proposta, o CMA-ES foi executado 1.000 vezes utilizando diferentes *seeds*, com MAXITER = 100, utilizando-se a Equação 50 para se determinar a estabilidade do método. As estatísticas resultantes dos melhores valores da função objetivo obtidos em cada execução são apresentadas na Tabela 4.7.

Tabela 4.7 – Análise de robustez considerando 1.000 execuções independentes do CMA-ES

Métrica	Valor
Média $\tilde{J}(P)$	0,6811
Desvio padrão	0,0031
Coefficiente de variação	0,46%
Mínimo / Máximo $\tilde{J}(P)$	0,6524 / 0,6833
IC 95%	0,6809 / 0,6812

Para a função $\tilde{J}(P)$, obteve-se um valor mínimo de 0,6524, máximo de 0,6833 e média igual a 0,6811, com desvio padrão amostral de 0,0031. O coeficiente de variação (razão entre o desvio padrão e a média da amostra) foi de apenas 0,46%, indicando baixa dispersão relativa. O intervalo de confiança de 95% (IC 95%) para a média foi calculado em

[0,6809; 0,6813]. Para a função base $J(P)$ (que não considera o termo de regularização nem a componente de SDI), os valores máximo e mínimo foram 0,9145 e 0,8474, respectivamente, com média de 0,8976 e desvio padrão de 0,0080. O coeficiente de variação foi de 0,89%, e o intervalo de confiança de 95% para a média foi [0,8971; 0,8981]. Esses resultados confirmam a robustez estatística do método frente à variabilidade das *seeds* iniciais.

O baixo desvio padrão e o coeficiente de variação inferior a 0,5% indicam que a abordagem proposta converge de maneira consistente para soluções de alta qualidade, apesar da natureza estocástica do método. Esses resultados confirmam que as melhorias observadas em cobertura e dominância não são significativamente sensíveis à inicialização do processo de aprimoramento.

Além disso, com relação à quantidade de iterações permitidas pelo método (gerações), testes adicionais com 1.000 iterações indicaram um ganho marginal inferior a 0,5% no valor de $\tilde{J}(P)$ em relação aos valores obtidos com 100 iterações, demonstrando que essa quantidade de gerações já coloca o algoritmo próximo da região ótima. Assim, optou-se por manter 100 iterações e investir em maior número de realizações independentes para análise de robustez estatística.

De forma geral, os resultados demonstram que a melhoria do ambiente de RF orientada a dados proposta apresenta desempenho consistente e robusto, demonstrando métricas estáveis na avaliação da função objetivo $\tilde{J}(P)$.

4.6.2 Melhorias de Cobertura e Interferência

Após a aplicação do vetor de ajuste de potência P a todas as células do cluster, a eficácia do método proposto é avaliada por meio dos ganhos obtidos em cobertura e dominância ao longo da área analisada. A área coberta é definida como o conjunto de elementos da malha espacial cuja célula servidora apresenta RSRP superior ou igual a -105 dBm (S_{\min}). Adicionalmente, considera-se que uma região apresenta dominância suficiente quando, além de satisfazer o critério de cobertura, a diferença de RSRP entre a célula servidora e a principal interferente é superior ou igual a 9 dB (Δ_{ref}).

A Figura 4.16 apresenta os mapas de cobertura e interferência correspondentes à configuração de referência e à configuração obtida após a aplicação do vetor de ajuste de potência gerado pelo método proposto. Para fins de avaliação qualitativa e quantitativa, o vetor P considerado foi aquele que apresentou o maior ganho no critério de cobertura com alta dominância, dentre as 1.000 execuções independentes realizadas.

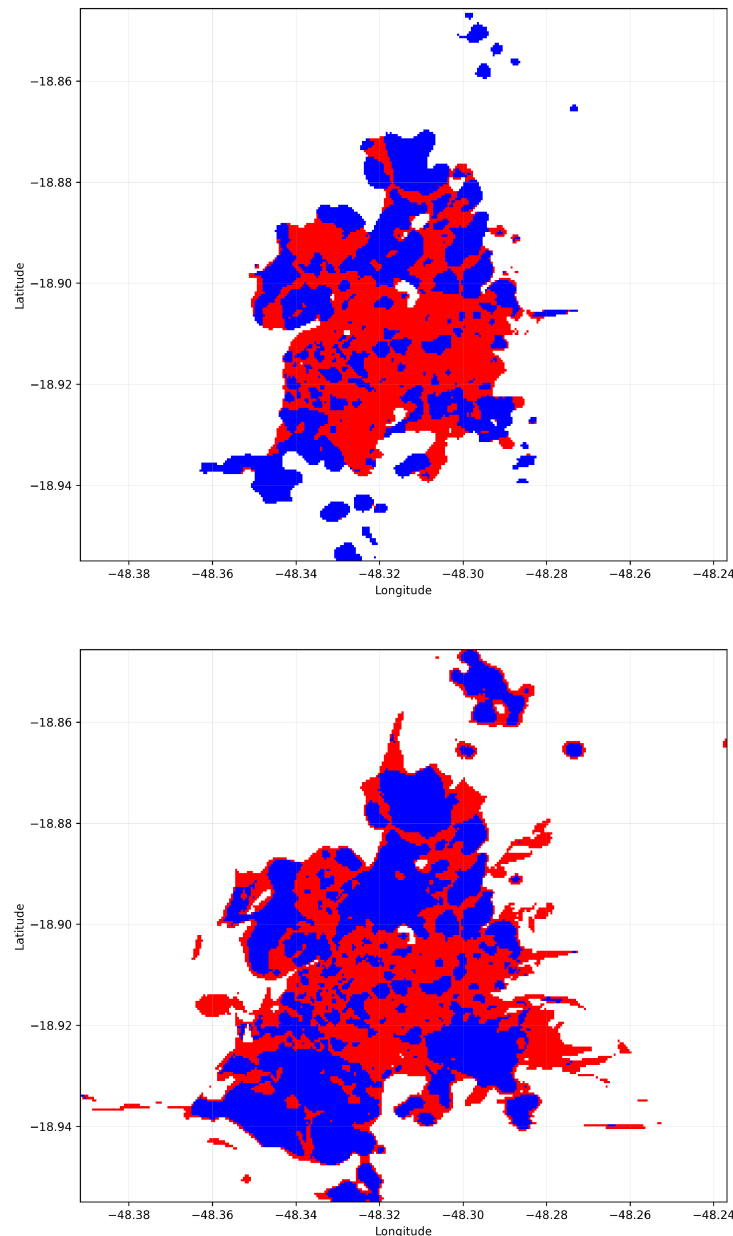


Figura 4.16 – Mapas de cobertura e interferência antes e após o aprimoramento da cobertura baseado em CMA-ES. Regiões em branco indicam ausência de cobertura ($\text{RSRP} < -105 \text{ dBm}$). Regiões em vermelho correspondem a pontos cobertos com dominância insuficiente ($\text{dominância} < 9 \text{ dB}$), enquanto regiões em azul indicam pontos cobertos com dominância adequada ($\text{dominância} \geq 9 \text{ dB}$).

Para aprofundar a análise do impacto da melhoria, foi gerado o mapa de variação de RSRP apresentado na Figura 4.17, obtido pela comparação do nível de sinal da melhor célula servidora antes e após a aplicação das correções de potência advindas do CMA-ES. Nesse mapa, regiões em verde indicam aumento de RSRP, regiões em amarelo indicam redução do nível de sinal e regiões em azul representam áreas sem variação significativa (alteração em módulo menor do que $0,1 \text{ dB}$). Essa visualização permite identificar padrões espaciais associados aos ajustes de potência aplicados.

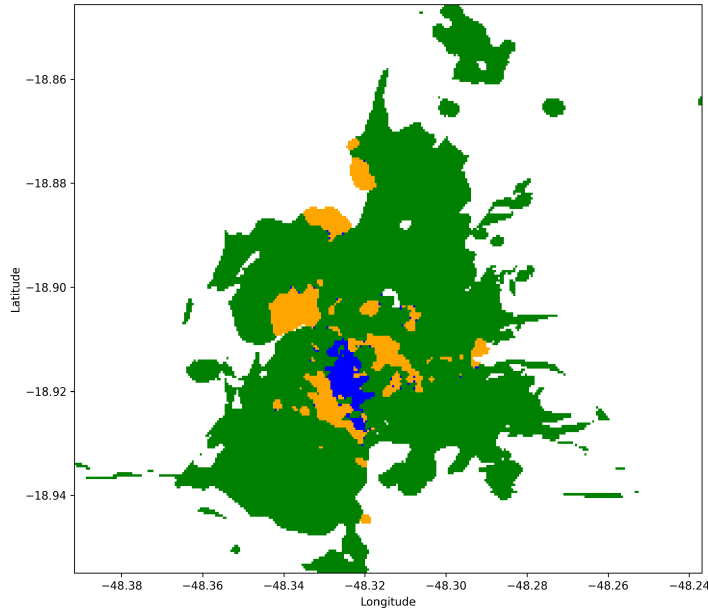


Figura 4.17 – Mapa de variação do RSRP após a correção das potências obtidas com CMA-ES. Regiões em verde indicam aumento do RSRP da célula servidora, regiões em amarelo indicam redução, e regiões em azul indicam ausência de variação em relação à configuração de referência.

A inspeção visual dos mapas revela dois efeitos principais: (i) uma melhoria significativa da cobertura nas regiões periféricas do cluster, decorrente do aumento de potência aplicado às células de borda, com a maior parte da nova área coberta apresentando dominância adequada; e (ii) o aumento da dominância na região central do cluster, obtido por meio da redução da potência de células mais centrais, ampliando a extensão das regiões bem dominadas no núcleo da rede, incluindo uma pequena região central onde houve poucos ajustes nas células, demonstrando uma manutenção nos níveis de RSRP.

4.6.3 Ganhos Quantitativos de Desempenho

A Tabela 4.8 resume o impacto quantitativo da melhoria proposta em termos de cobertura e dominância entre célula servidora e interferente. Os valores absolutos correspondem ao número de pontos da grade da A_T , enquanto os percentuais indicam a participação relativa de cada categoria em relação ao total de pontos analisados.

Tabela 4.8 – Distribuição de cobertura e dominância antes e após as adequações de potência

Categoria	Referência	Aprimorado	Ganho (%)
Sem cobertura	64.234 (81,4%)	57.285 (72,6%)	-10,8
Cobertura, baixa dominância	7.466 (9,5%)	10.177 (12,9%)	+36,3
Cobertura, alta dominância	7.192 (9,1%)	11.430 (14,5%)	+58,9

Observa-se uma redução de 10,8% na área sem cobertura, acompanhada por um aumento expressivo das regiões com dominância suficiente. Em particular, a fração da área

com boa dominância cresce de 9,1% para 14,5%, correspondendo a um ganho relativo bastante substancial de 58,9%. Também houve ganho de area com baixa dominância, correspondente a 36,3% com relação à situação original. Esses resultados corroboram quantitativamente as tendências observadas visualmente na Figura 4.16, evidenciando a eficácia do método proposto na melhoria simultânea da extensão de cobertura e das condições de interferência.

Além disso, também na quantificação dos ganhos de cobertura com e sem dominância, considerando apenas as *seeds* válidas nos critérios já mencionados, o modelo se mostrou muito robusto, apresentando baixos coeficientes de variação e intervalos de confiança de 95% estreitos, conforme mostrado na tabela 4.9.

Tabela 4.9 – Médias, dispersão e intervalos de confiança obtidos após 1.000 execuções.

Classificação do Elemento	Média	Desvio-padrão	Coefficiente de variação	IC 95%
Sem Cobertura	57.325	178	0,31%	[57.310 ; 57.333]
Cobertura Sem Dominância	10.467	238	2,27%	[10.453 ; 10.483]
Cobertura Com Dominância	11.100	161	1,45%	[11.094 ; 11.114]

De forma geral, os resultados demonstram que a estratégia de ajuste de potências buscando o aprimoramento da qualidade da cobertura com método unicamente orientado por dados proposta é capaz de promover melhorias consistentes e robustas nas métricas utilizadas para quantificar a qualidade de cobertura de uma área.

4.6.4 Aspectos Computacionais e Considerações de Implantação

O custo computacional do método de proposto foi avaliado a partir do tempo de execução do algoritmo CMA-ES em múltiplas execuções independentes. Cada processo foi executado por 100 gerações. Considerando o tamanho padrão da população do CMA-ES, isto é, a quantidade de soluções candidatas geradas a cada geração, conforme recomendada por (HANSEN, 2023) (valor também utilizada por padrão na biblioteca *pycma*, usada na implementação em Python deste trabalho):

$$\lambda = 4 + \lfloor 3 \ln K \rfloor \quad (51)$$

onde $\lfloor \cdot \rfloor$ representa a função piso (arredondamento para baixo) e, com $K = 24$ células, resulta em uma população de 13 indivíduos (ou vetores P) por geração, totalizando 1.300 avaliações da função objetivo por execução. No cenário avaliado, cada execução do CMA-ES demandou, em média, 18,5 s em uma estação de trabalho convencional, com desvio padrão de 0,2 s ao longo de 1.000 execuções independentes, conforme resumido na Tabela 4.10.

Tabela 4.10 – Desempenho computacional ao longo de 1.000 execuções independentes do CMA-ES

Métrica	Valor
Número de execuções	1.000
Gerações por execução	100
Avaliações da função objetivo por execução	1.300
Tempo médio de execução (s)	18,48
Desvio padrão (s)	0,18
Coefficiente de variação	0,98%
IC 95% para a média (s)	[18,47 ; 18,49]

O coeficiente de variação menor que 1% e o estreito intervalo de confiança indicam comportamento computacional estável e previsível, apesar do caráter estocástico do algoritmo. O tempo total do experimento, considerando as 1.000 execuções independentes, foi de aproximadamente 5,13 horas no *hardware* doméstico convencional mencionado na Seção 4.2.3. Porém, ressalta-se que a quantidade elevada de execuções foi demandada por rigor metodológico, visando a obtenção de estimativas estatisticamente fortes e válidas. A baixa variabilidade observada (coeficiente de variação inferior a 1%) e o intervalo de confiança de 95% extremamente estreito para o tempo médio de execução indicam uma boa estabilidade do algoritmo. Em aplicação prática, uma quantidade de execuções entre 5% e 10% dessa quantidade de 1.000 execuções é suficiente para a obtenção de bons resultados, o que totalizaria um tempo aproximado de execução entre 15 e 30 minutos em *hardware* convencional, considerando-se um *cluster* do mesmo tamanho.

Sob a perspectiva de implantação prática, o método proposto é concebido como parte de um fluxo de processamento orientado por dados de ponta a ponta, no qual a aquisição de medições, a extrapolação de cobertura e a etapa de aprimoramento da cobertura são executadas de forma integrada. As fases de coleta de dados e de modelagem do ambiente de rádio, que dependem de medições do tipo MDT, *crowdsourcing* ou *drive tests*, bem como de modelos de interpolação computacionalmente intensivos, constituem os componentes mais onerosos do processo e devem ser atualizadas sempre que uma nova correção for realizada. Assim, embora o CMA-ES em si apresente relativo baixo custo computacional, o tempo total necessário para a execução do *pipeline* completo da metodologia ARCADE pode ser significativamente influenciado pelas etapas de aquisição de dados e modelagem do ambiente de propagação.

4.6.5 Comparação dos Resultados Obtidos com Trabalhos Relacionados

Para a avaliação da contribuição científica de uma proposta metodológica é importante se estabelecer uma comparação direta de seus resultados quantitativos com os de trabalhos correlatos. Contudo, conforme evidenciado na Tabela 2.2, a baixa aderência de trabalhos

com premissas suficientemente similares às da metodologia ARCADE limita o escopo dessa comparação. Dentre os trabalhos avaliados, aquele que apresenta maior aderência metodológica é o de (FAGEN; VICHARELLI; WEITZEN, 2008).

Os resultados obtidos pela metodologia ARCADE podem ser contextualizados com os de (FAGEN; VICHARELLI; WEITZEN, 2008), que apresenta uma abordagem automatizada de otimização de cobertura com controle de sobreposição e interferência para redes móveis. Embora os contextos e métricas adotados sejam distintos, a comparação é pertinente, pois ambos os trabalhos compartilham o objetivo central de maximizar simultaneamente a cobertura e reduzir a interferência entre células de um mesmo *cluster* através de ajustes nos níveis de potência transmitida pela RBS.

Do ponto de vista metodológico, as abordagens diferem de forma fundamental. O método de (FAGEN; VICHARELLI; WEITZEN, 2008) pressupõe o acesso a informações de projeto da rede, como parâmetros de enlace e modelos de propagação, que são utilizados para calcular analiticamente um coeficiente de cobertura $c_i = \beta_i/\alpha_i$, definido como a razão entre a área de melhor servidor na presença de vizinhos e a área de cobertura isolada de cada RBS. O processo de aprimoramento da qualidade de cobertura é conduzido por um método iterativo baseado em Newton-Raphson, que ajusta os níveis de potência de transmissão (ajustes esses similares aos realizados na metodologia ARCADE) com base nesse modelo matemático explícito. Em contraste, o método proposto nesta tese é estritamente orientado por dados (*data-driven*): não requer qualquer informação de projeto, modelo analítico de propagação ou parâmetro de configuração da rede.

No que diz respeito aos resultados quantitativos, (FAGEN; VICHARELLI; WEITZEN, 2008) reporta, para uma rede de 12 RBSs com três setores cada, uma redução de 42% na área com sobreposição de cobertura e um decréscimo marginal de 3% na área total coberta. Por sua vez, a metodologia ARCADE, aplicada a um *cluster* com 8 RBSs, apresentou um ganho relativo de 58,9% na área com alta dominância e de 36,3% na área com baixa dominância, com uma redução de 10,8% na área sem cobertura. A Tabela 4.11 resume esse comparativo.

Tabela 4.11 – Comparativo de melhoria de cobertura e qualidade entre o ARCADE e o modelo proposto em (FAGEN; VICHARELLI; WEITZEN, 2008).

Método	RBSs	Células	Aumento da Área com Qualidade	Aumento da Área de Cobertura
Fagen, Vicharelli e Weitzen (2008)	12	36	42%	-3%
ARCADE	8	24	59%	36%

Embora as métricas e premissas dos dois métodos sejam bastante distintas, os ganhos obtidos são comparáveis em magnitude, demonstrando que, mesmo operando sem infor-

mações de projeto ou modelos analíticos de propagação, a metodologia ARCADE é capaz de produzir resultados competitivos quando comparada a métodos que utilizam ajustes de potência orientados por modelo para a melhoria da qualidade de cobertura.

4.7 Ponderações sobre Eficiência Energética e Emissão de Carbono

Atualmente, uma atenção especial deve ser dada às questões relacionadas ao consumo de energia e à emissão de carbono. Em especial, aplicações de IA generativas têm demandado altíssimo consumo energético devido à quantidade e ao tamanho dos *datacenters* necessários para treinamento e inferência. Esse cenário intensifica preocupações relacionadas à sustentabilidade ambiental, à eficiência energética e ao custo operacional desses sistemas em larga escala. Nesse contexto, o objetivo desta seção é realizar uma comparação não-exaustiva entre os métodos tradicionais de *tuning* de uma rede móvel celular com a metodologia ARCADE. Como será mostrado a seguir, os métodos serão comparados etapa a etapa, considerando fases similares do processo.

Cabe ressaltar, contudo, que não existe um procedimento único ou rigidamente padronizado para a *tuning* tradicional de RAN. Na prática, os fluxos operacionais podem variar entre operadoras e até mesmo entre equipes técnicas dentro de uma mesma organização. Assim, a comparação apresentada nesta seção baseia-se em uma estrutura genérica, representativa das práticas mais comumente adotadas no setor, sem a pretensão de esgotar todas as variações possíveis.

4.7.1 Metodologia para Comparação de Emissão de Carbono

Durante o processo de aprimoramento da rede, considerando tanto a metodologia tradicional quanto o ARCADE, são claramente identificadas três grandes etapas distintas: a aquisição das amostras, a extrapolação dos dados para uma generalização da cobertura por célula e, por fim, a etapa de análise dos dados e o *tuning* do sistema em si.

A figura 4.18 representa essas etapas, evidenciando os métodos usados em cada uma delas e ressaltando qualitativamente e de forma preliminar o potencial de consumo de energia e/ou emissão de carbono de cada uma delas. As caixas em vermelho representam maior potencial de emissão de carbono, enquanto as caixas em amarelo e verde representam, respectivamente, potenciais médio e baixo em termos de impacto nas emissões.

A seguir, cada etapa será analisada quantitativamente para se estimar o consumo energético associado e o correspondente potencial de emissão de carbono, de modo a permitir uma comparação entre a abordagem tradicional e a metodologia ARCADE.

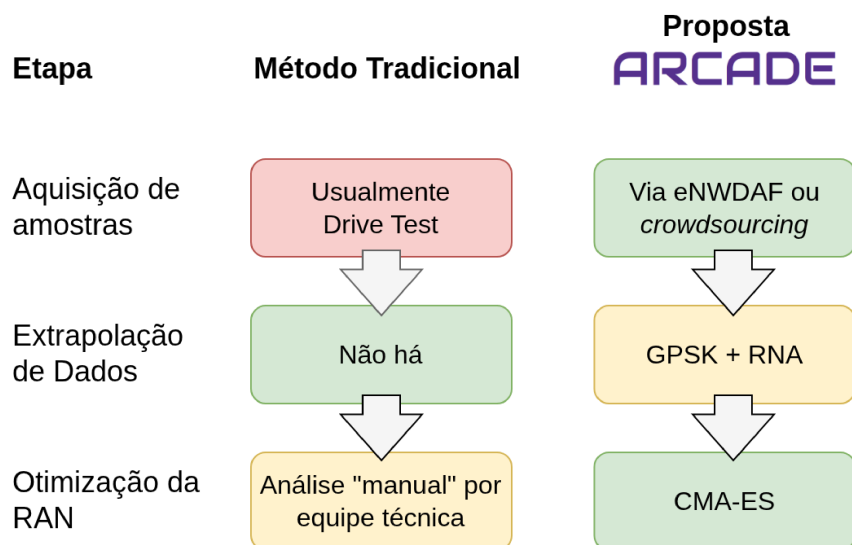


Figura 4.18 – Diagrama preliminar comparativo das etapas do processo de *tuning* de RAN. Figura elaborada pelo autor.

4.7.2 Emissão de Carbono na Etapa de Aquisição de Amostras

Tradicionalmente, o processo de aquisição de dados para aprimoramento da RAN é realizado com a coleta de dados via *drive tests*, que têm uma maior pegada de carbono associada ao consumo de energia para o deslocamento do veículo que contém os equipamentos que realizam essa coleta de dados. Há diversas variáveis a serem analisadas no processo de DT: tipo do veículo utilizado (carro, motocicleta, equipamentos embarcados em transportes públicos), classe do veículo (a combustão, elétrico ou híbrido), tipo de combustível utilizado no caso de veículo a combustão (etanol, gasolina ou óleo diesel), estilo de condução, potência e eficiência do veículo, horário, tráfego e temperatura ambiente, além de outros fatores.

Para uma análise básica, será considerado um veículo a combustão abastecido com gasolina C (E27), de médio consumo (12 km/L) em ciclo misto, típico de carros compactos a médios brasileiros (INMETRO, 2025), a uma temperatura de 20° C. Esse tipo de veículo ainda é o normalmente utilizado no cenário brasileiro, embora haja uma tendência de aumento no uso de carros híbridos ou elétricos, o que pode, num futuro próximo, reduzir as emissões. Nesse cenário, as emissões de escapamento, também chamadas de *Tank-to-Wheel* (TTW), são da ordem de 190 g CO₂/km, enquanto a pegada de carbono completa poço-à-roda, ou *Well-to-Wheel* (WTW) – que inclui a produção e distribuição do combustível – é de aproximadamente 230 a 245 g CO₂e/km (KAWAMOTO et al., 2019; CARVALHO, 2011; MERA et al., 2023). Esses valores são conservadores e alinhados com a frota brasileira, considerando-se o combustível exclusivamente gasolina C. Em cenários reais de DT, a emissão pode variar em ±15% conforme o tráfego e o estilo de condução. Neste trabalho, será considerada a média WTW de 237,5 g CO₂e/km rodado.

Para o caso da cidade de Uberlândia (MG), das quais foram obtidos os dados de DT, a

estimativa de quilômetros rodados para obtenção de um mapa de cobertura representativo de toda a área urbana, utilizando o padrão “rua sim, rua não” nas vias locais (amostragem alternada que não exige o percorrimto de todas as ruas, mas sim vias alternadas para coincidir com a direção das ruas e garantir representatividade espacial), é de aproximadamente 1.800 km. Esse número foi encontrado através de fontes que determinam a extensão total da malha viária urbana, que passou de 2.500 km em 2012 (ASSUNÇÃO, 2012) para 3.000 km em 2018 (Diário de Uberlândia, 2018). Será adotada uma estimativa conservadora de 3.000 km de malha viária total para a cidade. Em seguida, aplica-se o padrão de DT típico em uma RAN: percorrem-se integralmente as vias de hierarquia superior (arteriais e coletoras, que será considerado 25% da malha total) e adota-se amostragem alternada (“rua sim, rua não”) nas vias locais (75% restante), equivalente a 50% do comprimento dessas vias. O cálculo resulta em:

$$(0,25 \times 3.000) + (0,75 \times 3.000 \times 0,5) \approx 1.875 \text{ km},$$

arredondado para 1.800 km considerando eventual otimização de rotas, vias não percorridas, falha na amostragem em bairros de mais difícil acesso, dentre outros. Esse valor é estimado, mas está muito alinhado às práticas de campo em campanhas de DT para *tuning* de um sistema celular.

Aplicando o fator de emissão adotado de 237,5 g CO₂e/km ao percurso estimado de 1.800 km para um *drive test* completo de mapeamento de cobertura em Uberlândia, obtém-se uma emissão total de aproximadamente 427,5 kg de CO₂ emitidos no processo de um DT nessa cidade. Embora esse montante possa ser visto como moderado quando considerado isoladamente – equivalente ao carbono sequestrado por aproximadamente 19 árvores adultas ao longo de um ano inteiro (NOWAK et al., 2013) –, ele ganha relevância quando se multiplica pela frequência das campanhas de *tuning*, geralmente semestral ou anual, e pelo número de cidades atendidas por uma operadora, sobretudo considerando-se as operadoras nacionais.

A metodologia ARCADE, por sua vez, propõe que os dados de cobertura sejam coletados através de métodos decorrentes de uma evolução da arquitetura 3GPP, como o MDT, ou via *crowdsourcing*, conforme demonstrado na prática na Seção 4.1. Embora essas técnicas não sejam completamente isentas de emissão de carbono, já que qualquer método envolve intrinsecamente o consumo de energia elétrica, elas são substancialmente menos impactantes ao meio ambiente e, assim, contribuem para que o processo de aprimoramento da RAN seja mais sustentável.

4.7.3 Emissão de Carbono na Etapa de Extrapolação de Dados

No caso do ARCADE, o impacto de maior relevância se refere ao consumo de energia elétrica devido à carga de processamento no cálculo dos GPSKs e, sobretudo, das RNAs,

sendo esse último processo muito mais intensivo em uso de recursos computacionais do que o primeiro.

Comparativamente, o processo computacional do GPSK é proporcional a (CORMEN et al., 2009):

$$\mathcal{O}(N \cdot d) \tag{52}$$

onde:

- N é o número de amostras;
- d é a dimensionalidade dos dados.

Logo, no caso do GPSK, o custo computacional cresce de forma linear com o número de amostras e dimensões de análise. Não há nem uma retropropagação de erros nem um ajuste iterativo de pesos, o que simplifica bastante todo o processo.

Já as RNAs exige muito mais processamento porque envolve métodos computacionalmente custosos de ajustes de parâmetros (pesos e *biases*). Do ponto de vista de custo computacional, o esforço é proporcional a (CORMEN et al., 2009):

$$\mathcal{O}(N \cdot d \cdot L \cdot E) \tag{53}$$

onde, além de N e d , tem-se:

- L é o número de camadas de rede;
- E é a quantidade de épocas até se atingir a convergência da rede.

Em resumo, enquanto o GPSK é um processo que é resolvido com operações mais diretas e fechadas, as RNAs envolvem múltiplas camadas com funções não-lineares, sendo muito mais intensivo em operações de multiplicação de matrizes. Na prática, usando o *hardware* simples, como o mostrado na Tabela 4.2, os processos gaussianos são computados em poucos segundos por célula, enquanto as redes neurais artificiais levam horas de processamento para treinamento, como será mostrado adiante. Desta forma, para este trabalho, o esforço computacional – e seu respectivo consumo elétrico – do processamento GPSK será considerado desprezível com relação às RNAs.

Foi considerado fator de emissão para o Brasil de 38,5 kgCO₂e/MWh, de acordo com o Ministério da Ciência, Tecnologia e Inovação do Brasil (MCTI) (MCTI, 2024). A emissão de carbono é comparativamente baixa em relação a outros países, sobretudo os do G20, devido ao fato de o país gerar 79% de sua eletricidade a partir de fontes renováveis, predominantemente hidrelétricas, mas com um aumento ano a ano na geração solar e eólica (Climate Transparency, 2022).

A Tabela 4.12 apresenta os valores de kWh consumidos e de kgCO₂e emitidos pela RNA para cada PCI e o valor total durante o treinamento da rede. Não são considerados os treinamentos com GPSK, por se tratarem de um consumo irrelevante quando comparado ao consumido pelo treinamento da rede em RNA. Os valores de consumo estão estimados em seu patamar superior, uma vez que foram obtidos multiplicando-se o tempo de processamento pela potência nominal da fonte de 45 W.

Tabela 4.12 – Consumo energético e emissão de carbono associados ao processamento das RNAs por PCI

PCI	Tempo de Processamento (s)	Consumo de Energia (kWh)	Emissão de Carbono (kgCO ₂)
009	26.083,9	0,326	0,0126
010	39.055,8	0,488	0,0188
011	31.198,2	0,390	0,0150
165	37.493,6	0,469	0,0181
166	25.426,6	0,318	0,0122
167	23.538,9	0,294	0,0113
243	33.202,0	0,415	0,0160
244	43.466,3	0,543	0,0209
245	27.385,5	0,342	0,0132
246	23.524,2	0,294	0,0113
247	25.425,8	0,318	0,0122
248	31.204,8	0,390	0,0150
252	29.469,1	0,368	0,0142
253	24.714,7	0,309	0,0119
254	23.046,0	0,288	0,0111
333	28.740,2	0,359	0,0138
334	33.412,9	0,418	0,0161
335	22.046,0	0,276	0,0106
395	22.913,2	0,286	0,0110
396	26.063,9	0,326	0,0126
397	31.651,5	0,396	0,0152
404	35.906,2	0,449	0,0173
405	23.787,2	0,297	0,0114
406	38.814,6	0,485	0,0187
Total	70.7571,2	8,845	0,3405

O valor de 340 gCO₂ pode ser considerado um valor baixo em termos absolutos. Em termos de comparação, um chuveiro elétrico residencial comum de 5,5 kW (amplamente usado no Brasil) consome os 8,8 kWh da energia usada no processamento das RNAs no tempo equivalente de 1 hora e 36 minutos de uso contínuo.

4.7.4 Emissão de Carbono na Etapa de *Tuning* da RAN

Conforme apresentado na Seção 4.6.4, o custo computacional do CMA-ES é bastante baixo, equivalendo a 18,5 s por execução. Considerando 50 execuções por célula – de modo a se ter um conjunto estatisticamente consistente de resultados e possibilitar a seleção da melhor solução dentre elas –, e ainda levando-se em consideração o *cluster* de 24 células, o tempo total de processamento é da ordem de 6,2 h em um *hardware* doméstico convencional. Admitindo-se, de forma conservadora, o valor de 45 W (potência máxima da fonte de alimentação do equipamento) e o fator de emissão de 38,5 kgCO₂e/MWh (MCTI, 2024), estima-se, para todo o processamento, uma emissão de carbono total aproximada de 0,0107 kgCO₂e, ou o equivalente a 10,7 g de CO₂.

Em contraponto, a etapa de *tuning*, quando realizada manualmente por uma equipe especializada, também demanda várias horas de uso de *hardware*. O tempo total em horas vai, naturalmente, variar conforme o tamanho da equipe envolvida, os equipamentos usados e o nível de profundidade da análise. Em razão dessa variabilidade e da ausência de padronização metodológica, a quantificação precisa desse consumo torna-se imprevisível e difícil de estimar. Porém, pode-se presumir que sua ordem de grandeza seja em torno de algumas dezenas de horas de uso de equipamentos equivalentes aos utilizados pelo ARCADE.

Há também, obviamente, uma discrepância importante ao se comparar a melhoria do projeto realizada manualmente por uma equipe de engenheiros e a proposta no ARCADE, uma vez que a última se restringe a adequações nas potências dos canais de referência das células, enquanto a primeira envolve ajustes em diversos outros parâmetros de rede, inclusive físicos, como os azimutes e *downtilts* das antenas.

Portanto, embora o *tuning* viabilizado pelo ARCADE apresente uma emissão de carbono quantificável e muito baixa ($\approx 10,7$ gCO₂e), a comparação direta com o processo manual revela-se impraticável. As diferenças fundamentais de escopo – o CMA-ES limita-se a ajustes nas potências dos canais de referência, enquanto a abordagem humana envolve dezenas de parâmetros, inclusive físicos e dependentes de informação adicional externa –, e ainda considerando a alta variabilidade metodológica e de equipe, não é possível a realização de qualquer quantificação equivalente. Devido à impossibilidade de comparação metodológica justa e considerando que ambas as estratégias se situam na mesma ordem de grandeza de consumo de *hardware*, pode-se concluir por um empate técnico no que tange ao impacto ambiental associado à própria etapa de *tuning*.

4.7.5 Conclusão sobre Impacto da Metodologia ARCADE na Emissão de Carbono

Observa-se, atualmente, uma preocupação crescente – e bem fundamentada – em relação aos impactos das técnicas de IA sobre o consumo de energia e de recursos hídricos,

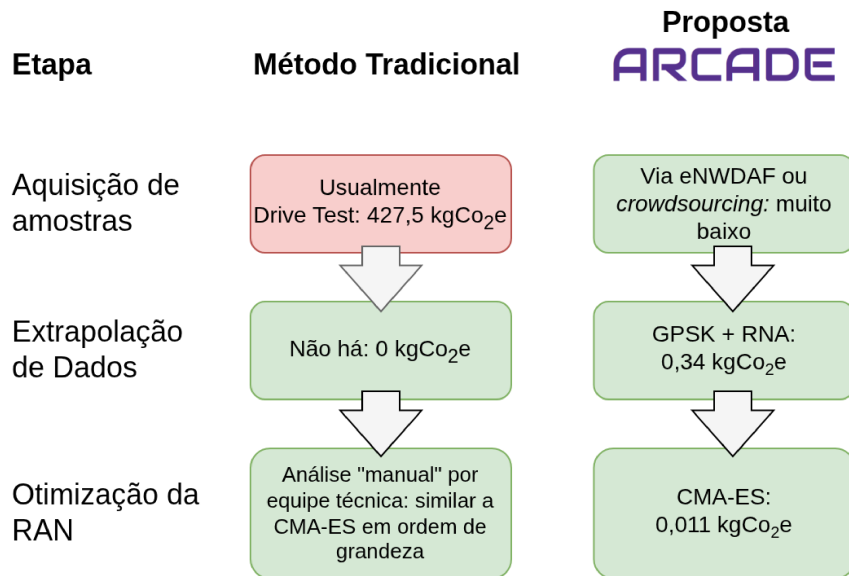


Figura 4.19 – Diagrama final comparativo das etapas do processo de *tuning* de RAN. Figura elaborada pelo autor.

bem como sobre o consequente aumento das emissões de carbono e dos impactos ambientais associados. O objetivo da Seção 4.7 foi a de realizar uma comparação do impacto nas emissões de carbono envolvidos na metodologia ARCADE com relação aos métodos tradicionalmente usados.

Essa comparação é complexa, devido às diferenças fundamentais de escopo na abordagem das etapas do processo de aprimoramento da cobertura da RAN. Porém, de uma forma mista qualitativa e quantitativa, e com a realização de uma comparação etapa a etapa, é possível verificar que o ARCADE não implica necessariamente em aumento significativo das emissões, podendo potencialmente até trazer uma redução das mesmas. O diagrama preliminar apresentado na Figura 4.18 pode ser atualizado conforme mostrado na Figura 4.19, com os dados calculados e estimados nesta seção.

Em síntese, a diferença entre o método tradicional de *tuning* de RAN e a proposta ARCADE concentra-se essencialmente na etapa de aquisição de amostras. Embora a coleta seja feita atualmente por meio de DTs, essa etapa poderia ser realizada via MDT ou *crowdsourcing* em qualquer uma das metodologias, eliminando por completo a discrepância observada quando se adota o DT no processo convencional. Dessa forma, não há diferença essencial entre as duas abordagens do ponto de vista de emissões de carbono. Consequentemente, pode-se afirmar que a abordagem ARCADE não implica impacto ambiental adicional significativo quando comparada aos métodos tradicionais de melhoria de rede.

4.8 Respostas às Questões de Pesquisa

As hipóteses e questões de pesquisa levantadas na Seção 1.3 foram comprovadas e respondidas ao longo do desenvolvimento deste trabalho. Isso foi realizado por meio da definição de conceitos no Capítulo 2, proposta da metodologia ARCADE no Capítulo 3 e da realização de análises quantitativas e experimentais no Capítulo 4. As respostas às principais questões de pesquisa são apresentadas a seguir:

□ **Q1:** *Como representar as amostras na geografia a ser analisada de uma forma contínua e organizada, de modo que caracterizem o ambiente de RF de forma acurada?*

R1: Para representar as amostras na geografia de forma contínua e organizada, propõe-se a construção de uma grade espacial regular composta por células de dimensões fixas de 50 metros por 50 metros, conforme mostrado na Seção 3.4 e, em especial, na Figura 3.4. Essa abordagem permite estruturar o espaço geográfico na forma de uma matriz bidimensional representando coordenadas de latitude e longitude, onde cada célula representa uma porção discreta e homogênea do ambiente físico. Essa abordagem discreta tem várias vantagens sobre a alternativa de um modelamento contínuo baseado em equações de interpolação, o que tornaria o modelo altamente complexo e de difícil manipulação.

Em cada célula da matriz será atribuída, de maneira exclusiva, uma e apenas uma amostra de potência do sinal de referência (ou RSRP) para cada PCI pertencente ao *cluster* em análise. O modelamento espacial obtido por meio dessa grade garante uma representação contínua do ambiente de RF, compatível com técnicas de visualização, extrapolação e análise de cobertura.

A escolha do tamanho do elemento da matriz (50 m × 50 m) foi motivada por um equilíbrio de densidade amostral. Na área analisada, tamanhos menores de grade poderiam levar a uma granularidade desnecessariamente pequena, impactando no esforço computacional para a extrapolação dos dados quanto na diminuição da densidade de amostras na área total. Por outro lado, elementos maiores tenderiam a mascarar variações locais importantes no ambiente de RF, como efeitos de sombra, obstruções por edifícios ou construções, bem como por outros fatores de origem topográfica ou morfológica, além de potencialmente gerar artefatos na representação da cobertura. As dimensões adotadas para o elemento da matriz se mostraram adequadas para captar variações espaciais relevantes em ambientes urbanos e suburbanos, ao mesmo tempo em que preservam a viabilidade computacional e a integridade estatística da base de dados. Cabe ressaltar que a decisão do tamanho do elemento da matriz tem que levar em consideração o tamanho da área total analisada, densidade amostral, morfologia, tamanho das células, além de outros fatores. Este trabalho considera que elementos de 50 m de aresta é geralmente bem adequado às aplicações

de macrocélulas em áreas urbanas e suburbanas, porém é interessante que se faça uma análise preliminar em outros casos de uso.

□ **Q2:** *Como, célula a célula, determinar-se quais amostras podem ser consideradas normais, quais são anomalias e quais correspondem a outliers?*

R2: A classificação das amostras em normais, anômalas ou *outliers*, célula a célula, é realizada com base em critérios estatísticos diretos e formais, conforme descrito em detalhes na Seção 3.4. O processo inicia-se com a definição do conjunto de amostras georreferenciadas pertencentes a cada célula da matriz espacial descrita anteriormente. Em seguida, para cada célula, determina-se o centróide da sua cobertura, calculado como o ponto médio das posições das amostras válidas, através de uma média aritmética simples, sem uso de qualquer ponderação.

Para cada amostra, calcula-se então a distância euclidiana ao centróide da célula correspondente usando-se o método de Haversine. Com base nesse conjunto de distâncias, obtêm-se a média d_m e o desvio padrão σ_m . A partir desses parâmetros, adota-se uma abordagem baseada em limiares múltiplos do desvio padrão para caracterizar o grau de aderência espacial de cada amostra em relação ao padrão central da célula.

A classificação segue a seguinte lógica:

- Amostras normais: distância ao centróide menor que $d_m + 3\sigma_m$;
- Amostras anômalas: distância entre $d_m + 3\sigma_m$ e $d_m + 6\sigma_m$;
- Amostras outliers: distância superior a $d_m + 6\sigma_m$.

Esse critério possibilita a identificação de erros relevantes de medição ou na base de dados, sobretudo aqueles que poderiam causar um desvio severo na modelagem baseada em RNA. Isso é evitado eliminando-se os *outliers* da base de amostras. Além disso, com a classificação das amostras, é possível delimitar-se as áreas de normalidade (A_N) e de anormalidade (A_A), de modo que os dados possam ser tratados de forma distinta em cada uma delas.

□ **Q3:** *Quais os melhores métodos para se realizar a extrapolação dos dados, inferindo-se valores ausentes em coordenadas onde não há amostras?*

R3: Dentre os diversos métodos avaliados para a extrapolação de dados em regiões sem amostras, o que apresentou melhor desempenho foi a utilização de processos gaussianos utilizando um *kernel* espacial do tipo RBF. Essa abordagem permitiu modelar a distribuição espacial do sinal, incorporando incertezas e variabilidades presentes nos dados reais, especialmente em contextos de propagação característicos de sinal celular, onde fenômenos como sombreamento, “canalização” e interferência causam variações locais não previsíveis.

A principal vantagem observada com o uso do GPSK com *kernel* RBF foi a sua capacidade de fornecer condições de contorno necessárias para evitar o *overfitting* no modelamento com RNAs e, ao mesmo tempo, preservar características locais de anomalias de cobertura, mantendo a integridade estrutural dos padrões observados durante a fase exploratória dos dados.

Para atingir esse equilíbrio entre suavidade e capacidade de representação das anomalias, foi necessário realizar um cuidadoso ajuste dos hiperparâmetros do modelo, sobretudo o termo de regularização adicionado à matriz de covariância *alpha* (α). O processo de calibração foi conduzido por meio de validação cruzada, utilizando métricas como o coeficiente de determinação R^2 e a raiz do erro quadrático médio (RMSE) como critérios de desempenho. Os melhores resultados foram obtidos quando ambos os indicadores atingiram valores ótimos, refletindo tanto boa capacidade preditiva quanto baixo erro esperado do modelo.

Além disso, a extrapolação foi auxiliada pela introdução de amostras aumentadas geradas na Área de Anormalidade A_A . A quantidade e posicionamento dessas amostras foram planejados considerando-se as distâncias entre as anomalias e a Área de Normalidade A_N . Essa estratégia foi essencial para manter a coerência do modelo em regiões limítrofes e evitar a propagação artificial de padrões que não condizem com a distribuição real do sinal.

Esse conjunto de abordagens possibilitou um mapeamento mais fidedigno da cobertura de sinal, especialmente em regiões com escassez de medições.

- **Q4:** *Como fazer com que as amostras anômalas não interfiram no modelo de extrapolação, exagerando na representação da cobertura, sem que sejam simplesmente descartadas?*

R4: A modelagem com GPSK mencionada na R3 viabiliza o aumento das amostras na A_A , de modo que é possível tanto o estabelecimento de condições de contorno adequadas (sobretudo delimitando as áreas sem cobertura nas bordas da Área Total A_T), quanto uma representação adequada das anomalias de cobertura causadas pelas amostras anômalas. A extrapolação através de GPSK e a metodologia apresentada para amostragem na A_A garantem que as anomalias sejam evidenciadas, porém de forma suave, ou seja, se estendendo através de curvas de decaimento gaussiano. Ainda assim, mantém-se as características de anomalia, sem que haja descarte de amostras relevantes.

- **Q5:** *Dado o mapa de cobertura contínuo do cluster de células, na forma de um banco de dados estruturado que representa a configuração do sistema, como quantificar a qualidade do serviço prestado por esse cluster na área avaliada, de forma a se poder avaliar seu grau de eficiência?*

R5: A qualidade do serviço na área avaliada pode ser avaliada com base em três termos bem definidos, conforme definido na Seção 3.7.1:

- Índice de Cobertura (C_i): quantifica, para cada elemento da matriz que representa a Área Total, a qualidade de cobertura no elemento i , refletindo a potência do sinal recebido nesse ponto pela célula servidora do mesmo, sendo que valores próximos a zero indicam baixo nível de cobertura e valores próximos à unidade representam níveis bons ou ótimos de cobertura;
- Índice de Dominância (D_i): quantifica, também para cada elemento da matriz, a dominância da servidora no ponto, ou seja, a diferença em dB da potência recebida pela servidora e a potência recebida pela interferente dominante (segundo maior RSRP no ponto), sendo valores próximos a zero indicativos de baixa dominância e valores próximos a um, demonstrativos de uma grande dominância da servidora, ou seja, baixa interferência da interferente dominante;
- Índice de Distribuição Espacial (SDI): mede o quão similares são os tamanhos das células. Essa característica é altamente desejada, para que haja uma distribuição uniforme de tráfego da rede, de modo que não se sobrecarregue células de maior cobertura em detrimento às de menor cobertura. A variância de SDI representa fisicamente o grau de diferença de suas coberturas no *cluster*.

Uma vez definidos esses três termos, foi proposta na Seção 3.7.1 uma função objetivo $\tilde{J}(P)$, função do vetor P de ajustes de potências das células, que inclui, além dos três termos acima citados, um quarto termo de regularização, que penaliza desvios elevados nos ajustes de potência individuais e favorece soluções fisicamente mais realistas e viáveis do ponto de vista operacional de uma rede real.

De maneira geral, a função objetivo $\tilde{J}(P)$ pode ser considerada uma forma consolidada e consistente de se quantificar a qualidade do serviço prestado em uma área, baseado nos termos de cobertura, dominância e homogeneidade da cobertura das células.

- **Q6:** *Qual abordagem de modelagem e avaliação de desempenho é mais adequada para permitir que o sistema determine ajustes em seus parâmetros, de modo a maximizar a qualidade do serviço na área atendida pelo cluster de células?*

R6: Foi proposto um método baseado na técnica CMA-ES, formulado como um problema de otimização não linear e não convexa, no qual os ajustes de potência das células pertencentes ao *cluster* são tratados como variáveis de decisão. A função objetivo $\tilde{J}(P)$ foi definida de modo a maximizar a qualidade do serviço na área analisada, conforme já mencionado na questão anterior.

O CMA-ES foi escolhido por sua robustez na busca em espaços de alta dimensionalidade e por sua rápida convergência e – conforme demonstrado nos experimentos

– baixa variância nos resultados gerados por *seeds* aleatórias, apesar de sua natureza estocástica. Dessa forma, o método viabiliza um mecanismo otimização não-exaustiva capaz de convergir para configurações paramétricas que melhorem significativamente a qualidade da cobertura na região de interesse, conforme demonstrado nos experimentos realizados.

O problema de aprimoramento da cobertura foi então abordado com essa técnica, visando maximizar $\tilde{J}(P)$, considerando uma ponderação dos termos que favorece o critério de maximização da área com cobertura e boa dominância. De fato, o termo D_i foi o que apresentou maior peso relativo dentre os termos componentes da função objetivo, após vários testes de calibração desses pesos.

Os experimentos apresentados mostraram que o método CMA-ES usado para se maximizar a função objetivo $\tilde{J}(P)$ resultou em um aumento significativo da qualidade de cobertura da área pretendida, de forma que essa abordagem mostrou-se eficaz na solução do problema proposto.

Considerações Finais

Ao mesmo tempo em que o tema de auto-organização de uma rede de acesso celular é bastante explorada em trabalhos acadêmicos ou mesmo implementada em sistemas proprietários fornecidos por *vendors* tradicionais de tecnologia, verifica-se que, em sua grande maioria, essas soluções são de configurações bastante complexas. Normalmente, a implementação desse tipo de sistema é custosa e trabalhosa, além de tender a ser proprietária de um *vendor* em especial, devido ao grande volume de integrações a sistemas específicos.

Em contraponto, a proposta do ARCADE traz uma abordagem simplificada, direta, agnóstica a fornecedor de tecnologia e de relativa simples implementação. Nos experimentos, mostrou-se uma solução eficaz para análise e detecção de anomalias de cobertura, sem o trabalho e custos adicionais que a implementação de sistemas análogos implica, observando-se a importante premissa que o sistema já tenha sido ajustado em um processo de *tuning* dos parâmetros físicos.

O desenvolvimento desse trabalho mostra que os principais objetivos inicialmente propostos foram atingidos. Na Seção 1.2, foi estabelecido como objetivo principal apresentar uma metodologia para aprimoramento da cobertura celular, visando uma futura implementação autônoma, contando apenas com dados georreferenciados de níveis de cobertura por célula, sem o auxílio de informações adicionais e sem a utilização de modelos empíricos de propagação. Para chegar ao modelo proposto, foi necessário estabelecer uma metodologia formal de análise, conforme apresentado na Seção 3.1, composta por módulos específicos para a aquisição de dados, o pré-processamento de dados, o aumento de amostragem, o treinamento do modelo e, finalmente, para a melhoria do *cluster*. O processo é sequencial, sendo que, para todos os módulos com exceção do primeiro e do último, todos os passos são feitos a célula a célula.

A etapa de aumento de amostragem e delimitação de condições de contorno de cobertura por célula se mostrou imprescindível para um posterior modelamento e generalização das coberturas. Os experimentos realizados mostraram uma boa efetividade nessa generalização, que forneceu mapas de cobertura muito compatíveis com os dados medidos e

coerentes com o observado na prática, sobretudo considerando que as anomalias tendem a causar o *overfitting* do modelo de forma a inviabilizar uma extrapolação dos dados diretamente por uma RNA.

Por fim, o método CMA-ES mostrou-se muito eficaz e eficiente para a tarefa de melhoria de desempenho, baseando-se na função objetivo desenhada e na calibração dos parâmetros de pesos da mesma. Foram obtidos ganhos muito significativos tanto de cobertura com dominância quanto de cobertura total. O método, apesar de sua natureza estocástica, apresentou pouca variação em seus resultados, usando *seeds* distintas. Também o esforço computacional demandado pela execução do código que implementou o método foi excepcionalmente baixo, principalmente considerando-se que o *hardware* utilizado para todos os experimentos foi uma máquina convencional, sem suporte a aceleração por hardware, como *Graphics Processing Unit* (GPU) ou dispositivos dedicados a processamento paralelo.

Como resultado geral, entende-se que a contribuição deste trabalho não se limita apenas à proposição de uma nova metodologia, mas também à demonstração de que é possível abordar o problema de aprimoramento da qualidade de cobertura de forma pragmática, com baixo custo computacional e sem dependência de integrações complexas. Ao não se utilizar modelos empíricos clássicos de propagação e de bases externas de engenharia, o método concentra-se exclusivamente nas informações efetivamente disponíveis na rede, explorando-as de forma a trazer resultados concretos e tangíveis. Essa característica confere à proposta não apenas viabilidade técnica, mas também potencial aplicabilidade prática em cenários operacionais, especialmente naqueles que demandam simplicidade, rapidez de implementação e independência tecnológica.

5.1 Principais Contribuições

Este trabalho apresenta várias contribuições de ordens práticas e metodológicas para o estudo, análise e diagnóstico da cobertura de redes celulares, especialmente no que se refere ao processo de coleta de dados da RAN, ao tratamento de dados de medições obtidas (independente da forma em que foram obtidas) e à definição de um método que permita a melhoria da qualidade de cobertura de um *cluster* de células que se sobrepõem, complementam e interferem entre si. As principais contribuições, nesse sentido, são as descritas a seguir:

- *Desenvolvimento de Modelos de Interpolação Espacial*: Foi proposto um método de interpolação geográfica dos dados de cobertura das células em um *cluster* baseado no aumento de dados via GPSK e RNAs. O foco foi manter um equilíbrio entre a sensibilidade do modelo e a capacidade de generalização, o que apresenta um desafio significativo. A abordagem proposta previne o descarte de dados de anomalias reais – que podem indicar falhas na cobertura de uma célula –, ao mesmo tempo que

evita o *overfitting*, permitindo uma expansão dos dados limitados a uma grande área geográfica e viabilizando a representação da sobreposição das coberturas.

- *Consolidação de Base de Dados de Ambiente de RAN Real*: Foi consolidada uma base de dados consistindo de amostras de RF georreferenciadas coletadas de uma rede real e operacional, consistente com o observado na prática em uma cidade média brasileira, com um misto de morfologias rural, suburbana e urbana densa, representando bem um caso típico do Brasil. A base foi analisada e tratada, removendo-se todos os dados inconsistentes (erros de coleta) e duplicados (mais de uma medida em uma mesma coordenada geográfica), sendo esses últimos consolidados como a média das medições. Também foi realizada uma normalização das coordenadas e da potência recebida, de forma que possa ser diretamente usada em modelos baseados em RNA. O refinamento da base de dados permite a realização de análises geográficas mais confiáveis e representativas da realidade da cobertura real de um sistema celular típico de uma cidade média brasileira. A base original utilizada, já com as coordenadas deslocadas para efeito de anonimização, está publicamente disponível na pasta /data no GitHub do projeto ARCADE em <<https://github.com/drcoliveira/ARCADE>>.
- *Proposição de Método para Aprimoramento de Rede de Acesso de Rádio*: Foi estabelecido um método para viabilizar a melhoria de um *cluster* de células de forma que não se necessite de informações sobre o sistema advindos de fontes externas. Para isso, foi definida uma função objetivo e os pesos de seus termos foram calibrados de modo a se ter uma boa eficiência da solução. O método, baseado na técnica de otimização CMA-ES, mostrou-se muito eficiente tanto em termos de resultados obtidos quanto em termos de custo computacional.
- *Ferramentas Computacionais*: Ao longo do trabalho, foram desenvolvidas diversas ferramentas em Python que realizam as etapas de leitura, tratamento, interpolação, visualização e análise estatística dos dados. Cada módulo foi desenvolvido de forma independente, tornando-os mais facilmente adaptáveis a diferentes contextos e facilitando sua utilização em outros casos de uso futuros. Os códigos usados estão publicamente disponíveis no GitHub do projeto ARCADE em <<https://github.com/drcoliveira/ARCADE>>.

5.2 Contribuições em Produção Bibliográfica

Essa seção apresenta a produção bibliográfica gerada através da concepção, análise e experimentos realizados ao longo do desenvolvimento desse trabalho. As contribuições geradas foram:

- ❑ Oliveira, D. R. C.; Moreira, R.; Rosa, P. F.; Silva, F. de O. *6G Architecture Design: Evolution or Disruption?*. In: II Workshop de Redes 6G (W6G 2022) do XL Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Fortaleza, Brasil, 2022. DOI: <<https://doi.org/10.5753/w6g.2022.223561>>. **Prêmio de Melhor Artigo do Workshop W6G 2022.**

- ❑ Souza Neto, N. V. de; Gonçalves, M. A.; Oliveira, D. R. C.; Molinos, D. N.; Moreira, R.; Silva, F. de O. *Evolved NWDAF Towards a Fully Distributed Artificial Intelligence in the 6G Network Architecture*. In: IV Workshop de Redes 6G (W6G 2024) do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Niterói, Brasil, 2024. DOI: <<https://doi.org/10.5753/w6g.2024.3378>>.

- ❑ Oliveira, D. R. C.; Silva, F. de O. *A Method for Cellular Coverage Modeling in the Presence of Anomalies Using Neural Networks and Gaussian Processes*. In: The 39th International Conference on Advanced Information Networking and Applications (AINA-2025), Barcelona, Espanha, 2025. DOI: <https://doi.org/10.1007/978-3-031-87763-6_28>.

- ❑ Oliveira, D. R. C.; Silva, F. de O.; Moreira, R. *ARCADE: A RAN Diagnosis Methodology in a Hybrid AI Environment for 6G Networks*. In: V Workshop de Redes 6G (W6G 2025) do XLIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Natal, Brasil, 2025. DOI: <<https://doi.org/10.5753/w6g.2025.9532>>.

- ❑ Oliveira, D. R. C.; Silva, F. de O. *Análise e Detecção de Anomalias de Cobertura em Redes Celulares*. In: XIX Workshop de Teses e Dissertações em Ciência da Computação (WTDCC 2025), Uberlândia, Brasil, 2025. Site: <<https://techweek.facom.ufu.br/wtdcc-2025>>. **Prêmio “Best paper” de pós-graduação do WTDCC 2025.**

- ❑ Oliveira, D. R. C.; Faria, D. R.; Silva, F. de O. *CMA-ES-Based Self-Optimization for Mobile Networks Without Network Planning Information*. Artigo submetido para avaliação à revista *IEEE Wireless Communications Letters*, em março de 2026.

5.3 Trabalhos Futuros

A pesquisa desenvolvida ao longo desta tese permitiu consolidar uma metodologia consistente para a análise, modelagem e melhoria de cobertura em redes móveis com base em dados reais e limitações das informações de entrada. Ainda assim, há diversas questões que podem ser aprofundadas, refinadas ou exploradas sob outras abordagens teóricas e experimentais, sendo algumas propostas identificadas apresentadas a seguir.

- *Extensão do Modelo para Cenários 5G SA*: Os dados disponíveis – e que permitiram o modelamento apresentado neste trabalho – foram baseados em uma rede LTE em uma banda considerada média, de 1.800 MHz. O 5G, sobretudo em sua evolução 5G *standalone*, traz oportunidades de refinamento da pesquisa, especialmente nas seguintes linhas:
- Inclusão de *beamforming* nas análises: as antenas mais modernas, sobretudo as usadas no 5G (embora não restritas a essa geração), têm a capacidade de adequar seus diagramas de radiação de acordo com a localização espacial dos usuários. O uso de *beamforming* nas antenas demandará uma investigação mais profunda e, eventualmente, adequações nos modelos matemáticos usados na modelagem e aprimoramento das redes de acesso;
 - Inclusão do caso de uso de *Massive Multiple Input Multiple Output* (mMIMO): muito comum no 5G, mas também presente em algumas redes 4G, as antenas com capacidade mMIMO implementam múltiplos feixes espaciais independentes, viabilizando um aumento significativo da eficiência espectral das RBSs. Há, igualmente, um impacto no modelamento tanto do ambiente de RF quanto do método de aprimoramento do sistema, que devem ser avaliados e adequados a esse contexto;
 - Validação dos modelos em frequências mais altas: em virtude de suas maiores larguras de banda e conseqüente capacidade de tráfego, as faixas atualmente mais usadas no 5G (e futuramente no 6G) são mais altas que as de banda média e baixa do 4G, variando de 3,5 GHz – muito utilizada no 5G – à ainda incipiente faixa de 6 GHz. Há oportunidades de investigação para verificação se os modelos são compatíveis também para essas faixas e certamente a necessidade de nova calibração dos pesos da função de otimização para cada faixa em específico;
 - Novo modelamento considerando *Frequency Range 2* (FR2): no 5G, foram introduzidas novas faixas, denominadas ondas milimétricas, oficialmente chamadas FR2. São frequências muito altas, entre 24 e 71 GHz, que operam em condições de propagação muito diferentes das ondas em *Frequency Range 1* (FR1). Muito provavelmente um novo modelo deverá ser desenvolvido para esse caso em específico, especialmente em ambientes de altíssima densidade, como é um dos casos de uso mais típicos desse tipo de faixa.
- *Extensão do Modelo para Suporte a Redes Heterogêneas*: conforme mencionado na Seção 3.2, as redes heterogêneas não são suportadas pelos modelos definidos neste trabalho. O conceito de sistemas com múltiplos níveis envolve células de camadas inferiores sendo totalmente sobrepostas por células de camadas superiores, de forma

que a definição de cobertura com dominância como principal atributo na função de otimização não pode ser aplicado diretamente aos casos de micro, pico e femtocélulas. Desta forma, no estado atual da metodologia ARCADE, amostras de dados vindas de células de camadas inferiores devem ser simplesmente ignoradas para que não haja inconsistência no modelamento do ambiente de RF. Uma proposta de trabalho futuro pode envolver uma classificação automática (i.e., sem a intervenção de sistemas externos) das células em duas ou mais camadas e prever o caso de redes heterogêneas na melhoria do sistema.

- *Extensão do Modelo para Detecção de Anomalias de Cobertura:* embora o modelo já preveja a melhoria da qualidade de cobertura da rede de uma forma geral, pode ser muito útil que o mesmo também identifique e apresente regiões de anomalias de cobertura, tipicamente causadas por *overshooting* de células. Essas anomalias podem ser identificadas como artefatos, concentração de amostras anômalas ou inconsistência de cobertura. A delimitação dessas regiões pode auxiliar os engenheiros ou técnicos responsáveis pelo desempenho da rede a avaliarem alternativas de melhoria baseadas em outras técnicas que não o ajuste da potência da célula, como alteração de configurações de *downtilt* ou azimute de antenas. Para isso, técnicas como DBSCAN podem ser usadas para agregar áreas específicas, caracterizando uma região anômala.
- *Agregação de Novas Grandezas aos Dados de Entrada e Modelamento:* na base de dados de entrada, além das coordenadas geográficas e RSRP, também há diversas grandezas a serem exploradas, dependendo do método de coleta. Por exemplo, informações de RSRQ, *timestamp*, tipo de evento (no caso de sinalização do tipo MR), EARFCN, além de outros dados que podem ser obtidos dos SOs das UEs (no caso de amostras obtidas por *crowdsourcing*), podem ser agregadas aos modelos de forma a aprimorá-lo ou adicionar novas funcionalidades à metodologia.

Referências

- 3GPP. *3GPP TS 36.211 V17.6.0: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation*. [S.l.], 2024. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/36_series/36.211/36211-i01.zip>.
- 3GPP. *3GPP TS 36.214 V17.6.0: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements*. [S.l.], 2024. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/36_series/36.214/36214-i00.zip>.
- 3GPP. *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2*. [S.l.], 2024.
- 3GPP. *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication Management; Self-Organizing Networks (SON); Concepts and requirements (Release 18)*. [S.l.], 2024.
- 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*. [S.l.], 2024. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/36_series/36.331/36331-i40.zip>.
- 3GPP. *Vocabulary for 3GPP Specifications*. [S.l.], 2024. Release 18. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/21_series/21.905/>.
- 3GPP. *3GPP TS 38.211 V17.6.0: NR; Physical Channels and Modulation*. [S.l.], 2025. Disponível em: <https://www.3gpp.org/ftp/Specs/archive/38_series/38.211/38211-i50.zip>.
- ABOUELMAATI, D. et al. Empowering Next-Gen Networks: AI-Driven Autonomy in O-RAN and SON Architectures. **IEEE Access**, v. 13, p. 186903–186936, 2025. ISSN 2169-3536. Disponível em: <<https://doi.org/10.1109/ACCESS.2025.3626452>>.
- AKBARI, I. et al. How reliable is mdt-based autonomous coverage estimation in the presence of user and bs positioning error? **IEEE Wireless Communications Letters**, v. 5, n. 2, p. 196–199, abr. 2016. Disponível em: <<https://doi.org/10.1109/LWC.2016.2516989>>.
- ASSUNÇÃO, M. A. d. **Indicadores de mobilidade urbana sustentável para a cidade de Uberlândia, MG**. Dissertação (Mestrado) — Universidade Federal de Uberlândia, Uberlândia, 2012. Disponível em: <<https://doi.org/10.14393/ufu.di.2012.172>>.

- CARVALHO, C. H. R. de. **Emissões relativas de poluentes do transporte motorizado de passageiros nos grandes centros urbanos brasileiros**. Brasília, 2011. Disponível em: <<https://repositorio.ipea.gov.br/handle/11058/1443>>.
- CHEERLA, S.; RATNAM, D. V.; BORRA, H. S. Neural network-based path loss model for cellular mobile networks at 800 and 1800 MHz bands. **AEU - International Journal of Electronics and Communications**, v. 94, p. 179–186, set. 2018. ISSN 14348411. Disponível em: <<https://doi.org/10.1016/j.aeue.2018.07.007>>.
- CIOCARLIE, G. F. et al. Detecting anomalies in cellular networks using an ensemble method. In: **Proceedings of the 9th International Conference on Network and Service Management (CNSM 2013)**. Zurich, Switzerland: IEEE, 2013. p. 171–174. ISBN 978-3-901882-53-1. Disponível em: <<https://doi.org/10.1109/CNSM.2013.6727831>>.
- Climate Transparency. **Climate Transparency Report 2022: Brazil**. 2022. Acessado em 14 de novembro de 2025. Disponível em: <<https://www.climate-transparency.org/countries/americas/brazil>>.
- CORMEN, T. H. et al. **Introduction to Algorithms**. 3rd. ed. Cambridge, MA: MIT Press, 2009. 1312 p. ISBN 9780262533058.
- CORRÊA, F. **Proposta de um método de identificação de anomalias em redes móveis com base na representação dimensional de KPIs usando PCA e clustering**. Dissertação (Mestrado) — Universidade Federal de Uberlândia, Uberlândia, MG, Brasil, fev. 2020. Disponível em: <<http://doi.org/10.14393/ufu.di.2020.264>>.
- DAUNI, P. et al. Implementation of Haversine formula for school location tracking. **Journal of Physics: Conference Series**, v. 1402, n. 7, p. 077028, dez. 2019. ISSN 1742-6588, 1742-6596. Disponível em: <<https://doi.org/10.1088/1742-6596/1402/7/077028>>.
- DENG, J. et al. A Digital Twin Approach for Self-optimization of Mobile Networks. In: **2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)**. Nanjing, China: IEEE, 2021. p. 1–6. ISBN 978-1-7281-9507-0. Disponível em: <<https://doi.org/10.1109/WCNCW49093.2021.9420037>>.
- Diário de Uberlândia. Por mais e melhores ciclovias. **Diário de Uberlândia**, July 2018. Disponível em: <<https://diariodeuberlandia.com.br/noticia/17392/por-mais-e-melhores-ciclovias>>.
- DREIFUERST, R. M. et al. Optimizing Coverage and Capacity in Cellular Networks using Machine Learning. In: **ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. Toronto, ON, Canada: IEEE, 2021. p. 8138–8142. ISBN 978-1-7281-7605-5. Disponível em: <<https://doi.org/10.1109/ICASSP39728.2021.9414155>>.
- ERICSSON. **Ericsson Mobility Report November 2025**. 2025. Disponível em: <<https://www.ericsson.com/4aca6f/assets/local/reports-papers/mobility-report/documents/2025/ericsson-mobility-report-november-2025.pdf>>.
- FAGEN, D.; VICHARELLI, P.; WEITZEN, J. Automated Wireless Coverage Optimization With Controlled Overlap. **IEEE Transactions on Vehicular**

Technology, v. 57, n. 4, p. 2395–2403, jul. 2008. ISSN 0018-9545, 1939-9359. Disponível em: <<https://doi.org/10.1109/TVT.2007.912329>>.

FAN, S.; TIAN, H.; SENGUL, C. Self-optimization of coverage and capacity based on a fuzzy neural network with cooperative reinforcement learning. **Journal of Wireless Communications and Networking**, v. 2014, n. 1, p. 57, 2014. Disponível em: <<https://doi.org/10.1186/1687-1499-2014-57>>.

GHOSH, R. **Wireless Networking and Mobile Data Management**. Singapore: Springer Singapore, 2017. ISBN 978-981-10-3940-9 978-981-10-3941-6. Disponível em: <<https://doi.org/10.1007/978-981-10-3941-6>>.

HANSEN, N. **The CMA Evolution Strategy: A Tutorial**. arXiv, 2023. ArXiv:1604.00772 [cs]. Disponível em: <<http://doi.org/10.48550/arXiv.1604.00772>>.

HANSEN, N.; OSTERMEIER, A. Completely Derandomized Self-Adaptation in Evolution Strategies. **Evolutionary Computation**, v. 9, n. 2, p. 159–195, jun. 2001. ISSN 1063-6560, 1530-9304. Disponível em: <<https://doi.org/10.1162/106365601750190398>>.

HAYKIN, S. S. **Neural networks and learning machines**. 3. ed. ed. New York Munich: Prentice-Hall, 2009. ISBN 978-0-13-147139-9.

HOLMA, H.; TOSKALA, A. **LTE for UMTS: OFDMA and SC-FDMA based radio access**. Chichester, U.K: J. Wiley & Sons, 2009. ISBN 978-0-470-99401-6.

_____. **LTE for UMTS: Evolution to LTE-Advanced**. [S.l.]: Wiley, 2011. ISBN 978-0470660005.

HU, Y. et al. **Reflection-Driven Self-Optimization 6G Agentic AI RAN via Simulation-in-the-Loop Workflows**. arXiv, 2025. ArXiv:2512.20640 [cs]. Disponível em: <<https://doi.org/10.48550/arXiv.2512.20640>>.

HUANG, Y. et al. AI Empowered Modeling, Closed-Loop Optimization, and Field Trials of RAN Digital Twin. **IEEE Network**, v. 40, n. 1, p. 154–164, jan. 2026. ISSN 0890-8044, 1558-156X. Disponível em: <<https://doi.org/10.1109/MNET.2025.3565716>>.

Huawei Technologies Co., Ltd. **eRAN Cell Management Feature Parameter Description**. Shenzhen, China, 2018.

INMETRO. **Tabela PBE Veicular 2025**. 2025. Programa Brasileiro de Etiquetagem Veicular (PBEV). Atualizada em 24 de novembro de 2025 com dados de 895 modelos. Disponível em: <<https://www.gov.br/inmetro/pt-br/assuntos/avaliacao-da-conformidade/programa-brasileiro-de-etiquetagem/tabelas-de-eficiencia-energetica/veiculos-automotivos-pbe-veicular>>.

ISLAM, M. N. ul; MITSCHELE-THIEL, A. Cooperative fuzzy q-learning for self-organized coverage and capacity optimization. **2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)**, p. 1406–1411, 2012. Disponível em: <<https://doi.org/10.1109/PIMRC.2012.6362568>>.

JESKE, M. et al. Enhancing Reference Signal Received Power Prediction Accuracy in Wireless Outdoor Settings: A Comprehensive Feature Importance Study. **IEEE Transactions on Antennas and Propagation**, v. 73, n. 10, p. 8022–8037, out. 2025. ISSN 0018-926X, 1558-2221. Disponível em: <<https://doi.org/10.1109/TAP.2025.3576492>>.

JIANG, W. et al. Coverage Prediction in Mobile Communication Networks: A Deep Learning Approach With a Tabular Foundation Model. **Internet Technology Letters**, v. 8, n. 3, p. e70034, maio 2025. ISSN 2476-1508, 2476-1508. Disponível em: <<https://doi.org/10.1002/itl2.70034>>.

KAWAMOTO, R. et al. Estimation of CO₂ emissions of internal combustion engine vehicle and battery electric vehicle using LCA. **Sustainability**, v. 11, n. 9, p. 2690, 2019. Disponível em: <<https://doi.org/10.3390/su11092690>>.

KHAN, F. **LTE for 4G Mobile Broadband: Air Interface Technologies and Performance**. [S.l.]: Cambridge University Press, 2009. ISBN 978-0-521-88221-7.

KLAINÉ, P. V. et al. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. **IEEE Communications Surveys & Tutorials**, v. 19, n. 4, p. 2392–2431, 2017. ISSN 1553-877X. Disponível em: <<https://doi.org/10.1109/COMST.2017.2727878>>.

LI, J. et al. Self-optimization of coverage and capacity in lte networks based on central control and decentralized fuzzy q-learning. **International Journal of Distributed Sensor Networks**, v. 8, n. 8, p. 878595, 2012. Disponível em: <<https://doi.org/10.1155/2012/878595>>.

LIU, G. et al. 6G autonomous radio access network empowered by artificial intelligence and network digital twin. **Frontiers of Information Technology & Electronic Engineering**, v. 26, n. 2, p. 161–213, fev. 2025. ISSN 2095-9184, 2095-9230. Disponível em: <<https://doi.org/10.1631/FITEE.2400569>>.

MCTI. **Fator de emissão de CO₂ na geração de energia elétrica no Brasil em 2023 é o menor em 12 anos**. 2024. <<https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/02/fator-de-emissao-de-co2-na-geracao-de-energia-eletrica-no-brasil-em-2023-e-o-menor-em-12-anos>>. Publicado em 22 de fevereiro de 2024.

MERA, Z. et al. **Comparison of the Life-Cycle Greenhouse Gas Emissions of Combustion Engine and Electric Passenger Cars in Brazil**. [S.l.], 2023. Disponível em: <<https://theicct.org/publication/comparison-of-life-cycle-ghg-emissions-of-combustion-engines-and-electric-pv-brazil-oct23/>>.

MORADI, R.; BERANGI, R.; MINAEI, B. A survey of regularization strategies for deep models. **Artificial Intelligence Review**, v. 53, n. 6, p. 3947–3986, ago. 2020. ISSN 0269-2821, 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-019-09784-7>>.

MOYSEN, J.; GIUPPONI, L. From 4g to 5g: Self-organized network management meets machine learning. **Computer Communications**, v. 129, p. 248–268, set. 2018. Disponível em: <<https://doi.org/10.1016/j.comcom.2018.07.015>>.

NGUYEN, H. X. et al. Digital Twin for 5G and Beyond. **IEEE Communications Magazine**, v. 59, n. 2, p. 10–15, fev. 2021. ISSN 0163-6804, 1558-1896. Disponível em: <<https://doi.org/10.1109/MCOM.001.2000343>>.

NOKIA. **Nokia AVA helps Hutchison 3 increase network efficiency and improve the customer experience in Indonesia**. 2018. Acessado em 22 de fevereiro de 2025. Disponível em: <<https://www.nokia.com/about-us/news/releases/2018/11/05/nokia-ava-helps-hutchison-3-increase-network-efficiency-and-improve-the-customer-experience-in-indonesia/>>.

_____. **AI and Analytics: AVA Data Suite**. 2025. Acessado em 18 de fevereiro de 2025. Disponível em: <<https://www.nokia.com/ai-and-analytics/ava-data-suite/>>.

NOWAK, D. J. et al. Carbon storage and sequestration by trees in urban and community areas of the United States. **Environmental Pollution**, v. 178, p. 229–236, 2013. Disponível em: <<https://doi.org/10.1016/j.envpol.2013.03.019>>.

NWELIH, E.; ISABONA, J.; IMOIZE, A. L. Optimization of Base Station Placement in 4G LTE Broadband Networks Using Adaptive Variable Length Genetic Algorithm. **SN Computer Science**, v. 4, n. 2, p. 121, dez. 2022. ISSN 2661-8907. Disponível em: <<https://doi.org/10.1007/s42979-022-01533-y>>.

OJO, S.; IMOIZE, A.; ALIENYI, D. Radial basis function neural network path loss prediction model for LTE networks in multitransmitter signal propagation environments. **International Journal of Communication Systems**, v. 34, n. 3, p. e4680, fev. 2021. ISSN 1074-5351, 1099-1131. Disponível em: <<https://doi.org/10.1002/dac.4680>>.

Ookla. **About Ookla**. 2025. <<https://www.ookla.com/about>>. Acessado em 27 de maio de 2025.

_____. **Ookla Cell Maps**. 2026. Accessed: 2026-01-22. Disponível em: <<https://www.ookla.com/cell-maps>>.

OpenSignal. **About Us - OpenSignal**. 2025. <<https://www.opensignal.com/about-us>>. Acessado em 27 de maio de 2025.

Opensignal. **Our Approach to Mobile Network Experience Measurement**. 2026. Accessed: 2026-01-22. Disponível em: <<https://www.opensignal.com/our-approach>>.

PCTEL. **IBflex Brochure**. 2020. Accessed: 2025-03-24. Disponível em: <<https://www.pctel.com/wp-content/uploads/2020/07/IBflex-Brochure-web.pdf>>.

QIU, Y.; BOSE, A. Machine Learning for Modeling Wireless Radio Metrics with Crowdsourced Data and Local Environment Features. In: **2025 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting (AP-S/CNC-USNC-URSI)**. Ottawa, ON, Canada: IEEE, 2025. p. 507–510. ISBN 979-8-3315-2367-1. Disponível em: <<https://doi.org/10.1109/AP-S/CNC-USNC-URSI55537.2025.11266258>>.

RAPPAPORT, T. S. **Wireless communications: principles and practice**. 2. ed., 18. printing. ed. Upper Saddle River, NJ: Prentice Hall, 2009. (Prentice Hall communications engineering and emerging technologies series). ISBN 978-0-13-042232-3.

RASMUSSEN, C. E.; WILLIAMS, C. K. I. **Gaussian processes for machine learning**. 3. print. ed. Cambridge, Mass.: MIT Press, 2008. (Adaptive computation and machine learning). ISBN 978-0-262-18253-9.

RAZAVI, R.; KLEIN, S.; CLAUSSEN, H. A fuzzy reinforcement learning approach for self-optimization of coverage in LTE networks. **Bell Labs Technical Journal**, v. 15, n. 3, p. 153–175, dez. 2010. ISSN 10897089. Disponível em: <<https://doi.org/10.1002/bltj.20463>>.

REKKAS, V. et al. Machine Learning-Based Radio Environment Maps for 4G/5G Networks. **IEEE Transactions on Antennas and Propagation**, v. 73, n. 12, p. 10529–10543, dez. 2025. ISSN 0018-926X, 1558-2221. Disponível em: <<https://doi.org/10.1109/TAP.2025.3607974>>.

SAMSUNG. **Vista - Network Automation**. 2025. Acessado em 22 de fevereiro de 2025]. Disponível em: <<https://www.samsung.com/global/business/networks/products/network-automation/vista/>>.

SCHÖLKOPF, B. **Learning with kernels: support vector machines, regularization, optimization, and beyond**. Cambridge, Mass: MIT Press, 2002. (Adaptive computation and machine learning). ISBN 978-0-262-19475-4 978-0-262-25693-3 978-0-585-47759-6.

scikit-learn developers. **Gaussian process regression (GPR)**. 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html>. Accessed: 2025-09-04.

_____. **ConstantKernel — scikit-learn documentation**. 2024. Acesso em: mar. 2026. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.ConstantKernel.html>.

_____. **RBF — scikit-learn documentation**. 2024. Acesso em: mar. 2026. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.RBF.html>.

SEMENCHIKOV, D. et al. Evolutionary algorithms in high-dimensional radio access network optimization. In: **Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO Companion)**. Lille, France: [s.n.], 2021. p. 1834–1841. Disponível em: <<https://doi.org/10.1145/3449726.3463216>>.

SEZIA, S.; TOUFIK, I.; BAKER, M. P. J. **LTE, the UMTS long term evolution: from theory to practice**. Chichester, U.K: Wiley, 2009. ISBN 978-0-470-69716-0.

SINGH, Y. Comparison of Okumura, Hata and COST-231 Models on the Basis of Path Loss and Signal Strength. **International Journal of Computer Applications**, v. 59, 2012. Disponível em: <<https://doi.org/10.5120/9594-4216>>.

SIVERTSEN, K. N.; LIOU, A.; MICHELSON, D. G. Depth and Rate of Fading on Fixed Wireless Channels Between 200 MHz and 2 GHz in Suburban Macrocell Environments. **IEEE Transactions on Antennas and Propagation**, v. 58, n. 10, p. 3353–3362, out. 2010. ISSN 0018-926X. Disponível em: <<https://doi.org/10.1109/TAP.2010.2055783>>.

SKOCAJ, M. et al. Cellular Network Capacity and Coverage Enhancement with MDT Data and Deep Reinforcement Learning. **Computer Communications**, v. 195, p. 403–415, nov. 2022. ISSN 01403664. Disponível em: <<https://doi.org/10.1016/j.comcom.2022.09.005>>.

TANG, Z. et al. **Multi-Output Gaussian Process-Based Data Augmentation for Multi-Building and Multi-Floor Indoor Localization**. arXiv, 2023. ArXiv:2202.01980 [cs]. Disponível em: <<https://doi.org/10.48550/arXiv.2202.01980>>.

TSOULOS, V.; TSOULOS, G. V.; ATHANASIADOU, G. Machine Learning-Based in-Car KPI Predictions for Cellular Networks. In: **2025 14th International Conference on Modern Circuits and Systems Technologies (MOCASST)**. Dresden, Germany: IEEE, 2025. p. 1–6. ISBN 979-8-3315-3914-6. Disponível em: <<https://doi.org/10.1109/MOCASST65744.2025.11083923>>.

WANG, M.; HANDURUKANDE, S. B. Anomaly detection for mobile network management. **International Journal of Next-Generation Computing**, v. 9, n. 2, p. 80–98, jul. 2018. Disponível em: <<https://doi.org/10.47164/ijngc.v9i2.141>>.

YING, X. An Overview of Overfitting and its Solutions. **Journal of Physics: Conference Series**, v. 1168, p. 022022, fev. 2019. ISSN 1742-6588, 1742-6596. Disponível em: <<https://doi.org/10.1088/1742-6596/1168/2/022022>>.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks. **International Journal of Forecasting**, v. 14, n. 1, p. 35–62, mar. 1998. ISSN 01692070. Disponível em: <[https://doi.org/10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7)>.

ZHANG, H. et al. Deep Learning for Radio Map Estimation in Wireless Communication Networks. In: **Proceedings of the 2025 3rd International Conference on Artificial Intelligence, Systems and Network Security**. Xiangtan China: ACM, 2025. p. 318–326. ISBN 979-8-4007-1934-9. Disponível em: <<https://doi.org/10.1145/3797161.3797212>>.

Apêndices

Resultados Completos do Processo de Calibração dos Pesos

Este apêndice apresenta os resultados completos obtidos em cada fase do processo de calibração dos pesos do modelo de aprimoramento de cobertura. O objetivo é permitir a análise detalhada dos conjuntos de parâmetros avaliados ao longo das diferentes etapas do procedimento experimental descrito no Capítulo 4.

Cada seção a seguir corresponde a uma fase do processo de calibração, contendo as respectivas combinações de parâmetros testadas e os indicadores de desempenho associados.

A.1 Fase 1: Análise Preliminar

As tabelas abaixo representam os resultados obtidos durante a Fase 1 de calibração dos pesos. Valores em branco foram reprovados e correspondem aos que não atingiram o critério de valores superiores a 12,5% da área com boa cobertura e boa dominância, enquanto valores em amarelo foram reprovados por apresentarem percentual maior que 25% de ajustes fora do intervalo de -5 a +5 dB. Os valores em verde foram levados à avaliação na próxima fase.

Tabela A.1 – Resultados para $\lambda = 0,1$

Cobertura com dominância				Qtde. PCI com $ \Delta P > 5$ dB			
w_{dom}	$w_{SDI} = 0,1$	0,3	1	w_{dom}	$w_{SDI} = 0,1$	0,3	1
0,5	9101	9078	7607	0,5	1	1	0
1	9225	9516	9174	1	2	1	1
2	10645	10413	9586	2	4	3	2
3,5	11197	11596	10544	3,5	6	6	6
5	11943	11346	11398	5	7	7	7
7	11167	11943	11780	7	7	7	6
10	12108	11458	11287	10	12	11	7

Tabela A.2 – Resultados para $\lambda = 0,3$

Cobertura com dominância				Qtde. PCI com $ \Delta P > 5$ dB			
w_{dom}	$w_{SDI} = 0,1$	0,3	1	w_{dom}	$w_{SDI} = 0,1$	0,3	1
0,5	7361	7323	7357	0,5	0	0	0
1	7590	7587	7534	1	0	0	0
2	9112	8963	8906	2	1	1	1
3,5	9540	9517	9287	3,5	1	1	1
5	9636	10273	9776	5	2	3	1
7	10362	10971	10944	7	3	4	5
10	10784	11635	10886	10	5	7	6

Tabela A.3 – Resultados para $\lambda = 1$

Cobertura com dominância				Qtde. PCI com $ \Delta P > 5$ dB			
w_{dom}	$w_{SDI} = 0,1$	0,3	1	w_{dom}	$w_{SDI} = 0,1$	0,3	1
0,5	7258	7211	7256	0,5	0	0	0
1	7278	7224	7287	1	0	0	0
2	7372	7311	7338	2	0	0	0
3,5	7509	7447	7447	3,5	0	0	0
5	7676	7667	7642	5	0	0	0
7	9017	8989	9025	7	1	1	1
10	9194	9232	9260	10	2	2	1

A.2 Fase 2: Primeiro Refinamento

As tabelas abaixo representam os resultados obtidos durante a Fase 2 de calibração dos pesos. Valores em amarelo, reprovados, apresentaram percentual maior que 25% de ajustes fora do intervalo de -5 a +5 dB considerando-se mediana entre as *seeds* ou apresentaram menos de 80% das *seeds* válidas nesse critério. Os valores em verde foram levados à avaliação na próxima fase.

Tabela A.4 – Resultados para $\lambda = 0,1$ **Cobertura com dominância (mediana)**

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	10914	10744	
3,5	11783	11580	11070
7			11937

Cobertura sem dominância (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	10154	10206	
3,5	10317	10427	10376
7			10419

Qtde. PCI com $|\Delta P| > 5$ dB (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	5	4	
3,5	7	7	6
7			7

Aprovado (mín. 4 de 5 seeds válidas)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	S (5/5)	S (5/5)	
3,5	N (1/5)	N (1/5)	S (5/5)
7			N (1/5)

Tabela A.5 – Resultados para $\lambda = 0,3$ **Cobertura com dominância (mediana)**

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		9823	
7	10937	10918	10709
10	11484		11297

Cobertura sem dominância (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		8969	
7	10143	10147	10205
10	10205		10165

Qtde. PCI com $|\Delta P| > 5$ dB (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		2	
7	4	4	4
10	6		5

Aprovado (mín. 4 de 5 seeds válidas)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		S (5/5)	
7	S (5/5)	S (5/5)	S (5/5)
10	S (4/5)		S (5/5)

A.2.0.1 Fase 3: Segundo Refinamento

As tabelas abaixo representam os resultados obtidos durante a Fase 3 de calibração dos pesos. Todos os valores foram aprovados de acordo com os mesmos critérios estabelecidos na Fase 2, sendo os quatro maiores no desempenho de cobertura com dominância foram passados à Fase 4. Além disso, foi observado que nas quatro amostras estão representados casos de λ iguais a 0,1 e 0,3, w_{SDI} iguais a 0,1 e 1,0, e w_{dom} iguais a 3,5; 7,0 e 10,0, ou seja, há uma alta representatividade dos pesos nas amostras que evoluíram à Fase 4.

Tabela A.6 – Resultados para $\lambda = 0,1$ **Cobertura com dominância (mediana)**

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	10867	10745	
3,5			11101
7			

Cobertura sem dominância (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	10170	10199	
3,5			10362
7			

Qtde. PCI com $|\Delta P| > 5$ dB (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	5	4	
3,5			6
7			

Aprovado (mín. 8 de 10 seeds válidas)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
2	S (10/10)	S (10/10)	
3,5			S (10/10)
7			

Tabela A.7 – Resultados para $\lambda = 0,3$ **Cobertura com dominância (mediana)**

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		9787	
7	10927	10916	10685
10	11503		11278

Cobertura sem dominância (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		8986	
7	10115	10143	10209
10	10187		10165

Qtde. PCI com $|\Delta P| > 5$ dB (mediana)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		2	
7	4	4	4
10	6		5

Aprovado (mín. 8 de 10 seeds válidas)

w_{dom}	$w_{SDI} = 0,1$	0,3	1
5		S (10/10)	
7	S (10/10)	S (10/10)	S (10/10)
10	S (9/10)		S (10/10)

A.3 Fase 4: Validação Final e Consolidação dos Pesos

A tabela abaixo apresenta os resultados obtidos na Fase 4 do processo de calibração dos pesos. Nesta etapa, foram avaliados quatro conjuntos de parâmetros previamente selecionados: Conjunto 1 ($\lambda = 0,3$, $w_{dom} = 10$, $w_{SDI} = 0,1$), Conjunto 2 ($\lambda = 0,3$, $w_{dom} = 10$, $w_{SDI} = 1$), Conjunto 3 ($\lambda = 0,1$, $w_{dom} = 3,5$, $w_{SDI} = 1$) e Conjunto 4 ($\lambda = 0,3$, $w_{dom} = 7$, $w_{SDI} = 0,1$). A linha marcada em amarelo foi desclassificada pelos critérios estabelecidos, enquanto que a verde foi considerada a melhor configuração de pesos para a função otimização.

Tabela A.8 – Resultados da Fase 4

Conjunto	Cobertura com dominância	Cobertura sem dominância	PCI com ajuste > 5 dB	Seeds válidas
Conjunto 1	11367	10187	6	29/30 (97%)
Conjunto 2	11255	10225	6	26/30 (87%)
Conjunto 3	11114	10375	6	28/30 (93%)
Conjunto 4	10925	10112	4	30/30 (100%)