




Radiofrequency Signal Levels Prediction using Machine Learning Models

André Luiz da Costa Barcellos , Julio Cesar Duarte , and André Chaves Mendes 

Abstract—Mobile communication is rapidly evolving, generating a demand for networks with high quality and greater capacity. For efficient and reliable network design, it is essential to use accurate reception level prediction models to determine radio network coverage and prevent interference. Traditional models based on path loss propagation have limited accuracy, when used in urban environments, and, especially, when considering the devices' mobility. The main objective of this work is to propose a viable alternative to improve this accuracy with the use of machine learning techniques. The tests carried out in this work show that the use of the random forest technique together with attributes such as: geographic coordinates, distance, azimuth and antenna gain presents good results.

Index Terms—Radiofrequency signal levels, machine learning, path loss, mobile communication.

I. INTRODUÇÃO

A evolução da telefonia móvel celular está atraindo atenção significativa da academia e da indústria. Segundo Yijie Mao *et al.* [1], prevê-se que a sexta geração da tecnologia de comunicações móveis (6G) possibilitará a Internet das coisas, fornecendo serviços com maior *throughput*, ultra-confiabilidade, qualidade de serviço (QoS) heterogênea, conectividade massiva, e apoio a convergência de comunicações, sensoriamento, localização, computação e controle. Porém isso requer um uso mais eficiente dos recursos sem fio e poderosos meios para gerenciar a interferência.

Para o perfeito funcionamento da telefonia móvel, é necessária uma cobertura com sinais provenientes das diversas Estações Rádio Base (ERBs) das operadoras. Estes sinais se propagam até o terminal móvel por intermédio de ondas eletromagnéticas. Segundo Rappaport [2], os mecanismos de propagação da onda são atribuídos a fenômenos físicos como reflexões, difrações e dispersões, ou seja, o sinal que chega nas antenas do receptor é proveniente de diversos caminhos. Tais caminhos provocam atrasos diferentes e, por esta razão, contribuem de maneira positiva ou negativa para a composição do sinal recebido, dificultando, desta forma, a estimação destes valores, principalmente considerando o ambiente de prestação de serviço móvel. Os sistemas de telefonia móvel geralmente estão localizados em ambientes urbanos com vários prédios e construções que são obstáculos para o sinal proveniente do Transmissor (Tx) até o Receptor (Rx). Tais obstáculos provocam uma grande atenuação adicional nestes sinais, devido principalmente à difração. Por tais razões, é

muito difícil prever com precisão o nível do sinal recebido em um determinado ponto. Diferente dos canais estacionários e previsíveis, os canais de rádio são extremamente aleatórios e de difícil modelagem [2]. Para resolver este problema foram desenvolvidos modelos com o objetivo de calcular a atenuação dos sinais ao longo do caminho e assim poder estimar o nível recebido em um determinado ponto. Estes modelos são geralmente classificados como empíricos ou determinísticos.

Segundo Sarkar *et al.* [3], os modelos empíricos (ou estatísticos) são baseados na caracterização estatística do sinal recebido, são mais fáceis de implementar, requerem menos esforço computacional, e são menos sensíveis à geometria do ambiente. Já os modelos determinísticos têm como base conceitos da física e requerem uma grande quantidade de dados sobre a geometria, perfil do terreno, e localizações de construção. Dessa forma, necessitam do uso de mapas com alta resolução, que tem um custo elevado, para a identificação dos obstáculos. Os modelos determinísticos exigem mais poder computacional porém possuem uma maior acurácia [3]. A União Internacional de Telecomunicações (UIT), por intermédio de suas recomendações, sugere alguns métodos para estimar as atenuações de propagação como, por exemplo, a UIT-R-P.1546 [4] e a UIT-R-P.1812 [5]. Existem ainda outros métodos conforme sugerido por Sharma *et al.* [6] como os modelos Cost 231, Hata Cost e SUI. Estes modelos são utilizados para o cálculo das atenuações e baseiam-se em parâmetros como a distância Tx entre Rx, características do Tx, altura da antena receptora, dados do terreno, dentre outros.

Para avaliar a qualidade do serviço do sistema de telefonia móvel, o *3rd Generation Partnership Project* (3GPP), entidade responsável pela padronização de tecnologias de redes móveis, instituiu indicadores de qualidade que medem diversos parâmetros como falhas de chamadas, perda de sinais, congestionamento, potência do sinal recebido, e outros. Estes indicadores são chamados de *Keys Performance Indicator* (KPIs), dentre os quais destacam-se para este trabalho o *Reference Signal Received Power* (RSRP) que é definido como a média linear das contribuições da potência (em dBm) dos elementos de recurso que transportam sinais de referência da célula dentro da largura de banda do sinal de interesse. Para um bom mapeamento da área de cobertura da rede móvel, que é determinado pela região onde o nível de potência dos sinais recebidos RSRP é suficiente para o correto funcionamento do receptor, torna-se importante uma pesquisa de modelos de estimação cada vez mais precisos. Freitas [7] sugere um método baseado na caracterização de coeficientes de polinômios, através do método dos mínimos quadrados do algoritmo de *Levenberg-Marquardt* para o cálculo dos coeficientes com base em medidas de

André Luiz da Costa Barcellos, Julio Cesar Duarte and André Chaves Mendes are with Seção de Ensino de Engenharia de Computação, Instituto Militar de Engenharia - Urca, Rio de Janeiro - RJ, Brazil, e-mail:andrebarcellos@ime.eb.br

intensidade de campo na localidade onde deseja-se estimar os níveis. Estes coeficientes são utilizados para adequar os parâmetros dos modelos existentes aumentando a acurácia na localidade em que ele é aplicado. Em seu trabalho, as medidas de campo foram realizadas pela Universidade Federal Fluminense (UFF) no bairro de Higienópolis, localizado na Zona Norte do município do Rio de Janeiro. Recentemente tem surgido diversos trabalhos sugerindo o uso de técnicas de Aprendizado de Máquina ou *Machine Learning* (ML) para realizar a estimação destas atenuações como os de Moreta *et al.* [8], de Alimpertis *et al.* [9] e de Thrane *et al.* [10], que obtiveram bons resultados com a aplicação destas técnicas.

Conforme Mitchell [11], ML é o campo da Inteligência Artificial composto de técnicas que possuem a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência. Assim, a melhor forma de aprimorar o desempenho de um modelo de estimação dos níveis de potência recebida é utilizar a experiência adquirida pelas medidas realizadas em campo. Alimpertis *et al.* [9] utilizou o algoritmo de ML chamado *Random Forest* (RF) e, em sua conclusão, sugeriu como um possível trabalho futuro, estudar um híbrido de modelos tradicionais de propagação e ML.

Dessa maneira, o objetivo deste trabalho é utilizar as técnicas de ML para estimação dos valores de RSRP em um determinado ponto. O diferencial é a utilização de novos atributos como os parâmetros dos modelos determinísticos (distância, azimute em relação ao ponto de interesse e o ganho da antena transmissora para este azimute) além das coordenadas geográficas. Neste trabalho, os algoritmos de ML de regressão supervisionados: *Random Forest* (RF) [12], *Decision Tree* (DT) [13], *AdaBoost* (AB) [14] e *K-Nearest Neighbor* (KNN) [15] foram utilizados. O alvo a ser estimado é o nível de RSRP e, espera-se, com isso, uma diminuição dos erros nos valores estimados de RSRP. Para comprovação dos resultados e as vantagens da proposta foram utilizados dados obtidos em campo dos parâmetros de uma rede de telefonia celular de 4ª geração padrão *Long-Term Evolution* (LTE), conforme apresentado na seção III.

Este artigo está estruturado como se segue. Na Seção II é realizada uma revisão da literatura sobre modelos tradicionais de estimação de níveis de sinais e de cálculo das atenuações, além de trabalhos onde realiza-se esta estimação através do uso de técnicas de ML. A Seção III descreve os conjuntos de dados obtidos por medição de campo utilizados nos experimentos. Já a Seção IV trata da conceituação básica da metodologia no trabalho. Na sequência, os experimentos utilizados na avaliação da metodologia proposta são discutidos na Seção V. Finalmente, na Seção VI, abordam-se as considerações finais do trabalho desenvolvido e apontam-se trabalhos futuros.

II. REVISÃO DA LITERATURA

Nesta seção são abordados métodos das duas categorias de estimação, tanto os baseados em modelos tradicionais quanto aqueles baseados em técnicas de ML.

A. Modelos Tradicionais

Na literatura sobre cobertura de telefonia celular, há diversos modelos de estimação que utilizam a perda prevista

no caminho de propagação. A recomendação ITU-R 1546 [4] sugere um método para predições adequado à propagação de transmissão de um ponto para uma área na faixa de frequências de 30 MHz a 4 GHz e distâncias até 1000 km. O método é baseado na interpolação e extrapolação de dados empíricos de curvas de intensidade de campo derivadas em função da distância, altura da antena, frequência e o percentual do tempo.

A recomendação ITU-R 1812 [5] sugere outro método de predição adequado à propagação de transmissão de um ponto para uma área na faixa de frequências de 30 MHz a 6 GHz e distâncias entre 0,25 km e 3.000 km. Ela prevê níveis de sinal na mediana da distribuição de multi-percursos por determinadas porcentagens de tempo e de locais em que estes níveis são excedidos, fornecendo uma análise detalhada com base no perfil do terreno.

O modelo SUI, foi desenvolvido pela Universidade de Stanford e proposto para as faixas de frequência abaixo de 11 GHz divididos para três tipos de terrenos, nomeadamente A, B e C. O tipo A está associado à perda máxima de percurso e é apropriado para terrenos montanhosos com densidades de folhagem moderada a pesada. Já o tipo B é caracterizado por terrenos, na sua maior parte, planos com densidades de árvores de moderada a pesada ou terrenos montanhosos com densidade leve de árvores. Finalmente, o tipo C está associado à perda mínima de caminho e se aplica a terrenos planos com densidades leves de árvores. A equação básica de atenuação devida à propagação possui fatores de correção em função do tipo de terreno.

O modelo SUI corrigido mencionado neste trabalho trata-se do modelo SUI com os fatores de correção alterados conforme sugerido por Freitas [7], onde estes novos valores de coeficientes são calculados para cada cidade analisada em função dos valores medidos em campo. Tal cálculo permite que o modelo seja melhor adequado à estimação das atenuações de percurso nestas localidades.

O modelo Hata Cost, segundo Rappaport [2], foi criado pela *European Cooperative for Scientific and Technical* (EURO-COST) que formou o comitê de trabalho COST-231. Esse modelo foi projetado para ser usado na faixa de frequência de 1.500 MHz a 2.000 MHz e contém correções para ambientes urbanos, suburbanos e rurais. Sua simplicidade e a disponibilidade de fatores de correção o tornaram amplamente utilizado para o cálculo de atenuações devido a propagação.

Já o modelo Extended Hata é recomendado para faixas de 30 a 3.000 MHz e não considera o terreno, apenas as alturas do transmissor e receptor.

B. Modelos Baseados em ML

Alimpertis *et al.* [9] utilizaram a técnica RF com os parâmetros de entrada (coordenadas, data e hora, id da célula, modelo do dispositivo, frequência e distância) em medidas de campo realizadas por dispositivos móveis com uso de aplicativos. A métrica de desempenho usada foi a RMSE. Segundo os autores, o RMSE ficou em média em torno de 3,65 dB e foi menor quando considerado os atributos de hora, o que parece não ter uma relação direta com as características de propagação em si e desta maneira pode não ser tão eficaz para outras bases de dados ou até mesmo indicar algum viés.

Já em Moreta *et al.* [8], as medidas de campo foram realizadas com uso de um analisador de espectro que coletou os níveis de intensidade de campo dos canais de TVs digitais na localidade. Na metodologia proposta, a região analisada foi separada em sub-regiões, utilizando-se o algoritmo de ML não-supervisionado K-means. Depois, para cada sub-região, foram separados dados de treino e teste e só então aplicou-se os algoritmos (OK, RF, AB, KNN, *Support Vector Regression* (SVR), *Multilayer Perceptron* (MLP) e *Lasso Regression* (LR)) para a estimação dos valores de intensidade de campo. Os atributos utilizados foram as coordenadas geográficas dos pontos onde foram realizadas as medidas e as métricas de desempenho utilizadas foram a RMSE e MAE. Os valores médios de RMSE obtidos pelos algoritmos foram respectivamente: 8,203, 5,891, 6,188, 6,037, 6,252, 6,402 e 8,055 dB.

Thrane *et al.* [10] utilizaram a técnica de *Deep Neural Networks* (DNNs), com os atributos: coordenadas, dados relativos ao transmissor, distância entre Tx e Rx e imagens de satélites. As medidas de campo foram realizadas com uso de um equipamento de *drive-test* modelo TSMW [16] e a métrica de desempenho utilizada foi a RMSE. Segundo os autores, o modelo proposto obteve um ganho de 1 dB na frequência de 811 MHz e 4,7 dB na frequência de 2630 MHz, em comparação aos modelos tradicionais. Já o RMSE final foi em torno de 4,3 dB. A desvantagem é a necessidade de capacidade de processamento para o tratamento das imagens de satélites e o tempo gasto neste processo.

III. CONJUNTOS DE DADOS UTILIZADOS

Para a execução dos experimentos, fez-se uso de medidas de campo realizadas pela Anatel durante atividades de verificação de cobertura de operadoras em três cidades no interior do Brasil. Tais medidas são realizadas com uma plataforma de *drive-test* que consiste em um *scanner* do fabricante Rohde & Schwarz, modelo TSMW [16], conectado a três antenas omnidirecionais instaladas no teto de um veículo. Duas antenas para recepção do sinal LTE a ser medido e uma antena destinada a receber os sinais de GPS e marcar a localização exata onde as medidas foram realizadas. O download destes dados foi efetuado no site da Anatel na pesquisa pública do Sistema Eletrônico de Informações (SEI) [17]. As palavras chaves utilizadas nas pesquisas foram: *drive-test* e cobertura celular.

O uso de uma ferramenta profissional na coleta de campo em substituição a dados disponibilizados por aplicativos instalados em aparelhos celulares tem como principal vantagem a maior confiabilidade na precisão das medidas pois os equipamentos profissionais utilizam antenas externas omnidirecionais e são calibrados periodicamente por organismos certificadores, o que garante uma maior acurácia dos valores medidos. As medidas utilizadas neste trabalho foram efetuadas nas cidades brasileiras de Anápolis e Sidrolândia no estado de Goiás, assim como, na cidade de Camaquã no estado do Rio Grande do Sul.

A cada medida realizada pela ferramenta de *drive-test* são coletadas diversas informações do sinal recebido como: *Physical channel*, *Bandwidth*, *Physical cell identity*, Latitude, Longitude e *Reference Signal Received Power*.

Estes sinais são provenientes de várias ERBs transmissoras espalhadas na região analisada. Como este trabalho baseia-se na estimativa dos níveis de sinais em um determinado ponto distante do transmissor, torna-se importante, isolar um único transmissor de forma a analisá-lo separadamente. Assim, as técnicas estudadas aqui podem ser facilmente adaptadas para uso em outros estudos de propagação de vários outros serviços de telecomunicações como radiodifusão ou outras transmissões sem fio. Para a separação de um único transmissor foi escolhido o parâmetro *Physical cell identity* (PCI), onde considerou-se somente o PCI com o maior número de medidas.

Ao analisar as medidas de campo foi observado que, em vários momentos, devido ao trânsito local, havia mais de uma medida no mesmo ponto, nestes casos considerou-se para o nível da RSRP, a média destes valores. Foi verificado que a função de probabilidade da densidade de potência ou do inglês *Probability density function* (PDF) das cidades Anápolis e Camaquã aproxima-se a uma distribuição gaussiana já a PDF da cidade Sidrolândia se assemelha a uma distribuição *Rayleigh*, conforme a Fig. 1, onde no eixo X temos o nível de potência recebida em dBm e no eixo Y a frequência que este nível ocorre. Não foram considerados *outliers* nos dados devido a característica de variação dos níveis no ambiente móvel, que pode sofrer variações abruptas devido a propagação.

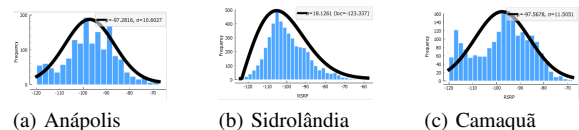


Fig. 1. PDF nas cidades.

IV. MÉTODO PROPOSTO

O método proposto baseia-se em tratar as medidas de campo selecionadas e levantar as características técnicas dos equipamentos de transmissão adicionando estas informações ao conjunto de dados como atributos nos algoritmos de ML. Ao unir no conjunto de dados elementos utilizados pelos modelos tradicionais àqueles utilizados por modelos baseados em algoritmos de ML, é possível diminuir o erro na estimação da RSRP.

A Fig. 2 mostra a metodologia básica utilizada neste trabalho para a avaliação do modelo proposto por intermédio da comparação do uso dos diversos modelos de estimação dos níveis de RSRP.

A etapa “Seleção de medidas” é onde obtém-se o conjunto de dados contendo as medidas efetuadas em campo. Já o “Pré-processamento” é uma etapa necessária para filtragem dos dados conforme descrito em detalhes na seção III. Na etapa “Cálculo de parâmetros” são incluídos atributos adicionais ao conjunto de dados que serão utilizados pelo modelo proposto. Assim é feito um levantamento das seguintes características de transmissão: o diagrama de irradiação da antena transmissora, as coordenadas do transmissor e o azimute da antena transmissora. Estas informações são extraídas através de pesquisa nos

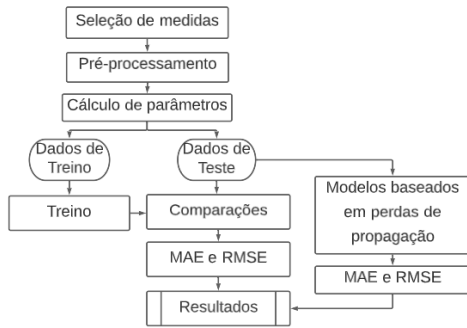


Fig. 2. Procedimento de estimação dos níveis de RSRP baseada em aprendizado de máquina.

arquivos de licenciamento das estações no sistema Mosaico da Anatel [18]. Com base em tais informações, são calculados os valores das distâncias e dos azimutes entre cada ponto de medida e o transmissor. Uma vez conhecido o valor do azimute é consultado o diagrama de irradiação da antena onde obtém-se o ganho da antena para aquele azimute. Os valores de distância, azimute e ganho devidamente calculados são adicionados ao conjunto de dados para serem utilizados como atributos dos modelos de ML. Cabe ressaltar que a informação do azimute além de ser importante para a determinação do ganho da antena transmissora, também indica o percurso principal do sinal e todos os obstáculos deste percurso, sendo considerado um atributo bem relevante. Ainda nesta etapa também são levantadas outras características necessárias para o uso das técnicas tradicionais como: a altura e inclinação da antena transmissora, a potência de transmissão, as perdas das linhas de transmissão e a frequência.

Nas etapas de “Dados de treino” e “Dados de teste” são selecionados de forma aleatória pontos para as fases de treinamento dos algoritmos de AM e de avaliação dos modelos de regressão gerados. Esta técnica de validação cruzada é conhecida como *holdout* onde os dados são divididos em dois conjuntos disjuntos. Tal separação deve permitir a variação da proporção entre os dados de treino e teste (entre 10% e 90%) para que durante os testes seja possível verificar a influência da quantidade de pontos de treino no valor do erro RMSE.

Na etapa de “Treino” seleciona-se o algoritmo de ML e realiza-se o treinamento, utilizando os dados de treinamento. Já na etapa “Comparações” estimam-se os níveis de RSRP com base no algoritmo treinado e compara-se o valor estimado com o valor real. Para esta comparação, são calculados os erros médio absoluto e a raiz do erro médio quadrático. Esta etapa é repetida para cada um dos algoritmos de ML utilizados: RF, AB, DT e KNN.

Na etapa “Modelos baseados em perda de propagação” faz-se a comparação entre os modelos baseados em ML, propostos neste trabalho, e os modelos tradicionais baseados nas perdas de propagação. Para isso, é necessário o cálculo dos níveis estimados em todos os pontos do conjunto de testes utilizando estes modelos, assim como a determinação dos erros.

Para o cálculo dos valores de RSRP pelos métodos tradicionais foi utilizado o software *HTZ Communication* [19], que foi alimentado com as informações das características de

transmissão levantadas na etapa anterior de “Cálculo dos parâmetros”. Devido a alguns dos modelos tradicionais testados utilizarem os dados do terreno e as características de cobertura do solo, também conhecidas como *clutter* (urbano, urbano denso, rural, estrada, floresta, lago e outros) para o cálculo das atenuações e estimação dos níveis de RSRP, foram carregados no software mapas disponíveis cuja resolução é de 30m, e quanto menor é este valor, mais detalhado é o mapa e maior é a resolução, com isso, mais preciso seriam os resultados de estimação. Já na etapa “Resultados”, são comparados os resultados dos diversos testes para avaliar o ganho no uso do modelo proposto em relação aos modelos tradicionais.

V. EXPERIMENTOS E RESULTADOS

Para avaliar a estimação baseada em modelos de ML foram criados dois experimentos. Um para o teste de desempenho, avaliação da seleção dos atributos e o impacto no RMSE do modelo em função do tamanho de conjunto de treino, e outro, como um teste para avaliar o modelo em condições mais reais de campo e comparação com os modelos tradicionais.

A. Teste de Desempenho

O teste de desempenho é baseado em testes realizados por Alimpertis *et al.* [9] com a intenção de avaliar qual é a melhor configuração do uso das técnicas de ML que apresenta a melhor acurácia nos níveis de potência estimados. Foram efetuados testes com todos os pontos disponíveis no conjunto de dados de cada uma das três cidades. Na Fig. 3, temos um resumo das etapas para este experimento.

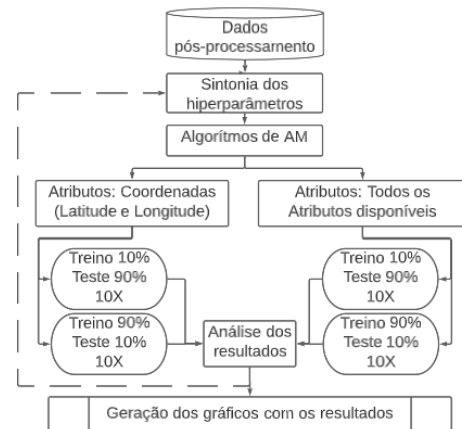


Fig. 3. Procedimento de estimação da RSRP do primeiro experimento.

Para este experimento, considerou-se uma variação na quantidade de elementos de treino no intervalo entre 10% e 90% dos dados, sendo utilizado o restante dos dados para o teste. Em cada proporção o processo é repetido 10 vezes com a escolha aleatória dos dados e o valor do erro final é dado pela média dos erros dos 10 testes. Estes testes foram realizados de duas maneiras diferentes. A primeira tendo como atributo somente as coordenadas geográficas e a segunda com todos os atributos disponíveis, ou seja, além das coordenadas geográficas, a distância do transmissor, o azimute do ponto

em relação ao transmissor e o ganho da antena transmissora no azimute do ponto.

Para maximizar o desempenho do modelo, com um menor erro RMSE possível, efetua-se uma busca em grade dos hiperparâmetros, com o uso da ferramenta *GridSearchCV* da biblioteca *scikit-learn* em linguagem de programação Python. Na tabela I tem-se os intervalos e o valor de cada um dos hiperparâmetros escolhidos. Para uma análise da contribuição individual de cada atributo gera-se a matriz de correlação entre os atributos disponíveis conforme Fig. 4

TABELA I
TABELA COM OS VALORES DOS HIPERPARÂMETROS.

Algor.	Tempo(s)	Hiperparâmetro.	Intervalo	Valor
RF	0,5	n_estimators	10 – 1000	250
		max_depth	2 – 200	100
		min_samples_split	2 – 128	2
		max_features	2 – 3	2
AB	0,4	n_estimators	10 – 1000	100
		learning_rate	0.1 – 2.2	0.1
		loss	[linear, square, exponential]	square
KNN	0,1	n_neighbors	1 – 20	6
		p	1 – 2	1
DT	0,1	max_depth	2 – 200	250
		max_features	2 – 3	2

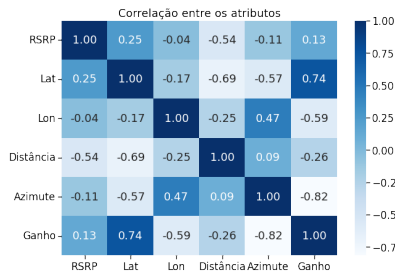


Fig. 4. Matriz de correlação entre atributos disponíveis.

Observou-se que os erros obedecem uma distribuição gaussiana, na Fig. 5 são mostradas as distribuições de cada método na cidade de Sidrolândia, onde no eixo X temos o erro em dB e no eixo Y a frequência que este erro ocorre, com isso o uso da métrica RMSE para avaliação dos resultados torna-se bastante coerente.

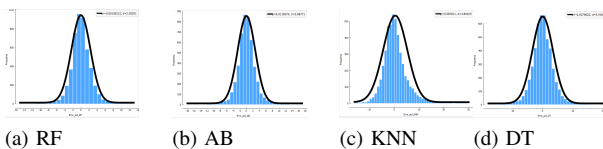
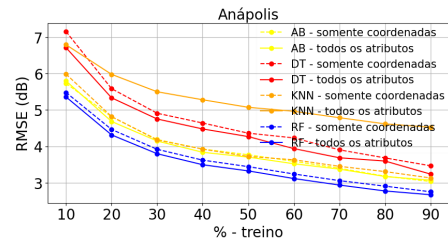


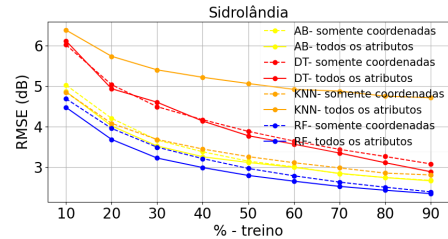
Fig. 5. Distribuição estatística dos erros de cada modelo de ML.

Os resultados do experimento podem ser visualizados nos gráficos da Fig. 6, onde as linhas pontilhadas representam os modelos tendo como atributos somente as coordenadas geográficas e as linhas contínuas representam os mesmos modelos tendo como atributos todos os parâmetros disponíveis.

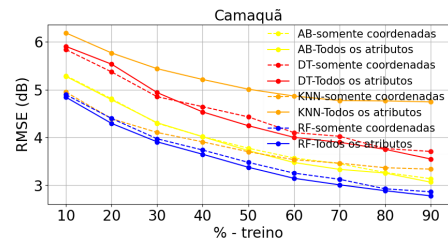
Analisando tais resultados vê-se que o RF apresenta erros menores em relação às outras técnicas. Isso se deve ao fato das diversas árvores de regressão funcionarem como uma



(a) Anápolis em relação ao PCI (235)



(b) Sidrolândia em relação ao PCI (221)



(c) Camaquã em relação ao PCI (320)

Fig. 6. Gráficos com os RMSE x parâmetros de entrada.

espécie de agrupamento das áreas medidas, sendo o resultado a média das diversas árvores, onde os erros são minimizados. O uso do RF com apenas as coordenadas geográficas já representa um enorme ganho no desempenho em relação os outros algoritmos e apresenta resultados ainda melhores quando utilizados atributos adicionais relativos à transmissão, tais como: a distância e o azimute entre Tx e Rx, e o ganho da antena de transmissão neste azimute. Outro ponto de destaque desta técnica é apresentar um bom desempenho mesmo com um número reduzido de amostras.

A técnica KNN, diferente do RF, teve um pior desempenho com a adição dos novos atributos se comparada com os resultados obtidos com somente as coordenadas. Isso explica-se pelo fato do KNN ser um modelo baseado em distâncias e que possui uma relação bem forte com as coordenadas espaciais. Conforme Faceli *et al.* [20] o problema de métodos baseados em distância é pressupor que os dados correspondem a pontos em um espaço d-dimensional (R^d), porém os demais atributos, por não possuírem essa característica, acabam divergindo os resultados ao mesmo tempo que funcionam como ruído.

B. Teste em Condições Reais de Campo

Este teste simula uma aplicação do modelo proposto para estimar a RSRP de vários pontos dentro de uma determinada área e compara o resultado com a estimativa realizada utilizando-se os modelos tradicionais para estes mesmos pontos. Dessa forma, a área foi dividida em sub-regiões para que

o treino e o teste fossem realizados em sub-regiões diferentes, adicionando-se uma certa distância entre os pontos de treino e teste, e evitando assim uma possível tendência do modelo em decorar os valores próximos durante a estimação.

Na Fig. 7 temos um resumo das etapas para este experimento.

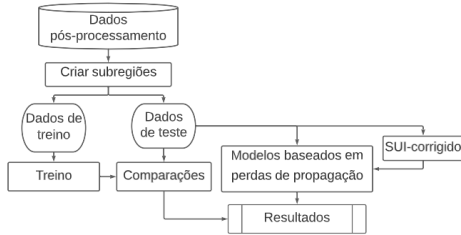
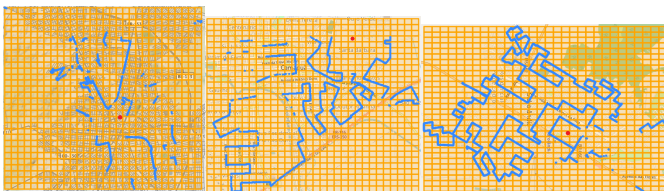


Fig. 7. Procedimento de estimação da RSRP do segundo experimento.

Na etapa de “Criar sub-regiões”, para cada conjunto de dados de cada cidade foi determinada uma região retangular no mapa onde foram tomadas as medidas e esta região foi dividida em 900 sub-regiões (30 linhas x 30 colunas). Algumas destas sub-regiões possuíam valores medidos de RSRP, e outras não, devido ao trajeto do veículo durante a tomada das medidas em campo. A Fig. 8 mostra este cenário para cada uma das cidades analisadas. Nesse cenário, as sub-regiões foram numeradas no sentido de leitura da esquerda para a direita e de cima para baixo, tendo sido incluída uma coluna na apresentação do conjunto de dados. Já as sub-regiões sem medidas foram suprimidas. Após a atualização do conjunto de dados, selecionou-se aleatoriamente sub-regiões para o treino e outras para teste na proporção de 80% e 20% respectivamente para a aplicação dos algoritmos de ML.

Na etapa “Modelos baseados em perdas de propagação” foram calculados, por intermédio dos métodos tradicionais REC.1812 UIT, REC.1546 UIT, Hata Cost, SUI e Extended Hata, os valores dos níveis de RSRP para todos os pontos utilizados para teste na etapa anterior utilizado-se o software *HTZ communications* [19]. Já para o SUI corrigido, foi utilizada a ferramenta CFtools do Matlab, para calcular os níveis estimados por intermédio das fórmulas do modelo SUI com os coeficientes corrigidos conforme Freitas [7].



(a) Cidade de Anápolis (b) Cidade de Camaquã (c) Cidade de Sidrolândia

Fig. 8. Medidas realizadas e a divisão em sub-regiões. O ponto em vermelho indica a posição do transmissor avaliado.

Na etapa de “Resultados” são apresentados o MAE e o RMSE de todos os modelos testados para comparação. As comparações entre os valores estimados de RSRP por intermédio dos métodos tradicionais e os encontrados nas medidas de campo para cada uma das cidades pode ser visualizado na

Tabela II, assim como no gráfico da Fig. 9. A comparação entre os valores estimados de RSRP e os valores encontrados nas medidas de campo utilizando-se as técnicas de ML, podem ser visualizados na Tabela III, assim como no gráfico da Fig. 10, onde, nas últimas colunas, foram exibidos os resultados com menor erro das técnicas tradicionais.

TABELA II
ERRO DOS VALORES ESTIMADOS PELOS MÉTODOS TRADICIONAIS (dB).

Cidade	Erro	1812	1546	Hata Cost	Ext. Hata	SUI	SUI Corrig.
Anápolis	RMSE	13.14	15.42	11.70*	11.39	17.50	11.54
	MAE	9.32	10.94	8.75*	8.19	12.24	10.01
Sidrolândia	RMSE	15.05	16.64	18.23	18.23	18.91	14.22
	MAE	10.50	12.81	13.15	13.16	15.25	11.16
Camaquã	RMSE	11.05	8.80	12.60*	13.45	10.10	8.73
	MAE	8.31	6.54	8.85*	9.93	8.31	6.55

* Embora tenha sido calculado para comparação, o método Hata Cost não foi concebido para frequências acima de 2 Ghz.

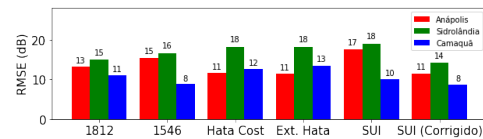


Fig. 9. Gráficos comparativo dos erros pelos métodos tradicionais.

TABELA III
ERRO DOS VALORES ESTIMADOS PELOS MÉTODOS BASEADOS EM ML (dB).

Cidade	Erro	Menor Erro					Modelo
		RF	AB	KNN	DT	Modelo	
Anápolis	RMSE	4.39	5.13	5.53	5.36	11.39	Ex-Hata
	MAE	3.42	3.72	4.13	3.92	8.19	Ex-Hata
Sidrolândia	RMSE	4.07	4.69	4.49	4.88	14.22	SUI corrigido
	MAE	3.14	3.38	3.42	3.59	10.50	1812
Camaquã	RMSE	5.20	5.56	4.80	6.77	8.73	SUI corrigido
	MAE	3.86	4.00	3.59	4.64	6.54	1546

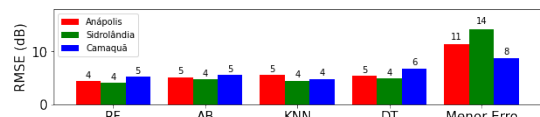


Fig. 10. Gráficos comparativo dos erros.

Conforme ilustrado nos gráficos da Fig.10, os resultados apresentados pelos modelos de ML possuem menores erros na estimação dos valores de RSRP que os melhores resultados utilizando os modelos tradicionais baseados nas perdas de propagação. Já no desempenho entre as diversas técnicas de ML, percebe-se uma grande vantagem da técnica RF em relação às outras.

VI. CONCLUSÕES E TRABALHOS FUTUROS

A comunicação sem fio está evoluindo rapidamente gerando uma demanda por redes com alta qualidade e maior capacidade. Para um projeto de rede eficiente e confiável é essencial o uso de modelos de estimação dos níveis de recepção precisos para a determinação da cobertura da rede de rádio e prevenção

de interferências. Os modelos tradicionais baseados nas perdas devidas à propagação possuem uma acurácia limitada quando utilizados em ambientes urbanos, sobretudo quando considera-se a mobilidade dos dispositivos atualmente.

O principal objetivo deste trabalho é propor uma alternativa viável para um menor erro na estimação dos valores de RSRP com o uso de técnicas de ML. A proposta não é substituir os métodos tradicionais e sim complementá-los. Podemos assim concluir que a técnica RF mostra-se adequada, principalmente quando considerados os parâmetros do azimute do ponto e ganho da antena transmissora mesmo com um número baixo de amostras de treino. O ganho com o uso do modelo proposto em relação aos modelos tradicionais foi da ordem de 11 dB na frequência de 1.837 MHz, na cidade de Sidrolândia, e de 7 dB a 3 dB na frequência de 2.665 MHz, nas cidades de Anápolis e Camaquã, respectivamente.

Esta técnica deve ser utilizada em softwares dedicados a esta estimação como um complemento aos modelos tradicionais, tornando-os mais baratos, eficientes e robustos. Aplica-se esta técnica também em futuras tecnologias como uma maneira de estimar as condições de um canal de comunicação onde o dispositivo aprenderá, com o passar do tempo, a estimar as perdas em um determinado ponto, permitindo o controle mais eficiente de parâmetros como potência e ganho das antenas em função do posicionamento do transceptor destino.

Como sugestão de trabalhos futuros podem ser citadas pesquisas em faixas de frequências mais elevadas utilizadas nos sistemas de 5ª geração de serviços móveis, como 3,5 GHz e até as ondas milimétricas de 26 GHz em regiões menores tendo em vista a diminuição do alcance.

REFERÊNCIAS

- [1] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Communications Surveys & Tutorials*, pp. 1–1, 2022.
- [2] T. Rappaport, *Comunicações sem fio: princípios e práticas*. Pearson Prentice Hall, 2009.
- [3] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, pp. 51–82, 2003.
- [4] ITU-R, "Method for point-to-area predictions for terrestrial services in the frequency range 30 mhz to 3 000 mhz," Recommendation 1546-6, International Telecommunication Union, Geneva, Aug. 2019.
- [5] ITU-R, "A path-specific propagation prediction method for point-to-area terrestrial services in the vhf and uhf bands," Recommendation 1812-6, International Telecommunication Union, Geneva, Sept. 2021.
- [6] P. K. Sharma and R. Singh, "Comparative analysis of propagation path loss models with field measured data," *International Journal of Engineering Science and Technology*, vol. 2, no. 6, pp. 2008–2013, 2010.
- [7] P. V. A. De Freitas, *Avaliação e Modelagem da Cobertura de Sinal na faixa de 3, 5 GHz para um Ambiente com Clutter*. MSc dissertation, Universidade Federal Fluminense, 2021.
- [8] C. Moreta, M. R. C. Acosta, and I. Koo, "Prediction of digital terrestrial television coverage using machine learning regression," *IEEE Transactions on Broadcasting*, pp. 1–11, 2019.
- [9] E. Alimpertis, A. Markopoulou, C. Butts, and K. Psounis, "City-wide signal strength maps: Prediction with random forests," in *The World Wide Web Conference, WWW '19*, (New York, NY, USA), p. 2536–2542, Association for Computing Machinery, 2019.
- [10] J. Thrane, D. Zibar, and H. L. Christiansen, "Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 ghz," *IEEE Access*, vol. 8, pp. 7925–7936, 2020.
- [11] Mitchell, *Machine Learning*. McGraw Hill, 1997.
- [12] A. Cutler, D. R. Cutler, and J. R. Stevens, *Random Forests*, pp. 157–175. Boston, MA: Springer US, 2012.

- [13] W.-Y. Loh, "Classification and regression trees," *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 1, pp. 14–23, 2011.
- [14] A. Kadiyala and A. Kumar, "Applications of python to evaluate the performance of decision tree-based boosting algorithms," *Environmental Progress & Sustainable Energy*, vol. 37, no. 2, pp. 618–623, 2018.
- [15] Y. Song, J. Liang, J. Lu, and X. Zhao, "An efficient instance selection algorithm for k nearest neighbor regression," *Neurocomputing*, vol. 251, pp. 26–34, 2017.
- [16] R. . Schwarz, "Universal radio network analyzers user manual," tech. rep. 1171.5960.32-8, Rohde & Schwarz GmbH & Co. KG, 2017.
- [17] Anatel, "SEI pesquisa pública." Disponível em: https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_processo_pesquisar.php?acao_externa=protocolo_pesquisar&acao_origem_externa=protocolo_pesquisar&id_orgao_acesso_externo=0. Acesso em: 28 maio, 2022.
- [18] Anatel, "Licenciamento de estações." Disponível em: <http://sistemas.anatel.gov.br/se/public/view/b/licenciamento.php>. Acesso em: 28 maio, 2022.
- [19] ATDI, "Htz communications v22.6.9 modelling all radio communications technologies from 10khz to 350 ghz." <https://atdi.com/products-and-solutions/htz-communications-for/>, 2020.
- [20] K. Faceli, A. C. Lorena, J. Gama, and A. C. P. L. F. de Carvalho, *Inteligência Artificial Uma Abordagem de Aprendizagem de Máquina*. LTC, 2019.



André Luiz da Costa Barcellos has a degree in Electrical Engineering (emphasis on Telecommunications) from University Estácio de Sá, in 2006. He is currently a graduate research student with the Instituto Militar de Engenharia (IME). He has experience in multidisciplinary areas, with an emphasis on telecommunications, data science, low power electric and propagation research



Julio Cesar Duarte has a degree in Computer Engineering from Instituto Militar de Engenharia (IME), in 1998, a Master's in Informatics from Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), in 2003, and a Ph.D. in Informatics from PUC-Rio, in 2009. He is currently a professor at the Graduate Program in Systems and Computing at IME. He has experience in multidisciplinary areas, with an emphasis on systems development, working mainly on the following topics: machine learning, deep learning, artificial intelligence, Portuguese natural language processing and malware analysis.



André Chaves Mendes holds a degree in Telecommunications Engineering (2001) from the University of São Paulo (Poli-USP), a Master's degree (2010) and a D.Sc. (2015) both in Teleinformatics and Automation from the Federal University of Rio de Janeiro (COPPE/UF RJ), Brazil. He is currently a professor at the Graduate Program in Systems and Computer Engineering (PPGSC) of the Military Institute of Engineering (IME-RJ), Brazil, and a post-doctoral researcher at the Autonomous Systems Engineering Department, University of Vigo, Spain, working mainly on the following topics applied to industry: Cyber-Physical Systems, Internet of Things, Intelligent Systems and Data Analytics.